Dissertation submitted to the
Combined Faculty of Natural Sciences and Mathematics
of the Ruperto Carola University Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by M.Sc, Valentina Carlini

Born in Rome, Italy

Oral examination: 11.05.2021

Mechanisms and limits of epigenetic perturbation inheritance in mammals

Referees: Prof. Dr. Sylvia Erhardt
Dr. Matthieu Boulard

Acknowledgements

The last four years at EMBL Rome were so rich and intense that I feel like I've been there forever. But everything comes to an end and it is time to say goodbye. I'm looking forward to start new adventures and now I want to thank all those who have contributed to this journey.

First, I wish to express my gratitude to my supervisor. Dear Jamie, thanks for giving me the opportunity to join your lab. At the beginning it was just the four of us, and I was popping into your office for every technical and insignificant issue, like that extra band on the PCR gel. With time I saw the lab growing and I really felt part of it. It was a period of tremendous progress both scientifically and personally and your support has been invaluable. You are a brilliant scientist and a great mentor and I feel very lucky to have met you.

I'm thankful to the members of my doctoral examination committee, Prof. Sylvia Erhardt, Dr. Matthieu Boulard, Prof. Frank Lyko, Prof. Edith Heard, for taking time to read my work.

Dear Sylvia, although we only met virtually, your insightful comments and suggestions were really valuable to find the right way to move forward in my project. Dear Matthieu, I want to thank you for being always available and eager to help both on the scientific and personal point of view. It made me feel reassured that I could always knocked at your door. Dear Prof Lyko, I would like to thank you for accepting to be in my examination committee and for always being available when I had questions. Dear Edith, it is an honor to have you in my committee. The year spent in your lab was instrumental for shaping my scientific interests and for giving me confidence to apply for PhDs programs. That period in Paris was really important for me to acquire self-awareness.

I also want to express my gratitude to Dr. Eileen Furlong and Dr. Claire Rougeulle for their suggestions and inputs at all stages of my research project during my thesis advisory meetings. They were really useful to find what was missing in my project.

I want to thank my lab mates for team up and share with me successes and failures in these years. Thanks Monica for our chats in the lab, for being my confidant and my reference for everything that has to do with embryos. Cristina, for helping me many times when I realised I couldn't cope on my own and for our walks in the mountains. Thank you Kristjan, for being the person I could ask for any doubt I had about techniques or a specific analysis like that CRISPR screen plot. Juliette, for being always kind with me and for bringing your smile in the lab. Ayele, I can not thank you enough for the many times I asked you to move my agar plates out from the warm room on weekends, without you the entire institute would be covered in bacteria carrying

recombinant CRISPR constructs by now. Thanks to Tafsut, you have been the best trainee one could ever ask for, unfortunately it did not last long before the pandemic started. Thanks to Marzia and Catrin, we haven't shared much time together yet, but you brought a new breeze in the lab.

Especially I want to thank all the people that makes EMBL Rome lively, sparkling and a great working environment. Thanks Yasmin, Sara, Sofia, Marina, Monika, Tom, Anastasiia, Irene and all the other amazing people of EMBL for accompanying me in this adventure. Thanks for all the summer parties, the debate clubs and the beer hours we shared. I wish I could properly celebrate one more time with you once it will be possible. Also, I'm thankful to all the people that contributed to this thesis with their expertise. Thanks Cora and Chris for the many hours spent at FACS sorting millions of cells and to all the other people in the facilities which support was instrumental for this thesis.

Finally, I would like to spend a few words to thank my family and friends that have been there with me all the time. Thank you to my soulmate, Nicola, the person that listens to me till late night trying to understand the difference between a Cas9 that binds and cuts from one that is inactive and only used to recruit something else. The one who learned that scientists are totally dedicated to their projects and that if you want to plan a weekend out you first have to deal with tissue culture duties. Thanks for your unwavering support and belief in me.

Thank you, mum and dad, you gave me the invaluable tool to devote myself to my studies and work, without having to worry about anything else. You imprinted my way. Thanks to my sister, Chiara who always put up with me when I was coming back home nervous after a tough day in the lab. I also appreciate all the support I received from the rest of my family and family-in-law, especially my grandma and my aunties. I'm also thankful to my lifetime friends who have been patient with me lately and got organized to meet up at odds hours to fit my needs.

This work would not have been possible without the contribution of each of you.

Abstract

Epigenetic systems contribute to genome regulation in health and disease, and underpin chromatin-based memory. The epigenome is therefore tightly regulated but evidence has emerged that altered epigenetic states (epialleles) can be induced in response to perturbations or environmental stimuli. Moreover, these can be mitotically or meiotically heritable in *yeast*, *worms* and *plants,* driving phenotypes independent of the genotype. Nevertheless, the prevalence and significance of epialleles in *mammals* remains unclear. To investigate the potential for transmission of acquired chromatin states in *mammals*, I optimised a modular and releasable *dCas9* system coupled with KRAB epigenetic repressor to programme epigenetic states to endogenous loci in an ESC model of development. With this tool, I was able to induce *de novo* heterochromatin domains comprising physiologically relevant levels of DNA methylation, H3K9me3, H4K20me3, with concurrent loss of H3K4me3, leading to absolute silencing of local genes at the single-cell level.

Despite the extant paradigm predicting that such major heterochromatin regions are epigenetically transmitted, here I observed that deposited heterochromatin domains exhibit only transient memory function, which is rapidly reverted with time and DNA replications in pluripotent cells. By loss of function genome-wide CRISPR screening coupled with my epigenetic memory assay, I found that *Dppa2* is specifically responsible for counteracting epigenetic memory of epialleles. DPPA2 is a small protein that together with DPPA4 binds to most GC-rich gene promoters and is exclusively expressed in pluripotent cells. I found that deletion of *Dppa2* enables robust epigenetic memory of programmed heterochromatin in ESC, without influencing cell identity. Furthermore unlike ESC, I observed that epigenetic memory is maintained in *wildtype* lineage-restricted cells, which do not express *Dppa2/4*, under selective conditions that favour the epiallele. This includes stable inheritance of epigenetic silencing at the tumour suppressor gene *p53* in cell subpopulations, in both *in vitro* and *in vivo* assays. This result provides a proof of principle that epimutation of genes that facilitate a selective advantage, such as *p53*, can be inherited during organogenesis *in vivo*, with implications for predisposition to diseases including cancer. I propose this reflects the synergistic influences of weak-acting epigenetic inheritance and positive epiallele selection. This may be relevant to multiple gene-environment contexts in mammals and has relevance to the concept of 'soft inheritance'.

Taken together, my data demonstrate that epigenetic memory of induced heterochromatin is limited by DPPA2 activity, and consequently epialleles are rapidly reset in pluripotent, but not lineage-restricted, cells. By ectopically forcing expression of *Dppa2/4* in these differentiated cells, I observed epigenetic reversion of the *p53*-repressed epiallele. Overall, this suggests that *Dppa2* acts as a 'surveyor' mechanism that can 'sense' epigenetic aberrations during pluripotent phases to guard against the subsequent inheritance of acquired epialleles.

Abstrakt

Epigenetische Systeme tragen zur Genomregulation in Gesundheit und Krankheit bei und unterstützen das chromatinbasierte Gedächtnis. Das Epigenom ist daher streng reguliert, aber es gibt Hinweise darauf, dass veränderte epigenetische Zustände (Epiallele) als Reaktion auf Störungen oder Umweltreize induziert werden können. Darüber hinaus können diese in Hefe, Würmern und Pflanzen mitotisch oder meiotisch vererbbar sein und Phänotypen unabhängig vom Genotyp steuern. Dennoch bleibt die Prävalenz und Bedeutung von Epiallelen bei Säugetieren unklar. Um das Potenzial für die Übertragung von erworbenen Chromatinzuständen in Säugetieren zu untersuchen, habe ich ein modulares und freisetzbares dCas9-System, gekoppelt mit dem epigenetischen Repressor KRAB, optimiert, um epigenetische Zustände auf endogene Loci in einem ESC-Entwicklungsmodell zu programmieren. Mit diesem Werkzeug war ich in der Lage, de novo Heterochromatin-Domänen zu induzieren, die physiologisch relevante Niveaus von DNA-Methylierung, H3K9me3, H4K20me3 mit gleichzeitigem Verlust von H3K4me3 aufweisen, was zu einer absoluten Silencing von lokalen Genen auf Ebene einzelner Zellen führt.

Trotz des bestehenden Paradigmas, das vorhersagt, dass solche großen Heterochromatin-Regionen epigenetisch weitergegeben werden, habe ich hier beobachtet, dass abgelagerte Heterochromatin-Domänen nur eine vorübergehende Gedächtnisfunktion aufweisen, die mit der Zeit und DNA-Replikationen in pluripotenten Zellen schnell rückgängig gemacht wird. Durch ein genomweites CRISPR-Screening mit Funktionsverlust, gekoppelt mit meinem epigenetischen Gedächtnis-Assay, fand ich heraus, dass *Dppa2* spezifisch für das Entgegenwirken des epigenetischen Gedächtnisses von Epiallelen verantwortlich ist. DPPA2 ist ein kleines Protein, das zusammen mit DPPA4 an die meisten GC-reichen Genpromotoren bindet und ausschließlich in pluripotenten Zellen exprimiert wird. Ich fand, dass die Deletion von Dppa2 ein robustes epigenetisches Gedächtnis von programmiertem Heterochromatin in ESC ermöglicht, ohne die Zellidentität zu beeinflussen. Darüber hinaus beobachtete ich, dass im Gegensatz zu ESC das epigenetische Gedächtnis in Wildtyp-Zellen, die *Dppa2/4* nicht exprimieren, unter selektiven Bedingungen, die das Epialle begünstigen, erhalten bleibt. Dies schließt die stabile Vererbung von epigenetischem Silencing am Tumorsuppressor-Gen *p53* in Zellsubpopulationen ein, sowohl in in vitro als auch in vivo Assays. Dieses Ergebnis liefert einen Grundsatzbeweis dafür, dass die Epimutation von Genen, die einen selektiven Vorteil

ermöglichen, wie *p53*, während der Organogenese in vivo vererbt werden kann, was Auswirkungen auf die Prädisposition für Krankheiten einschließlich Krebs hat. Ich schlage vor, dass dies die synergistischen Einflüsse von schwach wirkender epigenetischer Vererbung und positiver Epiallele-Selektion widerspiegelt. Dies könnte für mehrere Gen-Umwelt-Kontexte bei Säugetieren relevant sein und hat Bedeutung für das Konzept der "weichen Vererbung".

Zusammengenommen zeigen meine Daten, dass das epigenetische Gedächtnis des induzierten Heterochromatins durch die DPPA2-Aktivität begrenzt ist und Epiallele folglich in pluripotenten, aber nicht lineage-beschränkten Zellen schnell zurückgesetzt werden. Indem ich ektopisch die Expression von *Dppa2/4* in diesen differenzierten Zellen erzwang, beobachtete ich die epigenetische Reversion des *p53*-unterdrückten Epiallels. Insgesamt deutet dies darauf hin, dass DPPA2 als "Überwachungsmechanismus" fungiert, der epigenetische Aberrationen während pluripotenter Phasen "erkennen" kann, um die anschließende Vererbung von erworbenen Epiallelen zu verhindern.

Table of Contents

List of Tables

Abbreviations

5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydoxymethylcytosine
5mC	DNA 5-methylcytosine
ATAC-seq	Assay for Transposase-Accessible Chromatin using sequencing
BD	Bromodomain
BET	Bromodomain-extraterminal protein family
BFP	Blue fluorescent protein
C. elegans	*Caenorhabditis elegans*
CAF1	HP1-Chromatin assembly factor 1
CD	Chromodomain
CDS	Chromoshadow domain
CGI	CpG island
ChOR-seq	Chromatin occupancy after replication
CRISPR	Clustered regularly interspaced short palindromic repeats
CUT&RUN	Cleavage Under Targets and Release Using Nuclease
DAPI	4,6-diamino-phenylindole
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DOHaD	Developmental origins of health and disease
Dot1L	H3K79 methyl-transferase
DOX	Doxycycline
DOXwo	Doxycycline washout
Dppa2	Developmental pluripotency associated factor 2
Drosophila	*Drosophila Melanogaster*
DSBs	Double strand breaks
E.Coli	Escherichia Coli
EpiSC	Epiblast stem cells
ERV	Endogenous Retroviruses
ESC	Embryonic stem cell
Esg1	Embryonal stem cell-specific gene 1

FACS	Fluorescent activated cell sorting
FBS	Fetal bovine serum
FDR	False discovery rate
GFP	Green fluorescent protein
gRNA	Guide RNA
H3K27	histone H3 lysine 27
H3K4	histone H3 lysine 4
H3K9	histone H3 lysine 9
H4K20	histone H4 lysine 20
HAT	Histone acetyl-transferase
HCP	High CpG density promoter
HDAC	Histone deacetylases
HEK293T	Human embryonic kidney cell line
HMT	Histone methyl-transferase
HP1	heterochromatin Protein 1
IAP	Intracisternal A particle
ICM	Inner cell mass
ICP	Intermediate CpG density promoter
ICR	Imprinting Control Regions
IDE1	Induce definitive endoderm 1
iPS	Induced pluripotent stem cells
Kansl2	KAT8 Regulatory NSL Complex Subunit 2
KAP1	KRAB-associated protein 1
Kb	Kilobase
KDM	Lysine de-methylases
Kmt2d	H3K4 specific histone methyltransferase
KO	knockout
KRAB	Krüppel associated box domain
LCP	Low CpG density promoter
LIF	Leukemia inhibitory factor
LTR	Long terminal repeats
MAGeCK	Model-based Analysis of Genome-wide CRISPR-Cas9 Knockout
Mb	Megabase
MeCP2	Methyl-CpG-binding protein

ncRNA	Non-coding RNA
NGS	Next generation sequencing
NPC	Nucleosome Core Particle
p53	Tumour protein p53
PAM	Protospacer adjacent motif
PBS	Phosphate-buffered saline
PcG	Polycomb group proteins
PCNA	Proliferating cell nuclear antigen
PEV	Position Effect Variagation
PGC	Primordial germ cell
piRNA	Piwi RNA
PRC1/2	Polycomb repressive complex 1/2
PRE	Polycomb response element
PTM	Post-translational Modification
qPCR	Quantitative PCR
RFB	Replication fork barrier
RNA	Ribonucleic acid
RRA	Relative ranking algorithm
SAM	S-Adenosyl methionine
scFv	single-chain variable fragment
siRNA	Short interfering RNA
Smarcc1	SWI/SNF histone remodeller factor
TAD	Topologically associated domain
TALE	Transcription activator-like effector
TE	Transposable elements
TET	Ten-eleven translocation methyl-cytosine dioxygenases
TOM	tdTomato fluorescent protein
TSS	Transcriptional starting site
Untr.	Untransfected
WB	Western Blot
WT	wildtype
XCI	X chromosome inactivation
Zmym3	zinc finger MYM-type 3

"Chance events—injuries, infections, infatuations; (…) impinge on one twin and not on the other. Genes are turned on and off in response to these events, as epigenetic marks are gradually layered above genes, etching the genome with its own scars, calluses, and freckles."

- Siddhartha Mukherjee, The Gene: An Intimate History (2015)

1 Introduction

1.1 Epigenetics: from past to present

When Waddington coined the term *epigenetics* in 1942, broadly defining it as changes in phenotype without changes in genotype, there was very little understanding of its mechanism (Waddington, 1942a, b).

Historical ground-breaking studies by Muller in *Drosophila Melanogaster* (Muller, 1930) and McClintock in *maize* (McClintock, 1951) on position effect variegation (PEV) and transposable elements, respectively, provided the first examples of non-Mendelian inheritance. In the course of the following decades, these pioneering findings fostered more research that lead to Holliday's definition of epigenetic traits as -traits that are mitotically heritable without a change in DNA sequence- and he defined these changes "epimutations" (Holliday, 1987). Many examples of heritability due to the mitotic transmission of an epigenetic mark rather than a DNA variant come from the plant kingdom (Luo et al., 1995).

In mammals, the first and most striking epimutation observed is that responsible of the *agouti viable yellow* (A^{vy}) phenotype described by Morgan and colleagues in 1999 (Morgan et al., 1999), where the expression of the *agouti* gene depends on the epigenetic status of a retrotrasposable element located upstream its promoter. Three years later, Cooney and collaborators and Waterland and co-workers

independently observed that the *agouti* phenotype of the offspring was affected by the diet of the mother suggesting that the external environment could introduce epigenetic perturbations (Cooney et al., 2002; Waterland and Jirtle, 2003).

The notion that epigenetic systems allow cells with the same genotype to exploit different functions it is now known for more than 60 years (Nanney, 1958). Despite the numerous advantages that have been made in the field since these observations, the precise mechanisms by which replicating cells inherit the parental epigenetic states to maintain their identity is not completely unravelled yet. Also, whether similar mechanisms are involved in the inheritance of aberrant epimutations is not known.

Currently, the definition of the word 'epigenetics' is ambiguous (Greally, 2018). Here, I will refer to it as a malleable layer of molecular mechanisms that can respond to perturbations and environmental changes and stably influence chromatin structure and accessibility, ultimately modulating gene expression.

It is only in the past 20 years that epigenetic knowledge at the molecular level has been enhanced, thanks to increasingly sensitive techniques, such as genome wide chromatin profiling (Johnson et al., 2007). This lead to a burst of discoveries, marking the modern era of epigenetic research (Allis and Jenuwein, 2016). Shortly after the engineering of the CRISPR-Cas9 tool by Charpentier and Doudna groups in 2012, Cas9 has been further modified to disrupt its catalytic activity and to be exploited as a modulable docking platform to recruit specific epigenetic effectors to chromatin (Gilbert et al., 2013; Jinek et al., 2012; Qi et al., 2013). The simplicity and scalability of CRISPR-Cas9 system compared to previous epigenome editing tools has revolutionized the entire epigenetics field, providing a tool for locus specific deposition of epigenetic marks on demand.

The 'CRISPR revolution', together with the increase in the sensitivity of epigenomics techniques, provides nowadays unprecedent tools to investigate among others, the heritability of epimutations *in vitro* and *in vivo*.

1.2 Chromatin

1.2.1 Chromatin structure

In eukaryotic nuclei the genetic material is organized in a structured yet dynamic complex with proteins to form chromatin. Chromatin is composed by repeating subunits, the nucleosome core particle (NPC), in which 147 bp of DNA is left-handed wrapped 1.7 times around each nucleosome to achieve the first level of compaction, the nucleosome fiber. Around 30 bp of DNA is left out from the NPC and form the linker DNA. Nucleosomes consist in eight histone proteins (called histone octamer): H2A and H2B organized to form dimers and H3 and H4 organized in tetramers, composed on two of each histones.

Histones are mainly characterised by a globular domain but the remaining 20-30% of the mass is composed of a largely structurally undefined but evolutionary conserved N-terminus tail rich in lysine and arginine residues which makes it extremely basic. Thanks to its charge, this domain is responsible of inter-nucleosomal interactions and long-range fiber-fiber contacts to mediate condensation of chromatin into higher-order chromatin structures (Allan et al., 1982; Schwarz, 1994). Moreover, this tail is subjected to numerous post-translational modifications (PTMs) that modify its charge and, directly or indirectly altering tail interactions, elicits different chromatin states (Zheng et al., 2003).

Another layer of chromatin diversification is given by a number of histone variants that can replace the canonical ones at specific chromatin locations or during specific timepoints of the cell cycle. The replacement occurs mainly during S-phase, referred as replication-coupled (RC), when the new histones are deposited behind the replication fork barrier to fill in the gaps left by the redistribution of the parental histones (Henikoff and Smith, 2015). Deposition can also be replication-independent (RI) and involve local replacement of an existing nucleosome or subunit (Marzluff et al., 2002). For instance, the histone variant H2A.Z, which facilitates transcriptional competence, substitutes the canonical H2A by the activity of specific ATP-dependent chromatin remodelling complexes (Wu et al., 2005).

In addition to the four ‘core’ histones, another primary component of most eukaryotic chromatin are linker histones (H1 and H5), that are found with a stoichiometry of approximately 1:1 of histone:nucleosome (Li and Zhu, 2015). As its name implies, linker histones are associated with linker DNA, providing it with partial protection from nucleases. They also exerts many other roles including regulation of

gene expression (Shen et al., 1996), influence of nucleosome spacing on DNA (Blank and Becker, 1995) and promotion of folding and assembly higher order chromatin structure (Allan et al., 1986). Electron microscopy experiments initially, revealed that the nucleosome fiber is further assembled to form a bigger fiber with a diameter of 30 nm, so called 30 nm fiber (Finch, 1976). However, a number of recent studies conducted on frozen hydrated sections of yeast and mammalian cells provided evidence that chromatin in native conditions consists of irregularly folded 10-nm, and not 30 nm, fiber (Chen et al., 2016; Cai et al., 2018).

How chromatin further folds into higher-order structure is still largely under debate (Woodcock and Ghosh, 2010). The maximum level of compaction is reached in the metaphase chromosome (850 nm) in which the DNA is condensed 10.000 to 20.000-folds. Instead, during interphase, chromatin is additionally organised in structures ranging in the diameters from 120 nm/170 nm to 250 nm (Kireev et al., 2008). This more or less compaction of the nucleosome chain reflects different functional states of the chromatin (Kieffer-Kwon et al., 2017).

Organization of DNA into this higher-order structures, not only serves to compact the 1.8 mt long DNA into nuclei of ~10 nm diameter, but also to achieve high level of control over DNA transactions (DNA transcription, replication, repair and recombination). Therefore, folding and regulation of chromatin metabolism are strongly interconnected. On the other hand, these structures are meta-stable and can be disassembled and reassembled to allow rapid access of the DNA to chromatin regulators.

1.2.2 Chromatin compartments

At larger scales chromatin is organised in separated genomic regions in which DNA on the same or different chromosomes can interact with each other forming distinct compartments. Most of these compartments have been originally defined based on the differences in apparent chromatin compaction visible by microscopy (Heitz, 1928). The denser and dark foci of condensed chromatin stained with 4,6-diamino-phenylindole (DAPI) throughout the cell cycle were called heterochromatin to distinguish them from domains that were not stained after telophase called euchromatin. Generally, transcriptionally repressed genomic regions are heterochromatic while euchromatin is characterised by active and accessible regions.

Different varieties of heterochromatin can be defined by the combinations of histone post-translational modifications in constitutive (c-Het) or facultative (f-Het)

heterochromatin. A typical mark of c-Het is di- or tri-methylation of histone H3 at Lysine 9 (H3K9me2 or me3) while f-Het is usually enriched in tri-methylation at Lysine 27 (H3K27me3) (Filion et al., 2010). In both cases histones are usually hypoacetylated. Constitutive heterochromatin is typically concentrated at the nuclear lamina and at peri-centromeric regions, rich in 'satellite repeats', and generally has the crucial function to prevent expression of transposable and repetitive elements that can otherwise introduce genetic instability. On the other hand, facultative heterochromatin is considered to be more dynamic and can be formed at various chromosomal regions to lock cell-type specific cues in a development-dependent manner. Finally, euchromatic regions are characterised by methylation of Lysine 4 of histone H3 (H3K4me3) and acetylation of various other residues. These regions are rich in active genes and enhancer elements.

In most cell-types euchromatin is generally located at the nuclear interior while heterochromatin is segregated at the nuclear periphery. However, this partitioning is dynamic. For example, during cell differentiation many regions are repositioned from the nuclear lamina to the interior and vice-versa (Peric-Hupkes et al., 2010). Interestingly, targeting of a repressed region with an engineered transactivator protein VP16, not only induces its transcriptional activation but also the redistribution of the locus from the periphery to a more interior location of the nucleus across several micrometers in a couple of hours' time (Tumbar and Belmont, 2001).

High throughput chromosome conformation capture measurements have revealed that chromosomes are subdivided into contact domain in the range of kilobases (Kb) or megabases (Mb) length, so called topologically associated domains (TADs) (Nora et al., 2012; Sexton et al., 2012). TADs have been described to influence transcription by insulating regulatory sequences from interacting in neighbouring domains (Lupianez et al., 2015) and by preventing spreading of epigenetic marks acting as barriers between domains (Narendra et al., 2015).

1.3 Epigenetic systems in mammals

Histone variants and post translational modification of histone tails can affect the recruitment of proteins to and folding of chromatin. The most studied histone PTMs involve modifications of key lysine (K) or arginine (R) residues on the histones H3, H4, H2A and H2B such as methylation, acetylation, phosphorylation ubiquitination. In this chapter, I will mainly focus on the first two modifications. Acetylation has the effect of changing the overall charge of the highly positive histone-tail residues and in turn loosens nucleosomes interaction making the DNA more accessible. As a result, histone acetylation has been linked with transcriptional activation. On the other hand, none of the three lysine methylation states (mono- di- and tri-methylation) changes the electronic charge of the aminoacidic-histone tail. Thus, the function of histone methylation in mediating chromatin regulation is exerted only through enzymes that recognize this modification (Martin and Zhang, 2005). In turn, according to lysine residue position, histone methylation is associated with activation (eg. H3K4me3) or repression (eg. H3K27me3) of transcription.

Enzymes called 'writers', catalyse these chemical modifications and are usually highly specific for a given residue and sometimes can be redundant. According to the modification that they deposit these enzymes are called histone acetyl-transferases (HAT) or histone methyl-transferases (HMT); HMTs that specifically modify lysine are called KMT. These latter enzymes use intermediate metabolites such as acetyl-CoA and S-adenosylmethionine (SAM) as donors for histone acetylation and methylation respectively.

These modifications generate a so called 'histone code' and they can act as highly selective binding platforms for the association of specific regulatory proteins, the histone 'readers', that in turn direct distinct DNA-template programs (Strahl and Allis, 2000). Several conserved domains have been identified in these proteins to be responsible of binding specific modified residues. For example, the chromodomain (CD) and PHD domains recognise and bind methylated histone H3, being involved in regulation of gene expression. The bromo-domain (BD) instead targets acetylated histones and regulate transcription, repair, replication and chromosome condensation.

All chromatin marks known so far are reversible and many studies identified a number of enzymes involved in removing these modifications called 'erasers'. Similarly to the 'writers' counterpart, 'erasers' are classified as histone de-acetylases (HDAC) or lysine de-methylases (KDM).

Other epigenetic systems include methylation of cytosine at CpG on DNA and non-coding RNA. Combination of all the aforementioned epigenetic systems orchestrate complex gene expression plasticity and inheritance and are subject of intense research.

1.3.1 Heterochromatin components

1.3.1.1 H3K9 methylation

While fission yeast have a single H3K9 methyltransferase (Clr4/KMT1) that is responsible for all the three states of methylation (Nakayama et al., 2001), mammalian cells have several of them with different roles in diverse cellular events (Hyun et al., 2017). Among the different KMTs in mammals, SETDB1 catalyses H3K9 mono-methylation in constitutive heterochromatin, that acts as a substrate for SUV39H1/2 to promote di- and tri-methylation (Lachner et al., 2001; Loyola et al., 2009). Another KMT, G9a-GLP mono- and di-methylates H3K9 in euchromatic regions to promote gene repression (Tachibana et al., 2002).

Biochemical studies conducted by Müller and colleagues described a two-stages activation mechanism by which, after recognition of H3K9me3 by its CD, SUV39H1 undergo allosteric activation of a latent chromatin binding motif that exerts both an anchor function and stimulation of its methyltransferase activity on spatially close nucleosomes (Müller et al., 2016). This mechanism is at the basis of H3K9me3 spreading together with the interaction with the heterochromatin protein (HP1). In fact, SUV39H1, recruits the binding of HP1 in combination with HP1 own chromodomain that recognises methylated H3K9 (Lachner et al., 2001). H3K9me3 deposited by other KMTs, like G9a-GLP, is not able to recruit HP1 (Stewart et al., 2005). HP1 in turn recruits the binding of other chromatin modifiers including H3K9 methyltransferases and histone deacetylases that catalyse the spreading of heterochromatin in a DNA-sequence independent manner (Zhang et al., 2008). In addition to its CD, HP1 also have a C-terminal chromoshadow domain (CSD) that mediates intra- and inter-protein interactions. A current model suggests that the interaction via their CSD of two HP1s bound to H3K9me3 bring in close proximity adjacent nucleosomes thus promoting chromatin compaction (Canzio et al., 2013). In addition, the CDS can recruit also other chromatin remodelling and modifying factors that further cooperate to chromatin condensation (Platero et al., 1995).

1.3.1.2 DNA methylation

DNA methylation occurs at the fifth carbon of cytosines (5-methylcytosine (5mC)), predominantly in the context of symmetric CpG di-nucleotides (Wyatt, 1950). However, the mammalian genome is generally CpG poor except for discrete regions (on average 1kb in length) with high density of these dinucleotides, named CpG islands (CGIs). CGIs characterise around two-third of gene promoters (Gardiner-Garden and Fromme, 1987) and according to the density of CpG content, promoters are classified in low, intermediate and high CpG-density promoters (LCP, ICP, HCP) (Mikkelsen et al., 2007). These dense CGIs, like the ones found at HCP, are very rarely methylated (Bird et al., 1985), while studies in human and mouse showed that ICP are the most responsive promoters to DNA methylation regulation (Meissner et al., 2008; Weber et al., 2007).

There are three phases of DNA methylation: establishment (de-novo DNA methylation), maintenance and demethylation. Each of these phases are catalysed by specific enzymes called DNA methyltransferases (DNMT) that 'write' the modification, using SAM as methyl-donor, and the Ten-eleven translocation methylcytosine dioxygenases (TET) family with an 'erasing' activity.

In mammals there are two main *de novo* DNMTs (DNMT3A and B), together with the catalytically inactive DNMT3L, which acts as a cofactor and stimulates the DNA methyl transferase activity in the germline (Bourc'his et al., 2001; Okano et al., 1999; Okano et al., 1998). DNMT3 enzymes have both a 'writing' domain, the MTase, located at the C-terminus, and two chromatin-reading domains: the PWWP and the ADD (ATRX-DNMT3-DNMT3L). This last domain binds to unmethylated H3K4. When H3K4 is methylated, it repels the ADD domain that thus binds to the MTase domain inhibiting its own catalytic activity (Guo et al., 2015). Being as these two marks are mutually exclusive, DNA methylation is almost always associated with repression of transcription. On the other hand, the PWWD domain binds to H3K36me3, a mark that is typically found at gene bodies, where DNA methylation probably has a role in repressing intragenic cryptic promoters, and facilitating transcriptional elongation and co-transcriptional splicing (Greenberg and Bourc'his, 2019).

Maintenance DNA methylation is carried out by DNMT1 together with the multidomain ubiquitin E3 ligase UHRF1. DNMT1 by itself exist in an autoinhibitory configuration. UHRF1 binds specifically to hemimethylated CpG and recruits and activates DNMT1 (Ishiyama et al., 2017; Song et al., 2011). As a result, when DNMT1 auto-inhibition is released, it not only methylates the daughter DNA strand, but also binds H3 tails that have been previously ubiquitinated by UHRF1 (Nishiyama et al., 2013; Qin et al., 2015).

DNA de-methylation, instead, can occur either by passive DNA-replication dependent dilution or can be actively carried out by TET enzymes which progressively oxydize 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Ito et al., 2010; Ito et al., 2011; Kriaucionis, 2009).

DNA methylation has historically been associated with gene repression (Wolffe, 1999) but this epigenetic mark does not confer silencing per-se. The repressive activity might be exerted by the recruitment of other heterochomatin remodellers or by a decreased affinity of transcription factors to methylated promoters compared to the unmethylated counterparts, as shown in a recent study (Yin et al., 2017). However, the main target of DNA methylation in mammalian genomes are not genes but transposable elements (TE) and genome defence against TE has since long been proposed as a major driver of DNA methylation evolution (Yoder and Bestor 1997).

DNA methylation is also the main regulator of imprinted genes, a class of genes which are mono-allelically expressed according to the parent of origin. The maternal and paternal alleles are characterised by a differential methylation state at specific imprinting control regions (ICRs) which are usually of high CpG density. ICRs may function either as insulators recruiting proteins that prevent long range interaction between enhancers and promoters or involving functional non-coding RNA *in cis* (Reik, 2007).

1.3.1.3 Crosstalk between H3K9me and DNA methylation

Crosstalk between H3K9me3 and DNA methylation occurs at many levels in mammals and it is now clear that these two marks rely on each other to establish efficient chromatin function. Many loss-of-function genetic studies revealed how much these two marks are inter-connected. *Setdb1*, *G9a*, *Suv39H1* and *2* independent knockouts in mouse embryonic stem cells (mESC) revealed either global or locus specific loss of DNA methylation (Arand et al., 2012; Ikegami et al., 2007). Conversely, *Dnmt* triple-knockout shows a decrease in global H3K9me3 in human cancer cell lines but not in mESC (Espada et al., 2004; Tsumura et al., 2006).

Numerous are the examples of direct interaction between H3K9 and DNA methyltransferases found in literature. For instance, at pericentromeric heterochromatin, the SUV39H1-HP1 complex directly recruits DNMT3B (Lehnertz et al., 2003). Similarly, during DNA replication, G9a coordinates both H3K9 methylation and DNA methylation through recruitment of DNMT1 (Esteve et al., 2006).

In addition to direct interactions, other co-factors can act as a bridge between these two marks. For example, UHRF1 binds with its TTD domain H3K9me2/3 and this promotes DNA methylation by DNMT1 (Liu et al., 2013). However, mutation in the

TTD domain that abolish Uhrf1 capability to bind H3K9me2/3, had little effect on DNA methylation maintenance in vivo in mouse (Zhao et al., 2016). H3K9me3-Uhrf1-mediated interaction therefore promotes but is not essential for maintenance DNA methylation.

DNA methylation and H3K9me3 have the important role of silencing endogenous retroviruses (ERVs), a class of transposable elements that account to 10% of the mouse genome and that can be responsible of genetic instability. However, the relevance of these two marks on ERV's silencing depends of the developmental stage. In the early preimplantation embryo and in ESC, characterized by a global DNA hypomethylation, KRAB-associated protein 1 (KAP1, also called TRIM28) is targeted to ERVs by the zinc-finger proteins KRAB and recruits SETDB1 to deposit H3K9me3 (Matsui et al., 2010; Rowe et al., 2010). In absence of *Dnmt3a* and *b* and *Dnmt1* these retroelements are not re-activated in mESC, showing that DNA methylation is not needed for proviral silencing at this stage (Hutnick et al., 2010). Once H3K9me3 is established, silencing is superseded by DNA methyltransferases DNMT1 and DNMT3, and H3K9me3 is lost at later stages of development (Leung and Lorincz, 2012; Matsui et al., 2010; Rowe et al., 2010).

Although the evidence reported above indicates that H3K9me3 acts upstream of DNA methylation, the reverse is also true and DNA methylation deposition can be preparatory for H3K9me3. For example, the methyl-CpG-binding protein (MeCP2) contains a domain that binds to methylated DNA and is able to recruit histone de-acetylates and histone methyltransferases such as SUV39H1 to reinforce repressive chromatin states at targeted neuronal genes (Ballas et al., 2005; Fuks et al., 2003).

Several other crosstalks are known and when studying the effect of each regulatory mark, one has always to consider it as part of a complex interconnected mechanism.

1.3.1.4 Dependency of H4K20 methylation on H3K9me3

H4K20me3, together with H3K9me3 and DNA methylation is highly enriched at heterochromatin (Mikkelsen et al., 2007). H4K20 methylation is catalysed by many histone methyltransferases specific for each methylation state. SET8/KMT5A is responsible of H4K20 mono-methylation that, on its hand, is a substrate of SUV4-20H1/KMT5B and SUV4-20H2/KMT5C that catalyse di- and tri-methylation respectively (Fang et al., 2002; Schotta et al., 2004). SUV4-20H2 is recruited to heterochromatin through its interaction with HP1 which interacts with H3K9me3 (Schotta et al., 2004). *Suv39h1/h2* or *Hp1* mutants show global reduction of H4K20me3 therefore targeting H4K20me3 is strongly modulated by H3K9me3 (Yang et al., 2008).

H4K20 methylation plays a role in chromatin compaction through the L3MBTL1 (L(3)mbt-like 1) methylation binding protein. L3MBTL1 recognises and binds H4K20me1/me2 on two nucleosomes in proximity and interacts with histone H1 to induce chromatin condensation (Trojer et al., 2007). Confirming this role on nucleosome compaction, knockdown of SET8/KMT5A in human embryonic kidney cell line (HEK293) induced aberrant nuclei formation and chromatin decondensation (Houston et al., 2008), while in vitro H4K20me3-marked nucleosome array provided direct evidence of chromatin compaction (Lu et al., 2008).

1.3.2 Polycomb proteins

Polycomb group (PcG) proteins are among the main constituents of facultative heteochromatin and play an important role in orchestrating cell-differentiation and developmental programs, such as X-chromosome inactivation (XCI). PcG are mainly divided in two classes: Polycomb repressive complex 1 and 2 (PRC1 and PRC2) deputed to the deposition of H2AK119 ubiquitination and H3K27 methylation respectively (Schuettengruber et al., 2017).

Mammalian PRC2 is composed of four core proteins among which there are the histone methyltransferase EZH1/2 and the reader EED (Kuzmichev et al., 2002). Bearing both reading and writing activity in the same complex, PRC2 is thought to facilitate H3K27me3 spreading to adjacent loci in a positive feedback loop (Hansen et al., 2008; Margueron et al., 2009). PRC2 can be recruited at specific genomic regions by accessory proteins or by non-coding RNAs such as in the case of XCI. In this case, the expression of a long non-coding RNA Xist occurs early during development and induces the recruitment of PRC2 mediated H3K27me3 on the inactive X chromosome (Xi), required to stabilize Xi chromatin structure (Silva et al., 2003).

As H3K27me3 also H2AK19ub1 is associated with transcriptional repression and is deposited by the catalytic component of PRC1 complex, the ubiquitin E3 ligase RING1B. In mammals there are two main PRC1 complexes (Gao et al., 2012) all containing the RING1B protein but differing in the accessory subunits. The canonical PRC1 has a CBX subunit (CBX-PRC1) that contains a chromodomain specific for H3K27me3 and its recruitment on chromatin depends on this latter mark (Blackledge et al., 2014). Variant PRC1 instead is recruited independently of H3K27me3 and contains the ubiquitin binding protein RYBP (PRC1-RYBP). Recently, it was demonstrated that PRC1-RYBP induces propagation of H2AK119Ub1 via a positive feedback loop facilitated by H1-mediated chromatin compaction (Zhao et al., 2020).

Variant PRC1 complexes were also shown to drive nucleation of H3K27me3 (Blackledge et al., 2014). This context-dependent interaction between PRC1 and PRC2 can be thus considered cooperative rather than hierarchical binding (Schuettengruber et al., 2017).

1.3.3 Markers of active chromatin

It has long since been shown that acetylation of histone H3 and H4 neutralises the positive charge of lysine residues counteracting the formation of highly compacted chromatin structures (Garcia-Ramirez, 1995; Tse et al., 1998). HATs usually establish broad domains of histone acetylation that lead to partial decondensation of chromatin and marks regions of transcriptional competence. Consequently, promoters are more accessible for transcription initiation and chromatin unfolding also facilitates transcriptional elongation per se (Eberharter and Becker, 2002). Besides this 'passive' role in chromatin decompaction, histone acetylation can recruit proteins through their bromodomain which themselves can regulate transcription. These bromodomain proteins includes many HATs, chromatin remodelling complexes and the bromodomain-extraterminal (BET) family. This latter family has two BD domain and are responsible of recruiting general transcription factors to chromatin (Josling et al., 2012).

Other than acetylation, also some methylated residues on histone H3 correlate with transcriptional activation. For example, H3K4me3 predominantly localises at the 5' end of active genes and associates with RNA pol II, phosphorylated at serine 5 and thus instructed for transcriptional initiation (Eissenberg and Shilatifard, 2010). Another mark, H3K36me, instead, accumulates at the 3' end, as the H3K36me3-specific methyltransferase SETD2 interacts with the elongating form of RNA pol II, phosphorylated at serine 2 (Kizer et al., 2005).

Mammals encode for six complexes that catalyse H3K4me3 (SET1A/B and MLL1-4) each containing a SET methyltransferase domain and accessory proteins that are required both for the catalytic activity and for influencing the mono- di- or trimethylation state (Shilatifard, 2012). Each of these complexes show essential roles in positively regulating specific targets since individual MLL mutants are embryonic lethal (Howe et al., 2017).

How H3K4me3 links with active transcription is not yet completely understood and we do not know how SET/MLL complexes are recruited to chromatin. Despite

H3K4me3 correlation with active transcription, loss of H3K4me3 in mESC only induces mild changes in gene expression genome wide and at specific CpG island promoters (Clouaire et al., 2012). Therefore, recently, the growing idea is that H3K4me3 is not instructive for transcription but instead might be deposited as a result of it, influencing processes such as splicing, transcription termination and serving as memory of previous transcriptional state (Howe et al., 2017).

1.3.4 Non-coding RNAs

Diverse classes of RNAs have been proposed to be involved in transcriptional regulation and chromatin architecture. These include RNAs that do not code for any protein (non-coding RNA (ncRNA)) and usually function by recruiting complexes or by building self-reinforcing loops with other heterochromatin components.

Short ncRNAs comprise short interfering RNAs (siRNA) and piwiRNA (piRNA) and are usually shorter than 32 bp. For example, in mammalian germ cells, piRNA are involved in silencing transposable elements by targeting them for DNA methylation (Aravin et al., 2008; Carmell et al., 2007). Long ncRNAs instead are typically longer than 200 nt and have been found to interact with many chromatin modifiers such as PRC2, YY1 and CTCF (Mishra and Kanduri, 2019). The most studied lncRNAs comprise, other than Xist, that I already briefly described in section 1.3.2, KCNQ1 opposite strand/antisense transcript 1 (*Kcnq1ot1*) and HOX transcript antisense RNA (*HOTAIR*) and exert their function by interacting with PRC2 (Pandey et al., 2008; Rinn et al., 2007).

Therefore, many ncRNAs have repressive functions and act in combination with H3K9 methylation, Polycomb or DNA methylation. Nevertheless, in mammals, there is also a class of ncRNA that are transcribed from enhancer regions, namely enhancer RNA (eRNA), that can positively regulate transcription. They usually act as nascent RNAs and activate transcription of neighbouring genes *in cis* (Orom et al., 2010).

1.4 Mitotic stability (chromatin replication and epigenome maintenance)

Propagation of genetic information through cell division is ensured by semiconservative DNA replication. Nevertheless, DNA replication poses a challenge for the maintenance of chromatin states and it remains elusive how epigenetic information, carried by histone modifications, can be conserved during mitosis. During DNA replication, after the passage of the replication fork barrier (RFB) that disrupts histone-DNA contacts, nucleosomes are dissociated in H3/H4 tetramers and H2A/H2B dimers (Xu et al., 2010), old histones re-associate with daughter chromosomes and shuffle with newly synthetized ones (Kaufman and Rando, 2010).

During chromatin assembly, parental H3 and H4 maintain their PTM and are evenly segregated on leading and lagging DNA strands by distinct replication fork components, ensuring symmetrical histone-PTM composition in sister chromatids (Alabert et al., 2015; Petryk et al., 2018). In addition to this observation, a recent study demonstrated, by using a method to track parental histones in a locus specific manner, that old nucleosomes are locally redeposited only at repressive chromatin domains but are not preserved at active regions (Escobar et al., 2019). This is crucial for information passage from old to new histones for example to allow propagation of chromatin states responsible of cell identity during cell replication. The restoration kinetics of parental PTM levels are very heterogeneous and modification-specific. Re-establishment of the methylation levels usually occurs in a stepwise fashion by enzymes that progressively deposits mono- di- and trimethylation (Stewart-Morgan et al., 2020).

How are silent chromatin domains maintained though DNA replication and cell division?
De and Kassis propose two possible models each supported by recent publications: the *cis-recruiting mechanism* in which histone modifying enzymes are recruited by specific DNA sequence on the replicating chromatin and the *self-propagation mechanism* according to which histone modifying enzymes are recruited by the modifications themselves (De and Kassis, 2017). In *Drosophila Melanogaster*, Polycomb Response Elements (PREs) recruit Polycomb repressive complex 2 (PRC2) that trimethylates H3K27. Supporting the first model, PREs are required both for establishment and maintenance of the silent transcriptional state (Francis et al., 2004), and in the absence of them, modified histones get diluted each cell division (Coleman and Struhl, 2017; Laprell et al., 2017). On the other hand, recent studies showed that repressed transcriptional states can be established by tethering histone

methyl transferases at synthetic transgenes. In support of the *self-propagation* model, this repressed state can be maintained through several cell divisions in the absence of the tethered enzyme (Bintu et al., 2016; Ragunathan et al., 2015).

Consistent with the *self-propagation mechanism*, most well-characterized loci subject to epigenetic inheritance are associated with big domains (Reinberg and Lynne, 2018). A mathematical model predicts that 3D spatial compaction promotes self-propagation *in trans* by increasing long distance contacts and this might cooperate with cis-recruitment to achieve strong stability (Jost and Vaillant, 2018).

As mentioned before, it is known that a number of histone modification systems contains both a reading and writing modules thus, in principle, favours positive feedback loops. What remains elusive is the extent to which these positive feedback loops actually result in heritable epigenetic states. Also, chromatin states are usually a result of interactions between different histone-PTM, DNA methylation, ncRNAs and histone variants thus understanding epigenetic memory is rather complex and cannot be exempt from considering these crosstalks.

1.4.1 DNA methylation inheritance mechanisms

Propagation of DNA methylation though cell replication relies on a DNA template system and well-studied DNA methylation binding proteins (DNMT1/UHRF1) with high affinity for hemimethylated DNA. As the replication fork proceeds, DNMT1 is recruited at the RFB by direct interaction with the sliding clamp protein PCNA (proliferating cell nuclear antigen) and UHRF1 which specifically binds hemimethylated DNA via its SRA domain (Arita et al., 2008; Chuang et al., 1997) (Fig. 1.1.a). This semiconservative copying system ensures faithful inheritance of DNA methylation though cell replication. Although DNMT1 together with UHRF1 are the main DNA methylation maintenance players, also DNMT3A and B can be involved in 5mC propagation. In this case, a pool of DNMT3A and B already bound to nucleosomes might be involved in methylation of CpG previously missed by DNMT1 in a sort of backup mechanism (Jeong et al., 2009).

The most extreme example of inheritance of this mark was recently reported in a species of a pathogenic fungus, in which, after the loss of the de-novo DNMT more than 50 mya, propagation of the ancient patterns of DNA methylation was ensured by the maintenance DNMT under pression of evolution (Catania et al., 2020). However, as further described in paragraph 1.5.2, more complex organisms, such as mammals, undergo drastic DNA demethylation during early embryo development and primordial germ cell specification (Hayashi et al., 2007). Thus, a similar long-term

DNA methylation inheritance would not be possible outside of specific loci which retain this mark and are known as escapees DNA-methylation reprogramming is therefore considered a major barrier for transgenerational transmission of epigenetic information in mammals, and understanding the extent of this heritability is of great interest in the field.

1.4.2 Propagation of H3K9 methylation

Differently from DNA methylation, histone marks do not have a DNA template system to copy the modification on the newly deposited histones. However, similarly to the DNA methylation maintenance-machinery, a mechanism to ensure the recruitment of specific histone methyltransferases at the replication fork barrier has been described at pericentric heterochromatin. In this case, the HP1-Chromatin assembly factor 1 (CAF1) chaperone complex is recruited by the PCNA at the RFB and is required for feedback from old to new histones by engaging SetDB1 that monomethylates H3K9 before histone deposition (Loyola et al., 2009). Monomethylated H3K9 subsequently serve as a substrate for SUV39H1 to restore H3K9me3 in a stepwise manner and ensure propagation of this mark at pericentric heterochromatin (Fig. 1.1.b).

As described in section 1.3.1.1 many histone writers, such as SUV39H1, also have a reading function and this feedforward loop has been proposed to ensure self-propagation of these marks independently of the underlying DNA sequence (Allshire and Madhani, 2018). Nevertheless, many studies revealed that this read-write coupling mechanism is not always sufficient to guarantee epigenetic memory outside of its canonical context. Using similar approaches in yeast, two groups showed that H3K9me3 deposited by tethering the histone methyltransferase $Clr4^{SUV39H1}$ was rapidly erased upon release of the enzyme, unless the potential histone demethylase Epe1 was removed from the cells (Audergon et al., 2015; Ragunathan et al., 2015). These results show that active removal of the marks can counteract its stable inheritance. Alternatively, another study in fission yeast revealed that siRNAs can strength H3K9me3 memory in wild type cells that otherwise would have been erased by Epe1 (Yu et al., 2018). This suggest a role for siRNA and H3K9me3 coupled positive feedback loops to counteract epigenetic memory erasure.

Because H3K9me3 and DNA methylation are heavily interconnected by cross-reading systems, the faithful propagation mechanism of this latter one could help *cis* inheritance of H3K9me3 read-write mechanism. Experiments in mammalian somatic cells by transient targeting of engineered transcriptional repressors suggested that H3K9me-mediated repression is reversible, but the silent state was persisting for

many cell replications if in combination with DNA methylation (Amabile et al., 2016; Bintu et al., 2016). By contrast, wide H3K9me3 heterochromatin domain at endogenous murine regions obtained by HP1 locus-specific tethering was heritably propagated through multiple cell divisions, although in this case the role of DNA methylation in its maintenance cannot be ruled out (Hathaway et al., 2012).

Concurrently, these results illustrate that epigenetic memory through cell replication is dependent on a multi-layered and complex network of histone modification positive feedbacks, together with DNA methylation semiconservative propagation and possibly ncRNAs. Consequently, the potential of mitotic and meiotic epigenetic inheritance of these systems remains incompletely understood

1.4.3 Epigenetic memory of Polycomb marks

Similarly to H3K9 methylation, also propagation of polycomb marks is facilitated by a read-write coupling system. As described in section 1.3.2, the PRC1 subunits RYBP-RING1B and PRC2 subunits EED-EZH2 establish a positive feedback loop (Margueron et al., 2009; Zhao et al., 2020) (Fig. 1.1.c). This is at the basis of information passage from old to newly incorporated histones in replicating chromatin for H2AK119Ub1 and H3K27me3 respectively. Supporting this, by transient targeting of PRC2 on a reporter construct, a group showed that H3K27me3 was transmitted for several cell divisions in human fibroblast (Hansen et al., 2008). This ensures that H3K27me3 is propagated not only during chromatin replication but also outside S-phase to preserve chromatin structure and transcriptional programs. In mammals restoration of the parental levels of H3K27me3 is quite slow and generally is completed by G1 phase (Sharif and Koseki, 2017). In addition, in drosophila, Polycomb also employs specific DNA-sequence dependent anchoring factors and their disruption results in loss of PcGs protein and silencing through cell divisions (Coleman and Struhl, 2017; Laprell, 2017).

Interestingly, in plants, a mechanism by which PRC2 is recruited at the RFB by PCNA has also been proposed and links H3K27me3-mediated vernalization with DNA replication, strictly regulating the process of flowering after a long cold (Jiang and Berger, 2017). This process is biphasic and require distinct Polycomb components: a nucleation phase that is metastable and a propagation phase that is dependent on DNA replication to confer memory stability (Yang et al., 2017). Whether a similar two-step mechanism also occurs in mammals needs to be unravelled.

Although H3K27me3 is well known to induce stable repression trough cell divisions, for long it remained unclear whether it is the mark itself or PRC2 that remains bound

to it, that is the real driver of memory. By crossing sperm and oocytes deprived of either H3K27me3 or PRC2 in *C. elegans*, a study revealed that in the absence of PRC2, H3K27me3 provide short term-memory consistent with the passage of the mark between DNA replication, while PCR2 is required for long term inheritance of the signal (Gaydos et al., 2014). Whether PRC2 is itself inherited trough chromatin or is *de-novo* recruited by H3K27me3 remains unclear. By quantifying chromatin-associated proteome at different stages of the cell cycle it was recently reported that repressive modifiers usually remain associated with chromatin during S-phase. On the other hand chromatin activators are evicted from it (Ginno et al., 2018). This suggests another layer of regulation of transcriptional states propagation and give insights regarding the differential inheritance of active versus repressed chromatin states.

1.4.4 Is euchromatin inherited?

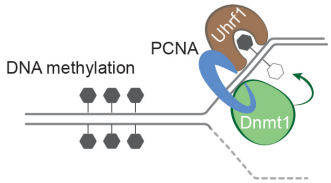
Compared to repressive chromatin, the mechanism underlying active chromatin propagation are less understood. The roots of this lack of knowledge lie in the still uncertain role of H3K4me3 as cause or consequence of transcription as already mentioned in section 1.3.3 (Howe et al., 2017). Collectively, the studies already discussed, showing that nucleosomes marked with active PTM are not redeposited locally after cell divisions (Escobar et al., 2019) and that positive chromatin regulators are not retained on chromatin during DNA replication (Ginno et al., 2018), argue against inheritance of active chromatin states. In addition, among all the bromodomain proteins that can read acetylated histones found so far, none of them show a positive feedback loop by stimulating the activity of the writer (Reinberg and Lynne, 2018).

In contrast to this, experiments of nuclear transfer in the amphibian *Xenopus Laevis,* revealed that the correct development of the embryo is limited by H3K4me3-transcriptional memory of the donor cell (Hormanseder et al., 2017). Importantly H3K4me3 from the donor somatic cell was retained after nuclear transplant on the myogenic gene MyoD in non-muscle cell lineage showing that, in this case, memory is independent on the transcriptional state (Ng and Gurdon, 2008). A form of cell memory is thus encoded by H3K4me3 to maintain gene expression stability in somatic cells.

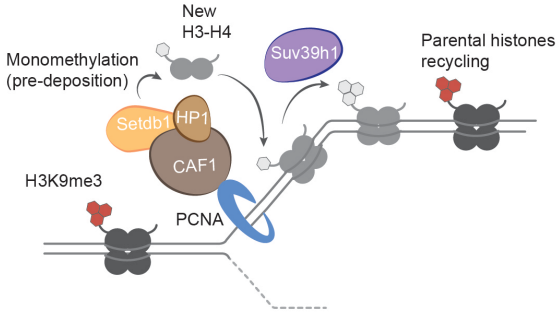
Investigating histone chromatin occupancy after replication (ChOR-seq), it was revealed that H3K4me3 was fully restored by the end of mitosis and this is the fastest rate for a tri-methylated state (Reveron-Gomez et al., 2018). This rapid recycling might enable fast restoration of transcription after replication and in turn

transcriptional resumption might drive self-reinforcing loop to recruit active chromatin marks and transcription factors (Stewart-Morgan et al., 2019). Therefore, transcriptional cues essential for cell identity and differentiation potential might be propagated by complex feedback loops but without the allosteric activation central in H3K9me3 and H3K27me3 propagation.

a. Model of DNA methylation propagation



b. Mitotic propagation of H3K9me3 at pericentric heterochromatin



c. Faithfull inheritance of H3K27me3

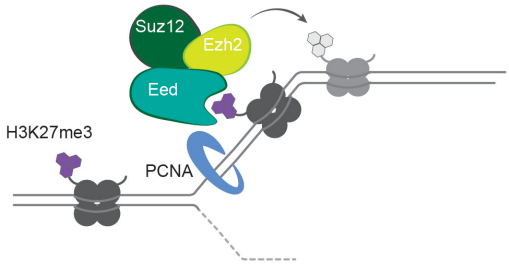


Fig. 1.1 Examples of mechanisms for propagation of epigenetic marks during DNA replication in mammals. (a) The well-established model for DNA methylation maintenance involves recruitment of the DNA methyltransferase 1 (DNMT1) at the replication fork barrier via direct interaction with the PCNA and with Uhrf1 that recognises the hemi-methylated DNA with its SRA domain. (b) Proposed model for H3K9me3 propagation at pericentric heterochromatin according to Loyola et al. 2009. The PCNA recruits the HP1-Chromatin assembly factor 1 (CAF1) chaperone complex at the RFB and engages SetDB1 that monomethylates new histones before their incorporation. H3K9me1 subsequently serves as a substrate for SUV39H1 to restore H3K9me3 in a stepwise manner. (c) H3K27me3 mark is propagated thanks to positive feedback loop established by the EED-EZH2 read-write coupling system.

1.5 Developmental epigenetics

All the cells of a multicellular organism contain the same DNA, yet they exhibit different functions and have very different phenotypes. Epigenetic mechanisms provide the necessary plasticity for complex genomes to orchestrate differential expression of genes according to environmental stimuli and developmental cues. During development, as the single cell embryo divides into numerous cells that form the foetus and subsequently a completely new individual, epigenetic systems and transcription factors are layered on the genome to progressively restrict cellular potential. Nonetheless, this information has to be reset cyclically to enable full potency in the gametes that will generate a new embryo in order to complete the biological life cycle. Thus, a balance is necessary, whereby epigenetic information is reversible, to enable recovery of cellular potential during reprogramming, but also provides memory to maintain cellular identity during development.

1.5.1 Epigenetic mechanisms during pluripotency and differentiation

Pluripotency states arise during development when the totipotent zygote undergoes successive cleavages forming the blastocyst. The embryo at this stage is characterised by the extraembryonic trophectoderm and the inner cell mass (ICM) (Rossant and Tam, 2009). The ICM is made of pluripotent cells that will give rise to all somatic lineages of the embryo and the germline. Notably, this pluripotency window persists only transiently *in vivo* and pluripotent cells are rapidly committed towards their fates. However Embryonic Stem Cells (ESCs) can be isolated from the ICM and cultured in the presence of specific inhibitors *in vitro*. Such small molecules inhibit crucial pathways, such as LIF (leukemia inhibitory factor)/STAT3 (signal transducer and activator of transcription 3) and the dual inhibition ('2i') of the MEK (mitogen activated protein kinase)-ERK and GSK3 (glycogen synthase kinase 3β) pathways, ultimately promoting self-renewal and inhibiting cell differentiation (Ying et al., 2008). In this so-called 2i/Lif condition, ground state pluripotency can be maintained almost indefinitely and ESCs are considered the best prototypes of pluripotent cells (Tee and Reinberg, 2014).

The plasticity of ESCs is generally attributed to hyperdynamic chromatin features. Indeed, in ESC a balance of factors prevents heterochromatin expansion and, on the other hand, contributes locally to the silencing of lineage-specific genes until differentiation is triggered (Gaspar-Maia et al., 2011). Overall, ESCs are characterised by reduced levels of heterochromatin. Importantly, the same chromatin architecture

observed in ESC was also found in eight-cell embryos and pluripotent epiblast cells (which are the source of ES cells) *in vivo* (Ahmed et al., 2010). This highlights the physiological relevance of the studies conducted in ESCs with the *in vivo* counterpart.

On the other hand, cell differentiation is a complex process that requires a series of ordered instructions to progressively commit the totipotent embryo into the different unipotent somatic cells that constitute the new organism. Starting from ESC, the pluripotency gradient of the initial phases of embryo development can been recapitulated *in vitro* by differentiation in in EpiSC that mirror the epigenetic state of the post-implantation embryo (Hackett and Surani, 2014; Nichols and Smith, 2009; Weinberger et al., 2016). This *in vitro* model has been extensively used to investigate the epigenetic mechanisms governing the differentiation programmes.

Progressive restriction of cellular potential is accompanied by a gradual constraint in chromatin accessibility, deposition of DNA methylation and redistribution of histone marks (Buecker et al., 2014). Levels of DNA methylation in the embryo go hand in hand with expression of the de novo DNA methyl-transferases, markedly increasing in post-implantation epiblast (Smith et al., 2014). As normal levels of methylation are restored, this mark is particularly important in keeping silencing of transposable elements and it was recently discovered a new class of DNA methyltransferase (DNMT3C) that has the specific role to maintain silencing of TEs during spermatogenesis (Barau et al., 2016).

Cellular transitions are also characterised by a progressive deposition and redistribution of histone marks. H3K27me3 is particularly important for the transcriptional silencing of key developmental genes and embryos gradually gain this mark from the 2-cell to the early blastocyst stage (Liu et al., 2016). Nonetheless, PRC2-null embryos undergo normal preimplantation development but die during gastrulation (O'Carroll et al., 2001). H3K27me3 is therefore essential for transcriptional repression in post-implantation embryo, suggesting that this mark is required for maintaining rather than initiating gene silencing (Riising et al., 2014).

H3K9me3 plays an central role in gene and transposon repression, such as long terminal repeats (LTRs), during early development and mostly substitutes the silencing activity mediated by DNA methylation that is globally depleted in preimplantation embryos (Cedar and Bergman, 2009). ChIP-seq at different stages of the mouse embryo revealed that, after an initial decrease in H3K9me3 in the zygote, it starts to be accumulated mainly at LTRs from the 2-cell stage. As differentiation proceeds, H3K9me3 progressively appears also at those genes involved in alternative cell fates, forming megabase-sized island resistant to gene activation (Soufi et al., 2012; Wang et al., 2018). Thus, during development, H3K9me3 is important also for repression of lineage-inappropriate genes and preservation of cell-type identity.

Embryos, lacking for the histone methyl transferase SETDB1 or the co-repressor KAP1 show over-expression of LTRs and generally arrest at the blastocyst stage, stressing the important role of this mark for proper development (Wang et al., 2018).

Differently from the other histone marks, H434me3 levels remain stable during embryo development and in different ESC states, but loss of MLL3 or MLL4 affect cell differentiation, suggesting and essential role for this mark in regulating expression of key developmental genes (Atlasi and Stunnenberg, 2017). Instead, at enhancer regions, H3K4me1 is dispensable for transcription but MLL3/4 are required for Pol II loading and gene expression (Dorighi et al., 2017). In pluripotent ESCs many of these developmental genes are marked both with H3K4me3 and H3K27me3 and thus are called 'bivalent genes' (Bernstein et al., 2006). According to the current model, the combination of both repressive and activating marks defines a poised state with H3K4me3 preventing permanent silencing by antagonising deposition of DNA methylation and H3K27me3 maintaining low expression levels. As differentiation proceeds, one of the two marks is lost leading to stable activation or repression of the promoter (Atlasi and Stunnenberg, 2017). Interestingly, it was recently found that this poised state is not exclusive to pluripotency, but also somatic cells can acquire de-novo bivalency suggesting that this state can, more generally, provide cells with a highly dynamic responsiveness to external signals (Weiner et al., 2016). However, during cell differentiation gene expression is mainly initiated by reprograming TF, also called pioneer transcription factors. Binding of these pioneer TFs to chromatin promotes deposition of histone marks at activated enhancers (Koche et al., 2011).

1.5.2 Epigenetic reprogramming in early embryo and primordial germ cells

Somatic signatures, acquired during cell differentiation, represent a barrier for acquisition of the potency needed to generate a new complete organism. This is why, in many metazoans, gametogenesis and early embryo development are accompanied by widespread resetting of the epigenome in order to erase these signatures and ensure naïve pluripotency. This suggest that most of the epigenetic memory that accompanied differentiation is lost in the gamete's lineage. Primordial germ cells specification starts from a population of epiblast cells in the postimplantation embryo. At first, these cells undergo replication-dependent dilution of DNA methylation while they migrate, but a second phase of active demethylation occurs as PGCs entry in the gonads and might involve TET3-driven oxidation of 5mC in 5hmC (Gu et al., 2011; Iqbal et al., 2011). This includes erasure of DNA at imprinting promoters and reactivation of the X chromosome (Hajkova et al., 2002).

Another extensive epigenetic reprogramming occurs after fertilization of the oocyte by the sperm, when the maternal and paternal pronuclei of the zygote undergo global de-methylation with different kinetics. Paternal genome is demethylated more rapidly than maternal one reflecting active or passive erasure respectively, and, in mouse, 5mC reaches its lowest level at the blastocyst stage to markedly increase again after implantation (Hackett and Surani, 2013; Smith et al., 2012) (Fig. 1.2).

DNA methylation reprogramming is thus considered to be a crucial important developmental event (Reik et al., 2001). However, most of the genome is not functionally regulated by DNA methylation, thus the precise reason why such demethylation event should be needed is unclear. The assumption is that demethylation is required for pluripotency, yet, pluripotent ESC can be maintained fully methylated (Habibi et al., 2013) and induced pluripotent stem cells (iPS) can acquire pluripotency without undergoing DNA demethylation (Milagre et al., 2017). Indeed, several other organisms such as *C. elegans* and *Drosophila*, including some mammals (i.e. rabbit), go through early development without extensive DNA methylation erasure (Beaujean et al., 2004). The exact function of global demethylation waves thus remains to be unravelled.

Both during gametogenesis and embryo preimplantation development, global DNA demethylation is accompanied by remodelling of other chromatin marks, including H3K9me3 and H3K27me3 (Canovas and Ross, 2016). Using sensitive techniques, H3K27me3 dynamics was investigated at different stages of mouse embryo development and extensive loss of this mark was observed at developmental genes up to the blastocyst stage while canonical H3K27me3 was re-established post implantation, including at bivalent promoters (Zheng et al., 2016) (Fig. 1.2).

Although DNA methylation and histone marks reprogramming occur genome wide, some specific loci are resistant to reprogramming such as some potentially hazardous transposable elements and these elements are marked for transcriptional repression. In mouse oocytes and testis, H3K9me3 controls levels of ERVs and LINE1 (Liu et al., 2014). Residual DNA methylation is retained at IAPLTR1 during zygotic and PGC reprogramming to protect genome stability, but also at >200 single copy loci (Hackett et al., 2013b). Specific factors might be involved in protecting these regions from epigenetic erasure, such as the maternal factor STELLA that binds H3K9me2 and blocks Tet3 mediated DNA demethylation (Nakamura et al., 2012) and KAP1/Zfp57 that are critical for maintenance of maternal and paternal imprints (Messerschmidt et al., 2012). Rare H3K9me3 or 5mC epialleles could therefore escape global epigenetic reprogramming and be inherited though generations.

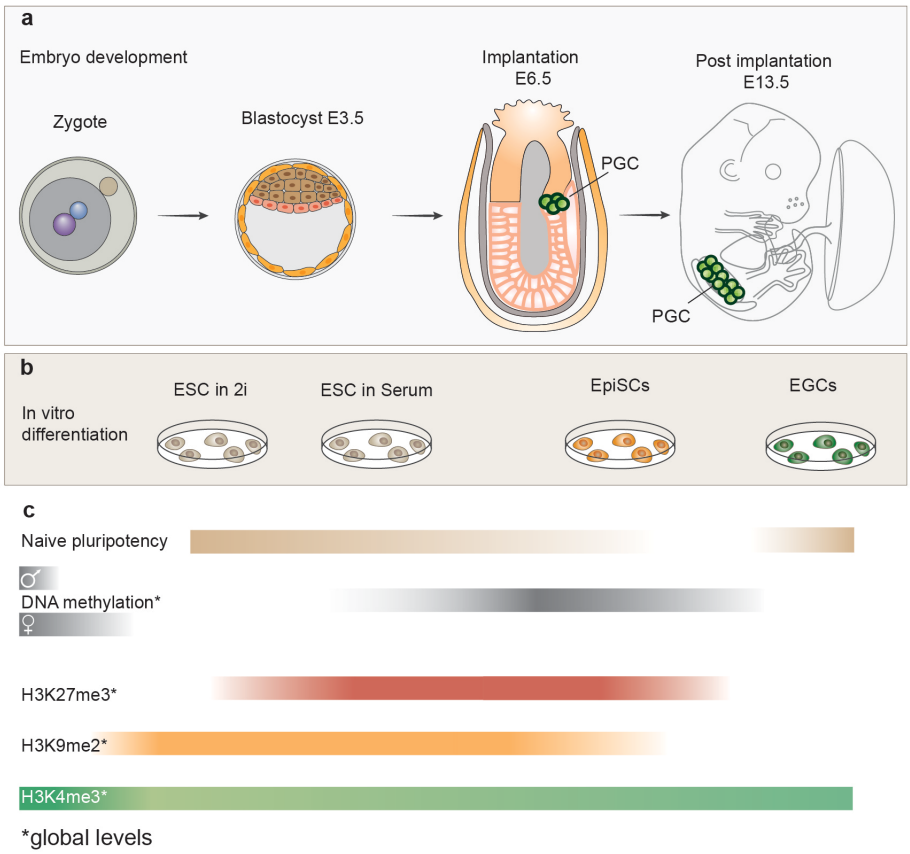


Fig. 1.2 Epigenetic dynamics during embryo development and ESC differentiation. (a) Schematic representation of the mouse embryo development. (b) Drawings represents the in vitro-equivalent of the different epigenetic states. (c) Gradient boxes indicate the global level of each mark, but individual loci, which are known to escape this general trend are not shown for simplicity. DNA methylation is globally erased in the paternal and maternal genomes in the zygote and is deposited again later in development. DNA undergoes a second wave of demethylation during primordial germ cells (PGCs) differentiation. On the other hand, ESCs in vitro are usually hypomethylated in 2i conditions and higher levels of DNA methylation are observed in Serum conditions and in primed cells (epiblast-derived stem cells (EpiSCs) genome-wide. Similarly, the repressive histone marks (H3K27me3 and H3K9me2) generally increase post-implantation and are reduced in 2i-cultured ESCs and Embryonic Germ cells (EGCs). Contrarily, H3K4me3 is present in broad domains in the oocyte but is restricted to transcription start sites after fertilization.

1.6 Inheritance of environmental cues

The notion that non-genetic factors are responsible for environment-driven long-lasting effects on phenotypic outcomes has long being recognised (Jablonka et al., 1992). Epigenetics is a valid candidate for this non-genetic carrier due to its dynamism that allows a fast response and possibility to be inherited. The importance of epigenetics in such environmental response is well characterised in plants as beautifully exemplified by the polycomb-based vernalization mechanism already described in section 1.4.3 (Jiang and Berger, 2017; Yang et al., 2017). Similar processes appear to take place also in animals although, such environmental induced epigenetic modifications, are often difficult to explain mechanistically (Miska and Ferguson-Smith, 2016). Nonetheless, several studies using different animal models suggested that effects induced by environmental exposure during early life can be inherited transgenerationally and are characterised by epigenetic changes. In this case, transgenerational epigenetic inheritance (TEI), therefore rely on the transmission the unit of epigenetic information called 'epiallele'.

According to the number of generations through an epiallele is transmitted and the parental origin of the triggering signal, Heard and Martienssen distinguish on intergenerational or transgenerational inheritance. In case the exposure occurs in a pregnant female, the F1 will be directly exposed in the uterus and the F2 through the F1's germline, and this are considered intergenerational effects. Only the inheritance in the F3 will be truly transgenerational. Alternatively, in case of paternal inheritance, the effects on the F1 generation is considered intergenerational because of the exposure through the F0's germline and transgenerational from the F2 generation onward (Heard and Martienssen, 2014).

Due to the complexity of such environmental-transgenerational mechanisms many studies are limited in finding a correlation rather than a causality of epigenetics underlying the inheritance of the acquired trait. Confounding effects might arise, for example, in the case of nutrient or stress insults that can lead not only to epigenetic changes but also to genetic mutations or alteration in gut microbiota composition that might be inherited in subsequent generations. It is therefore necessary to prove direct inheritance of the epigenetic modification through the germline to unequivocally link the transgenerational inheritance of the acquired trait to epigenetic mechanisms (Lim and Brunet, 2013).

1.6.1 Environmental epigenetic and transmission in 'animal model organisms'

Since the epigenetic mechanisms involved are often well conserved, studying epigenetic inheritance in lower metazoans helps the understanding of such a process also in mammals (Radford, 2018). Many environmental stimuli, such as temperature, starvation, diet and toxic challenges have been linked with epigenetic changes that can be inherited through several generations in many model organisms. One of the most striking and direct links between environment and epigenetic response in animals is that controlling sex determination in turtles according to the temperature during egg incubation. This mechanism has been recently molecularly dissected and involves the histone demethylase KDM6B which exhibits temperature dependent sex dimorphic expression in early embryos and directly controls H3K27me3-mediated silencing at male sex-determining genes (Ge et al., 2018).

C. elegans displays robust transgenerational epigenetic inheritance. Many examples involve inheritance of small-interference-RNAs, which have been shown to either suppress ectopic viral genome or the expression of genes involved in nutrition, in order to transmit protection against viral infections or starvation respectively (Rechavi et al., 2014; Rechavi et al., 2011). More recently, a case of transgenerational inheritance involving epigenetic marks was shown in response to temperature stress. Exposure to high temperatures induced changes in H3K9me3 at a specific reporter gene array that was transmitted for 14 generations, through both oocyte and sperm, after the environmental stimulus was released (Klosin et al., 2017). Together these studies show that, in this model organism, environmentally induced heterochromatin modifications may support a general mechanism for transgenerational epigenetic transmission of information.

Many are the examples of environmental TEI found in *Drosophila* involving heterochromatin formation. Environmental insults are directly translated in chemical modifications of transcription factor ATF2 that when phosphorylated, in turn of osmotic stress or heat shock, is released from DNA leading to heterochromatin formation that is transmitted transgenerationally in a non-mendelian fashion (Seong et al., 2011). Other examples involve transmission of PcG alteration in response to toxic stress during development (Stern et al., 2012). PRC2 has also been objective of an extensive study, which found that H3K27me3 epialleles induced by transiently enhancing chromatin contacts were inherited transgenerationally. Importantly the expression of these epialleles was modulated by a range of external stimuli mimicking naturally occurring environmental changes in the wild (Ciabrelli et al., 2017).

To date there are no exact evidences of wide epigenetic resetting between generations in *C.elegans* and *Drosophila* and this 'epigenetic continuity' might favour

environmental induced TEI. Contrarily, mammals undergo global erasure and reestablishment of epigenetic state each generation and this might represent a barrier for inheritance of environmental acquired epigenetic states.

1.6.2 Environmental epigenetic inheritance in mammals

Mammals are more complex organisms compared to *C.elegans* and *Drosophila*, therefore directly linking the external environment with epigenetic changes might be more challenging compared to the nematodes and fruit fly. However, the recent discovery that mammalian cells can directly sense the oxygen levels and translate the environmental changes into epigenetic changes is a strong example of such a mechanism. Two parallel studies showed that both histone demethylases KDM5A and KDM6A have low oxygen affinity and, in hypoxic conditions, their activity is disrupted, leading to aberrant accumulation of H3K27me3 (Batie et al., 2019; Chakraborty et al., 2019). Ultimately, oxygen directly affects cell differentiation through epigenetic alteration. However, neither of the studies investigated whether such epigenetic changes can be inherited through cell divisions.

The most widely used examples of TEI in mammals are the Agouti (A^{vy}) and Axin-fused ($Axin^{Fu}$) alleles which are subject of differential methylation of transposons inserted upstream of these genes. The metastable methylation status of these epialleles can be influenced by a rich methyl-donor diet in the pregnant dams (Cooney et al., 2002; Waterland and Jirtle, 2003). In fact, chromatin methylation is highly dependent on the intracellular abundance of its methyl-donor cofactor SAM, that, on its turn, depends on the catalysis of its obligatory precursor, the essential aminoacidic methionine (Met) (Ducker and Rabinowitz, 2017). Circulating quantity of methionine strictly depends on diet – e.g. vegan diets are particularly depleted in Met (Schmidt et al., 2016) - and circadian rhythms (Krishnaiah et al., 2017). These environmental fluctuations can thus ultimately affect the epigenome thanks to this metabolic-epigenetic link. Although SAM depletion induces global loss of H3K9me3, cells can activate adaptive responses to support heterochromatin stability. A recent study showed that in response of this depletion cells can redistribute H3K9me1 to retrotransposon elements and this promotes heterochromatin persistence upon metabolic recovery both *in vitro* and *in vivo* in mouse (Haws et al., 2020).

The concept that parents' experiences might affect the gametic epigenome and thus have a long-lasting effect on the phenotype of several successive generations is of great interest. The mechanism and the extent by which this can occur are instrumental to prevent major effects of parents' behaviour on progeny's health, but

are still largely unknown. Many studies, conducted in mouse model, have linked paternal diet, such as protein restriction or high fat diet, with altered offspring metabolism through inheritance of small RNAs called transfer RNA (tRNAs) via the sperm (Chen et al., 2016; Sharmaet al., 2016). Despite very few histones being retained in the sperm, where they are replaced with protamine, TEI has also been linked with inheritance of altered histone methylation in the sperm. By overexpressing the histone H3K4 de-methylase KDM1A (LSD1) during spermatogenesis in mouse, Siklenka and collaborators showed that the altered H3K4me2 pattern severely impaired offspring development and was inherited transgenerationally without changes in DNA methylation (Siklenka et al., 2015). Many other studies tried to investigate effects of other environmental insults in the parents, such as addiction to drugs, alcohol or tobacco, on foetal programming. Although effects of such behaviours are easily reflected in altered postnatal metabolisms in the next generation, defining such effects truly transgenerational epigenetic inheritance is unlikely given the global reprogramming which the mammalian germline is subjected to (Heard and Martienssen, 2014).

1.6.3 Environmental epigenetics and adaptive inheritance

Epigenetics as a fast and dynamic response to environmental changes poses the question whether epigenetic systems can be a source for organisms to quickly evolve in response to external stimuli. On this line of thinking, in 1980 Mayr coined the term “soft inheritance” (Mayr, 1980), concept subsequently revised by Richards as inheritance of randomly generated epialleles under the pressure of the environment. This ‘soft inheritance’ acts as a sculptor to engrave a malleable hereditary material, in contrast to the ‘hard’ genetic material (Richards, 2006). A striking example of adaptive inheritance is represented by the adaptive behavioural plasticity in *C. elegans*, in which, the exposure to attractive odorants, produce long-lasting olfactory imprints that enhance the ability for food finding and egg-laying rates for up to 40 generations (Remy, 2010). Although genetic variation and cultural-based transmission have been excluded, the molecular cues responsible of this adaptive response have not been unravelled yet and possibly involve epigenetic modifications (Wang et al., 2017).

In higher animal models, such soft inheritance is very rarely observed due to the wide epigenetic resetting in the germline. However, it has been extensively described in plants, such as in the case of the natural variation in plant symmetry that depends on a differentially methylated epiallele propagated for literally hundreds of years (Cubas et al., 1999).

According to Darwinian natural selection theory, random genomic variations can be fixed in a population if they provide a selective advantage to survive in the external environment (Darwin, 1859). Expanding this concept, a recent study conducted in a yeast species which lost the de-novo methyltransferase million years ago, suggested that DNA methylation has been maintained at high levels under the pressure of natural selection for so many years (Catania et al., 2020). Interestingly another study showed that yeasts can respond to external insults, as exposure to caffeine, by deposition of H3K9me heterochromatin at genes known to confer resistance to this compound when deleted (Torres-Garcia et al., 2020). Such epimutations are transmitted for subsequent generations but are transient and are lost when the external stimulus is released. This study reveals that epigenetic processes might help wild type organisms to adapt to disadvantageous environments, and favour the organism that acquired them to take over the whole population, following the natural selection law but without altering their genotypes.

1.7 CRISPR-Cas9-based tools serving epigenetics

The CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats – associated protein 9) system was discovered in bacteria as a naturally occurring defense mechanism against virus or phage infections (Barrangou et al., 2007). By targeting the endonuclease digestion of the exogenous DNA via complementarity of binding with a RNA molecule, the CRISPR-Cas9 system protects the integrity of the organism from the invading elements (Terns and Terns, 2011). Shortly after its discovery in prokaryotes, CRISPR-Cas9 was evicted from its natural context, simplified and repurposed in eukaryotic cells as a tool for efficient and precise genome editing exploiting its endonuclease activity (Jinek et al., 2012). In its simplicity lays the power of this tool that, other than the Cas9 protein needs only a guide RNA (gRNA) to be recruited to virtually any sequence of the genome as far as it is immediately followed by a protospacer-adjacent motif (PAM) (Fig. 1.3.a). This PAM sequence is constituted by a short nucleotide sequence (typically 2-6 bp) and is indeed crucial for the Cas9-PAM interacting domain to initiate contact with DNA and ultimately target site binding (Jinek et al., 2012; Sternberg et al., 2014). Many Cas9 orthologs have been isolated from different bacteria species differing mainly in the PAM repertoire. The most widely used Cas9 variant has been isolated by *Streptococcus pyogenes* (SpCas9) and recognises the PAM sequence 5′-NGG-3′ (Mojica et al., 2009), where 'N' represent any nucleotide. Given that the dinucleotide 'GG' has a frequency

of 5.2% in the human reference genome, there is approximately one PAM every 42 bp, thus spCas9 can be used to target hundreds of millions of sites.

1.7.1 CRISPR/(d)Cas9 mediated epigenome editing

The applicability of CRISPR-Cas9 tool has been further extended by engineering of a so called 'dead' version (dCas9) which, lacking its endonuclease activity, is exclusively used as a docking platform to recruit arrays of diverse epigenetic-effectors to precisely edit the epigenome (Gilbert et al., 2013; Qi et al., 2013) (Fig. 1.3.b). However, the epigenome editing tools anticipated the CRISPR-Cas era by the use of programmable enzymes fused to DNA binding domains such as zinc finger proteins (ZFPs) and transcription activator-like effectors (TALEs) (Perez-Pinera et al., 2012). Nonetheless, ZFPs engineering is very laborious and TALEs binding is sensible to DNA methylation (Valton et al., 2012), rendering these tools not optimal for large scale epigenome editing. On the other hand, the scalability and flexibility of dCas9 made its fortune.

The first studies reporting the application of CRISPR in transcriptional regulation comes from fusion of dCas9 to known functional activator, such as VP16, and repressor, such as KRAB (Krüppel associated box) domains (Cheng et al., 2013; Gilbert et al., 2013). These transcriptional regulation tools work by recruiting other transcription factors or chromatin modifying enzymes precisely at the target locus. For example, KRAB recruits KAP1/TRIM28 and HP1 proteins interfering with the position of the RNA polymerase II and promoting the binding of histone methyltransferases that ultimately deposits H3K9me3 (Groner et al., 2010). As a result, chromatin is locally compacted and the targeted genes are efficiently repressed (Gilbert et al., 2014). To date, dCas9 has been successfully fused to many different epigenetic 'writers' and 'erasers' including the histone acetyltransferase p300 (Hilton et al., 2015) or deacetylases (HDAC) (Kwon et al., 2017), the histone demethylase LSD1 (Kearns et al., 2015) but also to the DNA methyltransferase (Stepper et al., 2017) and TET enzymes to target removal of 5mC (Xu et al., 2016).

Thanks to the versatility of the Cas9 enzyme and gRNA, one of the most important advantage of the CRISPR/(d)Cas9 tool is to amplify the epigenetic effect (Fig. 1.3.c). This can be achieved by fusing multiple copies of an effector domain to the dCas9, by using multiple gRNAs to target several discrete DNA regions or by engineering the gRNA to recruit several effectors fused to an RNA binding-motifs (Zalatan et al., 2015). A fourth strategy was originally developed by Tanenbaum and collaborators for amplification of fluorescent signal and was effectively called SunTag system (Tanenbaum et al., 2014). The SunTag works by fusion to the dCas9 of a tail made of

a recombinant GCN4 peptide array. According to the number of GCN4 repeats the tail is composed of, as many effectors can be bound thanks to the fusion to a single-chain variable fragment antibody(scFv) which recognise the GCN4 epitope. Several studies showed the amplification power of this new system compared to the simple dCas9-effecor fusion, achieving 25 times better transcriptional activation with the dCas9-SunTag-VP64 (Tanenbaum et al., 2014) or higher deposition of DNA methylation with the dCas9-SunTag-DNMT3A (Pflueger et al., 2018). This system has also the potential to recruit multiple different epigenetic effectors simultaneously within the same locus to investigate the interaction of several epigenetic marks and their crosstalk.

The applications of the dCas9 tools, not only include transcriptional regulation and histone or DNA methylation modifications, but also extends to relocation of ncRNA by incorporating it in the gRNA sequence (Shechner et al., 2015), chromatin visualization by fusion of the dCas9 to fluorescent proteins (Chen et al., 2013) and DNA base editing to specifically correct single nucleotide mutations (Gaudelli et al., 2017). Other than its use in dissecting the epigenome and understanding the basics of epigenetic mechanisms, the CRISPR/Cas9 tool has also been employed for engineering cell identity, model disease mechanism and developing therapies.

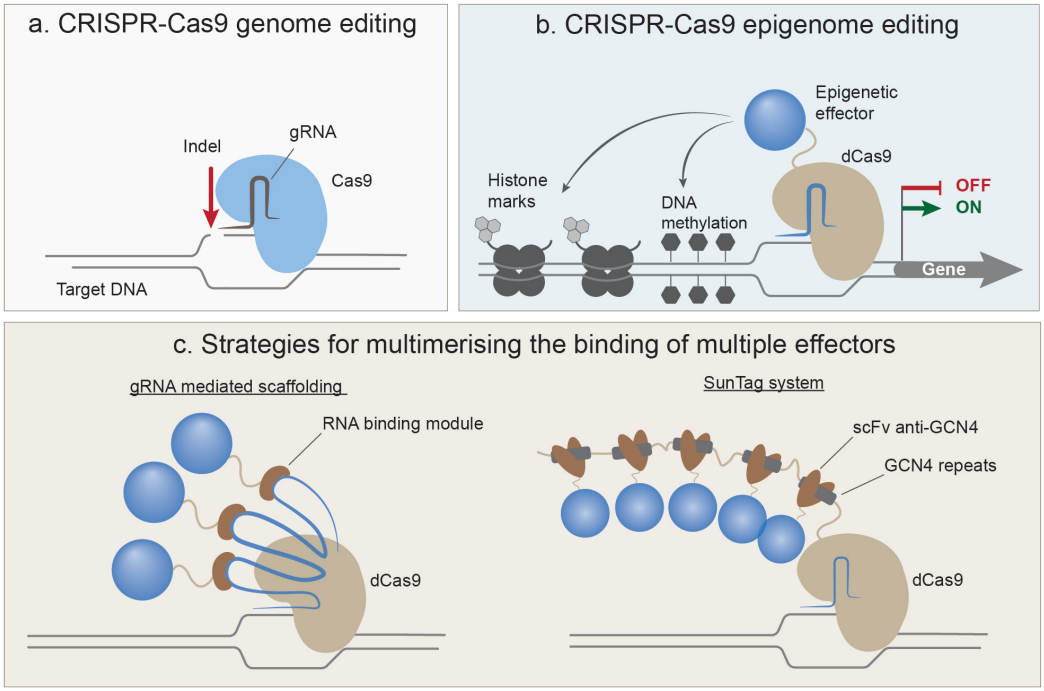


Fig. 1.3. Schematics of the CRISPR-Cas9 technology. (a) CRISPR-Cas9 genome editing: the endonuclease Cas9 is delivered on target DNA thanks to the complementarity of binding with a guide-RNA and induces double strand breaks which leads to indel mutations. (b) CRISPR-Cas9 has been further engineered for epigenome editing by disrupting Its catalytic activity (dead Cas9 (dCas9)). The dCas9 can be fused to nearly any protein domain and thus serve as a delivering system for DNA sequence-specific binding of epigenetic effectors which deposit epigenetic marks and ultimately control gene expression. (c) Different strategies for multimerising the binding of multiple effectors. On the left: dCas9 is coupled with an engineered gRNA containing hairpins that can be bound by RNA binding modules which are fused to epigenetic effectors. On the right: the SunTag system is composed of a dCas9 fused with a tail of CGN4 repeats that can be recognised and bound by multiple scFv domains linked with the epigenetic effectors.

1.7.2 CRISPR/Cas9 applied to high throughput genetic screens

The CRISPR/Cas9 gene editing activity in permanently knocking out gene expression facilitates the investigation of key players in specific biological pathways. Since its targeting is based on oligonucleotides, the main advantage of this system is its scalability and thus the possibility to virtually interrogate the role any gene in genome in a specific phenotype at the same time in a large pool (Shalem et al., 2015). Thanks to the commercialization of large pooled custom gRNA libraries and the possibility to identify, by next generation sequencing, selected library members in subpopulation of cells isolated according to a certain phenotype of interest, CRISPR/Cas9 tool is popularly used in forward genetic screens.

Oher than in its wild-type form dCas9 fused to epigenetic repressors or activators is also used to manipulate complex gene expression patterns and is combined to gRNA libraries for genome wide silencing or gain-of-function screens (Gilbert et al., 2014; Konermann et al., 2015).

2 Materials and Methods

2.1 Cloning

2.1.1 Gel electrophoresis

For size separation of DNA molecules, horizontal electrophoresis was used with agarose gels ranging between 0.8-2% in TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH ~8.3) supplemented with Sybr safe DNA gel stain (ThermoFisher). Samples were loaded with the appropriate volume of 6x loading dye (ThermoFisher) using GeneRuler 1kb plus ladder (ThermoFisher) to estimate DNA fragment size. Gel imaging was resolved using ChemiDoc imaging system (bio-rad).

2.1.2 Nucleic acids quantification

Concentration of nucleic acids was routinely estimated spectrometrically with Nanodrop ND-1000 by reading the absorbance at a wavelength of 260nm (A_{260}). Alternatively, for more precise quantitation and low inputs qubit fluorometer was used in combination with appropriate assay kit for DNA or RNA (ThermoFisher).

2.1.3 Plasmid construction

The pPB_TRE3G::dCas9-5XGCN4_*EF1a::*TetOn-Hygro, pPB_TRE3G::ScFv-KRAB-GFP_EF1a::Neo and pPB_TRE3G::ScFv-GFP_EF1a::Neo were constructed starting from plasmid pPlatTET-gRNA2 (Morita et al., 2016) (Addgene #82559) by PCR amplifying sequence of interest and cloning together with antibiotic resistance genes into pPiggyBac plasmids properly digested (Appendix Table 9.1). Similarly, the pPB_U6::gRNA_EF1a::BFP-Puro were cloned from Addgene plasmid #60955 into pPiggyBac recipient plasmid. Cloning was performed by homology arms recombination using In-fusion HD-Cloning (Takara # 639650) according to manufacturer instructions.

2.1.4 gRNA design and ligation

All gRNAs were designed using the GPP web portal (broad institute) for spCas9 PAM (NGG) searching. For gRNA cloning, the pPB_U6::gRNA_EF1a::BFP-Puro was digested with BlpI and BstXI (NEB) overnight at 37°C. Digested plasmid has been separated by gel electrophoresis on 1% agarose and gel extracted with (ZymoClean Gel DNA recovery #D4008). Reverse complement gRNA sequences (Appendix, Table 9.2) with appropriate overhangs were annealed at 10 μM final concentration with 10 mM Tris, pH 7.5 - 8.0, 60 mM NaCl, 1 mM EDTA, by heating at 95°C for 3 minutes and cooling down at RT for >30 minutes. Ligation of the annealed gRNA with the digested pPB_U6::gRNA_EF1a::Puro was performed with T4-DNA ligase (NEB #M0202S) for 1 hour at 37°C.

2.1.5 Bacteria transformation

Chemically competent bacteria DH5a *Escherichia Coli* (E.Coli) were transformed with 1-2 ul of the above Infusion or ligation reactions on ice for 30 minutes. After heat shock at 42°C for 45 seconds, the bacteria mixture was returned back to ice for 2 minutes and resuspended with 300 μl of room temperature S.O.C. rescue media. Bacteria were left recovering at 37°C for one hour prior to spreading on L-agar plate with appropriate antibiotic selection (Ampicillin 50 mg/ml). Bacteria colonies were left growing overnight at 37°C.

2.1.6 Bacteria culture and isolation of plasmids

Transformed bacteria colonies were incubated in 3 ml liquid Luria-Bertani (LB) culture media with appropriate antibiotic at 37°C overnight with shaking at 225 rpm. 100 µl of this culture was inoculated into 50 ml of LB containing antibiotic, left growing for further 16 hours and plasmid DNA extraction was performed with midi prep kit (ZymoPure Midi prep kit #D4200) following manufacturer instruction. Correct plasmid assembly was confirmed by diagnosis with appropriate restriction enzymes or sanger sequencing.

2.2 Mammalian cell culture

2.2.1 Routine cell culture and differentiation

Embryonic Stem Cells (ESCs) were routinely cultured in t2i/L media (NDiff (NB27 Takara #y40002) additionated with titrated 2i (1 µM PD0325901, 3 µM CHIR 99021, 1000 U/ml LIF, 1% FBS, 1%Pennicillin streptavidin) unless alternatively stated and incubated in humified atmosphere at 37°C and 5% CO_2. Cell passaging was performed every 2-3 days by washing cell colonies with PBS1X, dissociation with TripLE and plating appropriate number of cells on feeder-free gelatin coated plates. Culture media was changed daily. Cells were routinely tested for mycoplasma infection.

To induce endodermal differentiation $6x10^3$ naïve ESC were seeded per cm^2 of a gelatin-coated plate and maintained in t2i/L media. After 24hr cells were washed trice with PBS1X and culture medium replaced with endoderm medium (Borowiak et al., 2009)(RPMI (Thermo Fisher Scietific #12-633-012) supplemented with 0.02% FBS, 2mM L-Glutamine, 5 µM IDE1 and 1% Pennicillin streptavidin). Medium was changed every other day.

2.2.2 DNA transfection

DNA transfection was accomplished with Lipofectamine 3000 (Thermo Fisher Scietific #L3000001) unless otherwise stated. ES Cells were seeded at least 24 hrs in advance to be 70-90% confluent the day of transfection. Appropriate amounts of DNA were calculated according to manufacturer instructions. Media was changed after 6 hours and replaced with antibiotic selection containing medium.

2.2.3 Flow cytometry

Exploiting the fluorescence reporters, cells can be analysed and separated by Fluorescence activated cell sorting (FACS). For FACS, cells are gently dissociated in cell suspension by TrpLE, resuspended in PBS1X plus FBS 1% (FACS media) and filtered (BD, cup-Filcons #340632). FACS Aria III (Becton Dickinson) or Attune NxT Flow Cytometer (ThermoFisher) were used for sorting or analysis respectively. Data analysis was performed with FlowJo v10.5.3 (Tree Star, Inc.).

2.2.4 Generation of reporter cell lines

For the generation of the *Esg1*$^{-tdTomato}$ reporter cell line, *Stella*-GFP:*Esg1*-tdTomato (SGET) compound-reporter mESC line (Hackett et al., 2018) were used and the GFP cassette was disrupted by CRISPR-Cas9 targeting. To disrupt GFP expression, SGET ESCs were transfected with the spCas9 plasmid (pX459 Addgene #62988) and pPB_U6::gRNA_EF1a::Puro containing a gRNA complementary to the GFP sequence. After selection with puromycin for three days, GFP negative cells were isolated by Flow Cytometry and seeded at low density (1000 cells per 9.6cm^2) for single colony picking. After clonal expansion, homozygous GFP sequence disruption was confirmed by PCR genotyping and sanger sequencing.

For the generation of the *p53*$^{-tdTomato}$ reporter cell line, the tdTomato sequence was PCR amplified from a donor vector with ultramers carrying 180 bp overhangs complementary to the 3' end of the *p53* gene. The tdTomato dsDNA thus amplified was introduced by transfection in WT A9 mESC together with the spCas9 plasmid (pX459 Addgene #62988) carrying a gRNA sequence complementary for the *p53* 3'-end. After 3 days of puromycin selection and one day recovery, tdTomato positive (TOM+) single cells were isolated by Flow Cytometry into 96 well plate. The single cells were expanded and correct integration of tdTomato downstream the *p53* gene was confirmed by PCR genotyping and sanger sequencing. Importantly, integrity of

the *p53* gene on both alleles was confirmed by sanger sequencing and normal levels of *p53* mRNA expression were verified by qPCR.

2.2.5 Generation of knockout ESC lines

For generation of gene knockouts (KO) *Esg1*$^{-tdTomato}$ or *p53*$^{-tdTomato}$ ESC lines were transiently transfected with two spCas9 plasmids (pX459) carrying gRNAs targeting exon sequence for critical catalytic activity of the gene of interest (*Dppa2, Zmym3, Kmt2d, Smarcc1, Dot1L, Kdm3a*) (Appendix, Table 9.3). Transfected cells were selected with puromycin (1.2 ug/ml) for three days and subsequently seeded at low density (1000 cells per 9.6cm^2). After single colony picking and expansion, colonies with bi-allelic genetic deletion were screened by PCR genotyping and absence of the protein was confirmed by PCR genotyping or/and western blot (Appendix, Tables 9.4 and 9.5).

2.2.6 Epigenetic editing and memory experiment

Esg1$^{-tdTomato}$ or *p53*$^{-tdTomato}$ WT or KO reporter ESC lines were co-transfected with pPB_TRE3G::dCas9$^{-5XGCN4}$_*EF1a::*TetOn-Hygro, pPB_TRE3G::ScFv-KRAB^{-GFP}_EF1a::Neo, pPB_U6::gRNA_EF1a::BFP-Puro containing appropriate gRNA sequence and pPY_CAG_Pbase using 1:1:0.2:0.2 molar ratio, respectively. Alternatively the pPB_TRE3G::KRAB$^{-GFP-scFv}$_EF1a::Neo was replaced with a construct carrying only expression of *GFP*$^{-scFv}$ pPB_TRE3G::ScFv-GFP_EF1a::Neo as a control. Cells with successful integration of the 3 cassettes were consequently selected with Hygromycin (250 μg /ml) for 5 days, Neomycin (300 μg/ml) for 3 days and Puromycin (1.2 μg /ml) for 2 days and after two days recovery, expression of dCas9$^{-5XGCN4}$ and *KRAB*$^{-GFP-scFv}$ was induced with Doxycyclin (DOX) (1 ng/ml) for 7 days. After one week of DOX induction cells were dissociated with TrpLE, resuspended in FACS media and filtered. Effective induction of the epigenome editing tool was confirmed by expression of GFP and BFP by flow cytometry and BFP/GFP positive cells that successfully repressed the targeted reporter were sorted by flow cytometry as tdTomato negative (TOM-). TOM- cells were then cultured for further 3 or 7 days in the absence of DOX and memory of the reporter silencing was estimated by flow cytometry analysing tdTomato expression in cells that switched off the epigenetic editing tool gated as BFP+/GFP- cells.

2.2.7 *Dppa2/4* overexpression in endoderm cells

For *Dppa2* and *4* overexpression experiment the coding sequences of these two genes have been amplified starting from mouse cDNA and cloned into piggy-bac plasmids under the control of the CAG promoter and carrying the Bleomycin resistance gene separated by a P2A self-cleavable domain. Specifically Dppa2 has been cloned either singly or together with Dppa4 in the same open reading frame and separated by a T2A self-cleavable domain (pPB_CAG:Dppa2-P2A-Bleo or pPB_CAG:Dppa2-T2A-Dppa4-P2A-Bleo).

p53$^{-tdTomato}$ WT cells containing the dCas9$^{-5XGCN4}$ KRAB$^{-GFP-scFv}$ or alternatively the *GFP*$^{-scFv}$ and the *p53*_gRNA^{345up} were treated with DOX for five days. Cells that acquired the silenced *p53*-epiallele were flow sorted with FACS Aria III (Becton Dickinson) according to tdTomato negative expression and 6x10^3 cells were seeded back per cm^2 and maintained in ESC medium +DOX. After one day from sorting, medium was replaced with endoderm differentiation medium (RPMI (Thermo Fisher Scietific #12-633-012) supplemented with 0.02% FBS, 2mM L-Glutamine, 5 μM IDE1 and 1% Pennicillin streptavidin) containing Doxycycline. The day after, cells were washed out to remove DOX and transfected with either the pPB_CAG:Dppa2-P2A-Bleo or pPB_CAG:Dppa2-T2A-Dppa4-P2A-Bleo and the pPY_CAG_Pbase. After 12 hours, Zeocyn antibiotic (50 μg/ml) was added to the culturing medium. Four days after DOX washout (equivalently to five days of differentiation) cells were analysed with the Attune Nxt flow cytometer (ThermoFisher) and 1x10^5 cells collected for bisulfate pyrosequencing.

2.3 Embryo manipulation and analysis

2.3.1 Embryonic Stem Cells microinjection

Prior to microinjection *p53*$^{-tdTomato}$ reporter ESC were transfected with dCas9$^{-5XGCN4}$, *p53_ gRNA*^{345up} and *KRAB*$^{-GFP-scFv}$ or alternatively *GFP*$^{-scFv}$ and treated with DOX for 7 days. ESC microinjection was performed by the Gene Editing & Embryology Facility of EMBL Rome, using 3.5-day old embryo derived from natural mating of C57BL/6J mice. Injected embryos were implanted back in pseudo-pregnant foster mothers. All

animals employed and procedures were in accordance with the gold-standard Italian and European Union regulation guidelines and approved by the local ethical committee.

2.3.2 Collection of E10.5 embryos for Flow Cytometry analysis

After 7 days from the injection, at the embryonic development day 10.5, the deciduum was collected from the uterus and put into a 6 cm dish with cold PBS+10% FBS. Embryos were then extracted from the deciduum and moved to fresh PBS+10% FBS, placenta removed and cleaned from debrids and tissue fragments.

Individual embryos were moved in one well of a round-bottom 96 well plates (Corning #CLS3367) containing 50µl of TripLE and incubated at 37°C for 20 minutes and pipetted until the embryo is entirely dissociated in single cell. The single cell suspension was then diluted with 100µL of PBS+1%FBS and spin down at 1200rpm for 5 min. Cell pellet was then resuspended in 300µl of FACS medium (1xPBS+1%FBS) and filtered (BD, cup-Filcons #340632) for flow cytometry analysis with Attune Nxt.

2.4 CRISPR screen

2.4.1 Generation of cells for CRISPR screen

spCas9 sequence was cloned together with a GFP cassette separated by a self cleavable T2A domain, into a plasmid containing the safe harbour Rosa26 homology arms (*pBS_R26_CAG::spCas9-T2A-GFP*). In order to integrate the *CAG::spCas9-T2A-GFP* cassette into the Rosa26 safe-harbour, *Esg1*$^{-tdTomato}$ or *p53*$^{-tdTomato}$ cell lines were co-transfected with the *pBS_R26_CAG::spCas9-T2A-GFP* plasmid and two pX459 plasmids containing gRNAs against the Rosa26 locus. Cells were selected with puromycin (1,2 µg/ml) for 3 days and subsequently GFP+ single cells were sorted at FACS into gelatin-coated 96 well plate. After clonal expansion correct integration of the CAG::spCas9-T2A-GFP cassette into Rosa26 locus was verified by PCR genotyping. Two of these clones were subsequently transfected with *pPB_TRE3G::dCas9*-

^{5XGCN4}_EF1a::TetOn-Hygro and *pPY_CAG_Pbase* in a 1:10 molar ratio for random genome integration of the dCas9-^{5XGCN4} cassette. Transfected cells were thus selected with hygromycin (250 ug/ml), plated at low density (1000 cells per 9.6cm^2) for single colony picking and clones with successful integration of dCas9-^{5XGCN4} were functionally tested by transfection with *KRAB$^{-GFP-scFv}$* and gRNA and DOX induction.

2.4.2 Lentivirus library preparation and transduction

Lentivirus library for gRNA delivery was produced by the Generic and Viral Engineering Facility of EMBL Rome. Briefly, the pooled gRNA Brie library (Addgene #73632) was co-trasfected in Lenti-X HEK 293T cells together with *pPax2* and *pMD2.G* plasmids according to BSL2 guidelines. After 48 and 72 hours from transfection, lentiviral-containing supernatant was harvested and filtered through a 0.22 μm low protein-binding filter. Viral particles were then concentrated using Lenti-X according to manufacturer instruction and resuspended in NDIFF 227. Lentiviral activity was estimated by transducing ESC across a titration curve and identifying a titration ratio to obtain 30-50% infection efficiency.

To generate KO library cell line, 7x10^7 *Esg1$^{-tdTomato}$* ESC containing the CAG::spCas9-T2A-GFP and *TRE3G::dCas9$^{-5XGCN4}$* were transduced in t2i/L medium with the pre-determined volume of lentiviral particles to ensure ~50% efficiency (>400 fold gRNA coverage) following BSL2 biosafety guidelines. 24 hours after transduction cells were washed five times to remove any residual lentiviral particles and cells were cultured in t2i/L medium with puromycin (1,2 μg/ml) selection for 7 days. Cells were passaged before sub-confluence maintaining a minimum of >3.2x10^7 cells to ensure gRNA library coverage (>400 fold coverage) and medium was changed daily.

2.4.3 spCas9 inactivation

Once the *Esg1$^{-tdTomato}$* KO library cell line was generated the *spCas9-T2A-GFP* cassette was inactivated to avoid interference with the *dCas9$^{-5XGCN4}$*. At this purpose >1x10^8 cells have been transfected with two tracr:crRNA against two unique sequences in the spCas9 that differs from dCas9$^{-5XGCN4}$ using Xfect RNA transfection reagent (Takara #631450) according to manufacturer instructions. After five days from transfection cells that received frameshift mutation in the spCas9 sequence have been isolated by sorting GFP- cells with FACS Aria III (Becton Dickinson). 3x10^7 GFP- cells have been plated back in t2i/L 10%FBS and further expanded for three days.

2.4.4 Epigenetic editing and memory in ESC with CRISPR library

A total of $2x10^8$ *Esg1*$^{-tdTomato}$ KO library cell line, already carrying dCas9$^{-5XGCN4}$ *and in which spCas9-T2A-GFP* has been inactivated, have been transfected with *pPB_TRE3G::KRAB*$^{-GFP-scFv}$*_EF1a::Neo*, *pPB_U6::gRNA_EF1a::BFP-Puro* containing a gRNA against *Esg1* and *pPY_CAG_Pbase* using Xfect mESC transfection reagent (Takara #631320) following manufacturer guidelines. After 24 hours, t2i/L medium was changed adding Neomycin (300 μg/ml) selection and cells have been induced with DOX (100 ng/ml) for seven days. After one week, $3x10^7$ TOM- cells have been sorted by FACS and cultured back in t2i/L without DOX. At the same time ~$3x10^7$ TOM+ cells have been isolated at FACS for genomic DNA (gDNA) extraction as a control. After 3 days of DOX washout cells have been detached, $3x10^7$ cells have been plated back in culture while, the remaining cells have been resuspended in FACS medium and subjected to Flow Cytometry to sort the cells that maintained the *Esg1*$^{-tdTomato}$ reporter OFF (identified as TOM-). Simultaneously TOM+ cells have been isolated as control and the two distinct cell populations have been snap-frozen for subsequent gDNA extraction. After a total of 7 days of DOX washout the sorting experiment was repeated, TOM- and TOM+ populations have been isolated and cells stored for gDNA extraction.

2.4.5 Genomic DNA extraction and library preparation

Genomic DNA was isolated from purified populations by using DNeasy blood and tissue kit (Qiagen # 69504) following manufacturer instruction including RNAse step. Integrated gRNAs were amplified in multiple parallel reactions with 500 ng of gDNA each with custom primers containing the P7 flow cell overhangs (5'-CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT-3') including 8 bp barcode and P5 overhang (5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCATCTTTGTGGAAAGGACGAAACACCG-3') using the Q5 Hot Start High-Fidelity polymerase (NEB #M0494S) for 22-24 cycles. sgRNA amplicons were purified using SPRI beads (Beckman Coulter #B23318) following the instruction for double size selection with 0.5X and 1.2X bead volume to sample volume ratio. Purified fragments were checked and quantified with a tape station automated electrophoresis system, equal sample amounts were pooled together into a multiplexed library and sequenced on next generation single-end (SE-50) sequencer.

2.5 RNA preparation and analysis

2.5.1 RNA extraction

When cell number was higher than $1x10^4$, RNeasy kit (Qiagen) was used to perform RNA extraction following manufacturer instruction, adding b-mercaptoethanol to the lysis buffer when starting from frozen cell pellets. DNA digestion was performed on-column with RNase-free DNase (Qiagen), unless RNA preparation was subsequently used for retrotranscription (in which case the DNase digestion was performed immediately before adding the reverse-transcriptase).

In case of limiting cell numbers ($<1x10^4$), the PicoPure RNA isolation kit (Applied Biosystems) was used. Here, small cell pellets were resuspended in the extraction buffer and incubated for 30 minutes at 42°C. After preparation of the purification columns with conditioning buffer, the cell extract was mixed with 70% EtOH and loaded into the column and the flow through discarded upon centrifugation. After three washing steps, the RNA was retrieved by elution of the column in 30 µl of elution buffer.

2.5.2 Reverse transcription

1 µg of purified RNA was used for synthetizing cDNA using PrimeScript RT Reagent Kit (Takara). Briefly, RNA was incubated with gDNA eraser for 2 minutes at 42°C and subsequently the PrimeScript enzyme mix and random hexameres were added to the reaction and incubated for 15 minutes at 37°C following by 5 seconds at 85°C to stop the reaction. Importantly to check for contaminations, a control reaction in which the RNA was incubated with all the other components except the RT enzyme mix (-RT control) was performed.

2.5.3 Real-time quantitative PCR

cDNA obtained from retrotranscription of RNA and -RT controls were diluted and specific targets quantified by real-time quantitative qPCR using primers designed at exon-exon junctions to minimise amplification from contaminant DNA (Appendix, Table 9.6). The reaction was performed using SYgreen Blue Mix (PCRbio) and QuantStudio 5 (Applied Biosystems) thermal cycler.

2.6 Protein preparation and analysis

2.6.1 Protein extraction and Western Blot

Proteins were extracted from cell pellet by resuspending in RIPA buffer (Sigma #R0278) containing protease inhibitors (Roche 4693159001) and incubating at 4°C for 30 minutes. Cell lysis supernatant was collected after centrifuging at full speed and proteins were quantified with Rapid Gold BCA Protein Assay Kit (ThermoFisher #A53225), following manufacturer instructions.

To perform western blot, 10-20 μg of proteins were mixed with bolt LDS sample buffer (ThermoFisher #B0007) and bolt reducing agent (ThermoFisher #B0004), heated at 70°C for 10 minutes and loaded on 4-12% Bis-Tris gel (ThermoFisher, #NW04125BOX) for low molecular weight proteins or 3-8% Tris-acetate gel (ThermoFisher #EA0375BOX) for high molecular weight proteins. Proteins were separated with 200V electrophoresis using either MES running buffer (ThermoFisher, NP0002) or Laemmli buffer (10% glycerol; 75mM Tris; 0.1% SDS) for low or high molecular weight proteins respectively.

Proteins were transferred on PVDF membrane (ThermoFisher #IB24002) using the iBlot 2 transfer stack and the membrane was subsequently saturated with with 5% milk/PBS for one hour at room temperature. Primary antibody incubation was always performed overnight at 4°C in 0.5%milk/PBS/0.05%tween (Appendix, Table 9.5) and after three washes in PBS/0.05% tween HRP-linked secondary antibody incubation was carried on in 0.5%milk/PBS/0.05%tween for 1 hour at room temperature. After washing thrice with 0.5%milk/PBS/0.05%tween, detection was performed by incubating the membrane with Pierce ECL waster blot solution (ThermoScientific #32132) for 5 minutes prior imaging with ChemiDoc XRS+ system (BioRad).

2.7 Epigenetic analysis

2.7.1 Bisulphite pyrosequencing

DNA bisulfite conversion was performed directly starting from cell pellets (a maximum of $1x10^5$ cells per sample) using the EZ DNA Methylation-Direct kit (Zymoresearch #D5021) following manufacturer instructions and eluting the final product in 10 µl H_2O. Target genomic regions were PCR amplified using 1µl of converted DNA with biotin-conjugated bisulfite primers (Appendix, Table 9.7), using the PyroMark PCR kit (Qiagen #978703) with the following conditions:

1. 95°C for 15 min
2. 94°C for 30 sec
3. 56°C for 30 sec
4. 72°C for 30 sec
5. 72°C for 10 min

Repeat from 2-4 for 30-45 cycles

The PCR reaction was inspected by 1.5% agarose gel electrophoresis and Pyrosequencing was performed using the PyroMark Q24 advanced reagents (Qiagen, #970902). Essay conditions were generated using the PyroMark Q24 Advanced 3.0 software and appropriate volumes of Reagent, Substrate and single nucleotides were estimated for each reaction. 10 µl of the PCR reaction was mixed with streptavidin beads (GE Healthcare #17-5113-01) by shaking for 5 minutes at room temperature and after separation of DNA strands and release of samples into the Q24 plate (Qiagen) using PyroMark workstation (Qiagen), sequencing primers were annealed to DNA by heating at 80°C for 2 minutes and cooling down at RT for 5 minutes. Pyrosequencing was run on PyroMark Q24 advanced pyrosequencer (Qiagen) with target specific dispensation order (Appendix, Table 9.8). Results were analysed with PyroMark Q24 Advanced 3.0 software.

2.7.2 CUT &RUN

The newly developed CUT&RUN (Cleavage Under Targets and Release Using Nuclease) protocol (Skene and Henikoff, 2017) was used to detect proteins-DNA interaction and histone modifications. A total of $3x10^5$ cells per sample were pelleted and washed twice with wash buffer (20mM HEPES pH 7.5, 150mM NaCl, 0.5 mM

Spermidine containing protease inhibitor) and incubated with concavallin beads by rotating for 10 minutes at room temperature. After placing samples on a magnet stand, supernatant was pulled off, cells resuspended with Antibody buffer (wash buffer with 0.02% digitonin, 2mM EDTA) containing 0.5 ug of target-specific antibody and left rotating overnight at 4°C.

Samples were placed on a magnet stand to remove antibody buffer and washed twice with wash buffer containing 0.02% digitonin (Dig-wash buffer). When using mouse raised primary antibody an extra step by conjugating with a rabbit α-mouse secondary antibody was required due to the low affinity of ProteinA to mouse IgG. In this case incubation with 0.5 μg of rabbit α-mouse secondary antibody was performed for 1 hour at 4°C on a tube rotator. After one extra wash with the Dig-wash buffer, cells were incubated with 700 ng/ml of pA-MNase and left rotating at 4C for one hour followed by two more washes. MNase reaction was thus activated by adding 4mM $CaCl_2$ and incubating at 0C for 30 minutes and immediately stopped with 1X final concentration of STOP buffer (340mM NaCl, 20mM EDTA, 200mM EGTA, 0.02% Digitonin, 0.05 mg/ml glycogen, and RNaseA). Target chromatin was released by incubating at 37C for 10 minutes, centrifuging at full speed for 5 minutes in a refrigerating centrifuge and the supernatant collected after incubation on magnet stand. DNA was finally released from chromatin by incubation with 0.4% SDS and 0,5 mg/ml Proteinase K at 70°C for 10 minutes.

Purification and size selection of DNA was performed using SPRI beads (Beckman Coulter) following the instruction for double size selection with 0.5X and 1.3X bead volume to sample volume ratio.

2.7.2.1 CUT&RUN qPCR

CUT&RUN DNA fragments were diluted ten times with H_2O and 2 μl used to perform quantitative-PCR with SYgreen Blue Mix (PCRbio) and QuantStudio 5 (Applied Biosystems) thermal cycler. Primers were designed to amplify desired regions and control regions in which the mark is expected to be enriched (positive controls) or depleted (negative controls) (Appendix, Table 9.9). Relative abundance of histone marks was estimated comparing Ct-values of target regions to positive control regions.

2.7.2.2 CUT& RUN sequencing

Libraries were made starting from 10ng of CUT&RUN DNA fragments using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB #E7645S) using the following PCR program: 98°C 30s, 98°C 10s, 65°C 10s and 65°C 5min, steps 2 and 3 repeated fro 12-14 cycles depending on input DNA. Library samples were sequenced on the Nextseq Illumina sequencing system (paired-end 40 sequencing).

2.7.3 ATAC-seq

Prior to harvesting, cells were initially treated in culture medium with 200U/ml of DNase for 30 minutes at 37°C to digest degraded DNA possibly released from dead cells. After detaching cells with TryPLE, $5x10^4$ cells were counted and pelleted at 500 RCF at 4°C for 5 minutes. Supernatant was aspirated and cells resuspended in 50 µl of cold ATAC-Resuspension Buffer (10 mM Tris-HCl pH7.4, 10 mM NaCl, 3 mM MgCl2) with 0.1% NP40, 0.1% Tween20 and 0.01% Digitonin and incubated on ice for 3 minutes. Lysis was washed out using 1 ml of cold ATAC-Resuspension Buffer with 0.1% Tween20 and mixed. Nuclei were pelleted at 500 RCF for 10 minutes at 4°C. After removal of supernatant, nuclei were resuspended in 50 µl of transposition mixture (25 µl 2xTD buffer, 2.5 µl transposase (Illumina Tagment DNA Enzyme and Buffer Kit #20034197), 16.5 µl PBS1x, 0.5 µl 1% digitonin, 0.5 µl 10% tween20, 5 µl H_2O) and incubated at 37°C for 30 minutes in a thermomixer with 1000 RPM shaking. Reaction was cleanup with DNA clean and concentration kit (Zymo #D4014) following manufacturer instructions and eluted in 21 µl of elution buffer. 20 µl of this product was used for PCR amplification using Q5 Hot Start High-Fidelity polymerase (NEB #M0494S) and a unique combination of the dual barcoded primers P5 and P7 Nextera XT Index kit (Illumina #15055293) following the cycling conditions below:

1. 98°C for 30 sec
2. 98°C for 10 sec
3. 63°C for 30 sec
4. 72°C for 1 min
5. 72°C for 5 min

Repeat from 2-4 for 5 cycles

After the first 5 cycle, 5 µl of the pre-amplified mixture was used to determine additional cycles by qPCR amplification using SYgreen Blue Mix (PCRbio) and the above used P5 and P7 primers in a QuantStudio 5 (Applied Biosystems) thermal cycler. After qPCR amplification, manually assess amplification profiles plotting linear Rn versus cycle. the number of the additional PCR cycles to be performed equals to one-third of the maximum fluorescent intensity in this plot (Buenrostro et al., 2015). The identified number of extra PCR cycles were performed by placing the pre-amplified reaction back in the thermal cycler. Final cleanup of the amplified library was performed using the DNA clean and concentration kit (Zymo #D4014) and DNA amplicons eluted in 20 µl of H_2O.

2.8 Bioinformatics analysis

2.8.1 CRISPR-screen NGS analysis

The Model-based Analysis of Genome-wide CRISPR-Cas9 Knockout (MAGeCK, v0.5.9) tool (Li et al., 2014) was used for counting and analysing the sgRNA repesentation. The reads were first trimmed using cutadapt (v1.15) (*cutadapt -g TTGTGGAAAGGACGAAACACCG*) and quality checked using FastQC and then, the gRNAs counts were normalised to total reads within the sample (MAGeCK -count -norm-method total). In the end, the TOM- and TOM+ sorted samples were compared by identifying significantly enriched/depleted gRNAs between samples using the -test command in MAGeCK. FDR threshold of <0.2 was used to identify final candidate lists.

2.8.2 CUT&RUN-seq analysis

After adapter-trimming with TrimGalore (v0.4.3.1, -phred33 --quality 20 --stringency 1 -e 0.1 --length 20) and quality check, raw Fastq sequences were aligned to the custom mouse mm10 genome with the inserted tdTomato reporter using Bowtie2 (v2.3.4.2, -I 50 -X 800 --fr -N 0 -L 22 -i 'S,1,1.15' --n-ceil 'L,0,0.15' --dpad 15 --gbar 4 --end-to-end --score-min 'L,-0.6,-0.6'). Analysis of the mapped sequences was performed using seqmonk (Babraham bioinformatics, v1.46.0) by enrichment quantification of the normalised reads.

2.8.3 ATAC-seq analysis

Raw reads were first trimmed with with TrimGalore (v0.4.3.1, reads>20bp, quality >30) and then quality checked with FastQC (v0.72). Output files were aligned to custom mouse mm10 genome with the inserted tdTomato reporter using Bowtie2 (v2.3.4.3, Paired-end settings, fragment size 0-1000, --fr, allow mate dovetailing). Uninformative reads were removed with Filter BAM (v2.4.1, mapQuality>=30, isProperPair, !chrM) and duplicated reads were filtered with MarkDuplicates tool (v2.18.2.2). The mapped and filtered sequences were then analysed with seqmonk (Babraham bioinformatics, v1.46.0) by performing enrichment quantification of the normalized reads. Correlation plots were generated by comparing enrichment of reads at promoters in sample versus control conditions.

3 Exploiting precise and releasable epigenome editing to investigate epigenetic memory in ESC

3.1 Introduction

Models of environmentally induced epigenetic memory have been shown in many organisms such as plants, *C. elegans* and *Drosophila* that can sense environmental changes and 'translate' them into epigenetic changes. One striking example of how the epigenome can respond to changes in the environment has been shown very recently also in mammalian cells that react to oxygen levels through the H3K27 histone demethylase KDM6A and translate it into loss of H3K27me, ultimately influencing cell fate (Batie et al., 2019; Chakraborty et al., 2019).

An example of extremely long memory of environmentally induced epigenetic aberration comes from nematodes where it was found that changes in temperature induced changes in H3K9me3 and that this alteration was inherited for 14 generations (Klosin et al., 2017). The peri conceptional environment in which the early embryo develops is considered to be one of the main temporal windows during which the environment could, in principle, alter epigenetic (re)programming events (Radford, 2018). Furthermore, if such changes occur early in development they, in principle, can be propagated through the whole organism.

Indeed, the pluripotency window persists only transiently *in vivo* and pluripotent cells are rapidly committed towards their fates. However, pluripotency can be captured *in vitro* by culturing the Embryonic Stem Cells (ESCs), isolated from the ICM, in the presence of specific MEK, ERK and GSK3 inhibitors. However, the inhibition of these pathways in ESC maintained in 2i/Lif culturing conditions leads to global DNA

hypomethylation and a general decrease in H3K27me3 (Ficz et al., 2013; Habibi et al., 2013; Leitch et al., 2013; Marks et al., 2012). Nonetheless, titrating the 2i inhibitor PD0325901 supports the restoration of high global levels of DNA methylation, without changes in cell identity compared to the full 2i/Lif condition (Gretarsson and Hackett, 2020).Importantly, in this culturing conditions, ESCs chromatin architecture appears hyperdynamic resembling the same state also found in eight-cell embryos and pluripotent epiblast cells (which are the source of ES cells) *in vivo* (Ahmed et al., 2010). This highlights the physiological relevance of the studies conducted in ESCs with the *in vivo* counterpart.

In such hyperdynamic epigenetic conditions it remains to be elucidated the extent of heritability of epigenetic perturbations during pluripotency. Indeed, as described in paragraph 1.5.2, during development, cells undergo massive changes in the epigenetic environment, including widespread DNA demethylation. However, a number of loci escape this erasure. For example, endogenous retroviruses (ERV) which are kept silenced throughout reprogramming by KAP1 (aka TRIM28)-KRAB-zinc-finger proteins-mediated recruitment of histone methyltransferases such as SETDB1 (Rowe et al., 2010) and the nucleosome remodelling and deacetylase (NURD) complex that removes H3K4me3 and H3K4ac. KRAB-mediated silencing may thus provide a mechanism to escape epigenetic reprogramming during pluripotency.

Nonetheless, it is of great interest to understand whether epigenetic changes occurring early in development can be propagated during the pluripotency window. Indeed, changes occurring in the cell precursors of all somatic lineages, can in principle be widely passed in all tissues of multicellular organisms, being potentially harmful.

In order to model such environmental-induced epigenetic changes *in vitro*, here I introduced precise perturbations via CRISPR-Cas9 epigenetic editing toolbox. As already explained in section 1.7.1, this tool is composed of a catalytically inactive Cas9 (dCas9) enzyme that, in this case, is exclusively used as a docking platform to recruit epigenetic effectors on a specific DNA sequence thanks to the complementarity of binding with a guide RNA (gRNA). Simply by changing the gRNA sequence, the system can be delivered on nearly any region of the genome as long as it contains a protospacer adjacent motif (PAM) sequence.

To date dCas9 has been used to recruit a plethora of different epigenetic factors either by fusion constructs or serving as anchors for multiple copies of antibody- or protein- tagged effectors such as the epigenetic cofactor KRAB (Gilbert et al., 2013), the transactivator VP16 (Cheng et al., 2013) or epigenetic modifiers as the DNA methyltransferase DNMT3a (Vojta et al., 2016) or the histone acetyltransferase p300

(Hilton et al., 2015), just to name a few. Some studies also investigated the stability of the changed epigenetic states. However, the results of such works are conflicting, with some reporting that these changes were lost after some days (e.g. (Kungulovski and Jeltsch, 2016) and others showing stable silencing of target gene expression with different degrees according to single or combinatorial deposition of the marks in cancer cells (e.g. (Amabile et al., 2016; Bintu et al., 2016). These apparently contradictory results underline that the inheritance or loss of an epigenetic domain depends on numerous factors including the underlying DNA sequence, the position in the nucleus and the chromatin context such as proximity with systems that propagate the marks or remove them.

By using a modulable dCas9-effector system, the aim of this chapter is to investigate the mitotic stability of epigenetic perturbations at endogenous loci in mouse embryonic stem cells which are here used as an *in vitro* model of pluripotency in preimplantation embryos.

3.2 Optimization of an epigenome editing tool to introduce epialleles

In order to precisely introduce epigenetic perturbations and follow the memory of the new epialleles, once the repressive stimulus is released, I needed a tool that allows (i) precise targeting, (ii) broad ON-target activity (i.e multiple histones modified), (iii) that is trackable and tunable and that (iv) can be dynamically turned ON and OFF on demand. During the first part of my doctoral project I therefore worked on optimising such a tool. Thanks to its modular and precise DNA sequence-specific activity CRISPR-dCas9 is the technology I was looking for. CRISPR-Cas9, indeed, has recently been engineered for recruitment of different protein domains on DNA using single-chain fusion constructs. However, the recruitment of just a single effector does not always induce great changes in gene expression (Cheng et al., 2013; Mali et al., 2013; Perez-Pinera et al., 2012). To increase the deposition of the desired modification, previous studies used synthetic reporters containing multiple binding sites for the same gRNA (Gilbert et al., 2013) or, alternatively, delivered multiple copies of dCas9-effector fusion using multiple nonoverlapping gRNAs (Cheng et al., 2013; Mali et al., 2013; Perez-Pinera et al., 2012). However, a more elegant and successful strategy to achieve signal amplification is the SunTag system developed by Tanebaum and collaborators (Tanenbaum et al., 2014). In this system, the dCas9 is fused with an array of GCN4 peptide repeats that can recruit an equal number of proteins fused to the counterpart GCN4-specific single-chain-

variable-fragment (scFv) antibody anchor. The original SunTag consists of ten copies of the GCN4 peptide separated by a short linker. However, I employed a slightly modified version of the system developed by Morita and co-workers (Morita et al., 2016) with a reduced number of GCN4 repeats (from 10 to 5) (hereafter named *dCas9-*^{5xGCN4}) and increased length of the linker to 22 amino acids, adjusted to avoid steric hindrance between the bound effectors.

Given its well-known role as transcriptional repressor, and its ability to recruit chromatin modifiers that ultimately introduce H3K9me3 and DNA methylation, which are thought most likely to induce a self-reinforcing epigenetic memory (Amabile et al., 2016), I employed the human KRAB ZNF10 repressive domain as the effector of choice.

I further optimised the epigenetic editing system as follows (Table 3.1 and Fig. 3.1):

i) To better exploit the modularity of the system I split the originally all-in-one vector containing the dCas9, the scFv-effector and gRNA (Morita et al., 2016) into three different plasmids, with only one of the three components. This also allows to easily combine different effectors and gRNAs of choice and control precise copy number balancing at the time of transfection.

ii) To quickly generate cell lines stably carrying the epigenetic editing system, I cloned the three constructs into PiggyBac vectors and delivered them in co-transfection with a piggy-bac transposase that randomly integrates the constructs in the genome (Ding et al., 2005). Furthermore, each plasmid contains an antibiotic resistance gene (Puromycin for the gRNA, Neomycin from KRAB and Hygromycin for *dCas9*$^{-5xGCN4}$) so that cells with stable integration of the exogenous DNA can be selected by appropriate antibiotic treatment.

iii) To finely tune the expression of the system both *dCas9*$^{-5xGCN4}$ and KRAB are placed under control of a Doxycycline (DOX) inducible promoter, the TRE3G promoter. The *dCas9*$^{-5xGCN4}$ plasmid also drives constitutive expression of the rTta transactivator, needed for DOX-mediated activation of the TRE3G promoter. Importantly, the epigenetic editing tool is reversible and its activity can be switched off by washing out DOX. Additionally, I cloned a d2 destabilization domain to both the components to confer 2-hour half-life in order for the system to be rapidly degraded upon removal of DOX.

iv) Since the system is inducible, in order to control its ON-OFF state, KRAB is fused to a green fluorescent protein (GFP) (hereafter named *KRAB*$^{-GFP-scFv}$). Additionally, the gRNA construct also contains a Blue Fluorescent Protein (BFP) (*gRNA*$^{-BFP}$) under control of a constitutive promoter. I conceived this dual fluorescent reporter to perform single-cell-analysis in desired subpopulation of cells by flow cytometry. For example, cells that have been successfully transfected will be BFP+,

and only upon doxycycline treatment will turn also GFP+ (GFP tracks for presence of both the effector, to which it is fused, and dCas9, which construct also contains the tTta transactivator, needed to turn ON the whole editing system). Therefore BFP/GFP double positive cells contain the complete system and can be isolated by fluorescence activated cell sorting (FACS). Ultimately, once doxycycline is removed, the cells that effectively switched off the system can be gated according to GFP- by flow cytometry analysis.

v) I employed an 'enhanced' gRNA with an AT-flip and extended stem loop to increase dCas9 occupancy time and therefore ON-target activity (Chen et al., 2013).

Table 3.1 Summary of the CRISPR Cas9 epigenome editing constructs

Name	Genome integration site	Main Promoter	Main construct description and remarks	Antibiotic Selection	Fluorescent reporter	Publica-tion of referen ce
Esg1$^{-tdTomato}$	5'UTR end of *Esg1*endogen ous gene on chromo-some 9	*Esg1* endogen ous promoter	tdTomato reporter separated with a T2A self-cleavable domain from the endogenous *Esg1*	Blasticydin	tdTomato in frame with T2A	
dCas9$^{-5XGCN4}$	Randomly integrated by Piggybac transposase	TRE3G Dox inducible promoter	Catalitically inactive sp-dCas9 fused with an array of five GCN4 repeats and a d2 destabilization domain. Also contains the rTtA trans activator under EF1a promoter.	Hygromycin	None	(Morita et al., 2016)
KRAB$^{-GFP-ScFv}$	Randomly integrated by Piggybac transposase	TRE3G Dox inducible promoter	KRAB epigenetic repressor fused to a single chain variable fragment of an antibody specific for GCN4. It is also fused to a d2 destabilization domain and a GFP fluorescent protein.	Neomycin	GFP (fusion with the main construct)	
GFP^{-ScFv}	Randomly integrated by Piggybac transposase	TRE3G Dox inducible promoter	GFP fluorescent protein fused to the ScFv domain. Control construct used to detect any non-epigenetic induced silencing due to binding of the DNA. It is also fused to the d2 domain.	Neomycin	GFP	
egRNA^{-BFP}	Randomly integrated by Piggybac transposase	U6 promoter	Target specific gRNA is inserted in an enhanced gRNA scaffold that is optimized for efficient binding of *dCas9*	Puromycin	BFP (under the control of Ef1a promoter)	(Chen et al., 2013)

In summary, upon transfection of the cells with the three constructs, the plasmids get stably integrated but the dCas9$^{-5xGCN4}$ and *KRAB*$^{-GFP-scFv}$ are not expressed until the addiction of DOX. On the other hand, once the epigenetic modification is deposited, memory of this alteration can be investigated by release of the editing tool mediated by DOX removal. The ON/OFF state of the system can be rapidly tracked by GFP

expression both using fluorescence microscopy and flow cytometry (Fig. 3.1.d). Importantly, over several experiments, I observed that the percentage of BFP/GFP double positive cells was well above 90% after 7 days of DOX induction (Fig. 3.1.d), implying that the cells consistently induced the epigenetic editing system, and that we could track epigenetic silencing specifically in those cells.

Furthermore, as the *KRAB*$^{-GFP-scFv}$ domain is anticipated to induce robust epigenetic silencing, cells have alternatively been transfected with a control construct (*GFP*$^{-scFv}$) in which only GFP is fused to scFv to check for any non-epigenetic-induced silencing of the reporter.

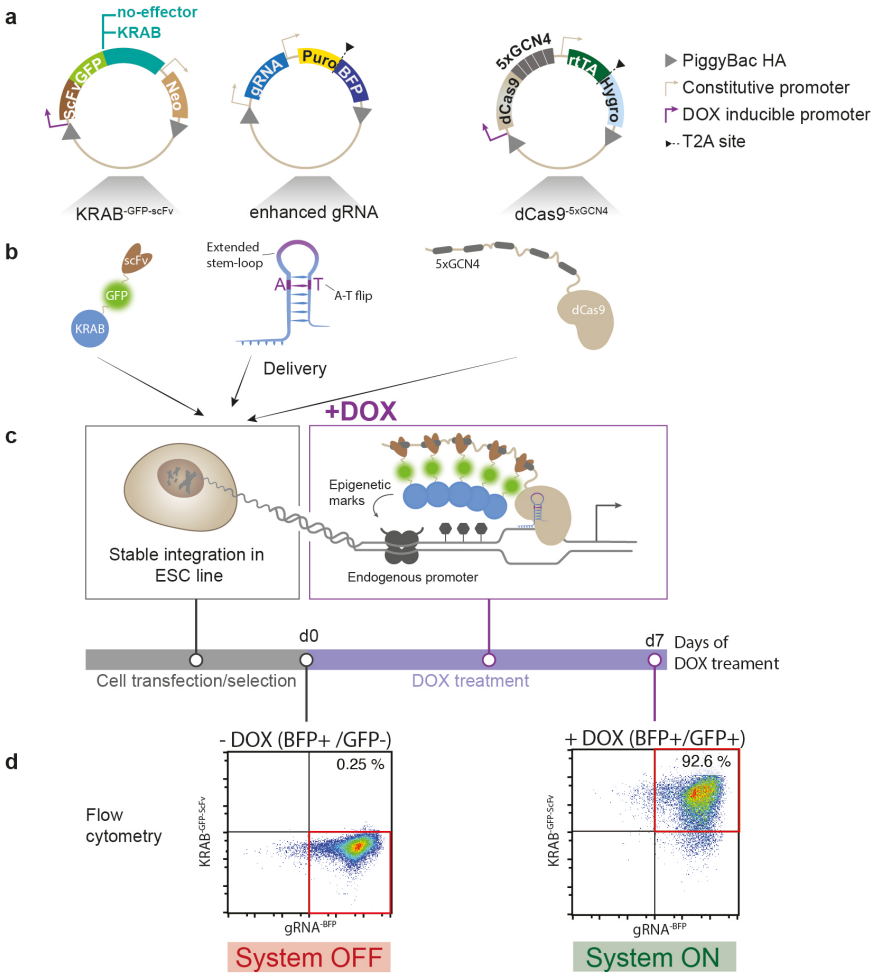


Fig 3.1. Optimization of an epigenome editing tool to introduce epialleles. (a) schematic representation of the constructs used to deliver the epigenetic editing tool into ESC. Symbols meaning is explained in the legend on the right. (b) Drawings represents the components of the epigenetic editing tool. From left to right: the effector (KRAB) fused to GFP and scFv; the enhanced gRNA with extended stem-loop and A-T flip (in purple); the dCas9 with the repetitive GCN4 tail. (c) Schematics of the workflow: from left to right, cells are transfected with the three constructs plus the PiggyBac transposase and selected with appropriate antibiotics to achieve stable integration. Following treatment with Doxycycline, the dCas9 $^{-5xGCN4}$ is recruited on the DNA thanks to the complementarity of binding with the gRNA and serve as a docking platform for the binding of up to five KRAB $^{-GFP-scFv}$ that in turn deposits the epigenetic marks. (d) Density plots, obtained by flow cytometry analysis, show gRNA $^{-BFP}$ and KRAB $^{-GFP-scFv}$ intensity in the cell population prior and upon DOX induction from left to right. Red boxes delimit the cells taken into consideration at each time point

3.3 A reporter for epigenetic silencing is employed to track epigenetic memory

In order to track the deposition of the epigenetic marks at the target sites, transcription of a fluorescent reporter gene is an easy readout that allows to follow memory at the single cell level by flow cytometry. Previous works, investigating the stability of epigenetic changes, have been carried out using synthetic reporters (Bintu et al., 2016; Amabile et al., 2016). However, these transgenes are either inserted in human artificial chromosomes (HAC) or carry ectopic promoters and repetitive cassettes for binding of TALE-tagged effectors thus artificially altering the chromatin structure.

Instead, I employed a reporter ESC line carrying mono-allelic integration of the tdTomato (TOM) fluorescent reporter replacing the coding sequence of the Embryonal stem cell-specific gene 1 (*Esg1*) (Fig. 3.2). This gene is widely expressed in ESC and has an intermediate CpG island (ICP) promoter which class of promoters is particularly sensitive to stable DNA methylation (Illingworth and Bird, 2009). This enables tracking the epigenetic state at an unaltered endogenous promoter and thus possibly repurpose a more physiological condition.

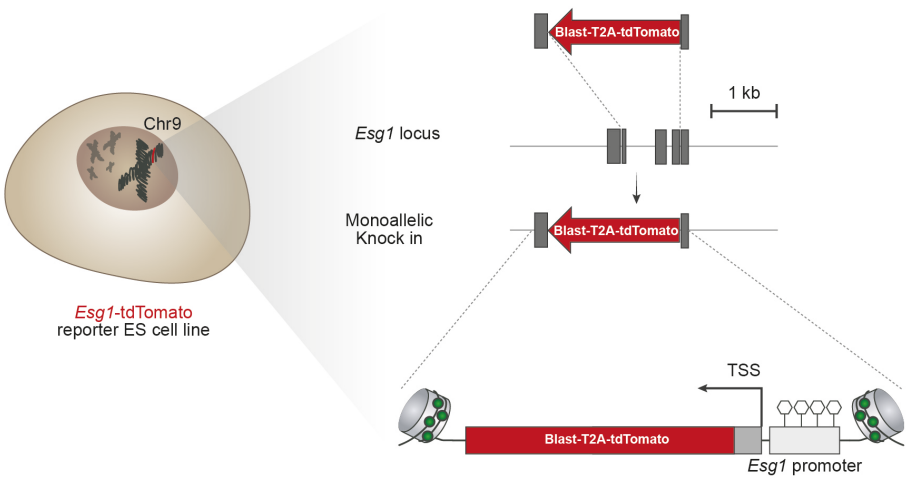


Fig 3.2. The endogenous reporter for epigenetic silencing. Top: schematics of the strategy to insert the coding sequence for tdTomato-T2A-Blast into the Esg1 gene replacing exons 1-3. The insertion was monoallelic. Bottom: schematic representation of the Esg1 reporter. White lollypops indicate unmethylated CpGs and green circles represent active histone marks. The black arrow indicates the position of the transcriptional start site (TSS).

3.4 Multiplexed Cas9-KRAB system efficiently deposits heterochromatin.

To induce the deposition of ectopic heterochromatin I employed the *Esg1*-reporter ES cell line carrying integration of dCas9$^{-5xGCN4}$ + *KRAB*$^{-GFP-scFv}$ (or alternatively *GFP*$^{-scFv}$ as a control) + a gRNA targeting a sequence 87 bp downstream the *Esg1* TSS (*gRNA*^{87dw}) generated as described in Fig. 1 and treated these cells for seven days with DOX (Fig. 3.1.c).

It is worth reminding that KRAB is a widely used effector for CRISPRi and it works by recruiting a co-repressor KAP1 that in turn is able to bind to histone methyl-transferases such as SETDB1, which deposits H3K9me3 (Schultz et al., 2002) and the NURD complex which removes histone H3 lysine-4 and acetyl groups. Ultimately, the KAP1-complex also induces DNA methylation by interacting with DNMTs (Quenneville et al., 2012). Additionally, H4K20me3 often colocalises with H3K9me3. Altogether the combination of H3K9me3, H4K20me3 and DNA methylation, and the absence of H3K4me3, are a hallmark of robust heterochromatin. Here ESCs are maintained in titrated 2iL (see Materials and methods) conditions which promote high global levels of DNA methylation while still maintaining cells in a pluripotent state (Gretarsson and Hackett, 2020).

At first, to estimate the efficiency of epigenetic editing by my multiplex system I performed CUT&RUN (Cleavage Under Targets & Release Using Nuclease) (Skene and Henikoff, 2017) for H3K9me3, H3K4me3 and H4K20me3 after 7 days of DOX treatment (Fig 3.3.a). Briefly, after binding of a histone-mark (or, alternatively, a transcription factor) by specific antibody in intact cells, the flanking DNA is cut thanks to an MNase protein fused to a peptide that recognises and binds a constant chain of the antibody. The DNA fragments are then released from the cells and can be analysed by quantitative real time PCR (qPCR) or next generation sequencing. qPCR amplification of two regions on the *Esg1* promoter located at 300 or 700 bp upstream the TSS (Fig. 3.3.b), shows approximately 5 to 10 folds ectopic deposition of H3K9me3 (Fig. 3.3.c) and up to almost 20 folds deposition of H4K20me3 (Fig. 3.3.d) in cells transfected with *KRAB*$^{-GFP-scFv}$ (in blue) compared to an untransfected control (in grey). Importantly no epigenetic modification was observed in the *GFP*$^{-scFv}$ control (Fig. 3.3.c and d). On the other hand, H3K4me3 that decorates the promoter in untargeted and *GFP*$^{-scFv}$ condition is almost completely erased upon targeting with *KRAB*$^{-GFP-scFv}$ (Fig. 3.3.e). Additionally, CUT&RUN sequencing confirmed the results obtained with CUT&RUN-qPCR, showing that epigenetic editing is specific to only the targeted locus, and only in *KRAB*$^{-GFP-scFv}$ and not in the *GFP*$^{-scFv}$ control. It also revealed that the deposition of H3K9me3 and H4K20me3 spreads for the whole *Esg1* locus covering in total around a 12 kb wide region (Fig. 3.4). Similarly, H3K4me3 is

completely lost only in *KRAB*$^{-GFP\text{-}scFv}$ for the whole locus. This might suggest that a positive feedback loop is in place to spread the ectopically-deposited marks.

I also quantified the deposition of DNA methylation at the targeted locus by pyrosequencing covering the whole promoter subdivided in three regions (Fig. 3.3.b). Of note, the *Esg1* promoter is originally hypomethylated as shown in the untransfected control in Fig. 3.3.f and the recruitment of the sole *GFP*$^{-scFv}$ does not affect its methylation state. Instead, when *KRAB*$^{-GFP\text{-}scFv}$ is recruited, I observed about a 40% increase in the total CpG methylation with the more distal CpG sites (-800 and -200 regions) - the further from the *gRNA*^{87dw} binding sites - being the most methylated compared to the region spanning the TSS. This result might seem counterintuitive but might as well reflect the dCas9 footprint or suggest that CpG sited around the TSS are the most refractory to DNA methylation (Illingworth and Bird, 2009).

Taken together these results show that upon single-gRNA tethering of an array of five *KRAB*$^{-GFP\text{-}scFv}$ an extensive de-novo landscape of repressive epigenetic modifications is deposited to a specific locus, whilst activating marks are erased. This extends to >10kb implying establishment of a robust heterochromatin domain across a significant genomic region, thereby enabling subsequent analysis of its function and memory.

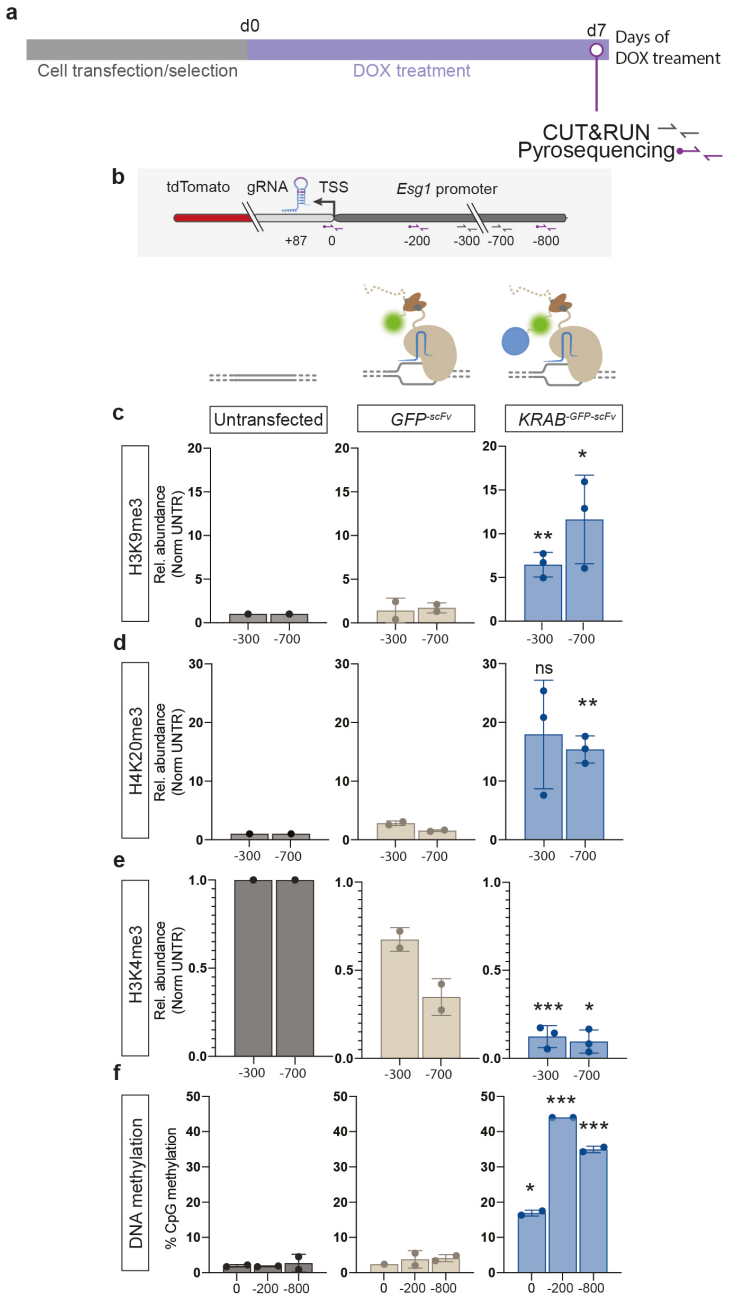


Fig 3.3. Targeted heterochromatin with a multiplexed Cas9-KRAB system. (a) Timeline of the experiment. (b) Schematics of the reporter: grey and purple arrows indicate the position of primer-pairs used for CUT&RUN-qPCR and Pyrosequencing, respectively. (c), (d), (e) Histograms show the relative abundance of each histone mark compared to a positive control region and the untransfencted control (grey) in GFP^{-scFv} (light brown) and KRAB$^{-GFP-scFv}$ (blue). (f) Histograms represent average of the percentage of CpG methylation of individual CpG sites at each region. Asterisks indicates Pvalues measured with one-tail unpaired t-test over two or three independent biological replicates (=p<0.05; **=p<0.01; ***=p<0.001).*

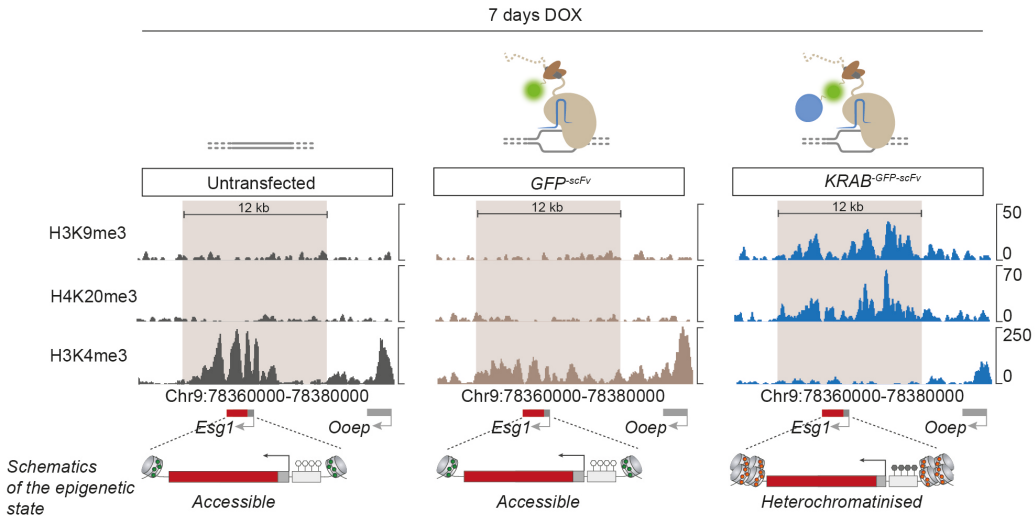


Fig 3.4. KRAB-induces deposition of a wide heterochromatin domain. From left to right are shown CUT&RUN tracks for each indicated mark in Untransfected (grey), GF $^{-scFv}$ (light brown) and KRAB$^{-GFP-scFv}$ (blue) after 7 days of DOX treatment. Brown boxes highlight the region of heterochromatin spreading induced upon editing. On the bottom: schematic representation of the reporter in each of the indicated conditions: white and grey lollypops indicate unmethylated or methylated CpG sites, respectively. Green and red circles represent active or repressive histones marks, respectively.

3.5 KRAB imposed heterochromatinization completely switches off transcription

DOX-mediated activation of the epigenetic editing tool can be tracked by means of fluorescence according to effector-GFP expression, while BFP, which controls for presence of gRNA, is constitutively active. Therefore, cells in which the epigenetic editing tool is completely functional can be identified according to GFP/BFP double expression. By fluorescent microscopy I observed that the reporter cell line strongly expresses tdTomato in -DOX conditions (Fig. 3.5.a). However, when the epigenetic editing system is turned on by addiction of DOX, and the cells turn green (GFP+), the *Esg1*-reporter is completely silenced (Fig. 3.5.a). To estimate the extent of the repressive transcriptional state induced by the newly established epigenetic domain, tethered with the *gRNA^{87dw}* (Fig 3.5.c), I analysed the expression of the reporter gene at the single-cell level by flow cytometry (Fig. 3.5.d). Notably, a consistent fraction of the cells (> 85% across four biological replicates) reached the completely off state after 7 days of KRAB tethering (>500-folds transcriptional repression), and were indistinguishable from a cell line not containing the tdTomato reporter (Fig. 3.5.d). Importantly the *GFP^{-scFv}* control induced only a mild repression, probably due to steric hindrance of the dCas9 binding proximally to the TSS, since no epigenetic marks were deposited (Fig. 3.3 and 3.4). The additional silencing imposed by *KRAB$^{-GFP-scFv}$* relative

to *GFP*$^{-scFv}$ is therefore epigenetically mediated. In fact, by using a *gRNA*^{475up} located further away from the TSS (475 bp upstream) (Fig. 3.5.c) I didn't detect any effect induced by *GFP*$^{-scFv}$ targeting, while, at the single cell level, the reporter targeted with *KRAB*$^{-GFP-scFv}$ shows a stochastic repression with a smaller percentage of cells (66%) reaching the complete OFF state (< 10^2) while another fraction of the population only achieves an intermediate level of silencing (10^2-10^4) or none (>10^4).

In summary, these data indicate that ectopically deposited heterochromatin marks strongly counteract transcription at the target locus, and this has a high penetrance at the single-cell level. Overall the epigenetic status of this promoter closely predicts the transcriptional activity of the gene, and can thus be used as a proxy for epigenetic memory.

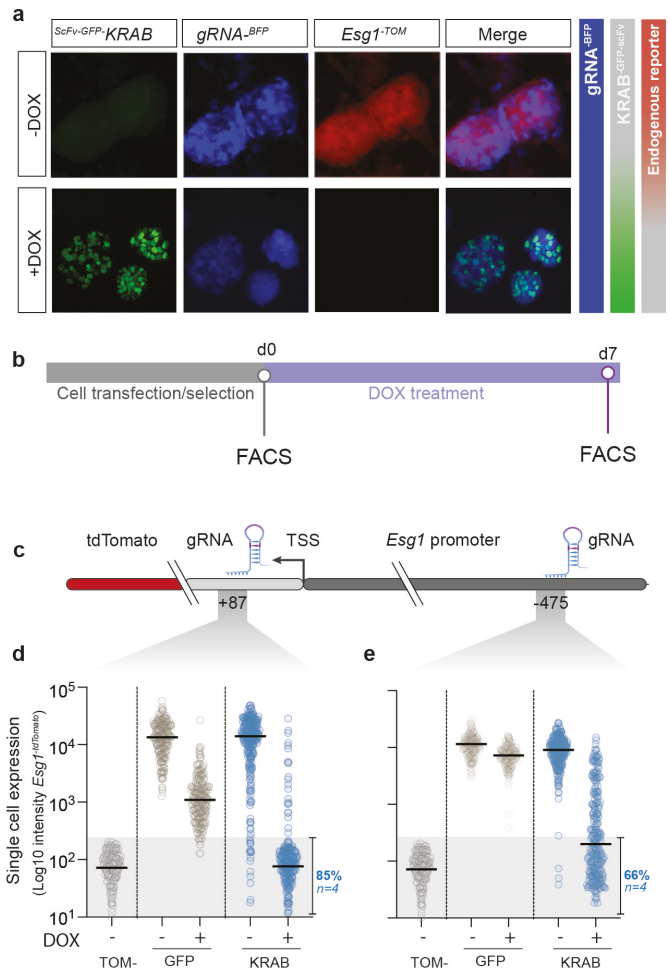


Fig 3.5. KRAB imposed heterochromatinization completely switches off transcription at the single cell resolution. (a) Epifluorescent microscope images of reporter ESCs in -DOX (top) or +DOX conditions (bottom). Schematics of the expression of each fluorophore is represented on the right: gRNA^{-BFP} is constitutively expressed in both conditions (blue box); when the epigenetic editing system is turned ON by DOX and KRAB$^{-GFP-scFv}$ is expressed (gradient green box) the Esg1$^{-tdTomato}$ reporter is silenced and the red fluorescence is lost in +DOX cells (gradient red box). (b) Timeline of the experiment.(c) Schematics of the Esg1$^{-tdTomato}$ reporter locus. Numbers refers to position of gRNAs in bases upstream (-) or downstream (+) the transcriptional start site (TSS). (d), (e) violin plots show Log10 intensity of the Esg1$^{-tdTomato}$ reporter at the single cell level (each dot is a single cell) in GFP^{-scFv} (light brown) and KRAB$^{-GFP-scFv}$ (blue) in - and + DOX conditions in one representative example. In grey is represented the basal level of fluorescence of a cell line not containing the tdTomato reporter. Percentage indicate the proportion of cells fully silenced (grey box) in the KRAB +DOX condition (average of four replicates). Black horizontal bars represent the median of fluorescence intensity in the population of cells.

3.6 Deposited heterochromatin is reverted upon removal of inducing signals

Having established a major heterochromatin domain, which existing paradigms suggest could self-propagate through cell division (Amabile et al., 2016; Hackett et al., 2013a; Rowe et al., 2010; Tang et al., 2015), the next question is whether the induced epi-mutation can be inherited across cell replications after release of the epigenetic editing system.

To do so, I sorted the TOM- cells that successfully acquired the epimutation after seven days of DOX treatment and re-plated them back in culture in absence of Doxycycline to release KRAB from the targeted locus (Fig. 3.6.a). I could confirm that the epigenetic editing system was effectively switched off by checking the expression of BFP and GFP by flow cytometry (Fig. 3.6.a). Importantly, the rapid depletion of the dCas9 system was enhanced by using a destabilised epigenetic effector and dCas9 variant with a 2-hour half-life. Furthermore, by gating for constitutive BFP positive cells, I could exclude dead cells that could confound the Tomato negative population (density plot in Fig. 3.6.a). To investigate the memory of the introduced epigenetic alteration, I analysed the reporter activity and its epigenetic state after 4 and 7 days after release of the epigenetic editing tool.

I observed that by 4 days of DOX washout, the endogenous reporter starts to lose silencing. However, a fraction of the GFP-/BFP+ population partially retains silencing while some other cells are completely reverted to the ON state, showing a bimodal distribution (Fig. 3.6.b). This single cell behaviour highlights the stochasticity of the memory triggered by this system. Although I filtered for cells with residual GFP expression I couldn't completely rule out if part of the memory observed at the early time point might be due to residual expression or leakiness of the system. However, after 7 days of DOX washout, epigenetic silencing is erased and the reporter expression is fully reverted ON (Fig. 3.6.b).

Consistently with the expression data, pyrosequencing revealed that DNA methylation in bulk cells is halved at the early time point investigated (4dwo) compared with the initial level of deposition observed after DOX treatment. Thus, DNA methylation is partially retained in a fraction of the population after 4 days of *KRAB*$^{-GFP-scFv}$ release. However, this partial memory does not last long and CpG methylation is almost completely lost from the *Esg1*-promoter after 7 days (7dwo) (Fig. 3.6.c).

To evaluate the persistence of the newly deposited histone marks on the reporter gene I performed CUT&RUN sequencing four or seven days after DOX removal (Fig. 3.7.a). I compared the histone marks enrichment at each timepoint with the +DOX condition and found that upon release of the triggering signal both H3K9me3 and

H4K20me3 are lost from the whole *Esg1* locus at both time points investigated. Consequently, endogenous levels of H3K4me3 are restored. Quantification of the histone marks abundance at the *Esg1* promoter by CUT&RUN-qPCR however, revealed a slight retention of H3K9me3 at the early time point (4dwo) while the levels of H4K20me3 and H3K4me3 are reverted back at the original state already at day 4 (Fig.3.7.b).

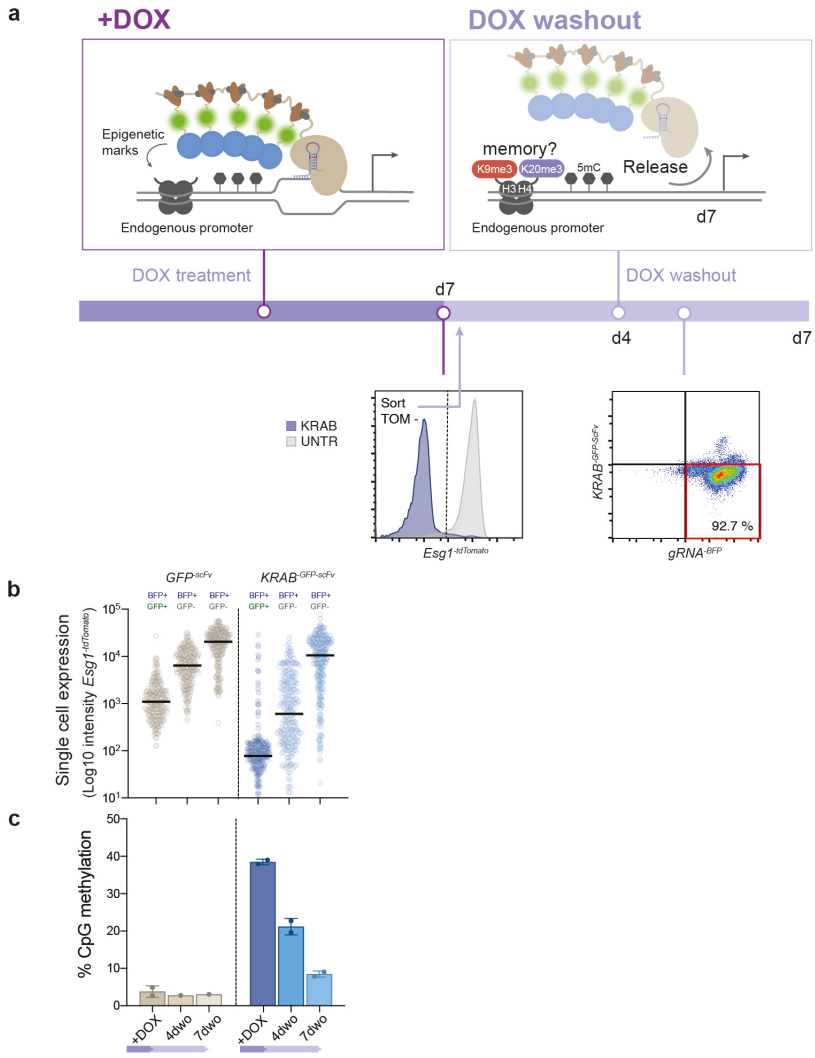


Fig 3.6. Loss of silencing and DNA methylation upon release of KRAB. (a) Timeline of the experiment: cells were treated for 7days with DOX to induce KRAB-mediated deposition of heterochromatin and cells that received the TOM- epiallele were flow sorted and put back in culture in the absence of DOX and analysed after 4 or 7 days. Density plot shows the intensity of gRNA^{-BFP} and KRAB$^{-GFP-scFv}$ upon DOX wo. Red box indicates the gate used to select BFP+/GFP- cells. (b) Violin plots show the distribution of Esg1 -tdTomato intensity (Log scale) in each condition (+DOX, 4days and 7days DOX wo) in GFP^{-scFv} (light brown) or KRAB$^{-GFP-scFv}$ (blue) over four biological replicates. Horizontal bars represent the median. (c) Histograms represent the average of %CpG methylation at region -200 (see schematics in Fig. 3.3) at each timepoint. Error bars indicate standard deviations between two independent biological replicates (dots).

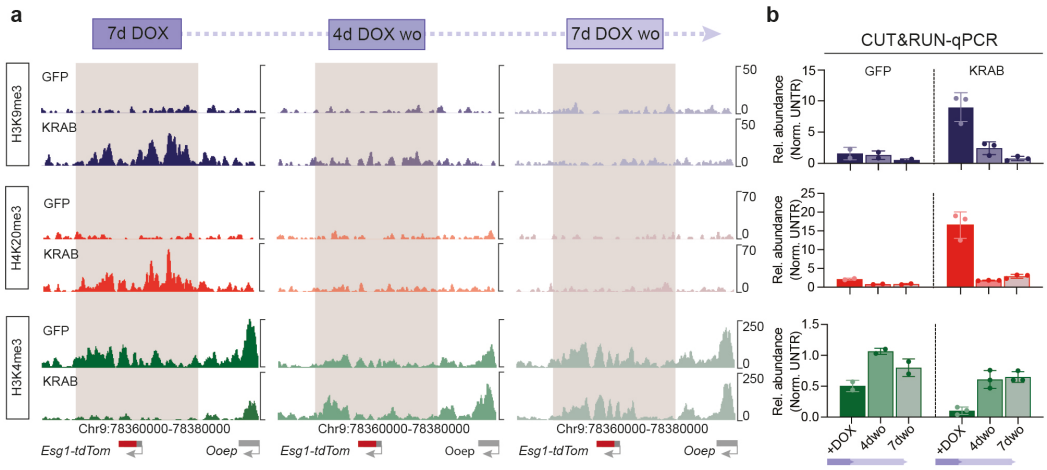


Fig 3.7. Deposited heterochromatin is erased in ESC upon removal of the triggering system (a) From left to right CUT&RUN tracks at +DOX, 4 and 7 days of DOX removal in GFP^{-scFv} or KRAB$^{-GFP-scFv}$ for H3K9me3 (dark blue), H4K20me3 (red) and H3K4me3 (green). Brown boxes highlight the region with heterochromatin spreading observed after DOX treatment. (b) CUT&RUN qPCR quantification of the relative abundance of each mark normalised to a positive control region and to the untransfected control. Error bars represent standard deviation between two or three biological replicates (each dot is a biological replicate).

Previous work, performed in human somatic cancer cell line similarly showed that KRAB mediated silencing has only a transient memory function (Amabile et al., 2016). However, when they combined the recruitment of KRAB with Dnmt3a and 3L they observed a synergistic effect of these effectors in stabilising transcriptional repression from a synthetic reporter gene.

To evaluate if the same cooperative effect occurs also with an endogenous locus in my physiologically-relevant ESC system, I induced combinatorial recruitment of single chain fusion catalytic domain of the DNA methyl transferase Dnmt3a with the cofactor Dnmt3L (*3a3L*$^{-GFP-scFv}$) and *KRAB*$^{-GFP-scFv}$ at the *Esg1-reporter* mediated by multiplex scFv thetering (Fig. 3.8.a). Upon 7 days of DOX treatment, I observed a stronger DNA methylation deposition when *3a3L*$^{-GFP-scFv}$ was combined with *KRAB*$^{-GFP-scFv}$ compared to single KRAB recruitment (Fig. 3.8.b). Importantly, CpG methylation returned back to the levels of KRAB-only when it was co-delivered with a catalytic inactive Dnmt3a3L (*Mut3a3L*$^{-GFP-scFv}$). Combination of *KRAB* with *3a3L* therefore has a synergistic effect on DNA methylation deposition.

However, single cell *Esg1-reporter* expression showed that the simultaneous recruitment of *3a3L* and *KRAB* did not have any additive effect on silencing after 7days of DOX treatment nor on memory of the repression after 7days of DOX washout (Fig. 3.8.c). On the contrary, the dual recruitment slightly reduced the extent of silencing induced by KRAB being tethered alone probably by interfering with the endogenous epigenetic machinery recruited on chromatin by KRAB.

Therefore, I hypothesised that in ESC, contrarily to somatic lines, KRAB is able to induce deposition of consistent levels of heterochromatin, including DNA methylation, per se, by recruitment of endogenous effectors and is able to completely turn off transcription. This luckily reflects the fact that ESCs, contrarily to somatic cells, express high levels of *de novo* DNMTs, that might explain the differences between this result and that reported by Amabile and co-workers (Okano et al., 1999; Okano, 1998). Thus, combinatorial recruitment of the two factors does not have any further synergistic effect on silencing nor on memory at this locus in ESC.

In summary, these data indicate that induction of a robust heterochromatic domain, and consequently extensive epigenetic silencing, is readily reversible from OFF>ON in an ESC model of early developmental stages. However, I found that DNA methylation and H3K9me3 is partially retained at the targeted locus 4 days after KRAB release, and this resulted in memory of the silenced epiallele in a fraction of the population. Thus, this slow rate of reversion may suggest a transient memory function (in effect a delay in reversion to ON).

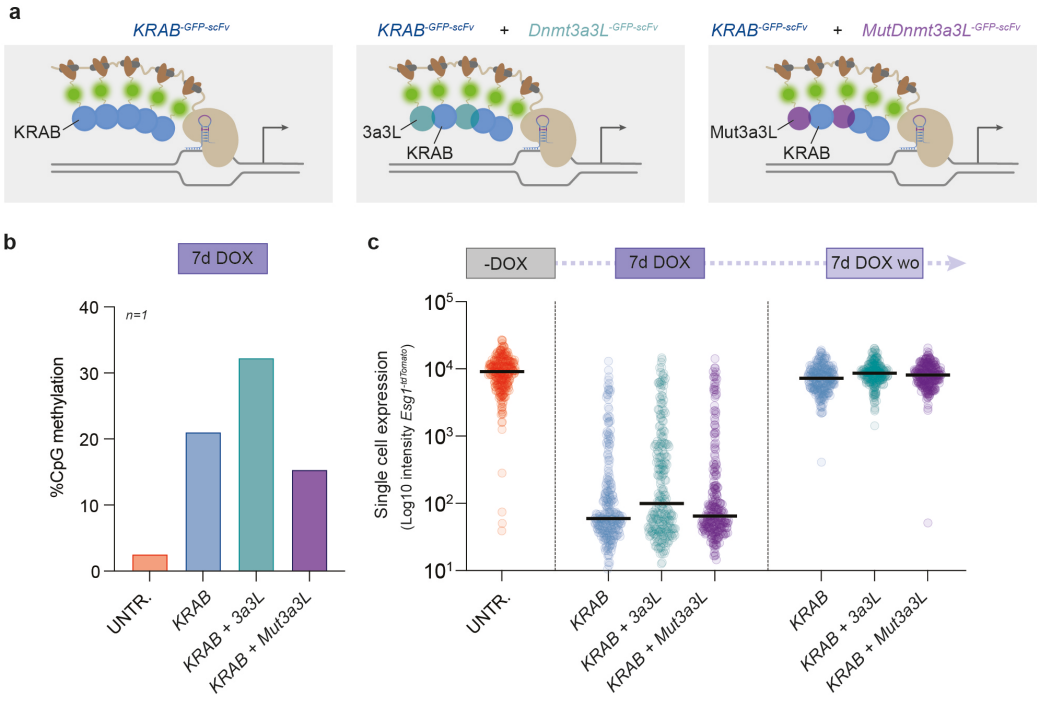


Fig 3.8. Combinatorial recruitment of Dnmt3a3L and KRAB does not have synergistic effect on Esg1-reporter silencing and memory (a) Schematics of the epigenetic tools used in the experiment. KRAB-GFP-scFv was either recruited alone or in combination with DNMT3 (3a3L-GFP-scFv) or the catalytically mutant Mut3a3L-GFP-scFv (b) Histogram plot shows the average of % DNA methylation at six CpG sites at Esg1 promoter. (c) Violin plots show the distribution of Esg1-tdTomato intensity (Log scale) in each condition (-DOX, 7days DOX and 7days DOX wo) in KRAB-GFP-scFv (blue), 3a3L-GFP-scFv (aqua green) and Mut3a3L-GFP-scFV (purple).

3.7 Induced heterochromatin mediates short term memory

In order to investigate whether deposition of the epigenetic marks is able to induce memory over a shorter time interval I compared the kinetics of KRAB-mediated epigenetic memory with a model of non-epigenetic induced silencing. For this purpose, I took advantage of the fact that *dCas9$^{-5xGCN4}$+ GFP^{-scFv}-only* at the TSS can impose a comparable level of silencing as *dCas9$^{-5xGCN4}$+ KRAB$^{-GFP-scFv}$* after 3days of DOX treatment (Fig. 3.9.a). This likely reflects steric hindrance exerted by the dCas9, which over the first few days has a greater repressive effect than heterochromatin per se. Nevertheless, in both cases equal gene repression is imposed but only cells transfected with *KRAB$^{-GFP-scFv}$* also acquire DNA methylation, H3K9me3, H3K20me3 and lose H3K4me3 (Fig. 3.9.b).

I induced equivalent silencing via either epigenetic (*KRAB$^{-GFP-scFv}$*) or non-epigenetic (*GFP^{-scFv}*) programming for 3 days. Cells that acquired quantitatively equal repression (TOM-) were flow sorted and cultured in the absence of doxycycline, to release the editing system. Additionally, to increase the speed of release of the epigenetic editing system, the *gRNA^{87dw}* was introduced only transiently so that it is diluted out with cell replications. This enables to exclude any confounding effects due to leakiness of the DOX promoter by essaying any memory function in the GFP-/BFP- fraction of the population (Supplementary Fig. 9.1).

Comparing the percentage of tdTomato silenced cells (in the GFP-/BFP- sub-fraction) over short time intervals, I observed that silencing is lost much faster in the *GFP^{-scFv}* control compared to epigenetic-induced repressed cells (*KRAB$^{-GFP-scFv}$*), where more than 50% of the cells remained silenced for more than 56 hours (Fig. 3.9.c). Importantly replication time was comparable in the two samples, excluding any confounding effect from replication dependent dilution of the marks (Fig. 3.9.d).

This data implies that induced heterochromatin has a short-term memory function that delays reactivation relative to an equivalently silenced reporter that does not carry DNA methylation, H3K9me3, H3K20me3 and loss of H3K4me3. Interestingly, my result is in accordance with a recent study in which, by proximity-dependent labelling of parental histones, they found that repressed but not active H3 domains were maintained for 48 hours by local redistribution of parental histones in mouse ESC (Escobar et al., 2019).

The loss of epigenetic repression is therefore time and cell division dependent and the following might be the contributing aspects: (i) factors that directly or indirectly safeguard epigenetic homeostasis and (ii) cell replication-dependent dilution of modified histones.

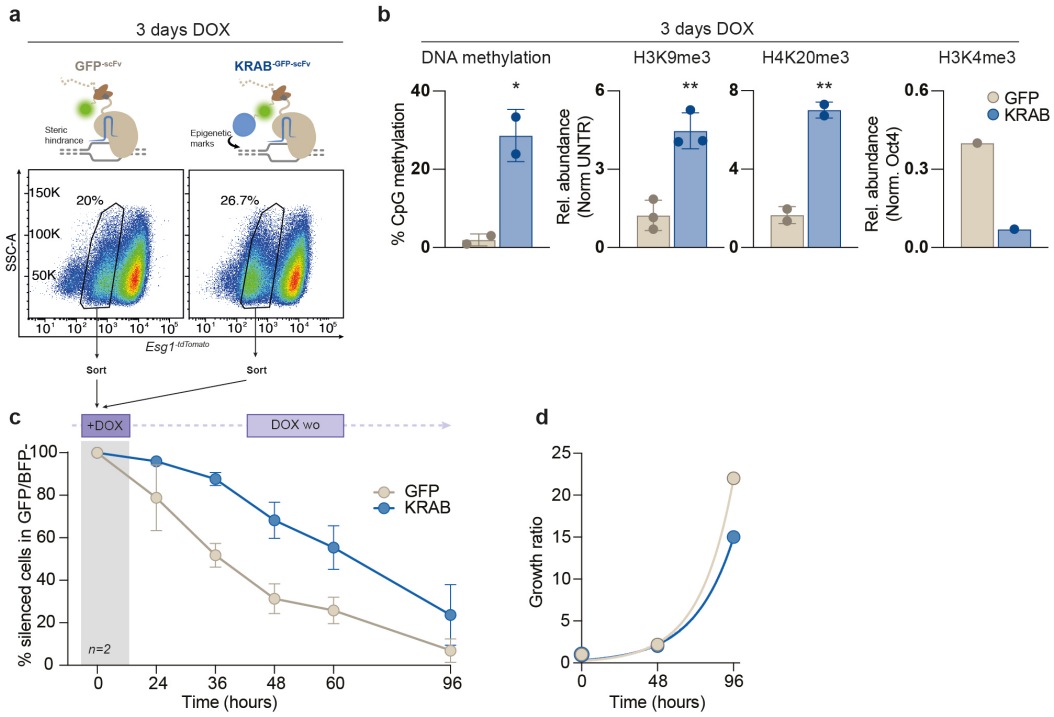


Fig 3.9. Induced heterochromatin mediates short term memory. (a) Density plots shows the intensity of Esg1$^{-dTomato}$ reporter plotted over side scatter area (SSC-A) by flow cytometry. Black boxes indicate the gates used to sort the cells with equal tdTomato repression after 3 days of DOX and subsequently used for the experiment (panel c). (b) Histograms shows % of CpG methylation or relative abundance nomalised to a positive control region and untransfected control for each of the marks indicated after 3 days of DOX. Statistics is calculated by one-tail unpaired t-test over two or three independent biological replicates (=p<0.05; **=p<0.01). (c) Time-curse of the % of silenced cells out of the GFP/BFP- population at 12 hours intervals up to 96 hours of DOX washout. Error is measured as standard deviation between two biological replicates. (d) Exponential curve shows growth obtained by measuring the ratio between the cell count at each timepoint over the initial number of cells at time 0. The difference observed in the growth between the two conditions does not account for the difference in epigenetic memory.*

3.8 Cell cycle inhibition slows down but does not block epigenetic erasure

In order to estimate the contribution of passive dilution of the epigenetic marks to the loss of inheritance that I observed, I treated the cells with a cell cycle inhibitor and analysed the reactivation kinetics of the reporter over short time intervals.

As described in the previous paragraph, cells were transiently transfected with the *gRNA^{87dw}*, treated with DOX for 3 days and KRAB-mediated silenced cells (TOM-) were flow sorted and put back in culture in the absence of doxycycline. 24 hours after DOX washout, cells were G2/M blocked by addition of the RO3066 inhibitor to the culture medium. Upon cell cycle blockage for further 24 hours, the cell cycle was

released by washing away the RO3306. I followed the reporter reactivation at short time intervals of 12 hours up to four days (Fig. 3.9.a). In the first phase of RO3306 treatment, despite cell cycle blockage, the percentage of tdTomato silenced cells underwent a ~20% decrease with a kinetics indistinguishable from cycling cells (Fig 3.9.b). Upon release of the cell cycle, after a total of 48 hours since DOX washout, treated cells underwent epiallele reactivation with a kinetics slightly slower compared to cells that were able to replicate the whole time. Thus, despite RO3306-treated cells going through a significantly smaller number of cell replications over four days (96 hours) (Fig. 3.9.c), both treated and untreated cells displayed a similar extent of epigenetic memory loss at the latest time point investigated.

Overall, I observed that inhibition of the cell cycle, partially slowed down, but didn't completely prevent the reporter reactivation, with a similar percentage of cells retaining reporter silencing after 4 days (96 hours) compared to untreated cells. This result suggests that other than cell replication dependent dilution of epigenetic marks, other 'active' factors contribute to erasure of epigenetic memory.

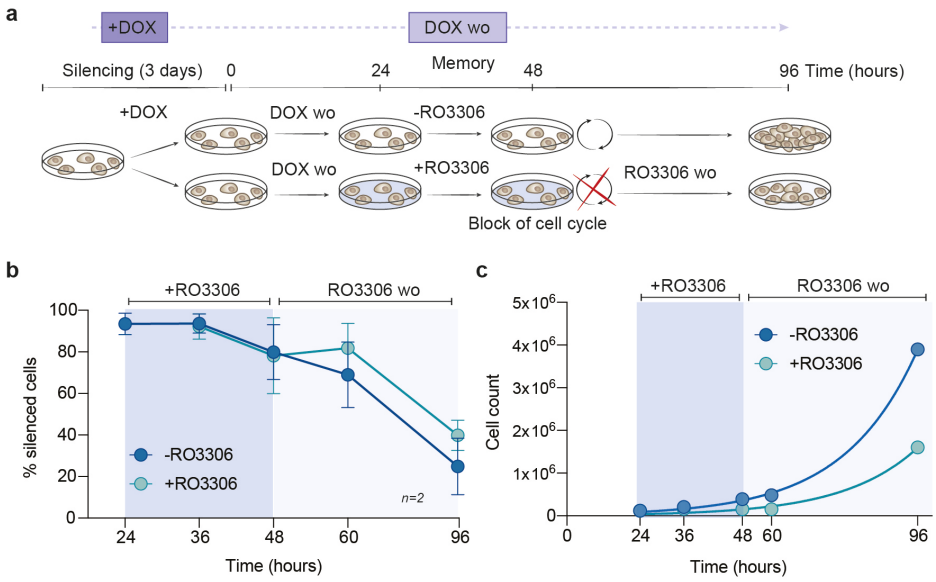


Fig 3.10. Cell cycle inhibition slows down but does not block epigenetic erasure. (a) Timeline of the experiment. briefly, cells are treated for three days with DOX to induce KRAB mediated silencing. TOM- cells are sorted (using the same gate as in Fig. 3.9.a) and plated back in culture in duplicates without DOX. After 24 hours of DOX washout one sample was treated with the cell cycle inhibitor RO3306 for further 24 hours. After a total of 48 hours from DOX washout the cell cycle was released and cells further analysed after 60 and 96 hours from release. In parallel cells were cultured in the absence of the inhibitor. (b) Time-course of the % of silenced cells in the + or -RO3306 conditions at every 12 hours intervals of DOX washout. Error is measured as standard deviation between two biological replicates. (c) Exponential curves represent cell count at each timepoint.

3.9 Discussion

Epialleles are defined as loci with the same DNA sequence that differ only in the epigenetic state within the same cell type and developmental stage. X-chromosome inactivation and imprinting are clear examples of epialleles that are stably transmitted in absence of the original signals in mammals. However, the extent to which epialleles can be inherited outside of endogenous developmental cues is less understood. Unravelling the potential for an epimutation to be maintained during the pluripotency window is of great interest since it could in principle be transmitted to all the downstream somatic lineages.

To study the memory of such epigenetic perturbations I optimised a releasable tool for ectopic deposition of heterochromatin on endogenous, otherwise active promoters, in pluripotent ESCs. By single gRNA tethering of a dCas9-multimerising system that recruits up to five KRAB effectors I obtained on-target deposition of a large heterochromatin domain of several kilobases in size. This domain is characterised by high levels of DNA methylation, H3K9me3, H4K20me3 and loss of H3K4me3. The strength of the heterochromatin domain obtained is indicative of the robustness of the dCas9-tethering system that I optimised. This is probably mediated by the employment of a modified gRNA scaffold that enhances occupancy time and on-target specificity and a docking platform to multimerise the binding of several effectors (Chen et al., 2013; Tanenbaum et al., 2014).On the other hand, the size of the domain achieved by just a single gRNA tethering reflects the fact that the ectopically deposited PTMs are able to recruit endogenous factors that compact chromatin supporting the spreading of such modifications though a read-and-write positive-feedback loop.

Despite big chromatin domains are thought to be eligible for transmission (Reinberg and Lynne, 2018), I observed that upon release of the triggering signal, the new epi-domain is not maintained through several cell replications and the locus reacquires the original epigenetic state. However, over a shorter period of DOX washout, I observed that, at the single cell level, the population showed a bimodal behaviour with some cells keeping memory of the repression and overall a partial retention of DNA methylation. This is not surprising since DNA methylation is the epigenetic modification with the most well characterised maintenance system.

Therefore, I investigated memory at a shorter temporal resolution. In order to do that, I exploited the fact that the sole tethering of GFP initially induced the same extent of repression compared to KRAB after 3 days of DOX induction but without any deposition of epigenetic marks. This gave me a valuable tool to compare the rate of reversal of silenced allele induced by only steric hindrance versus epigenetically mediated silencing. By analysing the reporter activity over short time points I

observed a much slower reactivation kinetics in cells that carried the epigenetically-silenced epiallele compared to non-epigenetically silenced cells. Interestingly, this data shows that the deposited heterochromatin has a short-term memory of roughly 48 hours.

Inheritance of chromatin states across DNA replication entails that parental histones and their PTMs must segregate to the same domain on daughter DNA strands and that the 'read and write' mechanisms must be engaged to restore appropriate levels of the mark on the newly incorporated histones. By proximity labelling of parental histones in ESCs it was recently shown that parental nucleosomes comprising repressive chromatin states are re-deposited locally and that the label was lost within 48 hours (Escobar et al., 2019). This study is in accordance with the extent of epiallele memory described with my system.

To understand the contribution of cell replication dependent dilution of the marks in the observed loss of memory, I used the cell cycle inhibitor RO3306 that blocks cells in the G2/M phase. Upon blockage for a short period, the cell cycle was released and I followed the reporter reactivation at short time intervals up to four days. Overall this experiment revealed a similar reactivation kinetics in the paused cells compared to non-paused cells. This suggests that heterochromatin loss is not completely a passive event and that other mechanisms might be involved in actively reverting the new epiallele to its unperturbed state.

Although it might seem counterintuitive that an epigenetic domain which is thought to self-maintain is rapidly lost, ES cells might have a 'safety' mechanism to prevent unauthorised inheritance of heterochromatin.

Such a mechanism has been previously reported in yeast where, by a releasable tethering of the histone methyltransferase *Clr4*, the authors found that targeted H3K9 methylation is actively removed upon release of the trigger. They also found that the responsible for this active removal was the putative histone demethylase *Epe1*. In fact, upon its deletion, ectopic H3K9 methylation was inherited through mitosis and even meiosis (Audergon, 2015 Ragunathan et al., 2015; Wang, 2017) and, revealing that *Epe1* protects the epigenome from inheriting altered epigenetic states.

Does a similar surveyor factor also exist in mouse ESC?

4 Identification of factors surveilling epigenetic memory in ESC

4.1 Introduction

Despite the extent of the introduced heterochromatin domain, I found that memory of the epi-mutations is lost over time, but cell-replication dependent dilution of the marks does not entirely explain the kinetics of reversion observed. I therefore speculated that there might be a “surveyor factor” for epigenetic homeostasis in ESC to maintain a “naïve” epigenome.

In order to unbiasedly identify such factor(s) that prevents the passage of epigenetic memory through replications in my system I applied a knockout screen. This is a powerful approach that enables to assess the functional role of genes through loss-of-function (LOF) genetics. This strategy exploits gRNAs to direct the binding of the Cas9 nuclease, which introduces double-strand-breaks (DSBs), in the DNA. When targeted on exon sequences this leads to DNA-repair induced frameshift-indels which most luckily disrupt the gene function. When applied at a genome-wide scale by using libraries of gRNAs this approach allows to interrogate the role of nearly every gene in the genome in the phenotype of interest (Hsu et al., 2014 Joung et al., 2017); (Shalem et al., 2015; Doench et al., 2016; (Koike-Yusa et al., 2014). The genes linked with the phenotype are identified by counting gRNAs (*knockout*) representation in the population of cells selected for that given phenotype, via next generation sequencing (NGS).

My epigenetic editing system coupled with the endogenous epigenetic memory reporter allows me to distinguish at the single cell level the population of cells that keep memory from those that revert back, using tdTomato fluorescence as a proxy.

Applying genome-wide genetic perturbation to this epigenetic editing assay, the aim of this chapter is therefore to identify the factor(s) that counteract epigenetic memory in ESC, as their knockout will be enriched in the population of cells that will keep an OFF transcriptional state.

4.2 Unbiased CRISPR screen reveals factors that safeguard epigenetic homeostasis in ESC

In order to identify the regulators of epigenetic memory in my model of epigenetic perturbation of endogenous chromatin I performed an unbiased genome wide CRISPR knockout screen. Schematics of the experimental design is shown in Fig. 4.1.a. I introduced a single-copy (nuclease active) Cas9 (wtCas9) fused to GFP and separated by a T2A (*Cas9*$^{-T2A-GFP}$) into the *Esg1*$^{-tdTomato}$ reporter cell line already carrying the *dCas9*$^{-5xGCN4}$ in the off state (*(off) dCas9*$^{-5xGCN4}$) and subsequently infected these cells with a pooled exon-targeting gRNA lentiviral library. After selection of the transduced cells and appropriate time for the knock out to be generated, the wtCas9 was inactivated by transiently transfecting the cells with a pair of gRNAs that recognize specific sequences in the active but not the catalytically inactive form of the Cas9 to induce its specific self-inactivation. Cells which successfully inactivated the *Cas9*$^{-T2A-GFP}$ were flow sorted according to loss of GFP (GFP-). The cell population is now composed of a heterogeneous pool of knockouts which only contains the *dCas9*$^{-5xGCN4}$ and have inactivated the *Cas9*$^{-T2A-GFP}$.

Next, I transfected these GFP- cells with *KRAB*$^{-GFP-scFv}$ and the *Esg1_ gRNA*^{87dw} and I turned on the epigenetic-editing system by DOX treatment. After seven days I isolated by FACS the cells that successfully switched off transcription and therefore acquired the silenced epiallele according to repression of tdTomato. These TOM- cells were subsequently cultured in absence of DOX to release *KRAB*$^{-GFP-scFv}$ tethering. After four or seven days from DOX withdrawal I flow sorted the population of cells that successfully retained the epigenetic-silenced state of the reporter (TOM-), which are predicted to be enriched for knockout of genes that delay (at four days) or completely block (at 7 days) epigenetic reversion. I additionally collected the cells that fully reverted to the ON state (TOM+) as control. Out of the TOM- I separated

two parallel populations: a wider or more stringent gate (bottom 2.5% TOM- population) was set to be able to identify factors that have a milder or stronger effect on memory reversal (Supplementary Fig. 9.2.a and b).

Finally, to identify the genes involved in antagonizing epigenetic memory I compared the gRNA frequency (indicative of frequency of KO in population) in TOM- populations with the complementary TOM+ population. Among the factors with false discovery rate (FDR) < 0.2, used to identify putative candidates, most have a role in transcriptional, translational or post translational regulation, with the latter potentially involved in deregulating the reporter expression downstream of transcription. Amongst the hits with the lowest FDR there are many factors with a well-known or predicted epigenetic function. I plotted the -log2 of the relative ranking algorithm (RRA) generated by our Model-based Analysis of Genome-wide CRISPR- Cas9 Knockout (MAGeCK) (Li et al., 2014) to identify significant genes, based on multiple targeting sgRNAs shifting their distribution (Fig 4.1.b and c and Supplementary Fig. 9.2.c and d).

Among the candidates with a well-characterised epigenetic activity there are the SWI/SNF histone remodeller factor *Smarcc1* (FDR 0.03), the H3K79 methyl-transferase *Dot1L* (FDR 0.16) and the H3K4 specific histone methyltransferase *Kmt2d* (FDR 0.13). This latter is enriched in both smaller and wider gates used but only at the earlier time point (Fig. 4.1.b and Supplementary Fig. 9.2.c).

Interestingly, the screen analysis revealed also candidates which have been linked with epigenetics, but not extensively unravelled yet. For example, the developmental pluripotency associated factor 2 (*Dppa2*) KO is enriched in both the smaller and the bigger gate with FDR 0.07 and 0.03 respectively but only in the long-term memory time point (Fig. 4.1.c and Supplementary Fig. 9.2.d). Other examples are the zinc finger MYM-type 3 (*Zmym3*) (FDR 0.03) and the KAT8 Regulatory NSL Complex Subunit 2 (*Kansl2*) (FDR 0.05).

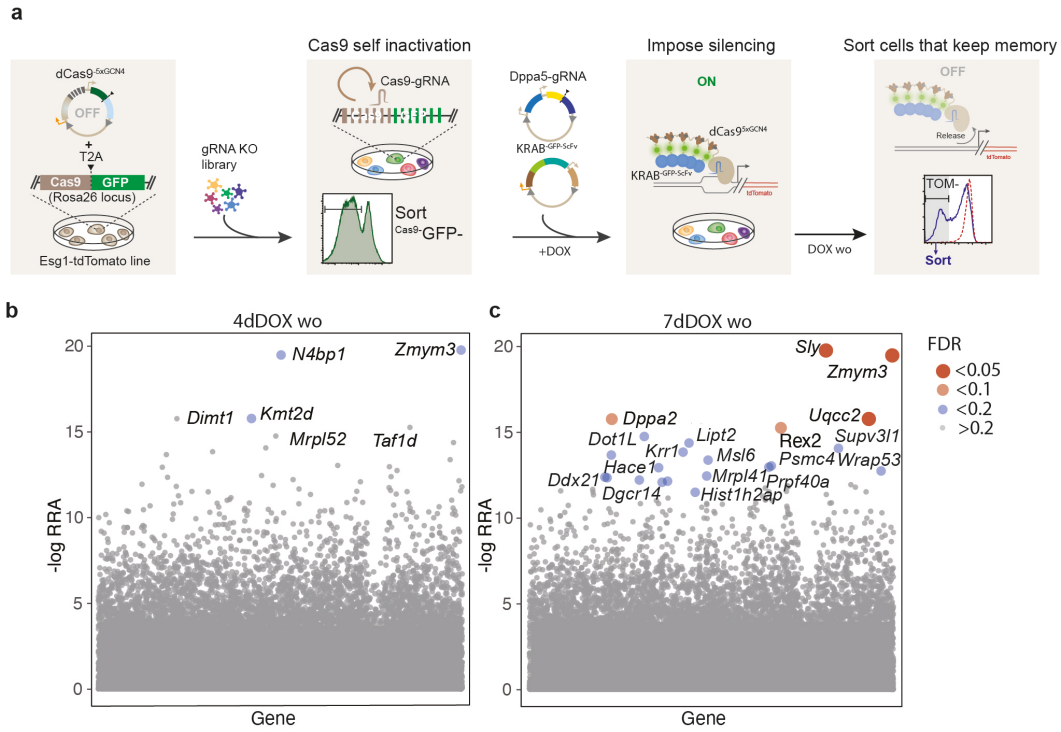


Fig 4.1. Unbiased CRISPR screen reveals factors involved in safeguarding against epigenetic memory. (a) Schematic representation of the workflow: Esg1$^{-tdTomato}$ reporter cell line carrying constitutive Cas9 -GFP and DOX inducible (off)-dCas9$^{-5xGCN4}$ has been infected with a gRNA lentiviral library. Self-inactivation of Cas9 has been achieved with a pair of gRNAs specific for its sequence and GFP negative cells have been isolated. Subsequent to the introduction of the gRNA^{-BFP} and KRAB$^{-GFP-ScFv}$ cells have been treated with DOX for 4 or 7 days. Ultimately, cells that retained the silenced epialleles have been sorted based on TOM- expression and DNA extracted and subjected to NGS. (b) and (c). Scatterplots show significant hits from the screen analysed by the -log relative ranking algorithm (RRA) generated by the Model-based Analysis of Genome-wide CRISPR Cas9 Knockout (MAGeK) for each gate used and each given timepoint. Size and color of the dots represent False Discovery Rate (FDR) as shown in the legend on the right.

4.3 Validation of the CRISPR screen candidates

In order to test the role of the candidates identified in the screen in antagonising epigenetic memory in pluripotent cells I generated clonal knockout ESC lines of each candidate and performed my assay for targeted epigenetic silencing on the *Esg1*$^{-tdTomato}$ reporter. *Dppa2*$^{-/-}$, *Kmt2d*$^{-/-}$, *Smarcc1*$^{-/-}$, *Dot1L*$^{-/-}$, *Zmym3*$^{-/-}$ ESC have been successfully generated at least in duplicates and clonally amplified.

At first, I assayed the reporter expression at the single cell levels in all knockout generated at several time points, prior and after DOX treatment (Fig. 4.2.a and b). Importantly, the deletion of the genes did not affect the basal expression of the reporter (-DOX condition), except in *Zmym3*$^{-/-}$ clones in which I observed a shift in *Esg1*$^{-tdTomato}$ expression. For the remaining, this implies gene deletion does not affect the reporter prior to induced epigenetic silencing. Moreover, upon DOX treatment for 7 days the reporter was efficiently repressed in all the KOs with roughly the same extent as the WT cells. Following removal of the triggering signal, *Smarcc1*$^{-/-}$ and *Dot1L*$^{-/-}$ reverted back to the active state with a similar kinetics to the WT control, suggesting these might be false hits from the screen. On the other hand, consistently with the result of the screen, in which *Kmt2d* KO was enriched only in the early time-point (Fig. 4.1.b), *Kmt2d*$^{-/-}$ cells showed full memory at 4 days of DOX washout while reverted almost to a fully ON state after 7 days (Fig. 4.2.a), over three independent clones (Fig. 4.2.b). This suggests that, counteracting rapid re-deposition of H3K4me3 by removal of the main H3K4 methyltransferase, changes the dynamics but does not completely restrict epiallele erasure, indicating redundant pathways are later engaged.

Of note *Zmym3*$^{-/-}$ cells show an intermediate behaviour with most of the cells that keep full memory at the earlier time-point of DOX removal but are completely reactivated after seven days (Fig. 4.2.a). However, I observed great variation between different clonal lines (Fig. 4.2.b) thus, although the role of ZMYM3 appears potentially interesting, it was not further investigated.

Remarkably, *Dppa2*$^{-/-}$ fully maintains epigenetic memory at the earlier time point. At the later time-point, it exhibits a bimodal distribution, with the majority of the cells still in the OFF state (Fig. 4.2.a). The same phenotype was observed in three independent clonal lines (Fig. 4.2.b).

In summary I found that KMT2D has a role in partially erasing epigenetic memory only at short time investigated, whilst DPPA2 completely counteracts memory of induced heterochromatin in pluripotent cells, for at least one week and multiple cell replications (Fig. 4.2.c).

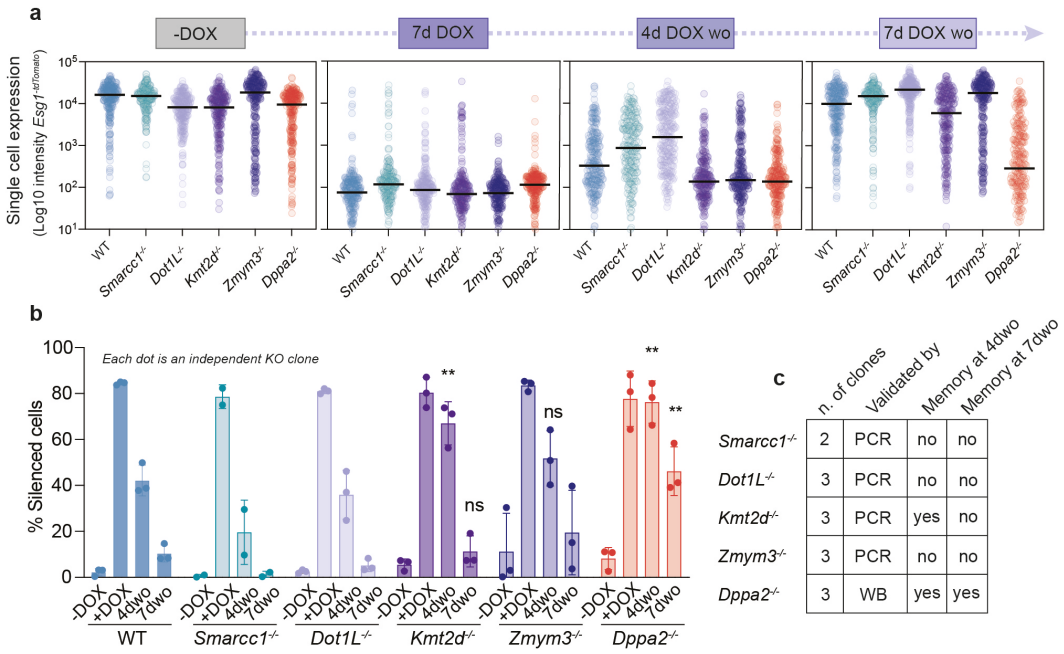


*Fig. 4.2. Validation of the CRISPR-screen candidates identifies Dppa2 and Kmt2d as putative regulators of epigenetic memory (a) Violin plots indicate expression of tdTomato reporter at the single cell level detected by flow cytometry at the timepoints indicated, in WT or individual knockout of the candidates enriched in the screen. Bars represent the median of fluorescence intensity in the population of cells. (b) Histograms show the percentage of Esg1 -tdTomato negative cells in WT or knockout cells at each individual timepoint. Error bars are standard deviation measured out of two or three independent knockout clones. Statistics is measured by one-tail unpaired t-test comparing knockout and WT (**=p<0.01). (c) Summary table.*

4.4 *Dppa2*$^{-/-}$ cells show a bimodal memory across extensive cell replication.

DPPA2 is usually found as a heterodimer with DPPA4 and these two proteins are expressed almost exclusively in the early embryos and the germline (Eckersley-Maslin et al., 2020). Very recently it has been shown that DPPA2-4 focally counteract aberrant de novo DNA methylation (Gretarsson and Hackett, 2020). Therefore, I investigated whether DPPA2 might be also involved in counteracting epigenetic memory in ESC.

I performed my epigenetic editing assay using three independent *Dppa2*$^{-/-}$ clonal lines, validated by western blot (Fig. 4.3.a), and three *wildtypes*. Importantly, knocking out *Dppa2* does not induce any significant deviation in *Esg1*$^{-tdTomato}$ expression from the WT demonstrating that *Dppa2* disruption does not affect the reporter expression per se (Fig. 4.2.a and 4.2.b). To confirm that the silenced epiallele can be inherited in absence of DPPA2 I performed time-course flow cytometry analysis of the *Esg1*$^{-tdTomato}$ reporter (Fig. 4.3.b). As already shown before (Fig. 4.2.a and b) WT cells rapidly lose memory, being completely reactivated by seven days, but at this time-point around 60% of *Dppa2*$^{-/-}$cells are still fully silenced (Fig. 4.3.b and c). Intriguingly, memory of the silenced state is maintained in a consistent fraction of the *Dppa2*$^{-/-}$ cells for 43 days after DOX washout and numerous cell replications, showing a bimodal distribution (Fig. 4.3.c). In addition, sorting the TOM- and TOM+ fractions out of this bimodal population after 13 days of DOX washout and culturing them separately, revealed that, while the TOM+ fraction remained positive, the TOM- started reacquiring the bimodal distribution after 8 days from sorting, suggesting a stochastic memory function (Fig. 4.3.d). Importantly *Dppa2 knockout* cells were phenotypically indistinguishable from WT cells and had the same replication rate (Fig. 4.3.e), in fact, pluripotency and self-renewal is not affected in this knockout (Eckersley-Maslin et al., 2020).

Taken together these data show that memory of introduced epimutations is not maintained in replicating ESC except when the *Dppa2* gene is knocked out. This suggests that DPPA2 might have a “surveyor” function to antagonise epigenetic perturbations during pluripotency.

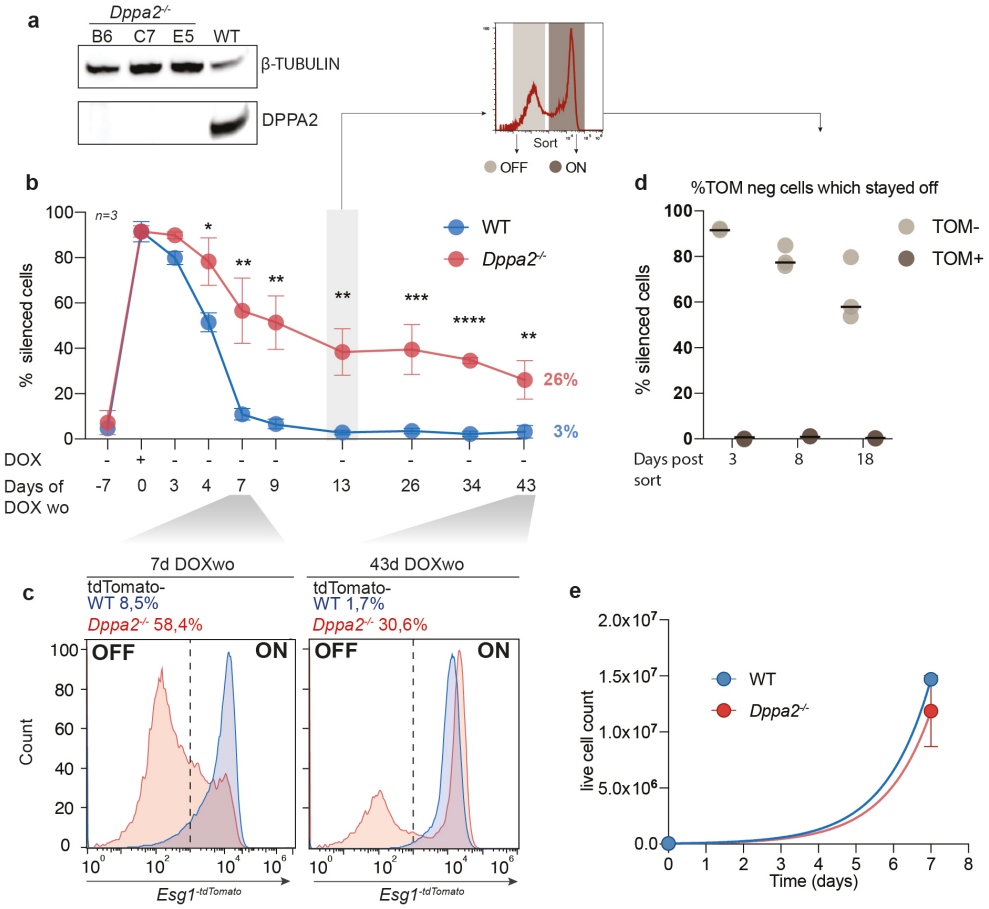


Fig 4.3. Dppa2$^{-/-}$ exhibits a bimodal memory across 43 days and extensive cell replication. (a) Western blot comparing DPPA2 protein in WT and knockout clones. β-TUBULIN is used as loading control. (b) Time-curse of % of silenced cells in Dppa2$^{-/-}$ vs WT at the timepoints indicated. Error bars are standard deviation measured out of three biological replicates. Statistics is measured by one-tail unpaired t-test between WT and knockout (=p<0.05; **=p<0.01; ***=p<0.001; ****=p<0.0001). Percentage on the right indicate the average fraction of silenced cells in knockout (red) or WT (blue) after 43 days of DOX washout. Grey box highlights the timepoint when cells have been sorted for the experiment as shown in panel d. (c) FACS histograms of tdTomato expression (logarithmic scale) in Dppa2 $^{-/-}$ (red) or WT (blue). Numbers on the top indicates percentage of tdTomato negative cells in each condition. (d) Percentage of silenced cells upon sorting of TOM- and TOM+ fraction of cells after 13 days of DOX wo and analysed at the timepoints indicated on the x axis. Bar indicates the media out of three biological replicates. (e) Exponential curves show the number of live cells at the times indicated.*

4.5 Epialleles are propagated in absence of DPPA2

To investigate which epigenetic marks are retained in *Dppa2*$^{-/-}$ cells I characterised the reporter's epigenetic state prior and after KRAB tethering. Pyrosequencing revealed that ablating *Dppa2* does not initially affect the deposition of DNA methylation by the epigenome editing system after 7 days of DOX treatment (Fig. 4.4.a). However, it shows that *Dppa2*$^{-/-}$ cells sorted for TOM- after 13 days of DOX removal (13dwo) significantly maintain most of the DNA methylation deposited while WT cells completely lose it.

On the other hand, CUT&RUN-qPCR shows that H3K9me3 is deposited both in WT and *Dppa2*$^{-/-}$ with similar extents at 7 days of DOX treatment (Fig. 4.4.b), while endogenous H3K4me3 is erased in both conditions at this time point (Fig 4.4.c). This data confirms that *Dppa2* deletion does not affect the deposition of the ectopic marks induced by the epigenetic editing system. As already shown in the previous chapter (Fig. 3.6 and 3.7), upon release of *KRAB*$^{-GFP-scFv}$, WT cells undergo a complete recovery of epigenetic state: H3K9me3 is lost and H3K4me3 is re-deposited (Fig. 4.4.b and c). Similarly, *Dppa2*$^{-/-}$ cells that keep the epiallele repressed (sorted for TOM minus) completely lose H3K9me3 after 13 days of DOX washout (Fig. 4.4.b). On the other hand, differently from WT, which reverts back to endogenous levels of H3K4me3, this mark is not reimpose in knockout cells (Fig. 4.4.c). Overall, whilst *Dppa2*$^{-/-}$ cells revert induced H3K9me3, they retain programmed DNA methylation and the absence of H3K4me3 which appears sufficient to maintain a silent transcriptional state at the reporter gene upon release of the triggering system.

This altered epigenetic state is also reflected by a reduction in chromatin compaction observed by ATAC-seq in the *Dppa2*$^{-/-}$ cells compared to WT cells after 13 days from release of KRAB (Fig. 4.4.d). Overall ATAC-seq correlation plot shows that deletion of Dppa2 does not alter the chromatin landscape outside of the epigenetically perturbed loci, with *Esg1*-reporter being among the most differentially accessible genes in *Dppa2*$^{-/-}$ vs WT (Fig. 4.4.e).

Taken together these data demonstrate that *Dppa2* knock out does not alter the epigenetic state of the cells in normal conditions. However, once epigenetic perturbation occurs, cells partially inherit these new epialleles. Interestingly this occurs probabilistically with a portion of *Dppa2*$^{-/-}$ cells reverting but the majority maintaining epigenetic memory. This implies the absence of DPPA2 shifts the balance between erasure and memory of induced epigenetic states towards memory, albeit incompletely. Overall, DPPA2 therefore acts as a surveyor for epigenetic states in ESC.

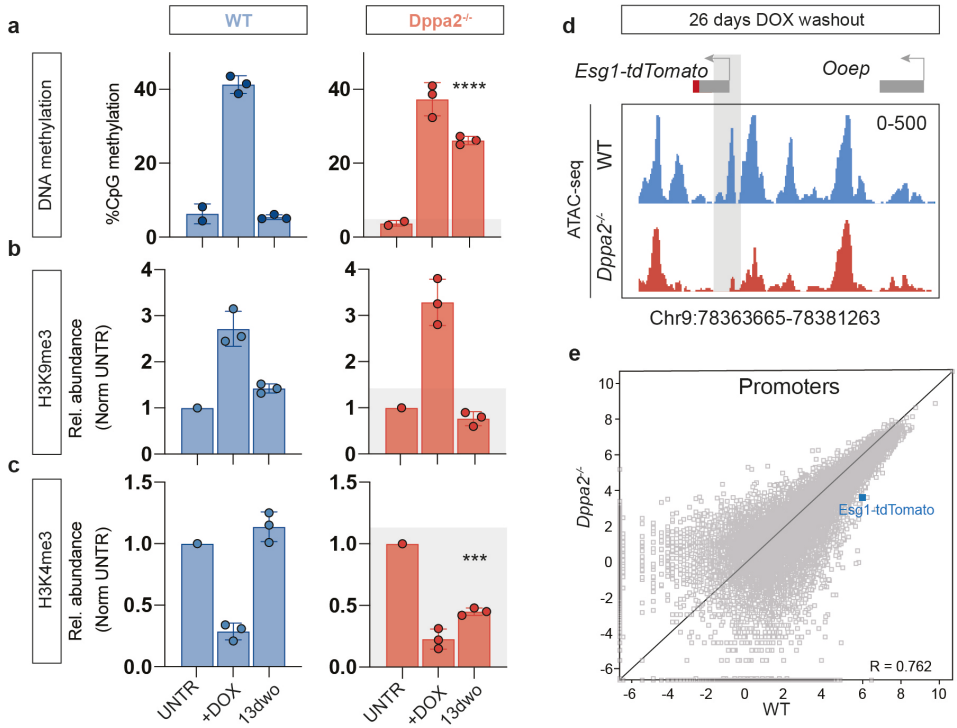


*Fig 4.4. Epialleles are propagated in absence of DPPA2 (a), (b), (c) Histograms represent %CpG methylation or abundance of histone marks relative to a positive region and untransfeced in WT (blue) or Dppa2$^{-/-}$ (red) cells transfected with KRAB$^{-GFP-scFv}$. Error bars represent standard deviation measured out of three independent clones. Statistics is measured by one-tail unpaired t-test between WT and knockout at the same timepoint (***=p<0.001; ****=p<0.0001). Grey box indicates the level of each mark in the WT sample at 13dwo. (d) ATAC-seq tracks showing chromatin accessibility at the Esg1$^{-tdTomato}$ locus after 26 days of DOX washout. Grey box highlights the position of the Esg1 promoter. (e) Correlation plot comparing accessibility of promoters in Dppa2$^{-/-}$ vs WT after 26 days of DOX washout.*

4.6 DPPA2 focally surveys for epigenetic perturbations at a subset of genes.

To understand wether DPPA2 has a focal function in regulating memory specifically at the *Esg1*-promoter or rather a global activity, I programmed heterochromatin-mediated silencing of a subset of targets and investigated the stability of the induced-repressive state upon release of the trigger. I selected targets according to their expression in ESCs and CpG density of their promoters, since DPPA2 is known to bind expecially CG rich sequences (Gretarsson and Hackett, 2020).

Among 33 regions initially tested, I followed up memory in *wildtype* and *Dppa2*$^{-/-}$ cells only of those targets which underwent a significant extent of silencing after tethering *KRAB*$^{-GFP-scFv}$ in the first place, judged by qPCR expression. Out of these, most of the regions at which I observed good level of silencing after DOX treatment are imprinting, tumor suppressors, CG+ promoters and the *Igll1* enhancer. Amidst these targets, I observed that most of the imprinting promoters (*Peg3, Mest* and *Plagl1*) and the *Igll1*-enhancer show almost complete memory even in the WT with a similar extent in the *Dppa2*$^{-/-}$(Fig. 4.5.a, c and e). This result is in line with the notion that the epigenetic environment at imprinting control regions is favourable to mantain memory of past epigenetic states (Barlow and Bartolomei, 2014), regardless of DPPA2 activity. Intringuily, *Jade1* (also known as *Phf17*) is an exception, and while it completely looses memory in *wildtype,* it significanlty retain silencing only in *Dppa2*$^{-/-}$. Interestingly, a differentially methylated region has only recently been mapped to *Jade1* promoter that is thus considered a candidate new ICR (Gigante et al., 2019).

At the other regions tested, as the tumor suppressor genes *Pten* and *Cdh1* and the GC+ promoters of *Adamts7* and *Greb1* genes, I could observe a trend in which *Dppa2*$^{-/-}$ cells always retain a sligth, although not significant, extent of memory compared to the *wildtype*. This trend in the average expression of the targets in the bulk population suggests two scenarios: either the expression level is representative of the whole population or, at the single cell level the population shows a bimodal behaviour with a small fraction retaining full memory masked by another subpopulation of the cells that revert back. In support of this latter hypothesis, I already observed a stochastic memory function in *Dppa2*$^{-/-}$ at the *Esg1*-reporter promoter when analysed at the single cell level by flow cytometry, with a fraction of the cells mantaining full memory and another portion of the population reverting back to the original expression state (Fig. 4.3.c). Furthermore, I observed that, among two biological replicates, *Greb1* showed high variability in terms of memory retention after 7d DOX washout, suggesting a partial effect of the *Dppa2*$^{-/-}$ on memory in at least one clone at this target.

To understand whether retention or loss of memory observed correlated with binding of DPPA2 at the targeted region I performed CUT&RUN-sequencing in *wildtype* cells against DPPA2 to ask where it is enriched. As shown in the sequencing traks in Fig 4.5.b,d and f, DPPA2 binding peaks at most of the regions investigated, but with different strenght showing only very mild binding at the *Igll1* enhancer, *Greb1* promoter and *Jade1* DMR, among which, these latter two are the targets that showed the strongest memory only in *Dppa2* knockout.

Taken together, these data suggest that *Dppa2* might not be a global surveyor of epigenetic memory buth rather control for epigenetic perturbations with different extents according to the targets. However the mechanism by which it exerts this surveyor function is not clear yet. My initial hypotesis was that DPPA2 migth survey for memory at binding sites. However, I observed that targets showing the stongest DPPA2 binding only partially ratain a silenced epiallele in *Dppa2*$^{-/-}$. This results nearly suggest an anticorrelation between DPPA2 binding and memory observed. In support of this, DPPA2 shows only a mild binding also at the *Esg1* promoter (Supplementary Fig.9.3.a), which was used in the screening that identified *Dppa2* as a regulator of epigenetic memory. Therefore the mechanism by which DPPA2 act still need to be unravelled.

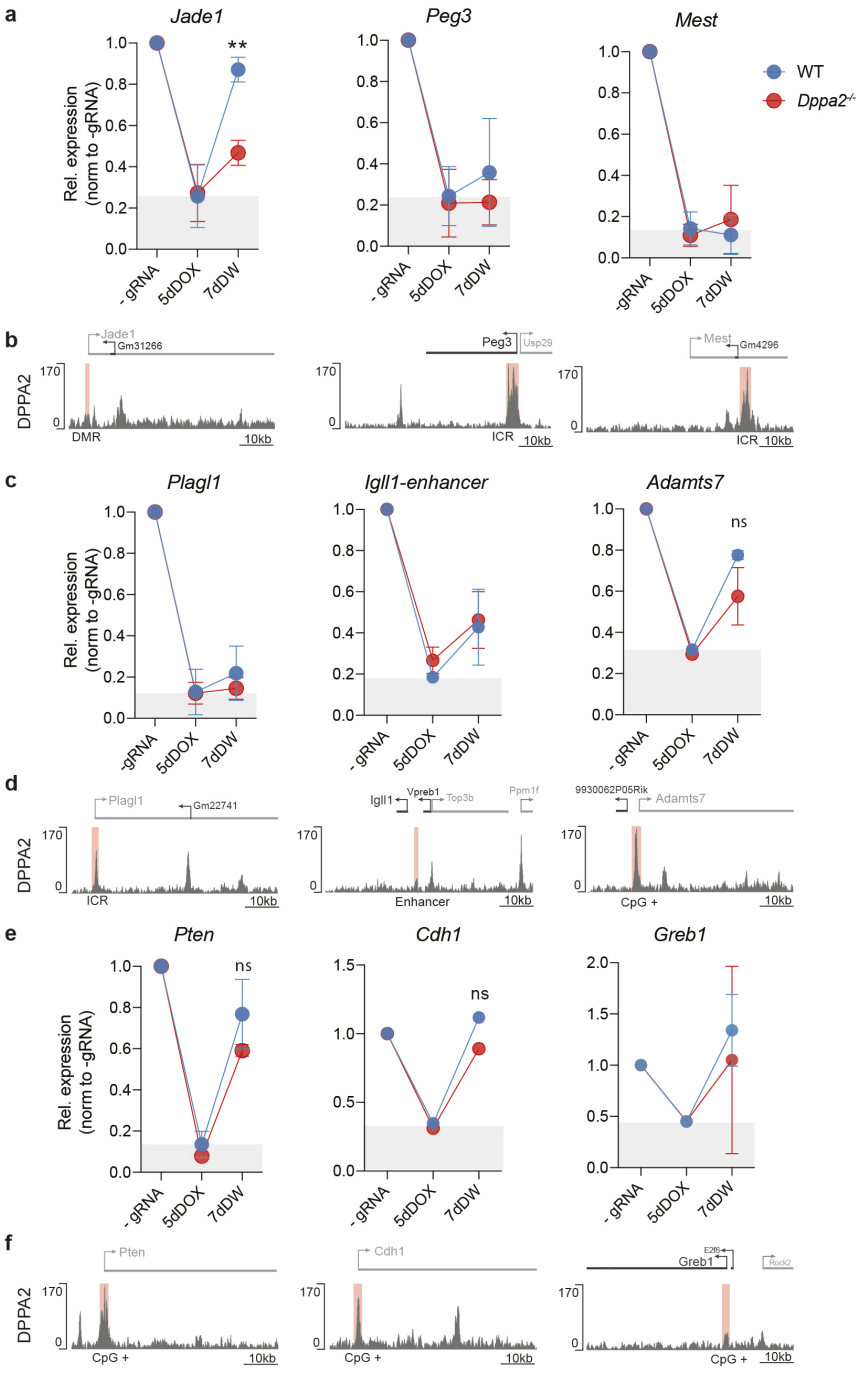


*Fig 4.5. DPPA2 focally surveys for epigenetic perturbations at a subset of genes. (a), (c), (e) Line-plots show expression of the indicated targets relative to the housekeeping gene Rplp0 and normalised to -gRNA control for each timepoint. Grey boxes represent the level of repression achieved after DOX treatment in the wildtype condition. Error bars are standard deviation out of two or three biological replicates. Statistics is measured by one-tail unpaired t-test between WT and knockout conditions (**=p<0.01). (b), (d), (f) Chromosome tracks represents the enrichment of DPPA2 in wildtype cells estimated by CUT&RUN. Each window is 60 kb wide. Red boxes highlight the position of each indicated feature.*

4.7 Discussion

My screen for factors erasing ectopically deposited epigenetic domains revealed three possible interesting candidates.

First, ZMYM3 is a transcriptional repressor, encoded on the X chromosome, which binds to DNA via its zinc finger domain. Very little is known about this protein and some insights into its role might be inferred from its paralogs. For example, ZMYM2 recruits the binding of other chromatin regulators on DNA such as the LSD1–CoREST–HDAC1 complex (Gocke and Yu, 2008). Analogously, also ZMYM3 might be involved in a similar role in ESC since it was shown to interact with the histone demethylase LSD1 in mouse spermatogonial stem cells (Hu et al., 2017). Although in this study they didn't show any effect of the absence of ZMYM3 on global H3K4 methylation genome wide by western blot, this finding does not exclude that ZMYM3 might affect H3K4 status at specific binding loci. Thus, in absence of ZMYM3, altered levels of H3K4me3 might be responsible for the observed maintenance of the silenced state.

Nonetheless, to validate this candidate I generated numerous Zmym3 knockout clonal cell lines and observed high inter-clonal variability. Some of these *Zmym3*$^{-/-}$ clones showed an impairment of expression of the reporter prior to epigenetic editing. This suggests that memory observed in these clones is not due to direct reversal of epigenetic state upon the epigenetic editing trigger is released but to an altered pre-existing epigenetic state.

Second, *Kmt2d*$^{-/-}$ is enriched exclusively in the early time-point screen suggesting that this protein might have a role in slowing down the reactivation kinetics of the silenced epiallele. This result is in accordance with the validation experiment in which I found that three independent clonal *Kmt2d*$^{-/-}$ lines showed full memory at 4 days of DOX washout while reverted almost to a fully ON state after 7 days. KMT2D (aka MLL2 or MLL4) is an H3K4 specific histone methyl-transferase and its deletion might impede rapid re-deposition of endogenous level of H3K4me3 subsequent to its erasure induced by epigenetic editing. However, the rapid reversal of the epigenetic state at the later time point suggests that other redundant H3K4 methyltransferases are engaged with time. Indeed, KMT2D has a partial functional redundancy with KMT2C (aka MML3) in mammalian cells, as co-deletion of both is required for global decrease of H3K4me1 (Lee et al., 2013).

Third, DPPA2 is a small DNA binding protein initially identified in a screen for pluripotency factors (Bortvin et al., 2003). This protein lacks any known catalytic activity but it can bind to DNA via its SAP domain, although a binding motif has not been found yet. In addition, it can associate with core histone H3 via its C-terminal domain and it is generally found on euchromatin (Madan et al., 2009; Masaki et al.,

2010). It is predominantly observed as a heterodimer with Dppa4 and this association leads to reciprocal stability, as single knockout induces a reduction in the amount of both proteins by western blot (Gretarsson and Hackett, 2020). However, Dppa4 was not among the hits of my screen, possibly suggesting that, due to the complexity of the workflow, the screen was not at saturation and I might have lost potential hits.

Candidate validation, using three independent *Dppa2*$^{-/-}$ ESC clonal lines, revealed that, not only do *knockout* cells maintain full memory in more than half of the population after 7 days of DOX washout, but also that the epiallele is inherited in a significant percentage of cells at least for 43 days and many cell replications. Importantly, differently from *Zmym3 knockout*, deleting *Dppa2* did not affect the reporter's expression prior to epigenetic editing, suggesting that *Dppa2* has a role specifically in counteracting epigenetic perturbations for a significant time.

By analysis of the chromatin state, I did not observe significant differences in the extent of DNA methylation and H3K9me3 deposition nor H3K4me3 loss in *Dppa2*$^{-/-}$ compared to *wildtype* cells upon epigenetic editing. This further indicates that DPPA2 deletion per se does not alter the epigenetic state of this promoter. After release of KRAB, I found that, *Dppa2*$^{-/-}$ sorted TOM- cells, differently from *wildtype*, keep most of the DNA methylation deposited. Instead, H3K9me3 is equally lost both in *wildtype* and *knockou*t cells implying that this mark is not required for keeping a silenced state at this promoter. Interestingly, levels of H3K4me3 remain low in absence of DPPA2 even after 13 days of KRAB disengagement in TOM- cells. Overall, at this time point chromatin is completely compacted in *Dppa2*$^{-/-}$ cells as revealed by ATAC-seq, and importantly this is specific for the *Esg1* promoter showing that Dppa2 knockout does not have major effects outside of the perturbed epiallele.

In summary, by characterising the silenced epiallele in *Dppa2 knockout* cells, I found that it carries memory exclusively of the past DNA methylation and H3K4me3 levels, while the deposited H3K9me3 is erased.

Other than its role in pluripotency, DPPA2/4 is recently emerging as an important regulator of bivalency controlling deposition of H3K4me3. Indeed, DPPA2 binding sites are enriched for H3K4me3 (Gretarsson and Hackett, 2020) and, importantly, DPPA2 interacts with the H3 Lys-4 histone methyltransferase KMT2d (Eckersley-Maslin et al., 2020). Having found *Kmt2d* as a hit of the screen reiterates the importance of loss of H3K4me3 in the memory of silenced states at this locus.

H3K4me3 and DNA methylation are mutually exclusive marks (Weber et al., 2007) and it has been proposed that methylation at Lys-4 of histones H3 might counteract the activity of de novo DNA methylation enzymes (Ooi et al., 2007). Therefore, the observed inheritance of DNA methylation at the *Esg1*-promoter might be promoted

by absence of H3K4me3. However, the cause-consequence relationship of memory of previous levels of these two marks remains to be elucidated.

A previous work in the lab, looking for regulators of DNA methylation reprogramming, found that DPPA2/4 has a focal role in counteracting aberrant de novo methylation during phases of epigenetic remodelling (Gretarsson and Hackett, 2020). My results further extend these findings and propose a key role of DPPA2 in the memory of induced or aberrant epigenetic states.

Three lines of evidence that i) reporter expression is unaffected prior to epigenetic editing in *Dppa2*$^{-/-}$; ii) equal acquired epigenetic states upon KRAB tethering between WT and *knockout*; iii) unaltered global chromatin accessibility outside of perturbed epialleles in *Dppa2*$^{-/-}$; imply that *Dppa2* has a role specifically to counteract epigenetic perturbations and its absence does not affect unaltered chromatin. DPPA2 might thus act as a surveyor to impede inheritance of altered epigenetic states in pluripotent cells.

When epigenetic perturbations are introduced at other regions in *Dppa2*$^{-/-}$ I observed different extents of memory at distinct loci. Some of these regions show only a partial retention of silencing, maybe indicative of a stochastic memory function in which the small population that inherited the epiallele is masked by the cells that lose it. This suggests that DPPA2 might not surveys for epigenetic memory equally in the whole genome but rather, its activity is greatly influenced by the surrounding epigenetic environment. Indeed, I found that the probability of memory is extremely context dependent, with imprinting genes keeping memory of perturbed states even in the *wildtype* condition. At these regions, the underlying DNA sequence or trans-acting factors might play in favor of epigenetic inheritance. By definition, imprints can maintain memory of past epigenetic states. Although not entirely surprising, mine is the first report, to my knowledge, that showed that these regions can also exhibit long lasting epigenetic perturbations once the trigger is released. This is extremely relevant since many imprints are associated with several disorders. For example, *Plagl1* is associated with transient neonatal diabetes mellitus. CRISPR-Cas9 epigenetic editing can be potentially employed to revert such imprinting diseases (Syding et al., 2020).

On the other hand, the case of *Jade1*, which maintains memory of the repressed epiallele exclusively in *Dppa2*$^{-/-}$ suggests that at this locus, similarly to what observed at the *Esg1* promoter, the epigenetic environment in *wildtype* cells is favorable but not sufficient to ensure memory. At these regions, the absence of *Dppa2* is instrumental to lock previous epigenetic states.

One hypothesis is that DPPA2 counteracts epigenetic perturbations at binding sites. However, by mapping DPPA2 footprint on the genome by CUT&RUN in *wildtype* cells I could not find a strong correlation between direct binding of this protein with strength of memory observed at specific targets in the *knockout*.

Therefore, the precise mechanism by which DPPA2 function with different degrees at different regions is still to be unravelled. *Esg1*-silenced epiallele stably retains perturbed levels of H3K4me3 and DNA methylation. One possibility is that DPPA2 indirectly controls level of these two marks and that individual targets respond differently upon altered H3K4me3 and DNA methylation in *Dppa2*$^{-/-}$.

Interestingly, my data also imply that H3K9me3 is dispensable to repress the transcription of the endogenous reporter. Indeed, I have not picked any H3K9me3 histone demethylase in the screen. To control if these candidates might have been lost due to the non-saturation of the screen, I deleted the histone demethylase KDM3A and applied the epigenetic editing assay in these cells (data not shown). Interestingly, *Kdm3a* knockout did not have any effect on memory of the epigenetic state, suggesting that H3K9me3 removal does not have a relevant role in maintaining a silenced state at this locus. Similarly, in *S. Pombe* it was recently demonstrated that H3K9me3 is not repressive per se (Duempelmann et al., 2019).

Overall, the major limitation of such a forward genetic screen is that of being unable to pick any proteins involved in redundant pathways. Thus, I cannot exclude that other factors might be involved in the erasure of memory of previous epigenetic states.

To note is that, *Dppa2/4* is exclusively expressed in early embryo and gametogenesis and this pattern suggests an exclusive role of DPPA2 as a surveyor for epigenetic perturbation exclusively in pluripotent cells and might explain the fact that memory of epigenetic perturbations has been previously observed in somatic cells (Amabile et al., 2016; Bintu et al., 2016).

Is the silenced state eligible to be maintained in the cells that undergo cell differentiation, when *Dppa2/4* is downregulated?

5 Epigenetic memory upon exit from pluripotency

5.1 Introduction

Although stable epigenetic silencing has been previously shown in mitotic somatic cells (Amabile et al., 2016; Bintu et al., 2016) my data indicates ESC rapidly reverse ectopic epigenetic states, suggesting pluripotent phases act as a blockade to heritable epigenetic memory. Also, I identified DPPA2 as a factor that might survey for propagation of epigenetic perturbations and this gene is exclusively expressed in pluripotent cells, suggesting that DPPA2 might act as a surveyor for epigenetic silencing specifically during pluripotency.

However, from a developmental view, pluripotency is a short transition window, and cells are rapidly committed towards their fates. Exit from pluripotency and priming for differentiation into somatic lineages is associated with genome-wide de novo DNA methylation (Kim and Costello, 2017). During differentiation, cells gradually acquire specific epigenetic signatures that restrict cellular potential. Most importantly these epigenetic marks need to be propagated through cellular replication in order to lock lineage-specific gene expression patterns.

It is well established that chromatin organization is overall more dynamic in pluripotent cells compared to their differentiated counterparts (Gaspar-Maia et al.,

2011). Indeed, lineage specification is generally accompanied by stabilizing epigenetic configurations.

In this context where mechanisms to propagate epigenetic signatures are in place, it is suggested that also ectopically introduced epigenetic changes can be stably maintained. The next aim of my project is therefore to investigate whether perturbed epigenetic states acquired during the pluripotency phase, which is suggested to be susceptible to induced epigenetic errors, can be propagated upon exit from pluripotency and potentially be maintained in several tissues and organs.

5.2 Ectopic silenced states can be propagated upon exit from pluripotency under selective advantage.

To investigate whether imposed epigenetic silencing, acquired during the pluripotency phase, can be maintained upon exit from pluripotency, I differentiated ESC towards definitive endoderm as an *in vitro* model of development (Borowiak et al., 2009).

However, the endogenous *Esg1*-reporter is silenced during endoderm differentiation as part of the normal developmental program. Therefore, to track the ectopically silenced state I generated a second reporter line by inserting tdTomato sequence under the control of the constitutively expressed *p53* promoter and separated by a T2A self-cleavable domain, without disrupting the gene function (Fig. 5.1). Indeed, *p53* is a key tumour suppressor and loss of function heterozygous *p53* deletions can be deleterious since mice can undergo cancer development in average 18 months (Donehower, 1996). As observed before, epigenetic memory might have a stochastic component therefore to be able to pick up memory in a small proportion of cells, being *p53* is a master regulator of cell cycle, I choose this target since it is expected to provide proliferation advantage in the cells where its function is abolished (McKinley and Cheeseman, 2017).

The experimental strategy is described in Fig. 5.2.a. $p53^{-tdTomato}$ cells carrying $dCas9^{-5xGCN4}$ + $KRAB^{-GFP-scFv}$ (or GFP^{-scFv}) + $p53\_\ gRNA^{345up}$ were induced with DOX and TOM- cells sorted after 5 days and put back in culture still in presence of DOX. Cells were thus cultured for a total of 7 days with DOX. Endoderm differentiation was induced by culturing the cells in presence of the small molecule IDE1 and was started from the sixth day so that DOX treatment covers the first 24 hrs needed for the cells to exit

naïve pluripotency. In parallel, ESCs were also maintained in culture in undifferentiated conditions following exactly the same experimental design.

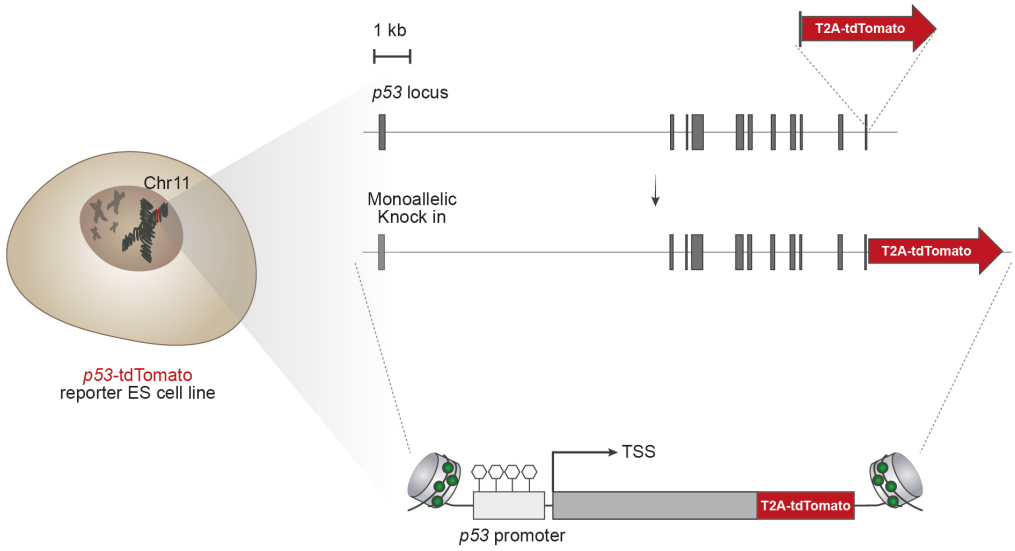


Fig. 5.1. Generation of p53 endogenous reporter for epigenetic silencing. Top: schematics of the strategy to insert the coding sequence for tdTomato-T2A at the 3' end of the p53 gene. The insertion was monoallelic. Bottom: schematic representation of the p53 reporter. White lollypops indicate unmethylated CpGs and green circles represent active histone marks. The black arrow indicates the position of the transcriptional start site (TSS).

To confirm that the new reporter equally responded to induced repression, flow cytometry analysis was conducted after 7 days of DOX treatment. Indeed, upon KRAB-targeting of the *p53*-reporter promoter with a gRNA binding 345 bp upstream the TSS ($gRNA^{345up}$), this new reporter underwent complete silencing (Fig. 5.2.b), recapitulating the same results obtained with the $Esg1^{-tdTomato}$ reporter. Once the targeting system is released, ESC completely reactivates the *p53-reporter* already after 4 days of DOX removal, consistent with results at $Esg1^{-tdTomato}$ in ESC (Fig. 5.2.c). Strikingly however, upon differentiation to definitive endoderm the majority of the cells keep full repression of the reporter gene both at a short and long time analysed (Fig. 5.2.d). Comparing the two timepoints of DOX removal in endoderm (Fig. 5.2.d) the percentage of repressed cells increases with time suggesting, as expected, a proliferation advantage of the *p53*-silenced (TOM-) cells over the tdTomato positive population.

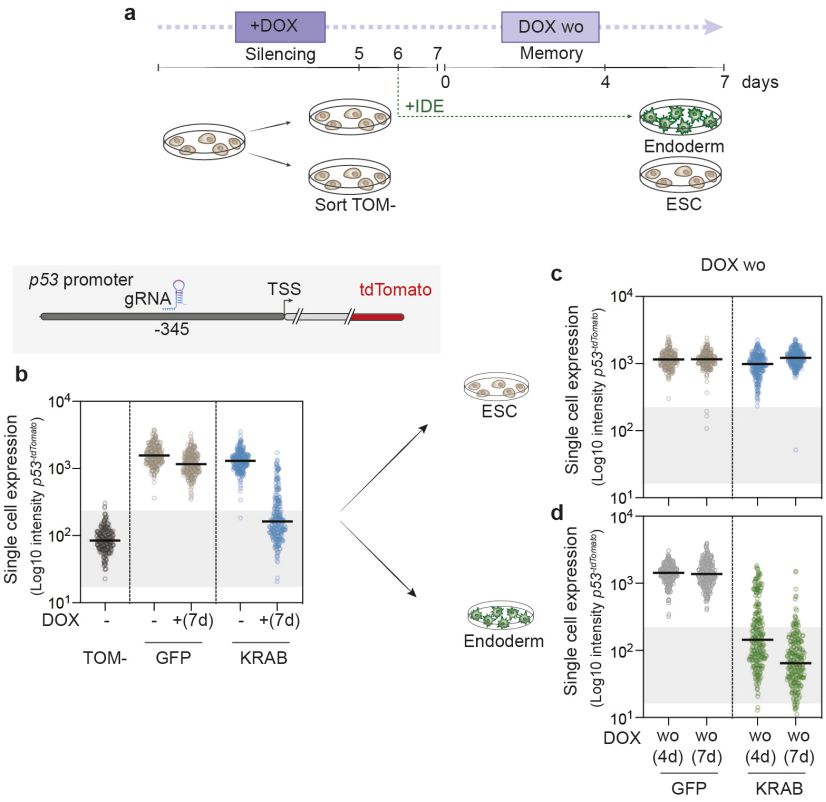


Fig 5.2. Ectopic silenced states can be propagated upon exit from pluripotency. (a) Timeline of the experiment: after 5 days of DOX induction, p53 $^{-tdTomato}$ negative cells are sorted and seeded back at low density in presence of DOX. Endoderm differentiation was induced with IDE1 after 24 hours from sort and DOX was removed after a total of 48 hours. In parallel cells were kept undifferentiated following the same experimental procedure. Cells were then analysed after 4 or 7 days of DOX removal. (b) Violin plots shows Log10 intensity of the p53 $^{-tdTomato}$ reporter in GFP^{-scFv} (light brown) and$^{KRAB\text{-}GFP\text{-}scFv}$ (blue) in minus and plus DOX (7 days) conditions. In grey is represented the basal level of fluorescence of a cell line not containing the tdTomato reporter. Black horizontal bars represent the median of fluorescence intensity in the population of cells. The schematics on the top illustrate the position of the gRNA used for the experiment. (c) and (d) expression of tdTomato after 4 or 7 days of DOX washout in ESC (c) and upon endoderm differentiation (d).

To rule out the extent of *p53*-silencing-mediated advantage, *p53*-epigenetically-silenced cells were equally mixed with non-silenced cells carrying a constitutive GFP reporter (CAG:GFP) and subsequently differentiated towards endoderm lineage (Fig. 5.3.a). By analysing GFP expression in ESC or Endoderm cells at 5 or 8 days after mixing (corresponding to 4 and 7 days of DOX washout, respectively) I observed that *p53*-silenced cells took over the population with time, reaching 95% of the population after 8 days but only in endoderm cells and not in ESC (Fig. 5.3.b and c). Endoderm cells with *p53*-silenced epiallele therefore have a strong advantage compared to non *p53*-silenced cells. Indeed I observed that *p53*-silenced cells

replicates at a faster rate and show a better viability compared to non-silenced cells upon endoderm differentiation but not in ESC (Fig 5.3.d,e and f).

Overall these data show that, contrary to ESC, perturbed epialleles can be potentially inherited in differentiated cells. However, I found that the memory of this perturbation might be facilitated by a selective advantage pressure. These data imply that an epigenetic perturbation that occurs during, or subsequent to, the pluripotency window of early development can be maintained upon cell differentiation. According to these findings, whether an epigenetic perturbation occurs during cell differentiation can be propagated, being possibly detrimental for the whole organism.

Indeed, epigenetic perturbations occurring early in life can persist long after the initial insult and are considered at the basis of the developmental origins of health and disease (DOHaD) paradigm (Walker et al., 2012). Numerous studies provide evidences for the DOHaD hypothesis linking epigenetic alterations due to exposure to environmental insults during gametogenesis and intrauterine life to the outcome of cancer and metabolic diseases (Radford et al., 2014; Trevino et al., 2020).

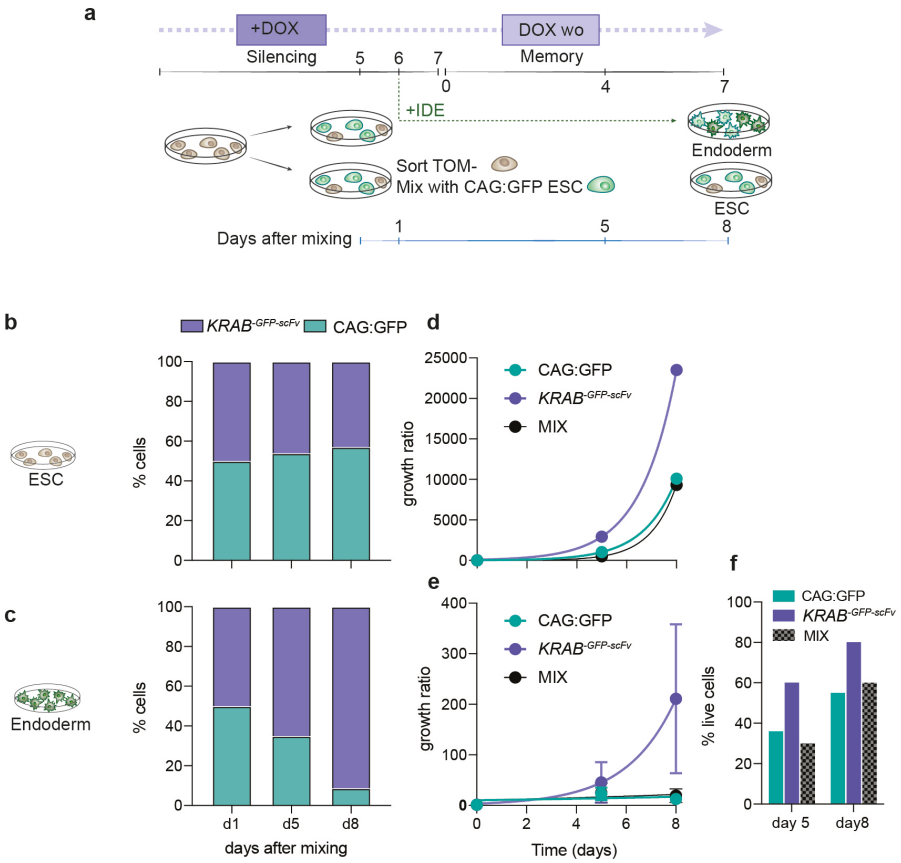


Fig 5.3. Epigetnetically p53-silenced cells have a selective advantage (a) Timeline of the experiment: after 5 days of DOX induction, p53 -tdTomato negative cells are sorted and mixed with a cell line containing a constitutive GFP reporter (CAG:GFP). Endoderm differentiation was induced in the mixed population with IDE1 after 24 hours from sort and DOX was removed after a total of 48 hours. In parallel cells were kept undifferentiated following the same experimental procedure. Cells were then analysed after 1, 5 or 8 days after mixing. (b), (c). Bars show proportion of the KRAB -GFP-scFv silenced cells (purple) or CAG:GFP cells (teal) in ESC (b) or endoderm differentiated cells (c) over the total population. (d), (e) Exponential curve shows growth obtained by measuring the ratio between the cell count at each timepoint over the initial number of cells at time 0 in ESC(d) or endoderm differentiated cells (e). (f) Bars indicate percentage of live cells in each sample.

5.3 *p53* silenced epiallele can be transmitted during *in vivo* embryo development.

To prove whether ectopic epigenetic states, acquired during the pluripotency window, can be transmitted during differentiation *in vivo* I injected KRAB-mediated *p53*-epigenetically silenced ESC into E3.5 blastocysts and implanted these embryos into foster mothers (Fig. 5.4.a). To investigate the memory of this silenced state upon post-implantation embryonic development, in absence of DOX, I compared $KRAB^{-GFP-scFv}$ treated cells with GFP^{-scFv} control. I collected embryos at E10.5, after post-implantation development, corresponding to 7 days of DOX washout, and chimerae were selected according to constitutive $gRNA^{-BFP}$ expression. By fluorescence stereomicroscope I found that both BFP and tdTomato are evenly distributed in all the embryonic tissues without any particular pattern (Fig. 5.4.b). To estimate the extent of memory of the silenced state, I further analysed the intensity of $p53^{-tdTomato}$ expression by flow cytometry in cells gated for BFP+. I observed a small but significant percentage of these cells retained memory of the silenced state (Fig. 5.4.c). However, I detected a lot of variability between samples, with some embryos showing no memory and others with up to 8% of silenced cells, this highlights the stochasticity of the memory event.

Considering that *p53* is a tumour suppressor gene and that *p53* knockout mice are susceptible to tumour development (Donehower, 1996). This data suggests that if an epigenetic perturbation occurs for a discrete time window during early development, it can be inherited further in development and transmitted in all somatic lineages, being potentially detrimental during later stages of life.

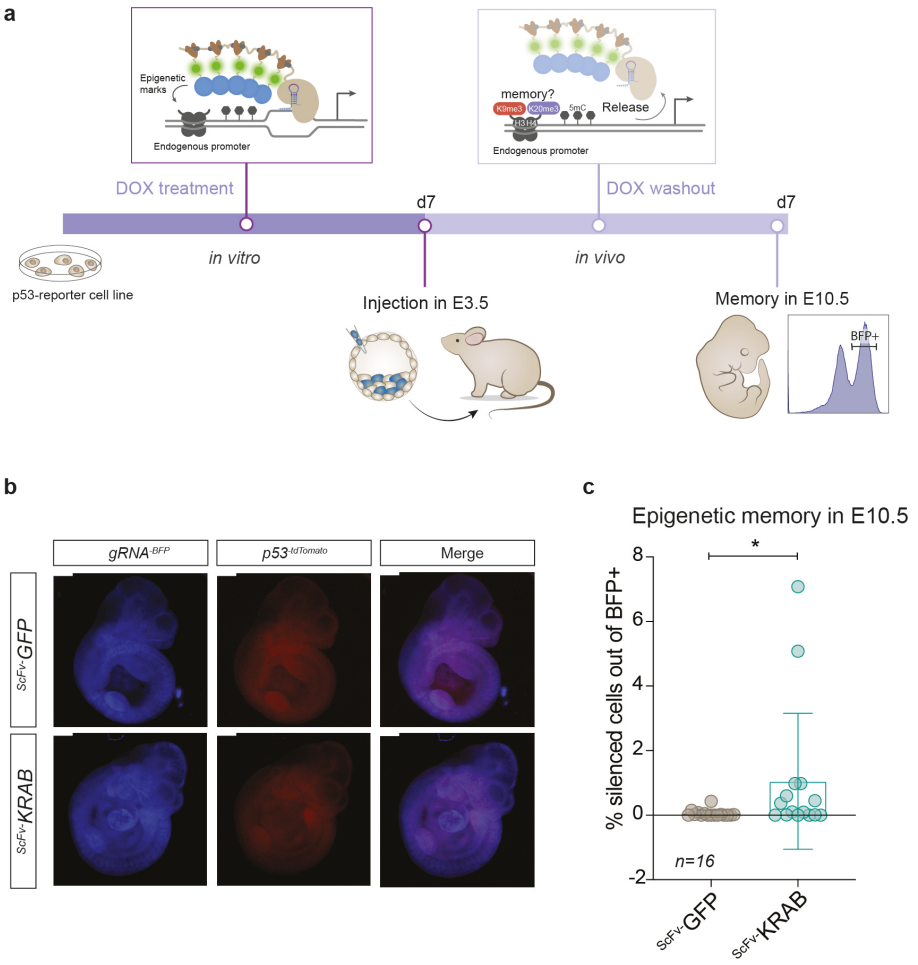


Fig 5.4. p53-silenced epiallele can be transmitted during in vivo embryo development. (a) Timeline of the experiment: after 7 days of DOX induction, KRAB-GFP-scFv or GFP-scFv cells were injected in E3.5 blastocyst and embryos were implanted back in fosters. After 7 days E10.5 embryos were collected and contribution of ESC to the chimerae has been estimated by constitutive gRNA-BFP expression. (b) Epifluorescence stereomicroscope images of E10.5 chimeric embryos. (c) % of p53-silenced cells in BFP+ population. Each dot represents single embryos out of two independent experiments. Statistics is measured by one-tail unpaired t-test (=p<0.05).*

5.4 Perturbed epigenetic states can be mitotically propagated upon exit from pluripotency.

To understand the chromatin features associated with memory of the repressive state at *p53*-reporter I used pyrosequencing to analyse DNA methylation and CUT&RUN to profile H3K9me3 and H3K4me3 comparing ESCs with Endoderm cells *in vitro*.

Analysis of DNA methylation state, revealed that upon DOX treatment, DNA methylation is strongly deposited at the targeted *p53* locus, reaching almost a 60% compared to a non-targeted control which is completely un-methylated (Fig. 5.5.a). Similarly to what observed at the *Esg1* locus (Fig. 3.6.c), upon DOX removal, DNA methylation is totally erased also at the *p53* locus in pluripotent ESC (Fig. 5.5.b). However, this mark is completely retained upon differentiation at both early and late timepoints of DOX washout (Fig. 5.5.c), suggesting that memory of DNA methylation, outside of its endogenous loci, is possible only upon specific differentiation cues and, probably, under selective advantages.

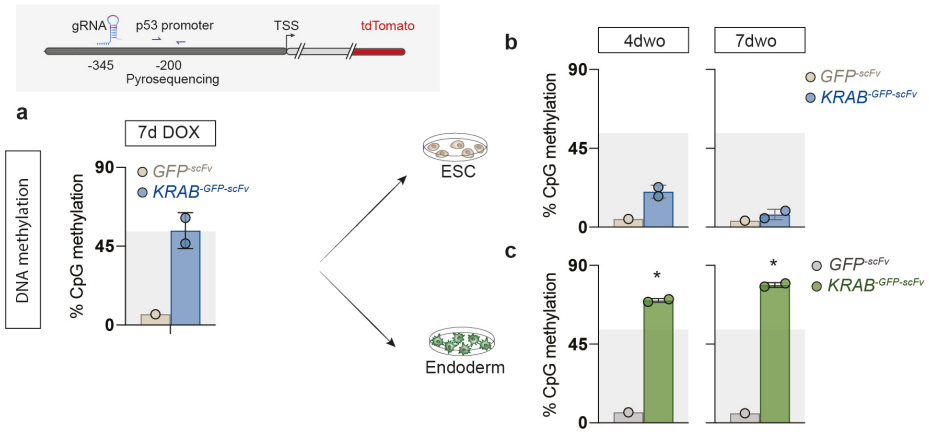


Fig 5.5. DNA methylation is transmitted upon endoderm differentiation. (a) On top: schematics of the reporter with position of the gRNA and primers used for pyrosequencing. % of CpG methylation in ESC after 7 days of DOX induction in GFP-scFv control (light brown) or KRAB-GFP-scFv (blue). Grey boxes represent the level of DNA methylation deposition after tethering with KRAB-GFP-scFv (7dDOX) (b) and (c) % of CpG methylation after 4 and 7 days of DOX washout in ESC (b) or Endoderm (c). Error bars represent standard deviation measured out of two biological replicates. Statistics is measured by one-tail unpaired t-test between GFP-scFv control and KRAB-GFP-scFv (=p<0.05).*

CUT&RUN-qPCR shows that, similarly to what observed at *Esg1* locus (Fig. 3.3), H3K9me3 is highly enriched (> 20 folds) also at this new promoter and concurrently H3K4me3 is erased upon KRAB tethering for 7 days (Fig. 5.6.a and d). However, I found that H3K9me3 is totally erased upon 7 days DOX removal in both ESC and endoderm (Fig. 5.6.b and c). Therefore, although *p53*-targeted endoderm cells completely maintain the silenced state, H3K9me3 is not inherited in these cells, suggesting that this mark is dispensable for transcriptional repression. On the other hand, the levels of the activating H3K4me3 mark, as expected, is re-established to the untargeted conditions in ESC (Fig. 5.6.e), but this mark is not re-deposited upon endoderm differentiation (Fig. 5.6.f). Although, I can not exclude that other epigenetic marks might be in place to hold the repressive state (i.e. H3K9me2), with my data, I can speculate that the combination of DNA methylation and absence of H3K4me3 is favourable for epigenetic memory of gene silencing upon exit from pluripotency.

To investigate more globally the epigenetic state of the locus I performed ATAC-seq analysing chromatin accessibility after DOX treatment for seven days and subsequently washing out the triggering system for further 7 days in ESC and upon Endoderm differentiation. I compared the differentially accessible regions between *KRAB*$^{-GFP-scFv}$ and *GFP*$^{-scFv}$ control at each time point. Correlation plots show that, despite all promoters have a strong correlation of chromatin accessibility between *KRAB*$^{-GFP-scFv}$ and *GFP*$^{-scFv}$, *p53* is among the most differentially accessible promoters upon DOX treatment and DOX washout only in differentiated cells but not in ESC (Fig. 5.7.a, b and c). In fact, ATAC-seq tracks reveal that, while being highly accessible in untargeted conditions, chromatin at *p53* transgene is strongly compacted and therefore inaccessible upon KRAB targeting. Importantly chromatin compaction is reverted back to the original state upon releasing of KRAB in ESC but memory of the epigenetic state is maintained upon endoderm differentiation (Fig. 5.7.a, b and c).

Remarkably in differentiated cells we observed that upon an extensive deposition of ectopic DNA methylation and H3K9me3 and consequent loss of H3K4me3, memory of H3K9me3 is not maintained but the cells only inherit DNA methylation and loss of K4me3. Interestingly, this is exactly the same combination of marks inherited in *Dppa2*$^{-/-}$ cells. Additionally, *Dppa2* is exclusively expressed in pluripotent cells in which I did not observe memory. This supports my hypothesis that DPPA2 might be a surveyor factor to counteract epigenetic perturbations specifically in pluripotency.

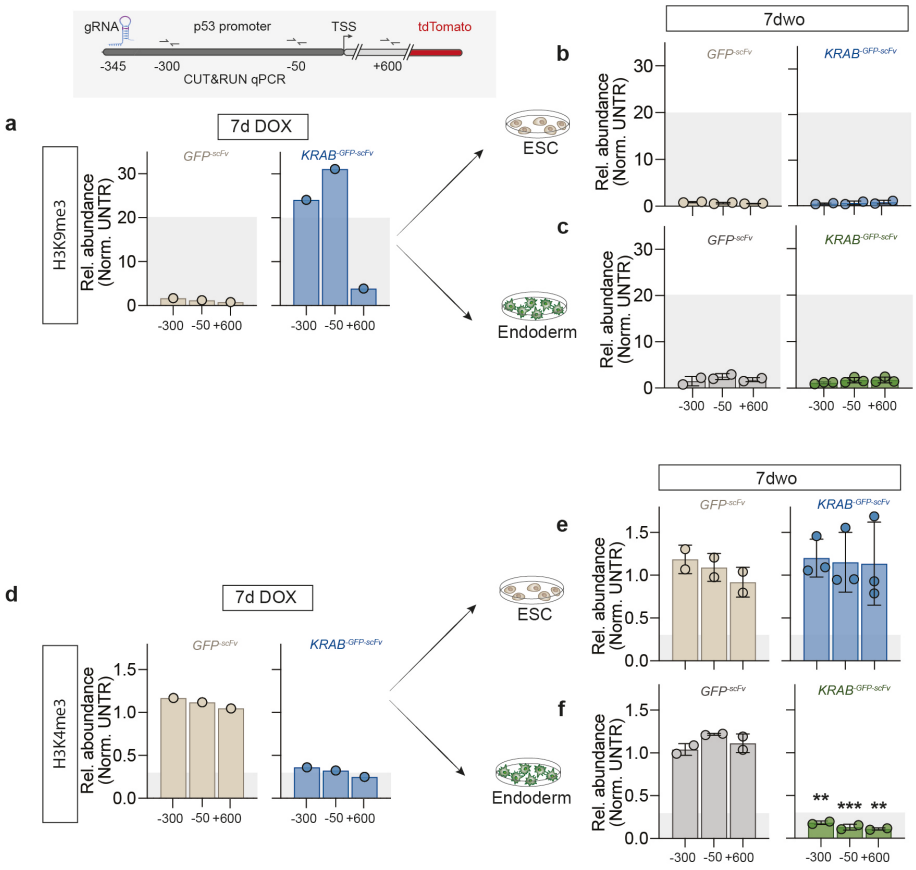


*Fig 5.6. H3K4me3 but not H3K9me3 perturbed states are inherited in endoderm cells. (a) On top: schematics of the reporter with position of the gRNA and primers used for CUT&RUN-qPCR (a) and (d) Histograms shows the relative abundance of each histone mark compared to a positive control region and the untrasfencted in GFP^{-scFv} (light brown) and KRAB$^{-GFP-scFv}$ (blue) after 7d of DOX and (b),(c),(e) and (f) after 7 days of DOX washout in ESC (b and e) and endoderm (c and f). Error bars are standard deviation between two or three biological replicates. Asterisks indicates Pvalues measured with one-tail unpaired t-test over two or three independent biological replicates (**=p<0.01; ***=p<0.001). Grey boxes indicate the average level at the three regions investigated of each histone mark after tethering with KRAB$^{-GFP-scFv}$ (7dDOX).*

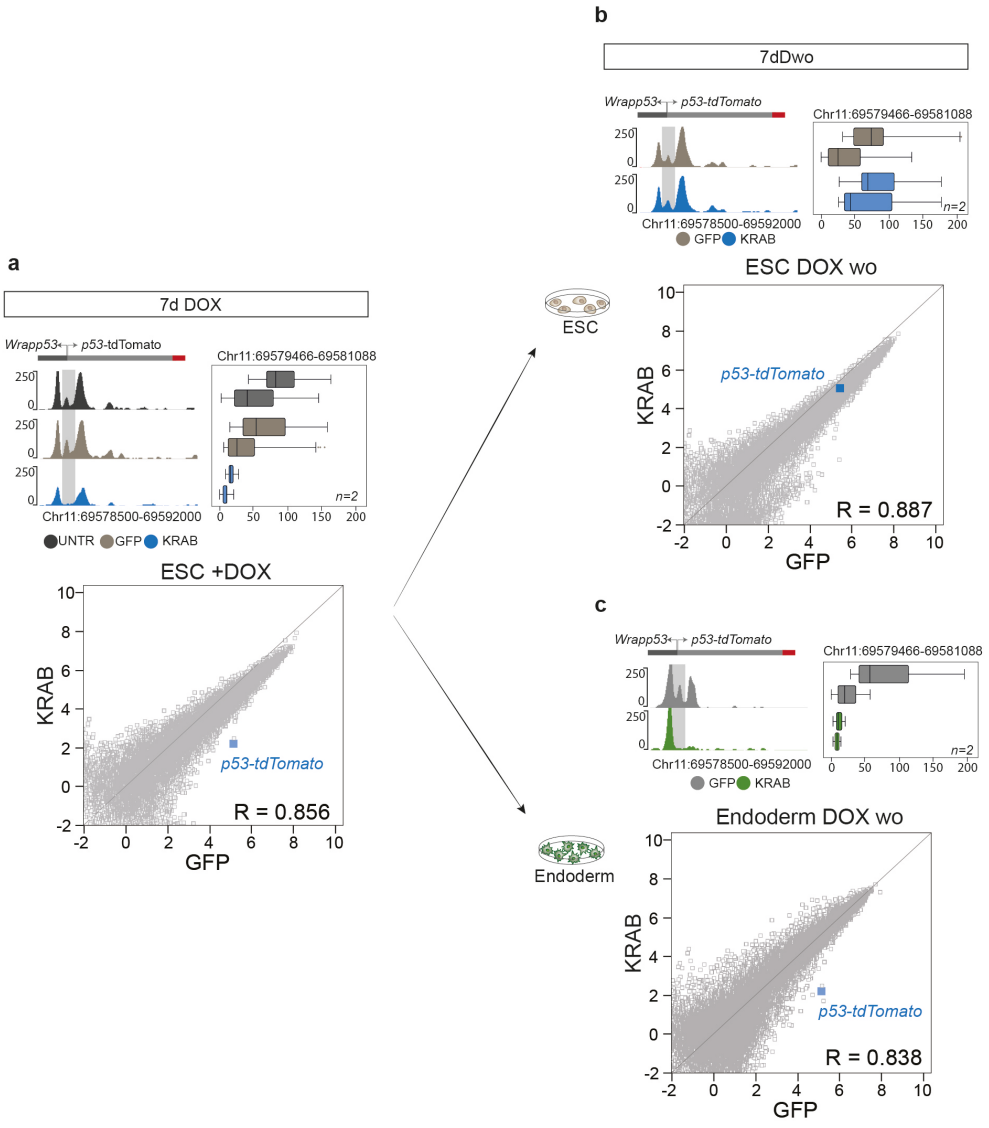


Fig 5.7. Endoderm cells mantain memory of chromatin compaction after release of KRAB (a), (b) and (c) on top ATAC-seq tracks showing chromatin accessibility at the p53$^{-tdTomato}$ locus after 7days of DOX (a) or 7 days of DOX washout (b (ESC) and c (Endoederm)). Grey box highlight the position of the p53 promoter. On the bottom: correlation plot comparing accessibility of promoters in KRAB$^{-GFP-scFv}$ vs GFP^{-scFv}. On the right of each panel: wiggle plot quantification of the region highlighted in grey (Chr11:69579466-69581088) in two biological replicates.

5.5 Memory in endoderm differentiated cells phenocopies memory in *Dppa2*$^{-/-}$

Endoderm cells undergo complete downregulation of *Dppa2* expression by 3 days of differentiation (Gretarsson and Hackett, 2020) and shows stable inheritance of the *p53*-silenced epiallele differently from the ESC counterpart. Furthermore, KRAB-mediated epialleles exclusively maintain memory of precedent states of DNA methylation and H3K4me3 both in *Dppa2*$^{-/-}$ and endoderm cells. This similarity suggests a link between differentiated and *knockout* cells.

Therefore, I asked if the same targets at which I observed memory in *Dppa2*$^{-/-}$ also retain epigenetic silencing in endoderm differentiated cells. I compared the extent of memory in *wildtype* ESC, *Dppa2*$^{-/-}$ ESC and *wildtype* endoderm cells after 7 days of KRAB release at the same targets investigated in paragraph 4.6 (Fig. 5.8). As already shown before, most imprinting genes selected (*Peg3*, *Mest* and *Plagl1*) retain full silencing in both *wildtype* and *Dppa2*$^{-/-}$ ESC. As expected, the same extent of memory is also observed in endoderm differentiated wildtype cells. Strikingly, *Jade1*, which maintained full memory only in the *knockout* cell line, shows the same phenotype in endoderm cells. Intriguingly, a similar trend in inheritance of the repressive state observed at some targets (*Pten, Adamts7, Greb1*) in *Dppa2*$^{-/-}$ compared to WT ESC is also reproduced in endoderm cells. On the contrary, epigenetic memory is equally lost at *Cdh1* in all three conditions. Interestingly, silencing at *Greb1* is significantly retained after DOX removal upon endoderm differentiation.

In summary, memory in endoderm cells closely phenocopies memory in *Dppa2*$^{-/-}$ knockout. This result indirectly suggests that inheritance of epigenetic perturbations in endoderm cells might be favoured by absence of DPPA2 and that this protein is a factor that 'surveys' for epigenetic alterations exclusively in pluripotent cells.

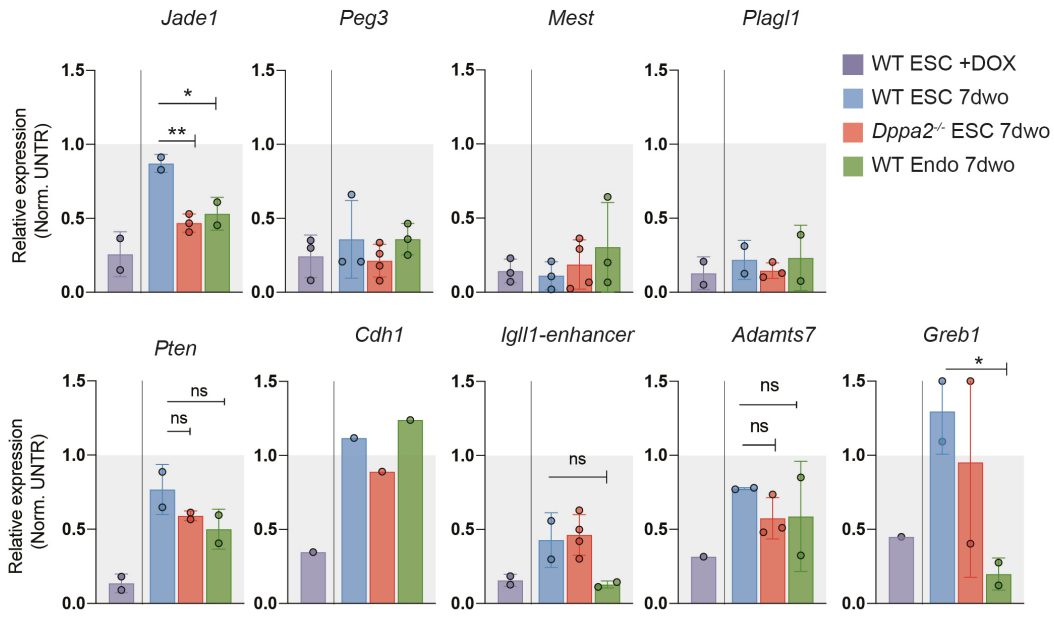


*Fig 5.8. Memory in endoderm differentiated cells phenocopies memory in Dppa2 knockout. Histogram plots show expression of the indicated targets relative to the housekeeping gene Rplp0 and normalised to untransfected control for each timepoint (grey box). Error bars are standard deviation out of two, three or four biological replicates. Statistics is measured by one-tail unpaired t-test between WT and knockout conditions (*=p<0.05; **=p<0.01).*

5.6 *Dppa2* overexpression during endoderm differentiation counteracts epigenetic inheritance

To prove whether DPPA2 has a protective activity against inheritance of the perturbed epialleles, I overexpressed this protein in *p53*-silenced cells while they undergo endoderm differentiation, condition in which *p53*-epiallele is normally maintained silenced. To achieve overexpression of DPPA2, the concurrent expression of DPPA4 is necessary to stabilise the first protein, as judged by western blot in HEK293T cells, which do not express DPPA2, transfected with either *Dppa2*-only or *Dppa2+4* overexpression constructs (Fig 5.9.a).

Briefly, upon programmed epigenetic editing at the *p53*-reporter promoter with *KRAB*$^{-GFP-scFv}$ (or *GFP*$^{-scFv}$ as a control) in ESC, endoderm differentiation was induced in cells carrying the silenced epiallele isolated at FACS for TOM-. The editing system was released by DOX washout after one day of endoderm differentiation and, at the same time, I transfected the cells to introduce the *Dppa2/4* overexpression construct (*Dppa2+4OE*). After 4 days (equivalent to 5 days of differentiation) I analysed the cells by flow cytometry to estimate the extent of *p53*-silenced epiallele inheritance in presence or absence of *Dppa2+4OE*. As expected, *p53*-silencing is maintained in KRAB untransfected cells (with only 45% of the cells reverting to TOM+ at this time-point) and *Dppa2+4OE* does not have any effect on *p53* expression per se in the GFP control (Fig. 5.9.b). Strikingly, I observed that *Dppa2+4* overexpression significantly reverts *p53* expression to an active state judged from the percentage of *p53*$^{-tdTomato}$ positive cells in this population, reaching 75% in two biological replicates.

Importantly, pyrosequencing revealed that reactivation of *p53* expression in *Dppa2+4OE* cells is epigenetically driven and corresponded to a significant decrease in CpG methylation at this promoter (Fig. 5.9.c) while *Dppa2+4* overexpression did not alter DNA methylation at this locus in the GFP control.

These data, although preliminary, show that DPPA2/4 overexpression is sufficient to reverse heritability of epialleles observed in lineage committed cells. This might also imply that the predisposition to inherit perturbed epiallele in endoderm cells might be due to absence of DPPA2 in these cells. Complementarily, this result strengthens the hypothesis that DPPA2 specifically counteract memory of epigenetic perturbations during pluripotency.

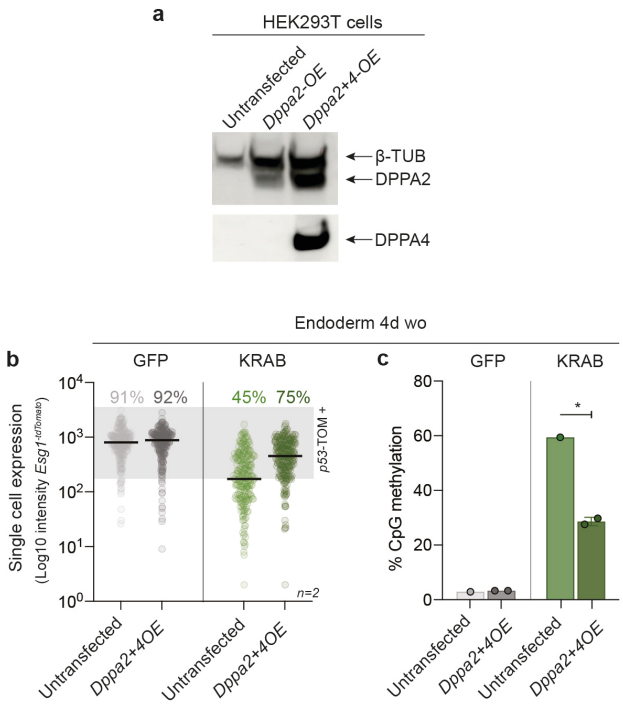


*Fig 5.9. Dppa2/4 overexpression during endoderm differentiation counteracts epigenetic memory. (a) Western blot shows expression of DPPA2 and 4 protein in HEK293T cells transfected with either Dppa2OE or Dppa2+4OE construct or in the untransfected control. β-TUBULINE was used as a loading control. (b) Violin plots represent expression of the p53-tdTomato reporter at the single cell level analysed by flow cytometry. Percentages indicate the fraction of p53-TOM+ population in each sample. Black horizontal bar is the median intensity in the population. (c) Histogram plot show % of DNA methylation at the p53 promoter. Error bars represent standard deviation between two biological replicates. Statistics is measured by one-tail unpaired t-test between WT and knockout conditions (*=p<0.05).*

5.7 Discussion

By releasable epigenetic editing of a second endogenous reporter, generated by in frame tdTomato tag downstream the *p53* gene, I further observed that my system is able to induce robust deposition of heterochromatin also at this promoter. Differently to what shown in pluripotent ESC, the same epiallele is stably transmitted upon definitive endoderm differentiation through many DNA replication events, upon release of the triggering signal. This suggests that the stable epigenetic landscape of differentiated cells is more favourable for epigenetic memory compared to the hyperdynamic epigenome characteristic of the pluripotency phases.

However, *p53* is a master regulator of cell cycle and apoptosis thus, when mutated, causes uncontrolled cell proliferation and genomic instability that is etiological for cancer development. Accordingly, I also observed that cells that acquire and maintain an epigenetically silenced *p53* allele proliferate faster upon endoderm differentiation and thus have a replicative advantage over the cells that revert to the active epiallele. This advantage is probably at the basis of the epigenetic inheritance observed.

This finding that, epiallele memory is favoured by a selective advantage in mammalian system is extremely relevant for the concept of soft inheritance as a carrier of adaptive evolution. The concept that epigenetic inheritance might be involved in environmental adaptation found many examples in nematodes and yeast. Interestingly, last year it has been reported that heterochromatin dependent epialleles provide an effective strategy for transient adaptation to external insults such as caffeine. Thus, the authors showed that phenotypic plasticity allows wild type cells to adapt to unfavourable conditions without undergoing genetic changes (Torres-Garcia et al., 2020). Another study also reported that DNA methylation has been propagated for millions of years in a species of fungi which lost the *de-novo* methyltransferase. They also propose that this extremely efficient example of epigenetic inheritance was favoured by natural selection (Catania et al., 2020).

Similarly, I found that epigenetic-silencing of the *p53* gene confers faster proliferation and resistance to apoptosis. Thus, the cells that inherit the repressive epiallele, also upon removal of the trigger, will have a strong advantage and will take over the population with time. The relevance of these findings is in the role that *p53* plays as a master tumour suppressor as this gene is found deregulated in 50% of human tumours (Hollstein et al., 1991). Indeed hypermethylation-mediated silencing of tumour suppressor genes is involved in nearly every cancer-relevant pathway and contributes to the high risk for the individuals that carry this epimutation to develop cancer (reviewed in (Hitchins, 2015)). Thus, if such an aberrant epigenetic state is acquired in the early embryo or in the germline it can be potentially inherited in all somatic tissues upon development. These epialleles are thus defined as

'constitutional epimutations'. Inheritance of such constitutional epimutations thus greatly enhance the risk of developing cancer. Strikingly I showed that *p53* silenced epiallele, acquired during the pluripotency phase, can be inherited also *in vivo*, where I found that up to 8% of the reporter cells that make up the E10.5 embryo maintained full repression. This partial memory, and high variability between embryos, highlight the stochasticity of the memory event. It remains to be elucidated the extent to which this mosaic memory increases the risk for these mice to develop cancer in adult life.

Analysis of the chromatin state of endoderm differentiated cells revealed that these cells maintain consistent levels of DNA methylation and do not reacquire H3K4me3 while undergoing complete loss of H3K9me3 upon 7 days since release of the trigger. Overall the chromatin state of these repressed epialleles appear to be fully compacted as shown by ATAC-seq. This chromatin arrangement is in complete agreement to what found in *Dppa2*$^{-/-}$ cells that maintains memory of KRAB-induced silencing. Indeed, *Dppa2* is rapidly downregulated in endoderm cells within 3 days of differentiation (Gretarsson and Hackett, 2020).

Thus, to confirm the role of DPPA2 as a surveyor of aberrant epigenetic states I set up a rescue experiment in endoderm cells carrying the silenced *p53* epiallele. I found that overexpression of the *Dppa2/4* heterodimer in these lineage-committed cells is sufficient to reverse inheritance of epigenetically silenced states, as judged by loss of DNA methylation at the *p53* promoter. This result, although preliminary, links the memory of epigenetic perturbation observed in Dppa2$^{-/-}$ cells with the same phenotype detected in endoderm cells.

6 Conclusions

Inheritance of epigenetic states is integral in the concept of epigenetics as envisioned by Waddington as a process to lock cell fates during development (Waddington, 1940). Yet, it is of great interest in the field understanding the extent to which epigenetic modifications can be inherited outside of this developmental program.

Epigenetics is considered to be at the interface between phenotype and the environment. Indeed, environmental insults can induce epigenetic mutations, sometimes resulting in functional alterations and diseases such as susceptibility to diabetes (Panzeri and Pospisilik, 2018) or cancer (Feinberg, 2018). An open debate is whether these epimutations can be transmitted mitotically and even meiotically across generations. Despite tremendous efforts in trying to understand how environmentally induced epigenetic memory works, the major challenge is to match phenotypes with specific epialleles. With the advent of epigenome editing tools, the ability to precisely perturb the epigenetic state of a promoter, provided a means to introduce desired perturbations that can be used as a model to reproduce environmentally induced epigenetic mutations. Thus, instead of promoting widespread epigenetic changes and then working backward to identify the epiallele behind the inherited phenotype observed, a reverse approach is to precisely introduce the epimutation responsible for a known phenotype and follow its heritability.

To explore this approach in my doctoral project, I combined a modular and releasable state-of-the-art CRISPR-dCas9 SunTag editing tool with the potent epigenetic cofactor KRAB and targeted two endogenous promoters tagged with fluorescent

reporters. As described in chapter 3, not only I obtained full transcriptional silencing, judged at the single cell level, but, for the first time, I showed that single gRNA tethering of an epigenetic editing tool is able to induce deposition and spreading of a wide heterochromatin domain, characterised by high levels of DNA methylation, H3K9me3, H4K20me3 and loss of H3K4me3, meaning that endogenous positive feedback loops have been engaged.

There is a ground body of evidence that epigenetic alterations can be inherited in mammals (Padmanabhan et al., 2013; Siklenka et al., 2015; Amabile et al., 2016, Bintu et al., 2016), although transgenerational epigenetic inheritance is thought to be rarer (Kazachenka et al., 2018). Given the robust positive feedback loop existing for some epigenetic marks, it is actually expected that they can propagate as far as they are included in big domains (Reinberg and Lynne, 2018; Jost and Vaillant, 2018). However, despite the extent of heterochromatinization achieved, here I observed that the chromatin domain deposited has only a transient memory function, and that it is rapidly reverted with time and DNA replications during pluripotency. In chapter 4 I described how I used a novel approach to unbiasedly identify factors that erase this memory by combining my assay for epigenetic editing with a genome-wide CRISPR-screen. To my knowledge this is the first time that a genome wide loss of function screen is coupled with a means to introduce epigenetic perturbations, to ultimately study mechanisms of epigenetic memory in mammals.

The result of the screen and subsequent validation suggests that the lack of heritability observed in pluripotent cells might be due to the activity of DPPA2, a small protein which is known to bind widely to many genes and transposable elements (Gretarsson and Hackett, 2020). I found that cells that lack DPPA2, do not show epigenetic alterations in unperturbed states at the investigated loci, but are eligible for propagation of an altered epigenetic state induced with the editing system.

Here, I propose that DPPA2 acts as a 'surveyor' factor that can 'sense' epigenetic aberrations and restores baseline epigenetic levels in embryonic stem cells. In support of this hypothesis I found that memory of past epigenetic states, acquired during pluripotency, can be maintained, under specific circumstances, once the cells are differentiated, as shown in chapter 5. Indeed, definitive endoderm differentiated cells do not express DPPA2 nor its heterodimer DPPA4 and this might partially explain the extent of memory observed.

Overall, the bimodal expression of the *Esg1*-reporter in *Dppa2*$^{-/-}$ at the single cell level and the mosaic memory observed in *p53*-reporter embryos *in vivo*, upon release of the trigger, suggests that epigenetic memory have a stochastic component. Several layers of complexity contribute to a given epigenetic state, including trans-acting factors, CpG methylation, histones modifications and higher order chromatin architecture. Inheritance or loss of a certain epigenetic perturbation therefore

depends on the combination of all of the above factors. Indeed, it has been suggested that epigenetic inheritance of transcriptionally silenced allele in heterochromatin, being generally accompanied by a stochastic switch between states, is imperfect (Ng et al., 2018; Saxton and Rine, 2019). My hypothesis is that DPPA2 activity, probably by controlling levels of DNA methylation and H3K4me3, tip the balance between memory and remodelling of epigenetic perturbations at a subset of genes in favour of this latter in pluripotent cells. On the contrary, in absence of this protein, such as in *Dppa2 knockout* or lineage committed cells, the stochastic switch favouring inheritance of the new epigenetic state occur more often and the balance is tilted towards memory (Fig. 6.1).

Indeed, by ectopically overexpressing these two proteins in endoderm cells carrying an epigenetically *p53* silenced epiallele, I found that these cells, that normally inherit memory of perturbed epialleles, revert to an unaltered epigenetic state. This result, not only strengthened the hypothesis that DPPA2/4 heterodimer tip the scale between memory and reprogramming but also that it guards against inheritance of epigenetic alterations.

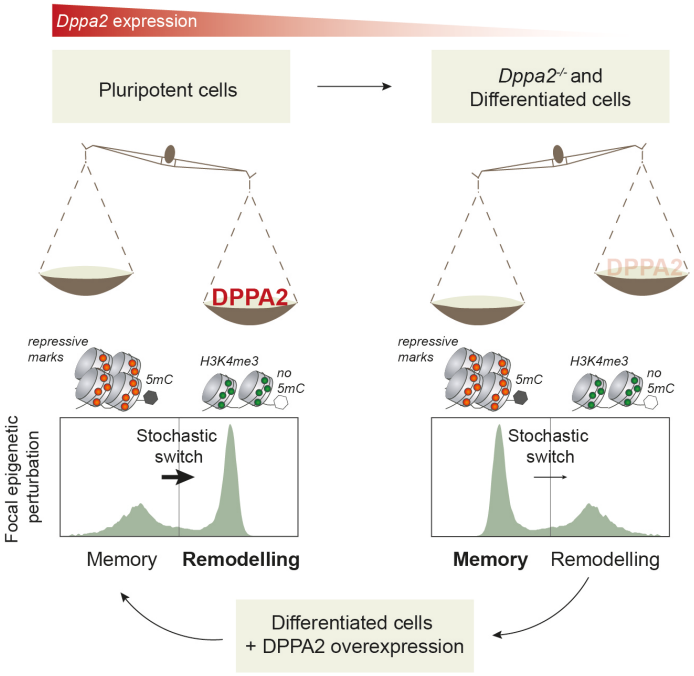


Fig 6.1. Model of the role of DPPA2 as a surveyor for epigenetic perturbations in pluripotent cells. Epigenetic inheritance of silenced states in heterochromatin is subjected to stochastic switches between memory and remodelling. DPPA2 is exclusively expressed in pluripotent cells where, probably by controlling levels of DNA methylation and H3K4me3, influences the stochastic switch between memory and remodelling of epigenetic perturbations at a subset of genes in favor of this latter. On the contrary, in Dppa2 $^{-/-}$ and endoderm differentiated cells, absence of DPPA2 favors inheritance of these epimutations by tilting the balance towards memory. Indeed, overexpression of Dppa2 in endoderm cells, which normally inherit the silenced state, reverts memory and bring back the scale towards remodelling. Epigenetic perturbations acquired during the pluripotency window, if not erased, can be propagated in the whole organism being possibly detrimental during later development and adult life. Therefore, DPPA2 might act as a surveyor against memory of epigenetic alterations at a subset of genes during this important timeframe.

7 Perspectives

7.1 A modular and releasable epigenetic editing system to investigate epigenetic phenomena

In my doctoral project I optimised an existing tool for epigenome editing by making it modulable, trackable, releasable and enhancing its on-target efficiency. To achieve this, I generated a three components system, fully modulable, by combining the state of the art dCas9 SunTag (Tanenbaum et al., 2014; Morita et al., 2016) with the epigenetic repressor KRAB and an enhanced version of the gRNA that contains an extended stem loop to stabilise the binding of the dCas9 and maximise on-target activity (Chen et al., 2013). Furthermore, I rendered the system trackable by fluorescent reporter tagging and releasable thanks to a DOX controlled promoter.

By using this combined system, I obtained deposition and spreading of a wide heterochromatin domain (>10kb), characterised by high levels of DNA methylation, H3K9me3, H4K20me3 and loss of H3K4me3, suggesting that endogenous positive feedback loops have been engaged. To my knowledge, no previous studies reported that such a big chromatin domain has been newly deposited by a single gRNA tethering.

Thanks to its modularity and on-target wide efficiency, this optimised epigenetic editing tool can be further employed to study many epigenetic phenomena by coupling with a suite of different epigenetic effectors to target nearly any region in the genome by simply changing the gRNA sequence. For example, important questions in the field, such as the causal relationship between combinations of

epigenetic marks and transcription or the role of functional DNA elements, such as transposable elements, in genome integrity, can be tackled.

7.2 Coupling of CRISPR mediated genome-scale perturbations and epigenome editing technologies

CRISPR screens have been applied to understand a variety of gene networks thanks to their potential to identify genes with a functional role in nearly any cellular or molecular process. Here, for the first time I combined a genome-scale loss of function CRISPR screen with a CRISPR-dCas9 tool to introduce epigenetic editing. This opens up incredible possibilities to combine these two powerful technologies to the unbiased research of factors that counteract or promote epigenetic perturbations in a range of mechanisms. Just to give an example, a genome wide LOF screen can be coupled with epigenetic editing at splicing junctions to investigate the role of epigenetic marks in regulating the splicing event.

The workflow of the screen might seem convoluted, due to the employment of both a catalytically active Cas9 to introduce genome editing and catalytically dead Cas9 to introduce epigenome editing. Therefore, I had the need to inactivate the wild type Cas9, once the genome-wide *knockouts* have been introduced, to impede undesired crosstalk with the dCas9 gRNA. Alternatively, this problem can be overcome by using two different Cas9 orthologs, one for genome editing and a different one employed for epigenome editing. Indeed, Cas9 from different bacterial species differs in the PAM repertoires and can be used in parallel without the risk of cross-interaction. Genome wide gRNA libraries from alternative Cas9 orthologs are already available such as for the *Staphylococcus Aureus* CRISPR-Cas9 variant (Najm et al., 2018).

7.3 Dppa2/4 potential to revert constitutional epimutations

Constitutional epimutations are defined as altered epigenetic states that arise early during development and thus can be inherited in the adult somatic tissues. These epimutations, when occurring at promoters of tumour-suppressor genes, are thought to be an initiating event for cancer progression (Hitchins et al., 2015; Lonning

et al., 2019). Specifically, in my project, I showed that *p53* ectopically-silenced epialleles, acquired during the pluripotency window, can be, to some extent, propagated in embryos, *in vivo*, once the repressive stimulus is released. Additionally, I found that overexpressing *Dppa2/4* counteracts inheritance of the *p53* silenced epiallele *in vitro.*

Although I have not investigated whether *Dppa2/4* ectopic expression has major effects on the epigenetic and transcriptional landscape of the cells, this heterodimer has the potential to be used as an agent to prevent heritability of such constitutional epimutations. More work needs to be done to identify any down effect of this ectopic overexpression, and eventually to fine tune *Dppa2/4* ectopic expression. Furthermore, it needs to be elucidated whether *Dppa2/4OE* has the same effect *in vivo* and a strategy to deliver ectopic expression of Dppa2 needs to be found. However, its employment as a prophylactic agent, to prevent inheritance of constitutional epimutations, is promising and should be further investigated.

7.4 Future directions

3D folding of chromatin into spatial compartments like TADs is expected to greatly contribute to maintenance of epigenetic memory, by spatial colocalization of epigenetic regulators (Jost and Vaillant, 2018). With my system, I obtained a wide deposition of a heterochromatin domain, way bigger than what reported before, but nowhere near covering the whole TAD in which the target promoter is located. Therefore, it would be interesting to understand whether, by increasing the size of the heterochromatin domain deposited, up to covering the whole TAD, can increase the probability of epigenetic inheritance, once the initial stimulus is released. Furthermore, understanding whether there is a threshold size of heterochromatin domain beyond which epigenetic memory is maintained, is of great importance. To do this, multiple gRNAs, delivered in a CARGO plasmid (Gu et al., 2018), will consent to target the *dCas9*$^{-5xGCN4}$ + *KRAB*$^{-scFv-GFP}$ to progressively wider regions and investigate the size-dependent effect of the deposited domain on the inheritance of the new epigenetic state.

A second line of experiments should be performed to have a mechanistic insight into DPPA2 activity as a surveyor for epigenetic aberrations. In fact, I could not identify a strong relationship between retention or loss of memory at a given locus, with DPPA2 binding at that site. One possibility is that DPPA2 binding at those promoters is

beyond detection. However, in order to understand the mechanism of DPPA2 as a surveyor for epigenetic aberration a second loss-of-function CRISPR-screen could be employed. Specifically, lentiviral gRNA-libraries can be used to introduce genome-wide *knockout*s in *Dppa2*$^{-/-}$ cells and then apply the epigenetic editing assay to investigate the factors that are involved in *Dppa2-knockout* mediated memory, according to shifts in the normal tdTomato-reporter expression.

One interesting aspect of this project is that with my system I generated a potential constitutional epimutation and showed that *p53* silenced epiallele can be inherited in post-implantation embryos *in vivo*. However, for lack of time, I did not investigate whether such epimutation confers higher predisposition to cancer, once the mice that carry the *p53*-silenced epiallele are subjected to risk factors. In order to investigate this, chimeric embryos, generated by injecting reporter ESC line carrying the silenced epialleles, should be implanted back in foster mothers and the pups let born and followed through life to see whether they develop tumours with a higher percentage compared to mice that do not carry the constitutive epimutation. Furthermore, whether it will be found that *p53* early epimutation inheritance is predisposing to tumour development, one could assess the role of *Dppa2/4* to protect against epigenetic aberration inheritance. By overexpressing *Dppa2/4* upon exit from pluripotency, for example by using a lineage specific promoter, it could be investigated whether the cells originated downstream are protected from cancer susceptibility.

8 References

Ahmed, K., Dehghani, H., Rugg-Gunn, P., Fussner, E., Rossant, J., and Bazett-Jones, D.P. (2010). Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. PLoS One *5*, e10531.

Alabert, C.B., T; Reveron-Gomez, N; Sidoli, S; Schmidt, A; Jensen, O.N; Imhof, A; Groth, A (2015). Two distinct modes for propagation of histone PTMs across the cell cycle. Genes Dev *29*, 585-590.

Allan, J., Harborne, N., Rau, D.C; Gould, H. (1982). Participation of core histone "tails" in the stabilization of the chromatin solenoid. Journal of Cell Biology *93*, 285-297.

Allan, J.M., T; Harborne, N; Bohm, L; Crane-Robinson, C. (1986). Roles of H1 domains in determining higher order chromatin structure and H1 location. J Mol Biol *187(4)*, 591-601.

Allis, C.D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nat Rev Genet *17*, 487-500.

Allshire, R.C., and Madhani, H.D. (2018). Ten principles of heterochromatin formation and function. Nat Rev Mol Cell Biol *19*, 229-244.

Amabile, A., Migliara, A., Capasso, P., Biffi, M., Cittaro, D., Naldini, L., and Lombardo, A. (2016). Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. Cell *167*, 219-232 e214.

Arand, J., Spieler, D., Karius, T., Branco, M.R., Meilinger, D., Meissner, A., Jenuwein, T., Xu, G., Leonhardt, H., Wolf, V.*, et al.* (2012). In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. PLoS Genet *8*, e1002750.

Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell *31*, 785-799.

Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., and Shirakawa, M. (2008). Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature *455*, 818-821.

Atlasi, Y., and Stunnenberg, H.G. (2017). The interplay of epigenetic marks during stem cell differentiation and development. Nat Rev Genet *18*, 643-658.

Audergon, P.N.C.B., Catania, S., Kagansky, A., Tong, P., Shukla, M., Pidoux, A.L., Allshire, R.C. (2015). Restricted epigenetic inheritance of H3K9 methylation. Science *348*, 132-135.

Ballas, N., Grunseich, C., Lu, D.D., Speh, J.C., and Mandel, G. (2005). REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. Cell *121*, 645-657.

Barau, J., Teissandier, A.l., Zamudio, N., Roy, S.p., Nalesso, V.r., Hérault, Y., Guillou, F., and Bourc'his, D.b. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science *354* 909-912.

Barlow, D.P., and Bartolomei, M.S. (2014). Genomic imprinting in mammals. Cold Spring Harb Perspect Biol *6*.

Barrangou, R., Fermaux., C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.N., Horvath, P. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science *315*, 1709-1712.

Batie, M., Frost, J., Frost, M., Wilson, J.W., Schofield, P., Rocha, S. (2019). Hypoxia induces rapid changes to histone methylation and reprograms chromatin. Science *363*, 1222–1226.

Beaujean, N., Hartshorne, G., Cavilla, J., Taylor, J., Gardner, J., Wilmut, I., Meehan, R., and Young, L. (2004). Non-conservation of mammalian preimplantation methylation dynamics. Curr Biol *14*, R266-267.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Bintu, L., Yong, J. Antebi, Y.E., McCue, K., Kazuki, Y. Uno, N., Oshimura, M., Elowitz, M.B. (2016). Dynamics of epigenetic regulation at the single-cell level. Science *351*, 720-724.

Bird, A., Taggart., M, Frommer, M, Miller, O.J, Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell *40*, 91-99.

Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L., Ito, S., Cooper, S., Kondo, K., Koseki, Y.*, et al.* (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. Cell *157*, 1445-1459.

Blank, T.A., Becker, P.B. (1995). Electrostatic mechanism of nucleosome spacing. J Mol Biol *252(3)*, 305-313.

Borowiak, M., Maehr, R., Chen, S., Chen, A.E., Tang, W., Fox, J.L., Schreiber, S.L., and Melton, D.A. (2009). Small molecules efficiently direct endodermal differentiation of mouse and human embryonic stem cells. Cell Stem Cell *4*, 348-358.

Bortvin, A., Eggan, K., Skaletsky, H., Akutsu, H., Berry, D.L., Yanagimachi, R., Page, D.C., and Jaenisch, R. (2003). Incomplete reactivation of Oct4-related genes in mouse embryos cloned from somatic nuclei. Development *130*, 1673-1680.

Bourc'his, D., Xu, G.L., Lin, C., Bollman, B., Bestor, T.H. (2001). Dnmt3L and the Establishment of Maternal Genomic Imprints. Science *294*, 2536-2539.

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell Stem Cell *14*, 838-853.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol *109*, 21 29 21-21 29 29.

Canovas, S., and Ross, P.J. (2016). Epigenetics in preimplantation mammalian development. Theriogenology *86*, 69-79.

Canzio, D., Liao, M., Naber, N., Pate, E., Larson, A., Wu, S., Marina, D.B., Garcia, J.F., Madhani, H.D., Cooke, R.*, et al.* (2013). A conformational switch in HP1 releases auto-inhibition to drive heterochromatin assembly. Nature *496*, 377-381.

Carmell, M.A., Girard, A., van de Kant, H.J., Bourc'his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev Cell *12*, 503-514.

Catania, S., Dumesic, P.A., Pimentel, H., Nasif, A., Stoddard, C.I., Burke, J.E., Diedrich, J.K., Cooke, S., Shea, T., Gienger, E.*, et al.* (2020). Evolutionary Persistence of DNA Methylation for Millions of Years after Ancient Loss of a De Novo Methyltransferase. Cell *180*, 263-277 e220.

Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet *10*, 295-304.

Chakraborty, A.A., Laukka, T., Myllykoski, M., Ringel, A.E., Booker, M.A., Tolstorukov, M.Y., Signoretti, S., Koivunen, P., Kaelin, W.G. (2019). Histone demethylase KDM6A directly senses oxygen to control chromatin and cell fate. Science, 1217–1222.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S.*, et al.* (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell *155*, 1479-1491.

Chen, Q.Y., M ; Cao, Z; Li, X ; Zhang, Y ; Shi, J ; Feng, G ; Peng, H ; Zhang, X ; Zhang, Y ; Qian, J ; Duan, E ; Zhai, Q ; Zhou, Q. (2016). Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. Science *351*, 397-400.

Cheng, A.W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T.W., Rangarajan, S., Shivalila, C.S., Dadon, D.B., and Jaenisch, R. (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res *23*, 1163-1171.

Chuang, L., Ian., H.I., Koh, T.W., Ng, H.H., Xu, G., Li, B.F. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF. Science *277*, 1996-2000.

Ciabrelli, F., Comoglio, F., Fellous, S., Bonev, B., Ninova, M., Szabo, Q., Xuereb, A., Klopp, C., Aravin, A., Paro, R.*, et al.* (2017). Stable Polycomb-dependent transgenerational inheritance of chromatin states in Drosophila. Nat Genet *49*, 876-886.

Clouaire, T., Webb, S., Skene, P., Illingworth, R., Kerr, A., Andrews, R., Lee, J.H., Skalnik, D., and Bird, A. (2012). Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. Genes Dev *26*, 1714-1728.

Coleman, R.T., and Struhl, G. (2017). Causal role for inheritance of H3K27me3 in maintaining the OFF state of a Drosophila HOX gene. Science.

Cooney, C.A.D., A.A; Wolff, G. (2002). Maternal Methyl Supplements in Mice Affect Epigenetic Variation and DNA Methylation of Offspring. The Journal of Nutrition *132*, 2393S–2400S.

Cubas, P., Vincent, C., and Coen, E. (1999). An epigenetic mutation responsible for natural variation in floral symmetry. Nature *401*, 157-161.

Darwin, C. (1859). On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life. London :John Murray,.

De, S., and Kassis, J.A. (2017). Passing epigenetic silence to the next generation. Science *356*, 28.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R.*, et al.* (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol *34*, 184-191.

Donehower, L.A. (1996). The p53-deficient mouse: a model for basic and applied cancer studies. Seminars in Cancer Biology *7*, 269-278.

Dorighi, K.M., Swigut, T., Henriques, T., Bhanu, N.V., Scruggs, B.S., Nady, N., Still, C.D., 2nd, Garcia, B.A., Adelman, K., and Wysocka, J. (2017). Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. Mol Cell *66*, 568-576 e564.

Ducker, G.S., and Rabinowitz, J.D. (2017). One-Carbon Metabolism in Health and Disease. Cell Metab *25*, 27-42.

Duempelmann, L., Mohn, F., Shimada, Y., Oberti, D., Andriollo, A., Lochs, S., and Buhler, M. (2019). Inheritance of a Phenotypically Neutral Epimutation Evokes Gene Silencing in Later Generations. Mol Cell *74*, 534-541 e534.

Eberharter, A., Becker, P.B. (2002). Histone acetylation: a switch between repressive and permissive chromatin. EMBO Rep *3(3)*, 224-229.

Eckersley-Maslin, M.A., Parry, A., Blotenburg, M., Krueger, C., Ito, Y., Franklin, V.N.R., Narita, M., D'Santos, C.S., and Reik, W. (2020). Epigenetic priming by Dppa2 and 4 in pluripotency facilitates multi-lineage commitment. Nat Struct Mol Biol *27*, 696-705.

Eissenberg, J.C., and Shilatifard, A. (2010). Histone H3 lysine 4 (H3K4) methylation in development and differentiation. Dev Biol *339*, 240-249.

Escobar, T.M., Oksuz, O., Saldana-Meyer, R., Descostes, N., Bonasio, R., and Reinberg, D. (2019). Active and Repressed Chromatin Domains Exhibit Distinct Nucleosome Segregation during DNA Replication. Cell *179*, 953-963 e911.

Espada, J., Ballestar, E., Fraga, M.F., Villar-Garea, A., Juarranz, A., Stockert, J.C., Robertson, K.D., Fuks, F., and Esteller, M. (2004). Human DNA methyltransferase 1 is required for maintenance of the histone H3 modification pattern. J Biol Chem *279*, 37175-37184.

Esteve, P.O., Chin, H.G., Smallwood, A., Feehery, G.R., Gangisetty, O., Karpf, A.R., Carey, M.F., and Pradhan, S. (2006). Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. Genes Dev *20*, 3089-3103.

Fang, J., Quin, F., Ketel, C.S., Wang, H., Cao, R., Xia, L., et al. (2002). Purification and functional characterization of SET8, a nucleosomal histone H4-lysine 20-specific methyltransferase. Curr Biol *12*, 1086–1099.

Feinberg, A.P. (2018). The Key Role of Epigenetics in Human Disease Prevention and Mitigation. N Engl J Med *378*, 1323-1334.

Ficz, G., Hore, T.A., Santos, F., Lee, H.J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.L., Walter, J*., et al.* (2013). FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. Cell Stem Cell *13*, 351-359.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J*., et al.* (2010). Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell *143*, 212-224.

Finch, J.T.K., A. (1976). Solenoidal model for superstructure in chromatin. ProcNatlAcadSciUSA *73(6)*, 1897-1901.

Francis, N.J., Kingston, R.E, Woodcock, C.L. (2004). Chromatin Compaction by a Polycomb Group Protein Complex. Science *306*, 1574-1577.

Fuks, F., Hurd, P.J., Wolf, D., Nan, X., Bird, A.P., and Kouzarides, T. (2003). The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. J Biol Chem *278*, 4035-4040.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Mol Cell *45*, 344-356.

Garcia-Ramirez, M., Rocchini., C, Ausio, J (1995). Modulation of Chromatin Folding by Histone Acetylation. J Biol Chem *270(30)* 17923-17928.

Gardiner-Garden, M., Frommer, M (1987). CpG Islands in vertebrate genomes. J Mol Biol *196*, 261-282.

Gaspar-Maia, A., Alajem, A., Meshorer, E., and Ramalho-Santos, M. (2011). Open chromatin in pluripotency and reprogramming. Nat Rev Mol Cell Biol *12*, 36-47.

Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. (2017). Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. Nature *551*, 464-471.

Gaydos, L.W., W; Strome, S (2014). H3K27me and PRC2 transmit a memory of repression across generations and during development. Science *345*, 1515-1518.

Ge, G., Ye, J., Weber, C., Sun, W., Zhang, H., Zhou, Y., Cai, C., Qian, G., Capel, B. (2018). The histone demethylase KDM6B regulates temperature-dependent sex determination in a turtle species. Science *380*, 645–648.

Gigante, S., Gouil, Q., Lucattini, A., Keniry, A., Beck, T., Tinning, M., Gordon, L., Woodruff, C., Speed, T.P., Blewitt, M.E.*, et al.* (2019). Using long-read sequencing to detect imprinted DNA methylation. Nucleic Acids Res *47*, e46.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C.*, et al.* (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647-661.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A.*, et al.* (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442-451.

Ginno, P.A., Burger, L., Seebacher, J., Iesmantavicius, V., and Schubeler, D. (2018). Cell cycle-resolved chromatin proteomics reveals the extent of mitotic preservation of the genomic regulatory landscape. Nat Commun *9*, 4048.

Gocke, C.B., and Yu, H. (2008). ZNF198 stabilizes the LSD1-CoREST-HDAC1 complex on chromatin through its MYM-type zinc fingers. PLoS One *3*, e3255.

Greally, J.M. (2018). A user's guide to the ambiguous word 'epigenetics'. Nat Rev Mol Cell Biol *19*, 207-208.

Greenberg, M.V.C., and Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol *20*, 590-607.

Gretarsson, K.H., and Hackett, J.A. (2020). Dppa2 and Dppa4 counteract de novo methylation to establish a permissive epigenome for development. Nat Struct Mol Biol *27*, 706-716.

Groner, A.C., Meylan, S., Ciuffi, A., Zangger, N., Ambrosini, G., Denervaud, N., Bucher, P., and Trono, D. (2010). KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. PLoS Genet *6*, e1000869.

Gu, B.S., T.; Spencley, A.; Bauer, M.; Chung, M.; Meyer, T.; Wysocka, J. (2018). Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. Science *359*, 1050-1055.

Gu, T.P., Guo, F., Yang, H., Wu, H.P., Xu, G.F., Liu, W., Xie, Z.G., Shi, L., He, X., Jin, S.G., *et al.* (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature *477*, 606-610.

Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., *et al.* (2015). Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature *517*, 640-644.

Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., *et al.* (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. Cell Stem Cell *13*, 360-369.

Hackett, J.A., Huang, Y., Günesdogan, U., Gretarsson, K.A., Kobayashi, T., and Surani, M.A. (2018). Tracing the transitions from pluripotency to germ cell fate with CRISPR screening. Nature Communications *9*.

Hackett, J.A., Sengupta, R., Zylicz, J.J., Murakami, K., Lee, C., Down, T.A., and Surani, A.M. (2013a). Germline DNA Demethylation Dynamics and Imprint Erasure Through 5-Hydroxymethylcytosine. Science *339*, 448-452.

Hackett, J.A., Sengupta, R., Zylicz, J.J., Murakami, K., Lee, C., Down, T.A., and Surani, M.A. (2013b). Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science *339*, 448-452.

Hackett, J.A., and Surani, M.A. (2013). Beyond DNA: Programming and Inheritance of Parental Methylomes. Cell *153*, 737-739.

Hackett, J.A., and Surani, M.A. (2014). Regulatory principles of pluripotency: from the ground state up. Cell Stem Cell *15*, 416-430.

Hajkova, P.E., S; Lane, N; Haaf, T, El-Maarri, O; Reik, W; Walter, J; Surani, M.A (2002). Epigenetic reprogramming in mouse primordial germ cells. Mechanisms of Development *117*, 15-23.

Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. Nat Cell Biol *10*, 1291-1300.

Hathaway, N.A., Bell, O., Hodges, C., Miller, E.L., Neel, D.S., and Crabtree, G.R. (2012). Dynamics and memory of heterochromatin in living cells. Cell *149*, 1447-1460.

Haws, S.A., Yu, D., Ye, C., Wille, C.K., Nguyen, L.C., Krautkramer, K.A., Tomasiewicz, J.L., Yang, S.E., Miller, B.R., Liu, W.H., *et al.* (2020). Methyl-Metabolite Depletion Elicits Adaptive Responses to Support Heterochromatin Stability and Epigenetic Persistence. Mol Cell.

Hayashi, K., Chuva de Sousa Lopes S.M., Surani, A.M. (2007). Germ Cell Specification in Mice. Science *316*, 394-396.

Heard, E., and Martienssen, R.A. (2014). Transgenerational epigenetic inheritance: myths and mechanisms. Cell *157*, 95-109.

Heitz, E. (1928). Das heterochromatin der moose. Jahrb Wiss Bot *69*, 762-818.

Henikoff, S., and Smith, M.M. (2015). Histone variants and epigenetics. Cold Spring Harb Perspect Biol *7*, a019364.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat Biotechnol *33*, 510-517.

Hitchins, M.P. (2015). Constitutional epimutation as a mechanism for cancer causality and heritability? Nat Rev Cancer *15*, 625-634.

Holliday, R. (1987). The inheritance of epigenetic defects. Science *238*, 163-117.

Hollstein, M.S., David; Vorgelstein, Bert; Harris, Curtis. (1991). p53 Mutations in Human Cancers. Science, 49-53.

Hormanseder, E., Simeone, A., Allen, G.E., Bradshaw, C.R., Figlmuller, M., Gurdon, J., and Jullien, J. (2017). H3K4 Methylation-Dependent Memory of Somatic Cell Identity Inhibits Reprogramming and Development of Nuclear Transfer Embryos. Cell Stem Cell *21*, 135-143 e136.

Houston, S.I., McManus, K.J., Adams, M.M., Sims, J.K., Carpenter, P.B., Hendzel, M.J., and Rice, J.C. (2008). Catalytic function of the PR-Set7 histone H4 lysine 20 monomethyltransferase is essential for mitotic entry and genomic stability. J Biol Chem *283*, 19478-19488.

Howe, F.S., Fischl, H., Murray, S.C., and Mellor, J. (2017). Is H3K4me3 instructive for transcription activation? Bioessays *39*, 1-12.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262-1278.

Hu, X., Shen, B., Liao, S., Ning, Y., Ma, L., Chen, J., Lin, X., Zhang, D., Li, Z., Zheng, C.*, et al.* (2017). Gene knockout of Zmym3 in mice arrests spermatogenesis at meiotic metaphase with defects in spindle assembly checkpoint. Cell Death Dis *8*, e2910.

Hutnick, L.K., Huang, X., Loo, T.C., Ma, Z., and Fan, G. (2010). Repression of retrotransposal elements in mouse embryonic stem cells is primarily mediated by a DNA methylation-independent mechanism. J Biol Chem *285*, 21082-21091.

Hyun, K., Jeon, J., Park, K., and Kim, J. (2017). Writing, erasing and reading histone lysine methylations. Exp Mol Med *49*, e324.

Ikegami, K.I., M, Suzuki, M, Tachibana, M, Shinkai, Y, Tanaka, S, Greally, J.M, Yagi, S, Hattori, N, Shiota, K. (2007). Genome-wide and locus-specific DNA hypomethylation in G9a deficient mouse embryonic stem cells. Genes Cells *12*, 1-11.

Illingworth, R.S., and Bird, A.P. (2009). CpG islands--'a rough guide'. FEBS Lett *583*, 1713-1720.

Iqbal, K., Jin, S.G., Pfeifer, G.P., and Szabo, P.E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proc Natl Acad Sci U S A *108*, 3642-3647.

Ishiyama, S., Nishiyama, A., Saeki, Y., Moritsugu, K., Morimoto, D., Yamaguchi, L., Arai, N., Matsumura, R., Kawakami, T., Mishima, Y.*, et al.* (2017). Structure of the Dnmt1 Reader Module Complexed with a Unique Two-Mono-Ubiquitin Mark on Histone H3 Reveals the Basis for DNA Methylation Maintenance. Mol Cell *68*, 350-360 e357.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129-1133.

Ito, S., Li, Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., Zhang, Y. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *333*, 1300–1303.

Jablonka, E., Lachmann, M., Lamb, M.J. (1992). Evidence, mechanisms and models for the inheritance of acquired characters. Journal of Theoretical Biology *158*, 245-268.

Jeong, S., Liang, G., Sharma, S., Lin, J.C., Choi, S.H., Han, H., Yoo, C.B., Egger, G., Yang, A.S., and Jones, P.A. (2009). Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA. Mol Cell Biol *29*, 5366-5376.

Jiang, D., Berger, F (2017). DNA replication–coupled histone modification maintains Polycomb gene silencing in plants. Science *357*, 1146-1149.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science *337*, 816-820.

Johnson, D.S., Mortazavi, A., Myers, R., Wold, B. (2007). Genome-Wide Mapping of in Vivo Protein-DNA Interactions. Science *316*, 1497-1502.

Josling, G.A., Selvarajah, S.A., Petter, M., and Duffy, M.F. (2012). The role of bromodomain proteins in regulating gene expression. Genes (Basel) *3*, 320-343.

Jost, D., and Vaillant, C. (2018). Epigenomics in 3D: importance of long-range spreading and specific interactions in epigenomic maintenance. Nucleic Acids Res *46*, 2252-2264.

Joung, J., Konermann, S., Gootenberg, J.S., Abudayyeh, O.O., Platt, R.J., Brigham, M.D., Sanjana, N.E., and Zhang, F. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. Nat Protoc *12*, 828-863.

Kaufman, P.D., and Rando, O.J. (2010). Chromatin as a potential carrier of heritable information. Curr Opin Cell Biol *22*, 284-290.

Kazachenka, A., Bertozzi, T.M., Sjoberg-Herrera, M.K., Walker, N., Gardner, J., Gunning, R., Pahita, E., Adams, S., Adams, D., and Ferguson-Smith, A.C. (2018). Identification, Characterization, and Heritability of Murine Metastable Epialleles: Implications for Non-genetic Inheritance. Cell *175*, 1259-1271 e1213.

Kearns, N.A., Pham, H., Tabak, B., Genga, R.M., Silverstein, N.J., Garber, M., and Maehr, R. (2015). Functional annotation of native enhancers with a Cas9-histone demethylase fusion. Nat Methods *12*, 401-403.

Kieffer-Kwon, K.R., Nimura, K., Rao, S.S.P., Xu, J., Jung, S., Pekowska, A., Dose, M., Stevens, E., Mathe, E., Dong, P.*, et al.* (2017). Myc Regulates Chromatin Decompaction and Nuclear Architecture during B Cell Activation. Mol Cell *67*, 566-578 e510.

Kim, M., and Costello, J. (2017). DNA methylation: an epigenetic mark of cellular memory. Exp Mol Med *49*, e322.

Kireev, I., Lakonishok, M., Liu, W., Joshi, V.N., Powell, R., and Belmont, A.S. (2008). In vivo immunogold labeling confirms large-scale chromatin folding motifs. Nat Methods *5*, 311-313.

Kizer, K.O., Phatnani, H.P., Shibata, Y., Hall, H., Greenleaf, A.L., and Strahl, B.D. (2005). A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. Mol Cell Biol *25*, 3305-3316.

Klosin, A., Casas, E., Hidalgo-Carcedo, C., Vavouri, T., and Lehner, B. (2017). Transgenerational transmission of environmental information in C. elegans. Science *356*, 320-323.

Koche, R.P., Smith, Z.D., Adli, M., Gu, H., Ku, M., Gnirke, A., Bernstein, B.E., and Meissner, A. (2011). Reprogramming factor expression initiates widespread targeted chromatin remodeling. Cell Stem Cell *8*, 96-105.

Koike-Yusa, H., Li, Y., Tan, E.P., Velasco-Herrera Mdel, C., and Yusa, K. (2014). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nat Biotechnol *32*, 267-273.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H.*, et al.* (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583-588.

Kriaucionis, S.H., N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in purkinje neurons and the brain. Science *324*, 929–930.

Krishnaiah, S.Y., Wu, G., Altman, B.J., Growe, J., Rhoades, S.D., Coldren, F., Venkataraman, A., Olarerin-George, A.O., Francey, L.J., Mukherjee, S.*, et al.* (2017). Clock Regulation of Metabolites Reveals Coupling between Transcription and Metabolism. Cell Metab *25*, 961-974 e964.

Kungulovski, G., and Jeltsch, A. (2016). Epigenome Editing: State of the Art, Concepts, and Perspectives. Trends Genet *32*, 101-113.

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

Kwon, D.Y., Zhao, Y.T., Lamonica, J.M., and Zhou, Z. (2017). Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC. Nat Commun *8*, 15315.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., Jenuwein, T (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120.

Laprell, F., Finkl, K., Muller, J (2017). Propagation of Polycomb-repressed chromatin requires sequence-specific recruitment to DNA. Science *356*, 85-88.

Lee, J.E., Wang, C., Xu, S., Cho, Y.W., Wang, L., Feng, X., Baldridge, A., Sartorelli, V., Zhuang, L., Peng, W.*, et al.* (2013). H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. Elife *2*, e01503.

Lehnertz, B., Ueda, Y., Derijck, A.A.H.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H.F.M. (2003). Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. Current Biology *13*, 1192-1200.

Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A.*, et al.* (2013). Naive pluripotency is associated with global DNA hypomethylation. Nat Struct Mol Biol *20*, 311-316.

Leung, D.C., and Lorincz, M.C. (2012). Silencing of endogenous retroviruses: when and why do histone marks predominate? Trends Biochem Sci *37*, 127-133.

Li, G., and Zhu, P. (2015). Structure and organization of chromatin fiber in the nucleus. FEBS Lett *589*, 2893-2904.

Li, W., Xu, H., Tengfei, X., Cong, L., Love, M.I, Zhang, F., Irizarry, R., Liu, J., Brown, M., Liu, S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biology *15*.

Lim, J.P., and Brunet, A. (2013). Bridging the transgenerational gap with epigenetic memory. Trends Genet *29*, 176-186.

Liu, S., Brind'Amour, J., Karimi, M.M., Shirane, K., Bogutz, A., Lefebvre, L., Sasaki, H., Shinkai, Y., and Lorincz, M.C. (2014). Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. Genes Dev *28*, 2041-2055.

Liu, X., Gao, Q., Li, P., Zhao, Q., Zhang, J., Li, J., Koseki, H., and Wong, J. (2013). UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. Nat Commun *4*, 1563.

Liu, X., Wang, C., Liu, W., Li, J., Li, C., Kou, X., Chen, J., Zhao, Y., Gao, H., Wang, H.*, et al.* (2016). Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. Nature *537*, 558-562.

Lonning, P.E., Eikesdal, H.P., Loes, I.M., and Knappskog, S. (2019). Constitutional Mosaic Epimutations - a hidden cause of cancer? Cell Stress *3*, 118-135.

Loyola, A., Tagami, H., Bonaldi, T., Roche, D., Quivy, J.P., Imhof, A., Nakatani, Y., Dent, S.Y., and Almouzni, G. (2009). The HP1alpha-CAF1-SetDB1-containing complex provides H3K9me1 for Suv39-mediated K9me3 in pericentric heterochromatin. EMBO Rep *10*, 769-775.

Lu, X., Simon, M.D., Chodaparambil, J.V., Hansen, J.C., Shokat, K.M., and Luger, K. (2008). The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. Nat Struct Mol Biol *15*, 1122-1124.

Luo, D., Carpenter, R., Vincent, C., Copsey, L., Coen, E. (1995). Origin of floral asymmetry in Antirrhinum. Nature *383*, 794-799.

Lupianez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R.*, et al.* (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012-1025.

Madan, B., Madan, V., Weber, O., Tropel, P., Blum, C., Kieffer, E., Viville, S., and Fehling, H.J. (2009). The pluripotency-associated gene Dppa4 is dispensable for embryonic stem cell identity and germ cell development but essential for embryogenesis. Mol Cell Biol *29*, 3186-3203.

Mali, P., Esvelt, K.M., and Church, G.M. (2013). Cas9 as a versatile tool for engineering biology. Nature Methods *10*, 957-963.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., 3rd, Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V.*, et al.* (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762-767.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A.*, et al.* (2012). The transcriptional and epigenomic foundations of ground state pluripotency. Cell *149*, 590-604.

Martin, C., and Zhang, Y. (2005). The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol *6*, 838-849.

Marzluff, W.F., Gongidi, P., Woods, K.R., Jin, J., and Maltais, L.J. (2002). The Human and Mouse Replication-Dependent Histone Genes. Genomics *80*, 487-498.

Masaki, H., Nishida, T., Sakasai, R., and Teraoka, H. (2010). DPPA4 modulates chromatin structure via association with DNA and core histone H3 in mouse embryonic stem cells. Genes Cells *15*, 327-337.

Matsui, T., Leung, D., Miyashita, H., Maksakova, I.A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M.C., and Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature *464*, 927-931.

Mayr, E. (1980). The Evolutionary Synthesis. (eds Mayr, E & Provine, W B), 1-48.

McClintock, B. (1951). Chromosome organization and genic expression. Cold Spring Harb Symp Quant Biol *16*, 16-47.

McKinley, K.L., and Cheeseman, I.M. (2017). Large-Scale Analysis of CRISPR/Cas9 Cell-Cycle Knockouts Reveals the Diversity of p53-Dependent Responses to Cell-Cycle Defects. Dev Cell *40*, 405-420 e402.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B.*, et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-770.

Messerschmidt, D.M., de Vires, W., Ito, M., Solter, D., Ferguson-Smith, A, Knowles, B.B. (2012). Trim28 Is Required for Epigenetic Stability During Mouse Oocyte to Embryo Transition. Science *335*, 1499-1502.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P.*, et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Milagre, I., Stubbs, T.M., King, M.R., Spindel, J., Santos, F., Krueger, F., Bachman, M., Segonds-Pichon, A., Balasubramanian, S., Andrews, S.R.*, et al.* (2017). Gender Differences in Global but Not Targeted Demethylation in iPSC Reprogramming. Cell Rep *18*, 1079-1089.

Mishra, K., and Kanduri, C. (2019). Understanding Long Noncoding RNA and Chromatin Interactions: What We Know So Far. Noncoding RNA *5*.

Miska, E.A., Ferguson–Smith, A.C. (2016). Transgenerational inheritance: Models and mechanisms of non–DNA sequence–based inheritance. Science *354*, 59-63.

Mojica, F.J.M., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology *155*, 733-740.

Morgan, H.D.S., H.G.E.; Martin, D.I.K; Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. Nature genetics *23*, 314-318.

Morita, S., Noguchi, H., Horii, T., Nakabayashi, K., Kimura, M., Okamura, K., Sakai, A., Nakashima, H., Hata, K., Nakashima, K.*, et al.* (2016). Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. Nat Biotechnol *34*, 1060-1065.

Muller, H.J.A., E. (1930). The frequency of translocations produced by X-rays in Drosophila. . Genetics *15*, 283-311.

Müller, M.M., Fierz, B., Bittova, L., Liszczak, G., and Muir, T.W. (2016). A two-state activation mechanism controls the histone methyltransferase Suv39h1. Nat Chem Biol *12*, 188-193.

Najm, F.J., Strand, C., Donovan, K.F., Hegde, M., Sanson, K.R., Vaimberg, E.W., Sullender, M.E., Hartenian, E., Kalani, Z., Fusi, N.*, et al.* (2018). Orthologous CRISPR-Cas9 enzymes for combinatorial genetic screens. Nat Biotechnol *36*, 179-189.

Nakamura, T., Liu, Y.J., Nakashima, H., Umehara, H., Inoue, K., Matoba, S., Tachibana, M., Ogura, A., Shinkai, Y., and Nakano, T. (2012). PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. Nature *486*, 415-419.

Nakayama, J.R., J.C; Strahl, B.D; Allis, C.D; Grewal, S.I.S. (2001). Role of Histone H3 Lysine 9 Methylation in Epigenetic Control of Heterochromatin Assembly. Science *292*, 110-113.

Nanney, D.L. (1958). Epigenetic control systems. Proc Natl Acad Sci U S A *44(7)*, 712-717.

Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science *347*, 1017-1021.

Ng, K.K., Yui, M.A., Mehta, A., Siu, S., Irwin, B., Pease, S., Hirose, S., Elowitz, M.B., Rothenberg, E.V., and Kueh, H.Y. (2018). A stochastic epigenetic switch controls the dynamics of T-cell lineage commitment. Elife *7*.

Ng, R.K., and Gurdon, J.B. (2008). Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription. Nat Cell Biol *10*, 102-109.

Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. Cell Stem Cell *4*, 487-492.

Nishiyama, A., Yamaguchi, L., Sharif, J., Johmura, Y., Kawamura, T., Nakanishi, K., Shimamura, S., Arita, K., Kodama, T., Ishikawa, F.*, et al.* (2013). Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. Nature *502*, 249-253.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J.*, et al.* (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381-385.

O'Carroll, D., Erhardt, S., Pagani, M., Barton, S.C., Surani, M.A., and Jenuwein, T. (2001). The polycomb-group gene Ezh2 is required for early mouse development. Mol Cell Biol *21*, 4330-4336.

Okano, M., Bell, D.W., and Haber, D.A. (1999). DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. Cell *99*, 247–257.

Okano, M., Xie, S; Li, En (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nature Genetics *19*, 219-220.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D.*, et al.* (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714-717.

Orom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q.*, et al.* (2010). Long noncoding RNAs with enhancer-like function in human cells. Cell *143*, 46-58.

Padmanabhan, N., Jia, D., Geary-Joo, C., Wu, X., Ferguson-Smith, A.C., Fung, E., Bieda, M.C., Snyder, F.F., Gravel, R.A., Cross, J.C.*, et al.* (2013). Mutation in folate metabolism causes epigenetic instability and transgenerational effects on development. Cell *155*, 81-93.

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell *32*, 232-246.

Panzeri, I., and Pospisilik, J.A. (2018). Epigenetic control of variation and stochasticity in metabolic disease. Mol Metab *14*, 26-38.

Perez-Pinera, P., Ousterout, D.G., and Gersbach, C.A. (2012). Advances in targeted genome editing. Curr Opin Chem Biol *16*, 268-277.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W., Solovei, I., Brugman, W., Graf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M.*, et al.* (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol Cell *38*, 603-613.

Petryk, N.D., M; Wenger, A; Stromme, C.B; Strandsby, A; Andersson, R; Groth, A (2018). MCM2 promotes symmetric inheritance of modified histones during DNA replication. Science *361*, 1389-1392.

Pflueger, C., Tan, D., Swain, T., Nguyen, T., Pflueger, J., Nefzger, C., Polo, J.M., Ford, E., and Lister, R. (2018). A modular dCas9-SunTag DNMT3A epigenome editing system overcomes pervasive off-target activity of direct fusion dCas9-DNMT3A constructs. Genome Res *28*, 1193-1206.

Platero, J.S., Hartnett, T., and Eissenberg, J.C. (1995). Functional analysis of the chromo domain of HP1. The EMBO Journal *14*, 3977-3986.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173-1183.

Qin, W., Wolf, P., Liu, N., Link, S., Smets, M., La Mastra, F., Forne, I., Pichler, G., Horl, D., Fellinger, K.*, et al.* (2015). DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. Cell Res *25*, 911-929.

Quenneville, S., Turelli, P., Bojkowska, K., Raclot, C., Offner, S., Kapopoulou, A., and Trono, D. (2012). The KRAB-ZFP/KAP1 system contributes to the early embryonic establishment of site-specific DNA methylation patterns maintained during development. Cell Rep *2*, 766-773.

Radford, E.J. (2018). Exploring the extent and scope of epigenetic inheritance. Nat Rev Endocrinol *14*, 345-355.

Radford, E.J., Ito, M., Shi, H., Corish, J.A., Yamazawa, K., Isganaitis, E., Seisenberger, S., Hore, T.A., Reik, W., Erkek, S.*, et al.* (2014). In utero effects. In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism. Science *345*, 1255903.

Ragunathan, K., Jih, G., and Moazed, D. (2015). Epigenetics. Epigenetic inheritance uncoupled from sequence-specific recruitment. Science *348*, 1258699.

Rechavi, O., Houri-Ze'evi, L., Anava, S., Goh, W.S.S., Kerk, S.Y., Hannon, G.J., and Hobert, O. (2014). Starvation-induced transgenerational inheritance of small RNAs in C. elegans. Cell *158*, 277-287.

Rechavi, O., Minevich, G., and Hobert, O. (2011). Transgenerational inheritance of an acquired small RNA-based antiviral response in C. elegans. Cell *147*, 1248-1256.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. Nature *447*, 425-432.

Reik, W., Dean, W., and Walter, J.r. (2001). Epigenetic Reprogramming in Mammalian Development. Science *293*, 1089- 1093.

Reinberg, D.V., Lynne D. (2018). Chromatin domains rich in inheritance. Science *361*, 33-34.

Remy, J.J. (2010). Stable inheritance of an acquired behavior in Caenorhabditis elegans. Curr Biol *20*, R877-878.

Reveron-Gomez, N., Gonzalez-Aguilera, C., Stewart-Morgan, K.R., Petryk, N., Flury, V., Graziano, S., Johansen, J.V., Jakobsen, J.S., Alabert, C., and Groth, A. (2018). Accurate Recycling of Parental Histones Reproduces the Histone Modification Landscape during DNA Replication. Mol Cell *72*, 239-249 e235.

Richards, E. (2006). Inherited epigenetic variation — revisiting soft inheritance. Nature Review Genetics *7*, 395-401.

Riising, E.M., Comet, I., Leblanc, B., Wu, X., Johansen, J.V., and Helin, K. (2014). Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. Mol Cell *55*, 347-360.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E.*, et al.* (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311-1323.

Rossant, J., and Tam, P.P. (2009). Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. Development *136*, 701-713.

Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P.V., Layard-Liesching, H., Verp, S., Marquis, J.*, et al.* (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. Nature *463*, 237-240.

Saxton, D.S., and Rine, J. (2019). Epigenetic memory independent of symmetric histone inheritance. Elife *8*.

Schmidt, J.A., Rinaldi, S., Scalbert, A., Ferrari, P., Achaintre, D., Gunter, M.J., Appleby, P.N., Key, T.J., and Travis, R.C. (2016). Plasma concentrations and intakes of amino acids in male meat-eaters, fish-eaters, vegetarians and vegans: a cross-sectional analysis in the EPIC-Oxford cohort. Eur J Clin Nutr *70*, 306-312.

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev *18*, 1251-1262.

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

Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., and Rauscher, F.J., 3rd (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes Dev *16*, 919-932.

Schwarz, P.M.H., J.C. (1994). Formation and stability of higher order chromatin structures. Contributions of the histone octamer. J Biol Chem *269(23)*, 16284-16289.

Seong, K.H., Li, D., Shimizu, H., Nakamura, R., and Ishii, S. (2011). Inheritance of stress-induced, ATF-2-dependent epigenetic change. Cell *145*, 1049-1061.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458-472.

Shalem, O., Sanjana, N.E., and Zhang, F. (2015). High-throughput functional genomics using CRISPR-Cas9. Nat Rev Genet *16*, 299-311.

Sharif, J., and Koseki, H. (2017). No Winter Lasts Forever: Polycomb Complexes Convert Epigenetic Memory of Cold into Flowering. Dev Cell *42*, 563-564.

Sharma, U., Conine C.C., Shea, J.M., Boskovic, A., Derr, A.G., Bing, X.Y., Belleannee, C., Li, X.Z., Fauquier, L., Moore, M.J., Sullivan, R., Mello, C.C., Garber, M., Rano, O.J. (2016). Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. Science *351*, 391-396.

Shechner, D.M., Hacisuleyman, E., Younger, S.T., and Rinn, J.L. (2015). Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. Nat Methods *12*, 664-670.

Shen, X.G., M.A. (1996). Linker histone H1 regulates specific gene expression but not global transcription in vivo. Cell *86(3)*, 475-483.

Shilatifard, A. (2012). The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annu Rev Biochem *81*, 65-95.

Siklenka, K., Erkek, S., Godmann, M., Lambrot, R., McGraw, S., Lafleur, C., Cohen, T., Xia, J., Suderman, M., Hallett, M.*, et al.* (2015). Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. Science *350*, aab2006.

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T.B., Webster, Z., Peters, A.H.F.M., Jenuwein, T., Otte, A.P., Brockdorff, N (2003). Establishment of Histone H3 Methylationon the Inactive X Chromosome Requires TransientRecruitment of Eed-Enx1 Polycomb Group Complexes. Developmental Cell *4*, 481-495.

Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife *6*.

Smith, Z.D., Chan, M.M., Humm, K.C., Karnik, R., Mekhoubad, S., Regev, A., Eggan, K., and Meissner, A. (2014). DNA methylation dynamics of the human preimplantation embryo. Nature *511*, 611-615.

Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature *484*, 339-344.

Song, J.R., O.; Bestor, T. H. & Patel, D. J. (2011). Structure of DNMT1–DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science *331*, 1036–1040.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell *151*, 994-1004.

Stepper, P., Kungulovski, G., Jurkowska, R.Z., Chandra, T., Krueger, F., Reinhardt, R., Reik, W., Jeltsch, A., and Jurkowski, T.P. (2017). Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase. Nucleic Acids Res *45*, 1703-1713.

Stern, S., Fridmann-Sirkis, Y., Braun, E., and Soen, Y. (2012). Epigenetically heritable alteration of fly development in response to toxic challenge. Cell Rep *1*, 528-542.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J.A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature *507*, 62-67.

Stewart, M.D., Li, J., and Wong, J. (2005). Relationship between histone H3 lysine 9 methylation, transcription repression, and heterochromatin protein 1 recruitment. Mol Cell Biol *25*, 2525-2538.

Stewart-Morgan, K.R., Petryk, N., and Groth, A. (2020). Chromatin replication and epigenetic cell memory. Nat Cell Biol *22*, 361-371.

Stewart-Morgan, K.R., Reveron-Gomez, N., and Groth, A. (2019). Transcription Restart Establishes Chromatin Accessibility after DNA Replication. Mol Cell *75*, 284-297 e286.

Strahl, B.D., Allis, C.D (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Syding, L.A., Nickl, P., Kasparek, P., and Sedlacek, R. (2020). CRISPR/Cas9 Epigenome Editing Potential for Rare Imprinting Diseases: A Review. Cells *9*.

Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H*., et al.* (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes Dev *16*, 1779-1791.

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

Tang, W.W., Dietmann, S., Irie, N., Leitch, H.G., Floros, V.I., Bradshaw, C.R., Hackett, J.A., Chinnery, P.F., and Surani, M.A. (2015). A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. Cell *161*, 1453-1467.

Tee, W.W., and Reinberg, D. (2014). Chromatin features and the epigenetic regulation of pluripotency states in ESCs. Development *141*, 2376-2390.

Terns, M.P., and Terns, R.M. (2011). CRISPR-based adaptive immune systems. Curr Opin Microbiol *14*, 321-327.

Torres-Garcia, S., Yaseen, I., Shukla, M., Audergon, P., White, S.A., Pidoux, A.L., and Allshire, R.C. (2020). Epigenetic gene silencing by heterochromatin primes fungal resistance. Nature *585*, 453-458.

Trevino, L.S., Dong, J., Kaushal, A., Katz, T.A., Jangid, R.K., Robertson, M.J., Grimm, S.L., Ambati, C.S.R., Putluri, V., Cox, A.R.*, et al.* (2020). Epigenome environment interactions accelerate epigenomic aging and unlock metabolically restricted epigenetic reprogramming in adulthood. Nat Commun *11*, 2316.

Trojer, P., Li, G., Sims, R.J., 3rd, Vaquero, A., Kalakonda, N., Boccuni, P., Lee, D., Erdjument-Bromage, H., Tempst, P., Nimer, S.D.*, et al.* (2007). L3MBTL1, a histone-methylation-dependent chromatin lock. Cell *129*, 915-928.

Tse, C.S., T; Wolffe, A.P; Hansen, JC (1998). Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III. Mol Cell Biol *18(8)*, 4629-4638.

Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S., Sakaue, M., Matsuoka, C., Shimotohno, K., Ishikawa, F., Li, E., Ueda, H.R.*, et al.* (2006). Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cells *11*, 805-814.

Tumbar, T., Belmont, A.S. (2001). Interphase movements of a DNA chromosome region modulated by VP16 transcriptional activator. Nature Cell Biology *3*, 134-139.

Valton, J., Dupuy, A., Daboussi, F., Thomas, S., Marechal, A., Macmaster, R., Melliand, K., Juillerat, A., and Duchateau, P. (2012). Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. J Biol Chem *287*, 38427-38432.

Vojta, A., Dobrinic, P., Tadic, V., Bockor, L., Korac, P., Julg, B., Klasic, M., and Zoldos, V. (2016). Repurposing the CRISPR-Cas9 system for targeted DNA methylation. Nucleic Acids Res *44*, 5615-5628.

Waddington, C.H. (1942a). Canalization of development and the inheritance of acquired characters. Nature *150*, 563–565.

Waddington, C.H. (1942b). The epigenotype. Endeavour *1*, 18–20

Walker, C.L.H., Shuk-mei (2012). Developmental reprogramming of cancer susceptibility. Nature Reviews *12*, 479-486.

Wang, C., Liu, X., Gao, Y., Yang, L., Li, C., Liu, W., Chen, C., Kou, X., Zhao, Y., Chen, J.*, et al.* (2018). Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. Nat Cell Biol *20*, 620-631.

Wang, X.M., Danesch. (2017). DNA sequence-dependent epigenetic inheritance of gene silencing and histone H3K9 methylation. Science *356*, 88-91.

Wang, Y., Liu, H., and Sun, Z. (2017). Lamarck rises from his grave: parental environment-induced epigenetic inheritance in model organisms and humans. Biol Rev Camb Philos Soc *92*, 2084-2111.

Waterland, R.A., and Jirtle, R.L. (2003). Transposable elements: targets for early nutritional effects on epigenetic gene regulation. Mol Cell Biol *23*, 5293-5300.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet *39*, 457-466.

Weinberger, L., Ayyash, M., Novershtern, N., and Hanna, J.H. (2016). Dynamic stem cell states: naive to primed pluripotency in rodents and humans. Nat Rev Mol Cell Biol *17*, 155-169.

Weiner, A., Lara-Astiaso, D., Krupalnik, V., Gafni, O., David, E., Winter, D.R., Hanna, J.H., and Amit, I. (2016). Co-ChIP enables genome-wide mapping of histone mark co-occurrence at single-molecule resolution. Nat Biotechnol *34*, 953-961.

Wolffe, A.P., Matzke, M.A. (1999). Epigenetics: Regulation Through Repression. Science *286*, 481-486.

Woodcock, C.L., and Ghosh, R.P. (2010). Chromatin higher-order structure and dynamics. Cold Spring Harb Perspect Biol *2*, a000596.

Wu, W.H., Alami, S., Luk, E., Wu, C.H., Sen, S., Mizuguchi, G., Wei, D., and Wu, C. (2005). Swc2 is a widely conserved H2AZ-binding module essential for ATP-dependent histone exchange. Nat Struct Mol Biol *12*, 1064-1071.

Wyatt, G.R. (1950). Occurrence of 5-Methyl-Cytosine in Nucleic Acids. Nature *166*, 237-238.

Xu, M.L., C; Chen, X; Huang, C; Chen, S; Zhu, B. (2010). Partitioning of Histone H3-H4 Tetramers During DNA Replication–Dependent Chromatin Assembly. Science *328*, 94-98.

Xu, X., Tao, Y., Gao, X., Zhang, L., Li, X., Zou, W., Ruan, K., Wang, F., Xu, G.L., and Hu, R. (2016). A CRISPR-based approach for targeted DNA demethylation. Cell Discov *2*, 16009.

Yang, H., Pesavento, J.J., Starnes, T.W., Cryderman, D.E., Wallrath, L.L., Kelleher, N.L., and Mizzen, C.A. (2008). Preferential dimethylation of histone H4 lysine 20 by Suv4-20. J Biol Chem *283*, 12085-12092.

Yang, H., Berry, S., Olsson, T.S.G., Hertley, M., Howard, M., Dean, C (2017). Distinct phases of Polycomb silencing to hold epigenetic memory of cold in Arabidopsis. Science *357*, 1142-1145.

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F.*, et al.* (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science *356*.

Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519-523.

Yoder, J.A., Walsh, C. P. & Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet *13*, 335–340.

Yu, R., Wang, X., and Moazed, D. (2018). Epigenetic inheritance mediated by coupling of RNAi and histone H3K9 methylation. Nature *558*, 615-619.

Zalatan, J.G., Lee, M.E., Almeida, R., Gilbert, L.A., Whitehead, E.H., La Russa, M., Tsai, J.C., Weissman, J.S., Dueber, J.E., Qi, L.S., *et al.* (2015). Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. Cell *160*, 339-350.

Zhang, K., Mosch, K., Fischle, W., and Grewal, S.I. (2008). Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. Nat Struct Mol Biol *15*, 381-388.

Zhao, J., Wang, M., Chang, L., Yu, J., Song, A., Liu, C., Huang, W., Zhang, T., Wu, X., Shen, X., *et al.* (2020). RYBP/YAF2-PRC1 complexes and histone H1-dependent chromatin compaction mediate propagation of H2AK119ub1 during cell division. Nat Cell Biol.

Zhao, Q., Zhang, J., Chen, R., Wang, L., Li, B., Cheng, H., Duan, X., Zhu, H., Wei, W., Li, J., *et al.* (2016). Dissecting the precise role of H3K9 methylation in crosstalk with DNA maintenance methylation in mammals. Nat Commun *7*, 12464.

Zheng, C.H., J.J. (2003). Structures and Interactions of the Core Histone Tail Domains. Biopolymers *68*, 539-546.

Zheng, H., Huang, B., Zhang, B., Xiang, Y., Du, Z., Xu, Q., Li, Y., Wang, Q., Ma, J., Peng, X., *et al.* (2016). Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. Mol Cell *63*, 1066-1079.

9 Appendices

9.1 Tables

Table 9.1 List of cloning primers used for epigenetic editing tool construction

Final Construct	Primer name	sequence (5'-3')	Donor vector	Recipient vector
pPB_TRE3G::ScFv-GFP_EF1a::Neo	TRE3G-scFv_Fw	AAGGTCTAGAGCTAGCCATGGGCCCCGACATCGTGA	pPlatTET-gRNA2 (Addgene #82559)	pPB_CAG rtTA-IP
	Linker_Neo:Rv	ATGATCTTTTGGTACCACCGCCTCCGGATCCG		
pPB_TRE3G::ScFv-KRAB-GFP_EF1a::Neo	SV40NLS-KRAB-scFValone.Fw	GATCCGGAGGCGGTGGTACCCGGACACTGGTGAC	pPB_TRE3G::KRAB-dCas9	pPB_TRE3G::ScFv-GFP_EF1a::Neo
	KRAB-scFValone.Rv	GATTATGATCTTTTGTTAGGGCTCTTCTCCCTTCTCCAAC		
pPB_TRE3G::ScFv-Dnmt3a3L-GFP_EF1a::Neo	SV40-Dnmt3a3L-scFValone.Fw	GATCCGGAGGCGGTGGTACCAACCATGACCAG	pET28-Dnmt3a3L-sc27	pPB_TRE3G::ScFv-GFP_EF1a::Neo
	Dnmt3a3L-scFValone.REV	GATTATGATCTTTTGGTACCCTAAAGAGGAAGTGAGTTTTGAG		
pPB_TRE3G::dCas9-5XGCN4_*EF1a::*TetOn-Hygro	TRE3G-dCas9_Fw	AAGGTCTAGAGCTAGCGTCGACACCGGGGCCC	pPlatTET-gRNA2 (Addgene #82559)	pPB_CAG rtTA-IP
	pPB_dCas9_Rv	TAGCCTCCCCCGTTTAAACGGATCTGCTAGCCCCT		
pPB_U6::gRNA_EF1a::BFP-Puro	pPB-U6_Fw	CCGGGCCCGCTCTAGAGATCCGACGCGCCATCT	pU6-sgRNA-EF1a-puro-T2A-BFP (Addgene #60955)	pPB_CAG rtTA-IP
	pPB-BFP_Rv	GGCACAAGCTTAATTAAGAATTCGTCGAGGGAC		

Table 9.2 List of gRNAs used for epigenetic editing

DNA target	gRNA sequence (5'-3')
Adamts7 promoter	TGCCTCGGCAGCTGGCTGAG
Cdh1 promoter	GGCTAGGATTCGAACGACCG
Esg1 gene body +87	CCACGCACGGCCCACAGCT
Esg1 promoter -475	GAGAAAAACCATCCAAACCA
Greb1 promoter	GCCCGCAAGACCCGGACGGG
Jade1 ICR	CGGGAGCAAGCCCCTACCAG
Mest ICR	CCGCAGCGATACAGGATCGG
p53 promoter -345	TATGTCAGATGCTGTAGTG
Peg3 ICR	GTGCACTAGACTGCCGACCC
Plagl1 ICR	CCACAGAGATGATCCACCCT
Pten promoter	CTGTCATGTCTGGGAGCCTG

Table 9.3 List of gRNAs used to generate clonal knockout lines

Gene target	gRNA sequence (5'-3')
*Dot1L*_gRNA#1	GTGTGTCCCATCAGTACTTG
*Dot1L*_gRNA#2	TGGCTGGCCGGAAACGTGGG
*Dppa2*_gRNA#1	GATAGATACCTGGTGGTGTG
*Dppa2*_gRNA#2	CGGGTGCCAAAGGAAAAAGG
*Kmt2d*_gRNA#1	AAATGGCTGTTGATCCCATG
*Kmt2d*_gRNA#2	CTCAGGTAGGTAGATTGGCA
*Smarcc1*_gRNA#1	TCAGCTAACACATTGAAGTG
Smarcc1_gRNA#2	TCTTTCCTCATGGAACAAAG
*Zmym3*_gRNA#1	CGAGAAGCTTCGATTCAGTG
*Zmym3*_gRNA#2	CCAACTTGGATACCCAGAGT

Table 9.4 List of primers used for PCR genotyping of clonal knockout lines

Gene Target	Primer direction	sequence (5'-3')	Amplicon size in the knockout (bp)
Zmym3	Forward	TTAGAAAGCCCAGGCCAGTAC	180
	Reverse	GTTCTGCTGCCGTGATTGCT	
Kmt2d	Forward	TCACCTGCTGTTCATGTCCA	286
	Reverse	CATCTCCCATGCGGTTCACT	
Dot1L	Forward	ATGGTCTTGTGAGCACTCCT	294

	Reverse	GACAGGGTTTCTCTGTAACC	
Smarcc1	Forward	TCAGCTCTCCTGAGCAGCTA	388
	Reverse	GAAACCTATTGGTCGGCAGC	
Dppa2	Forward	CGCCTGCTTCCTGAGGTTAT	400
	Reverse	CCTCTGCCTCCCAAGTGTT	

Table 9.5 List of antibodies used

Antibody	Use	Antibody Supplier	Catalog Number
Anti-DPPA2	Western blot/CUT&RUN	Sigma	AB4356
Anti-DPPA4	Western blot	R&D Systems	AF3730
Anti-H3K4me3	CUT&RUN	Diagenode	C15410003
Anti-H3K9me3	CUT&RUN	Abcam	ab8898
Anti-H4K20me3	CUT&RUN	Abcam	ab9053
Anti-Rabbit IgG	Western blot	Cell Signaling	7074
Anti-mouse IgG	Western blot	Abcam	ab6709
Anti-Goat IgG	Western blot	Fisher Scientific	PI31402

Table 9.6 List of qPCR primers for gene expression

Gene name	sequence (5'-3')
*Adamts7*_PCR_fw	CCTGCATCAGGTCTTGTGACT
*Adamts7*_qPCR_Rv	ACAGCTCATAGGACAGGAAG
*Cdh1*_qPCR_fw	GGGCAGAGTGAGATTTGAAG
*Cdh1*_qPCR_rv	GTCTGTCGCCACTTTGAATC
*Greb1*_qPCR_fw	AATGCCCTGCTTGGTTTC
*Greb1*_qPCR_rv	GATGTGGTTGGAGAACTC
*Igll1*_qPCR_Fw	CCATAGGCTTCCATCTAAGC
*Igll1*_qPCR_rv	ACTTGGGCTGACCTAGGATTG
*Jade1*_qPCR_Fw	GACAATGGCAGCTTGTCAAC
*Jade1*_qPCR_Rv	CCTAAACACCTCAGAAGGCT

*Mest*_qPCR_Fw2	CAAAATACTGCTTCGCTTGG
*Mest*_qPCR_Rv2	CTCTCATCCTTTGACTTTGG
*Plagl1*_qPCR_Fw	CACCTTCTTTCTGTCTGGGC
*Plagl1*_qPCR_Rv	GAGCCATGGCCTTTGGTTCT
*Pten*_qPCR_Fw	ATCACCTGGATTACAGACCC
*Pten*_qPCR_rv	ACTGAGGATTGCAAGTTCCG
*Rplp0*_qPCR_Fw	TCCAGAGGCACCATTGAAATT
*Rplp0*_qPCR_Rv	TCGCTGGCTCCCACCTT

Table 9.7 List of pyrosequencing primers

DNA target	Primer Type	sequence (5'-3')
P53 promoter	Forward	GGATAGGAAAGAGTATAGAGTTTAGAATAG
	Reverse	[Btn]ATTCTCCCTAAAATATTACCCTCAACA
	Sequencing	GATATAGAGGTAGGAGTT
Esg1 TSS 0	Forward	[Btn]GATTGGATTTGGAATATTTTTGGAT
	Reverse	ACACCCCCCACACAAATACTAAAACTTC
	Sequencing	CCTAAAATCTAATTCCTTAACAA
Esg1 promoter 200	Forward	GGTTTTGGGAATAGAAGTTAGTGTTG
	Reverse	[Btn]ACATCTAACAAAATAAAATCCCAATATCT
	Sequencing	AAGTTAGTGTTGGTGA
Esg1 promoter 800	Forward	GATAGAAGGGTATTAAATGTGTTAATGTT
	Reverse	[Btn]CCAACTCTCAAATTCCCACTTTTAT
	Sequencing	GGTATTAAATGTGTTAATGTTTTG

Table 9.8 Dispensation order for bisulfate pyrosequencing analysis

DNA target	sequence (5'-3')

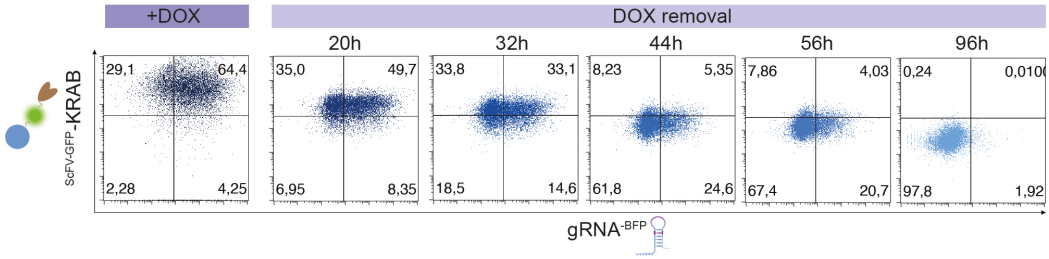
Esg1 TSS 0	GTAGTCGTGTTACTAGATGATCGTATATGTCTGTTATGGTATTATGTCAGTTCTGTAGTATGTTGGAG ATAGTATGTCTGTG
Esg1 promoter 200	ATTGACTGTATCGGTGATTAAGTGAGATGATCGTGAGTAGATTAAGATATAGTTGATTCTGAG
Esg1 promoter 800	AGATGATCGATATTATGCTAGAGAATATATGTGATAGTATATAATGTAGATGTATAGAGTGTGAGTA TATGAGTAGAGAGTGTAGTATGTCGAGTGTGTGTGAGTATTCTG
p53 promoter	ATTCGTATCGTTATATTGTTATAGTTATATTAGTATCGATTATATGTCG

Table 9.9 List of CUT&RUN-qPCR primers

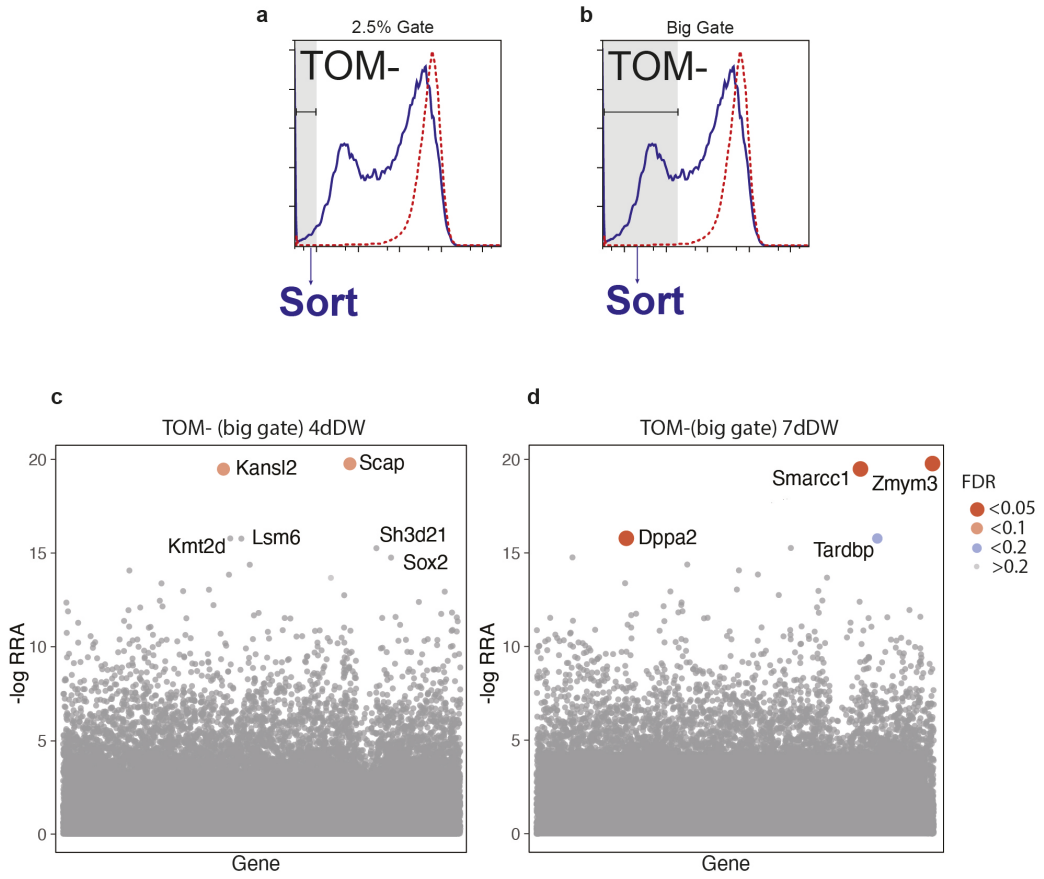
DNA target	Primer direction	sequence (5'-3')
p53 promoter 300	Forward	TGGCTACAAAGACTCTGT
	Reverse	CTATCCAGCTAGATAGTC
p53 promoter 50	Forward	TATCCAGCTTTGTGCCAG
	Reverse	AACTTTAGCCAGGGTGAG
p53 gene body 600	Forward	GCTTAAGACTTAAGACCC
	Reverse	TCTCATCCAGGAACGGAA
Esg1 promoter 700	Forward	GGGCATTAAATGTGCTAATGTTC
	Reverse	CTTCCTTGCATTGAAATAGTCG
Esg1 promoter 200	Forward	GGCAGCGAGGTCAAGAGTAG
	Reverse	CAGTCGCTGGTGCTGAAATA
Oct4 promoter	Forward	TGGGCTGAAATACTGGGTTC
	Reverse	TTGAATGTTCGTGTGCCAAT
Hbby promoter	Forward	TGCTCCAGTCTCAGGATTCA
	Reverse	CAGACATTTGGTGTCTCGGTA
H3K9me3_control1	Forward	CGTCTAGGTCCTCCCAATGA
	Reverse	TTGGAGCTCAGGAAAAAGGA
H3K9me3_control2	Forward	CCATCTGATCCAGGGGTAAA
	Reverse	GCAACTGTAGACGTGCCAGA
Gapdh promoter	Forward	AATTGGAGGAGGCTCAGAGG

	Reverse	GAAAGGGGCAGTGTCTCCTA

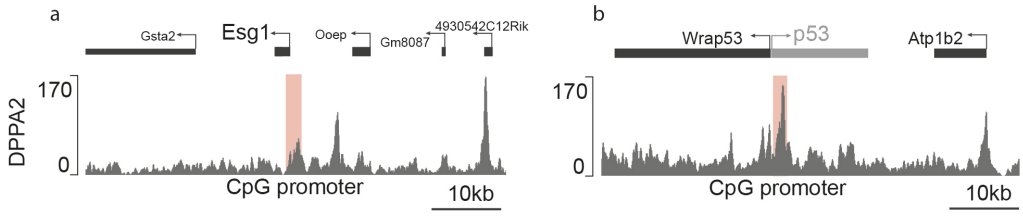
9.2 Supplementary figures



Supplementary Fig. 9.1. The epigenome editing system is progressively switched off upon DOX removal. Density plots show expression of GFP (KRAB $^{-GFP-scFv}$) and BFP (gRNA $^{-BFP}$) at the single cell level after DOX induction and at several timepoints of DOX washout in cells with stable integration of dCas9 -5XGCN4 and KRAB $^{-GFP-scFv}$ and transient transfection of gRNA -BFP .



Supplementary Fig. 9.2. Gate used to sort cells for the unbiased CRISPR screen and supplementary scatterplots. (a) and (b) histogram plots show distribution of cells according to tdTomato intensity. Grey box represents the gate used to sort tdTomato- cells. A more stringent gate was used to sort the bottom 2.5% TOM- cells (a) and a wider gate was used to include all tdTomato- cells designed according to a tdTomato+ sample (b). (c) and (d) Scatterplots show significant hits from the screen analysed by the -log relative ranking algorithm (RRA) from the cells sorted using the wider gate in b at each given timepoint. Size and color of the dots represents False discovery rate (FDR) as shown in the legend on the right.



Supplementary Fig. 9.3. DPPA2 binding at Esg1 and p53 regions. (a) and (b) Chromosome tracks represent the enrichment of DPPA2 binding in wildtype cells estimated by CUT&RUN. Each window is 60 kb wide. Red box highlights the position of the promoter of interest.