Regulation of heterochromatin by a pluripotency-associated long non coding RNA in mouse embryonic stem cells and in oocytes : implications for early embryogenesis

Laure Bernard

Sorbonne Université

Ecole doctorale Complexité du Vivant (ED515)

Epigénomique, Prolifération et Identité Cellulaire – Institut Pasteur

Regulation of heterochromatin by a pluripotency-associated long non coding RNA in mouse embryonic stem cells and in oocytes : implications for early embryogenesis

Par Laure BERNARD

Thèse de doctorat de Biologie

Dirigée par Pablo NAVARRO GIL

Présentée et soutenue publiquement le 27 Septembre 2021

Devant un jury composé de :

Présidente : Frédérique PERONNET, Directrice de recherche

Rapporteur : Antoine MOLARO, Chargé de recherche

Rapportrice : Amélie BONNET-GARNIER, Chargée de recherche

Examinateur : Pierre SAVATIER, Directeur de recherche

Examinatrice : Claire FRANCASTEL, Directrice de recherche

Directeur de thèse : Pablo NAVARRO GIL, Directeur de recherche

There is no ends, just new beginnings.

Abstract

The chromatin is a highly dynamic structure composed by: euchromatin, associated with transcriptionally active regions and heterochromatin, associated with gene repression in a cell-type-specific manner and at permanently silent regions such as pericentromeric regions. The control of heterochromatin, especially via histone H3 lysine 9 (H3K9) methylation, is essential for proper developmental progression. In mouse Embryonic Stem cells (mESCs), low levels of H3K9 methylation have been associated with pluripotency. In somatic cells, it has been shown to represent an epigenetic barrier for reprogramming back to pluripotency. How global H3K9 methylation levels are coupled with pluripotency remains largely unknown.

We identify an antisense long non-coding RNA to *Suv39h1* gene encoding a major H3K9 di- and tri-methylase. This antisense (*Suv39h1as*) is activated by the pluripotency transcription factor OCT4, leading to a downregulation of SUV39H1 expression. This genetic circuit couples heterochromatin control to pluripotency. *Suv39h1as* promoter was deleted by CRISPR/Cas9 technology to study the mechanistic basis of *Suv39h1* repression by its antisense as well as the functional consequences of the resultant increase in H3K9 methylation.

We first investigated the mechanism of *Suv39h1* regulation by *Suv39h1as*. We found that *Suv39h1as* controls the frequency of *Suv39h1* transcription and has no impact on *Suv39h1* mRNA stability. *Suv39h1as* establishes a complex local chromatin, by triggering euchromatin through the gene body but not at the promoter. Furthermore, a set of evidences suggest that other mechanisms might be involved in *Suv39h1* regulation. We hypothesized that Suv39h1as transcription can possibly interferes with *Suv39h1* transcription and/or modulates Suv39h1 isoforms. Nevertheless, mechanism has not been fully elucidated and remained in question.

We explored functional impact of SUV39H1 increase. First, subsequent H3K9me2/me3 increase has limited impact on mESCs. They are able to self-renew and differentiate. This is in line with studies of other H3K9 methyltransferases. Nevertheless, we also demonstrated that H3K9 methylation contributes to time the irreversible commitment into differentiation.

Finally, we studied the role of *Suv39h1as* during the early embryogenesis. *Suv39h1as* and *Suv39h1* have an anticorrelation pattern of expression during the very first step of the development, with a maximal *Suv39h1as* expression in the late oocyte. Moreover, SUV39H1 overexpression leads to early developmental default. We created a mouse line knock out for Suv39h1as expression. With Suv39h1as loss of expression, we observed an increase of SUV39H1 in the oocyte, with an increase of H3K9me3 specifically. This is inconsequent on females fertility and it questions whether these increases are maintained upon fertilization.

Résumé

La chromatine est une structure dynamique pouvant passer de l'état d'euchromatine, associée à une transcription active, à l'état d'hétérochromatine, associée à une absence de transcription. L'hétérochromatine peut être présente soit de manière variable selon le type cellulaire, soit de manière permanente à certaines régions comme les régions péricentromériques. L'hétérochromatine, caractérisée entre autres par la méthylation de la lysine 9 de l'histone 3 (H3K9), est étroitement contrôlée pendant le développement précoce chez la souris. Dans les cellules souches embryonnaires de souris, un niveau bas de méthylation de H3K9 est constitutif de la pluripotence, d'autant plus qu'il augmente au cours de la différentiation, où il est considéré comme établissant une barrière à un retour à l'état pluripotent. Cependant, le lien entre la méthylation de H3K9 et la pluripotence n'a encore jamais été étudié.

Tout d'abord, nous avons identifié un ARN long non codant transcrit en anti-sens du gène codant pour l'histone méthyltransferase SUV39H1, propageant la méthylation de H3K9 à l'hétérochromatine. Cet ARN non codant (*Suv39h1as*), contrôlé par le facteur de pluripotence OCT4, réprime l'expression de *Suv39h1*. Ce réseau génétique permet de coordonner la pluripotence et l'état de l'hétérochromatine. En supprimant les promoteurs de *Suv39h1as* grâce à la technologie CRISPR/Cas9, nous avons pu confirmer notre modèle et étudier de manière plus précise le mécanisme de régulation et l'impact fonctionnel de cette régulation.

Pour le mécanisme de régulation, nous avons montré que l'expression de *Suv39h1as* n'impacte pas la stabilité du messager de *Suv39h1* mais diminue la probabilité qu'il soit activement transcrit. En effet, *Suv39h1as* agit sur la fréquence de transcription en établissant notamment un contexte chromatinien particulier. Il favorise l'apposition de marque d'euchromatine sur le corps du gène *Suv39h1* et à l'inverse les défavorise sur son promoteur. Ces caractéristiques sont similaires à celles établie pour le couple sens/anti-sens *Xist* et *Tsix*. Cependant, plusieurs éléments suggèrent que ce n'est pas le seul mécanisme impliqué. Nous avons donc émis l'hypothèse que la transcription de *Suv39h1as* interfère avec celle de *Suv39h1* et/ou régule le niveau d'expression de ses différents isoformes. Cependant, ce modèle n'a pas été démontré.

Nous avons aussi exploré l'impact fonctionnel de *Suv39h1as* dans les cellules souches de souris. La surexpression de SUV39H1 induit bien une augmentation du niveau de H3K9me2 et H3K9me3 mais a très peu d'effet sur les cellules souches. Elles se renouvellent et se différentient sans aucuns problèmes. C'est tout à fait en accord avec les études réalisées sur d'autres histone méthyltransferases comme SUV39H2 et SETDB1. Cependant, nous avons

observé que le niveau de méthylation de H3K9 permet de coordonner la spécification irréversible au cours de la différentiation.

Enfin, afin d'étudier le rôle de *Suv39h1as* au cours du développement précoce, nous avons créé une lignée de souris knock-out pour *Suv39h1as*. En effet, *Suv39h1as* et *Suv39h1* sont exprimés de manière anti-corrélée lors des premières étapes du développement, avec une expression maximale dans l'oocyte tardif. De plus, plusieurs études ont démontré que la surexpression précoce de SUV39H1 provoquait un défaut au cours du développement. En l'absence de *Suv39h1as*, nous avons pu observer une surexpression de SUV39H1 dans l'oocyte tardif, associée à une augmentation de H3K9me3 et non de H3K9me2. Néanmoins, la suppression de l'expression de SUV39H1 n'impacte pas la fertilité des femelles, posant la question du maintien de cette surexpression après la fécondation.

Acknowledgements

First of all, I would like to thank the jury members for their time and consideration. I would also express my gratitude to the member of my Thesis Committee Members, Maxime Wery, Claire Francastel and Cosmin Saveanu, for following my work all along my thesis and for their useful critiques and suggestions.

Then, I would like to give my deep and specials thanks to Pablo Navarro Gil, my thesis supervisor, that gave me the chance to do my master internship and my PhD in his lab. Those almost four years were really fulfilling and fruitful. Thank you for the liberty and the scientific opportunities you gave me. Thank you for your time, your patience and all the things you explained me. Thank you for letting me the chance to explore my crazy ideas and start the mouse work, even though it was not planned at all at the beginning. Thank you for your instructive critics and scientific analysis that help me to build my scientific critical thinking. And above all, deep thanks for your precious support in my project to start over with medical school, especially with the opportunity to have a smooth transition with the few months of post-doc in the lab. This was possible for me thanks to you.

I would like to also thank all the past and present members of the lab. Although the past year and half was difficult between strikes and lockdown, it was a great time to work with all of you and I am sincerely happy to have met every one of you. Firstly, many thanks to Victor. Thank you for your time and your patience in supervising me during my master internship. It was greatly fruitful to discover the field and the project with you. I wish you all the best for your new career as double doctor. Special thanks to Agnès. Thank you Agnès for your support, your kindness and your dedication. Thank you for your daily investment to keep the lab functional. Thank you for your time and our coffee breaks at the 25, our personal as well as scientific discussions. They were really fructuous. And many thanks for your help more specifically in my project. Thank you, Amandine, for your sparkling presence in the office. It was a real pleasure to chat with you every day, to share personal and scientific life. Thank you for staying cheerful, empathic and smiling despite everything. I know the times were difficult for you, but it was a real daily support for me. I am sure you will have all the best you deserve soon. Thank you for the last buddy of the 6B office, Nicolas. Thank you for your help after the lockdown. It has been a pleasure to share coffee and yellow key with you. Congratulations for your new post, you really deserve it. Thank you to you Almira, the new comer in the 6B office. Thank you for your efficient bio-informatic analysis and for keeping your smile despite teleworking. Thank you also for your kindness and for having answered my statistical analysis

questions. Special thanks to Philippe. Sharing the office with you during my master internship was an experience. I was also sincerely happy to have you as a bench neighbor. Between balloon explosions and discussion about the last trendy TV shows, thank you for your presence, your advices and your help for many of my scientific questions. I wish you all the best for the new chapter of your life, I will miss you. Thank you Sandrine and Michel. Thank you for coming from your underground lab to the new luminous lab on the 6th floor. Being mixed with your group have been really positive. Thank you for the discussions and to have answered all my questions. Thank you for your help for the mouse side of the project, in the experimental design as well as in the realization of the experiment. Special thanks to Sandrine that support me with my first steps in the animal facility. It was not easy for me and it has been possible thanks to you. Thank you to Rémi, Inma, Anna and Nicola for their daily help in the lab, answering my scientific and technical questions as well as sharing happy lunch chatting. Thank you Inma for being always enthusiastic and encouraging, especially at the beginning of my PhD. Thank you Rémi for the help and support during the writing of the thesis. Thank you Alexandra for your kindness and sharing your expertise in FISH with me. I also spare thoughts to some of the past member of the lab, Nick for his help in analyzing data, Eslande, Thaleia and David. Finally, thank you Corinne for your help with the administrative work. Thank you for your efficiency and your dedication, it was a real relief for me.

I would also like to warmly thank the whole Stem Cells and Developmental Biology department, creating a fun and welcoming working environment. Thank you to Han Li and her teams for welcome me during my first steps in the Pasteur Institute. Thank you Aurélie, Matthieu, Coralie for your presence, sharing lunches at the canteen and our chatting in the corridor. Thank you Elsa for your empathy, your stimulation, your kindness, your good mood and your time. You were a real and critical support during those three years, both personally and scientifically. Thank you Carole for your time, our lunched and your precious help in the western blot, it was a real boost in my project. Thank you to many friends that have been surrounding me. Thank you Lison for your help in my first step in R. Thank you Mathilde and Laure for the happy lunches during those almost four years. Thank you Lisa, Alexandre, Marie-Alix and Eve for being present through the good and the tough moments of those years, especially the last few months. Thank you to Carole and Marc for our chatting while cooking the dinner and our game nights.

I am especially and deeply grateful to Quentin. Thank you for your kindness, your patience and your unlimited support. Thank you to have always been there and believing in me

despite my doubts and my questions. Thank you for your generosity and your help even on the scientific side, with you I am stronger and a better person.

Finally, all my gratitude goes to my parents. Thank you to my father Olivier, for listening all my doubts and questions. Thank you to my mother, Orly. Thank you for your caring and unconditional support even when you disagreed. You always believed in me, even when I didn't envision it. You made all this possible. Thank you.

Table of content

Tables and figures list

Introduction

I) Early mouse development, pluripotency and mouse Embryonic Stem Cells derivation

1) Mouse pre-implantation development

Embryogenesis starts with fertilization, when the two highly different gametes - oocyte and sperm - meet. The resulting one-cell embryo – or zygote – undergoes stereotyped developmental processes namely cleavage, compaction, cavitation and implantation in the uterus[1]. Termed "preimplantation development"[2], this early period involves a tightly regulated series of lineage specifications events coordinated by highly dynamic signaling and transcription factors (TFs) networks.

Upon fertilization, the two specialized gametes undergo extensive remodeling to form the diploid zygote. Before the first division of the embryo, development is divided in pronuclear stages[3,4] (PN0 to PN5): immediately after fertilization, during PN0 the paternal genome undergoes major chromatin remodeling. It loses protamines and is re-packaged by maternal nucleosomal histones to form the paternal pronucleus (PN). Meanwhile, sperm entry has triggered meiosis completion of the maternal genome that begins to form the maternal PN. PN1 is characterized by small PN located at the periphery of the embryo. At PN2, both pronuclei increase in size and begin their migration toward the center of the zygote, they are in G1 phase. During PN3, the pronuclei become larger and continue their migration to the center of the embryo. During this stage, both paternal and maternal genomes start to replicate. The replication is still ongoing when the PN reach the center of the zygote, corresponding to PN4 stage. Finally, during PN5, genomes are apposed to each other in post-replicative G2 phase. The PN stages end with the syngamy or the fusion of the two PN and beginning of the first cell division[3].

During the first three days of development – depicted on **Figure 1**, the embryo undergoes four cell divisions and forms 16 cells, called blastomeres[1]. As soon as the first division happens, maternally inherited mRNAs and proteins begin to be degraded while the zygotic genome begins to be gradually transcribed. This is called the zygotic genome activation (ZGA) or maternal zygotic transition (MZT) and lasts until the morula stage [5,6]. At the 8-cell stage, cells start to compact, increasing cell-cell contacts and promoting cell adhesion[7]. Although 8-cell stage blastomeres are considered identical and equipotent, accumulating evidence shows that they actually differ on several aspects including DNA methylation status,

histone modifications enrichment and TFs expression[8]. Notably, the first five cleavages of embryo occur at constant volume, *i.e.* without any cell growth, and consist of subsequent divisions of the initial egg cytosol. When the embryo reaches the 16-cells stage, the outer blastomeres are apico-basally polarized while the inner ones are apolar. This is when the first lineage specification occurs, with the inner cells giving rise to the Inner Cell Mass (ICM) and the outer cells differentiating to form the extraembryonic trophectoderm (TE)[9,10]. The morula undergoes another round of cell division to form the blastula or 32-cell stage.

At E3.5, between the 16 and 32 cells stages, the trophectoderm cells introduce some fluid inside the embryo to form a cavity called blastocoel[11]. From this step, the embryo is called blastocyst and keeps proliferating to reach over 128 cells at E4.5. During this time, cells from the ICM start to specify and physically segregate between the hypoblast or Primitive Endoderm (PrE) and the epiblast (EPI)[12]. The PrE cells become apposed to the blastocoel surface. The embryo then implants along the uterine horn, directed by the presence of regionalized EGF receptors in the uterus[1]. Meanwhile, the embryo cells keep proliferating, migrating and differentiating until forming the egg cylinder[1]. After gastrulation, the trophectoderm gives rise to the extraembryonic ectoderm that will form the placenta while the epiblast will form the three germ layers – ectoderm, mesoderm and endoderm - of the embryo. The PrE contributes to additional extraembryonic tissues such as the yolk sac.

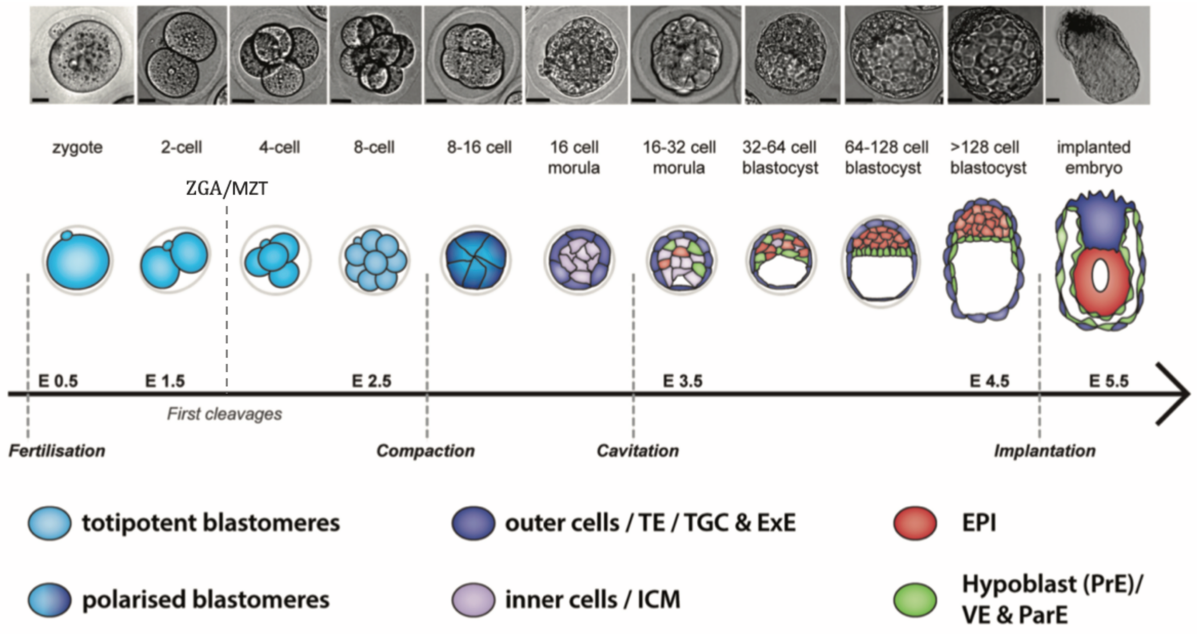


Figure 1: **Early stages of mouse development.**
Pictures and schematic representation of mouse early development steps. ZGA means Zygotic Genome Activation, MZT stands for Maternal to Zygotic Transition. (adapted from Piliszek et al.[2])

After blastocyst implantation, the founder cells of gametes, called primordial germ cells (PGC)[13], are specified in the epiblast at E6.25[14]. In parallel, sexually bipotential structures known as genital ridges arise from the mesoderm-derived coelomic epithelium. The primordial gonad forms when the monolayer becomes thicker and multilayered, and the PGCs migrate into it[15]. They reach the genital ridge around E10.0 – E11 and the first sex differences appear at E12.5[16]. Gametogenesis is mainly driven by the Bone Morphogenetic Protein (BMP) pathway[17] and expression of PGC specific genes such as *Fragilis*, *Stella* and *Blimp1*[17,18]. As early as E7.5 - E8.5, PGCs express pluripotency network genes and undergo fundamental epigenetic changes[19,20]. They start sex-specific gametogenesis with sexual specific genes starting to be expressed. The sexual differentiation is mostly orchestrated by the *Sry* gene[15]. In the XY gonad, *Sry* triggers upregulation of *Sox9* and *Fgf9* which activate the male pathway and repress the feminizing signals *Wnt4/Rspo1* and *ß-Catenin*. Interestingly, the sexual fate of the PGCs is determined by the sexual identity of the fetal gonad and not by the PGCs' own identity[15].

While all these pre-implantation events are occurring, the zygote experiences a progressive loss of developmental capacity[1] through successive and dynamic cell fate specifications. The first cell fate decision is the divergence of the TE and ICM fates in the morula. This is mainly driven by the Hippo pathway, known for its cellular role in integrating physical cues from the environment (cell-cell contact, physical stress, …)[21]. Notably, this signaling pathway is conserved between mice and human[21]. In outer cells, the pathway is repressed leading to nuclear translocation of the YAP1 cofactor that will form a complex with the TF TEAD4 and enhance expression of the TE markers GATA3 and CDX2[22]. Notch signaling is also involved in TE specification through the interaction between NICD and RBPJ that promotes TE genes expression, in particular CDX2[23]. Interestingly, CDX2 inhibits transcription of ICM-specific TFs such as OCT4[24]. However, it is not clear if and how Notch and Hippo signaling are coordinated. TEAD4 also regulates the expression of other TE TFs including *Eomes* and *Elf5* which form a positive feedback loop on *Tead4*[25]. In inner cells, hippo pathway is active through the phosphorylation of YAP by LATS1/2 kinase preventing its translocation in the nucleus[26] and expression of TE genes leading to ICM specification. The ICM identity relies on expression of TFs such as OCT4, SOX2 and SALL4[10]. *Oct4* is expressed from 2-cells stage until 16-cells stage where it becomes restricted to the ICM[24]. OCT4 interacts with SOX2 to enhance transcription of core pluripotency TFs such as NANOG[1]. OCT4 also represses *Cdx2* expression[24], all together inhibiting fates other than ICM. In addition, SALL4 maintains *Oct4* expression, contributing to ICM identity[27]. Interestingly, other epigenetic actors

such as the long non-coding RNA lincGET were found to be important for the 2-cells stage and ICM specification.. This phenomenon will be more detailed in the chapter III-3.

The second cell fate decision corresponds to the sorting of ICM cells in epiblast and hypoblast. The main driver of this specification is the antagonism between NANOG and GATA6. Although, from the 8 to the 32-cell stage NANOG and GATA6 are co-expressed, they gradually become mutually exclusive, forming a "salt and pepper" distribution[12]. In addition, the Fibroblast Growth Factor (FGF) pathway plays key role in this segregation. EPI progenitors secrete FGF4, that reinforces *Gata6* expression in cells expressing the receptor FGFR2 and that commit to the PrE fate[29]. In parallel, ICM cells are progressively spatially re-organized with NANOG negative, PDGFRα-GATA6 positive cells becoming adjacent to the blastocoel. This re-positioning is thought to also be DAB2-dependent[30]. These cells will form the PrE. FGFR activation also leads to ERK phosphorylation that will stimulate a feedback loop via DUSP4 in PrE cells[31]. Last, apoptosis was recently shown to act as a selective mechanism contributing to proper segregation of EPI/PrE cells[2].

2) Focus on the female germ cell development

Successful embryogenesis relies on viable and competent gametes. Their development depends on several factors including multiple support cell types contained in the reproductive organs. In mammals, ovarian development can be divided in two steps : 1) growth of ovary and establishment of the founding primordial follicle pool and 2) maturation of follicles and release of a fertilizable egg starting at puberty[15].

Once the PGCs reach the primordial gonads and sexual differentiation begins, PGCs proliferate with incomplete cytokinesis. These uncomplete divisions create syncytia called cysts or nests that remain connected by intercellular bridges allowing sharing of organelles and mitochondria. At E13.5, germ cells start their meiosis, stopping at the diplotene phase of prophase I. Between E17.5 and 5 days after the birth, female germ cells cysts break down to become primordial follicles (PMF). This step is associated with massive loss of germ cells[32]. Each PMF is composed of one prophase-arrested oocyte surrounded by a flattened epithelium that will later differentiate into granulosa cells. PMFs form a fixed ovarian quiescent follicle pool. From this reserve, PMF can possibly remain quiescent, be activated and start to grow or die directly in their dormant state[15].

PMF activation is marked by histological changes as oocyte growth begins and the flattened granulosa cells become cuboidal[33]. PMF activation is controlled by the

Phosphatidylinositol-3-Kinase (Pi3K) intracellular pathway[34] but extra-oocyte signals have recently been found to be important. Among them, mTORC1 signaling in granulosa cells is required for oocyte differentiation[35]. Activated PMFs become primary follicles.

When primary follicles further develop into secondary ones, they contain mid-grown oocytes surrounded by two or more layers of granulosa cells[33]. Additional layers of somatic cells, called theca cells encapsulate the follicle. They differentiate into an interna layer that becomes increasingly vascularized and an externa layer. As the follicle continues to develop, fluid-filled cavities arise and fuse to form an antrum, giving rise to antral follicles[36]. The antrum separates two different granulosa cell types: cumulus granulosa cells, adjacent to the oocyte and mural granulosa cells that line the follicle wall and are a major source of hormones important for the following steps[15].

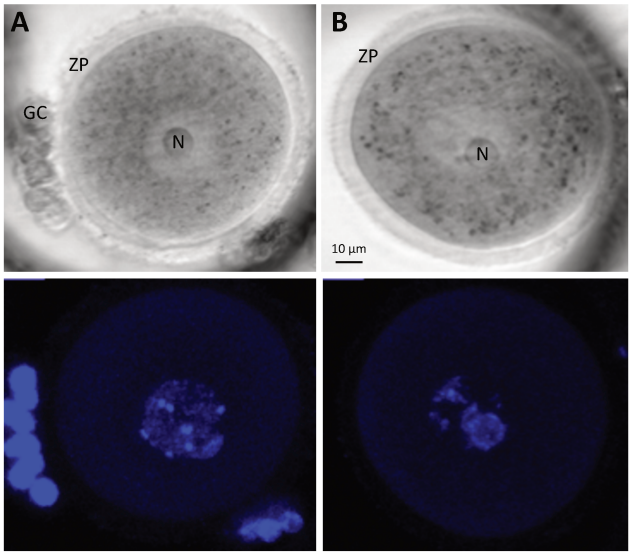


Figure 2: **Two stages of oocyte development.** Differential Interferance contrast (top panels) and Hoescht stained pictures (in blue – bottom panels) of **(A)** a Non Nucleolar Surrounded Germinal Vesicle oocyte or **(B)** a Surrounded Germinal Vesicle oocyte. On the transmitted light picture, we can observe some granulosa cells (GC), the Zona Pelucida (ZP) and the Nucleolus (N). Pictures taken by Sandrine Vandormael-Pournin.

Only a subset of antral follicles develops until pre-ovulatory stage, the remaining ones undergo atretic degeneration[15]. Selected follicles need to be hormonally competent and responsive to mature until ovulation. From that perspective, hormonal stimulation with Pregnant Mare's Serum Gonadotropin (PMSG) can be used to increase the number of dominant follicles, and addition of human Chorionic Gonadotropin (hCG) allows ovulation of those dominant follicles[37]. To achieve a proper ovulation, several features are required in a tight order. The first event is the meiotic resumption[38]: prophase-arrested oocytes are characterized by a large nucleus known as "germinal vesicle" (GV)[39]. During folliculogenesis, oocytes grow in size but remain in the GV stage, more mature ones are presenting a particular DNA organization with "surrounded nucleolus" (SN), and the immature ones are called "non-surrounded nucleolus" (NSN) **(Fig.2)**. When the follicle is selected for ovulation, meiosis resumes with the GV break down allowing completion of the meiosis I and extrusion of the first polar body. The oocyte then undergoes the second meiotic division until metaphaseII[39]. Meiosis will only be completed upon fertilization and extrusion of the second polar body[40]. During oocyte

maturation, there is an acute transcriptional activity to synthesize and store mRNAs and proteins that will be fundamental for the very first steps after fertilization[41].

During meiosis resumption, cumulus expansion, or cumulus mucification, creates a particular extracellular matrix to form cumulus-oocyte complex[40]. Subsequently, follicle rupture releases the cumulus-oocyte complex that contains a fertilizable oocyte. Residual follicle cells undergo reprogramming and terminal differentiation to create the corpus luteum[15]. The corpus luteum is essential for establishing and maintaining gestation. However, in absence of fertilization, it is eliminated through programmed cell death and forms a scar known as corpus albicans[15].

3) Mouse embryonic stem cells origins and maintenance

Studies of pluripotency and developmental processes are very important to understand embryogenesis. Derivation and *in vitro* propagation of mouse embryonic stem cells (mESCs) provided a very useful tool in that perspective. However, their derivation from embryos raise several questions: what is the origin of ESC? How similar are they to cells of the embryo? Is the pluripotency variable?

a) Mouse embryonic stem cells are derived from the inner cell mass

"Teratocarcinomas are bizarre tumors containing multiple tissue types, including differentiated structure"[42]. The first cells successfully propagated *in vitro* were extracted from teratocarcinomas. In the 129 mouse strain, they had the particularity to be maintained through serial transplantation, keeping the ability to generate all types of tumors. This suggested the presence of proliferative and undifferentiated cells called Embryonal Carcinomas cells (ECs)[43]. In 1975, Martin & Stevens[44] successfully co-cultured ECs *in vitro* with a feeder layer of mitotically inactivated fibroblasts. Interestingly, ECs can form aggregates that resemble early embryos, called embryoid bodies. Martin & Stevens further noticed that certain ECs injected in blastocysts can contribute to chimeric mice[44], thus proving that ECs can produce all somatic lineages.

In a second step, some cells lines were derived from the embryo itself, more particularly from the cells of the ICM[45,46]. Those cells were co-cultured with a layer of feeders in addition to serum-containing medium and presented unlimited expansion capacity. When grafted in adult mice, those cells formed multi-differentiated teratocarcinomas, and, when injected in blastocysts, they formed chimeric mice, contributing to all tissues including germ lines without forming tumors[47]. Those cells presented "capacity for indefinite capacity of generation of two

classes of progeny: daughter cell with equivalent proliferative and developmental potential and daughter specified for differentiation"[42] and were thus called ESCs. Their derivation was however quite variable, inefficient and only possible from the 129 strain. 129 mice are poor breeders and variable in behavioral and cognitive studies[1] and female cells derived from this strain tend to lose an X chromosome. It thus became necessary to derive ESCs from other mouse strains. The first attempt was to derive cells from diapaused embryos. Diapause corresponds to a developmental arrest in case of deleterious environmental conditions: cells synchronize at the epiblast stage and keep dividing. Diapaused embryos allowed a higher derivation efficiency for some permissive strains[45,48].

The developmental timing of derivation was also critical as ESCs derived from E3.5 ICM could not form colonies *in vitro* while E4.5 derived ESCs could[49]. Transcriptional studies showed that the ICM-derived ESCs are quite similar to preimplantation epiblast, also called Primitive ectoderm[48,50]. Some attempts were made to establish cell lines from post-implantation epiblast: these were mainly dependent on culture conditions, and the resulting cells were called Epiblast Stem cells (EpiSCs)[51]. Those cells are pluripotent though they have a predisposition for differentiation. Their transcriptional profile shows an heterogeneity quite similar to the one of primitive streak cells in the embryo[52]. Although EpiSCs can generate teratocarcinomas in adult mice, they cannot contribute to chimaeras when introduced in early embryo[1]. ESCs and EpiSCs are morphologically quite similar with identical nuclear to cytoplasm ratio and similar prominent nucleoli. However, EpiSC expand in 2D as an epithelium compared to ESCs and both cell types are very different from signaling pathways and epigenetic perspectives[53]. Finally, ESCs can also be differntiated into PGCs, they are then assimilated to embryonic germs cells but their DNA methylation profile differs importantly[53].

Curiously, male and female derived-ESC are very different[54]: female ESCs exhibit lower global DNA methylation levels, a difficulty to exit pluripotency and are less prone to differentiate. Those features are thought to be linked to the activation of both X chromosomes[55]. Indeed, in addition to the necessity to inactivate an X chromosome to exit pluripotency and initiate differentiation, expression of the pluripotency network is X dosage-dependent thus higher in female ESCs. Moreover, the female ESCs are intrinsically unstable in culture[54], perhaps due to their low levels of DNA methylation.

b) Defining mouse embryonic stem cells culture conditions

It was clear from different ESC derivation attempts, that culture medium was primordial for the maintenance of pluripotency and self-renewal. The first step was to understand that cells

can be cultured to a certain extent in media pre-incubated with feeders (conditioned media) instead of a feeders layer[56,57]. This suggested that a cue important for ESC maintenance was produced by embryonic fibroblasts. Few years later, the cytokine Leukemia Inhibitory Factor (LIF) member of the Interleukin family was found to be the critical signal[58,59]. Using LIF allowed the ESC derivation and long-term culture from 129 strain as well as the germ-line competence in chimeric mice[58,59]. Interestingly, LIF is produced at the fibroblast surface, explaining the relative inefficiency of conditioned media[60].

LIF binds to a heterodimeric receptor composed by LIF receptor (LIFR) and glycoprotein 130 (GP130)[61] **(Fig. 3)**. The complex' cytoplasmic domain is constantly bound by the family Janus tyrosin kinase (JAK1/2/3) that is normally inhibited[62]. Ligand binding causes autophosphorylation of the JAKs inhibitory domain. Activated JAKs will phosphorylate the LIF bipartite receptor allowing binding of other targets such as Signal Transducer and 1ctivator of Transcription 3 (STAT3)[63] that will also be phosphorylated. Phosphorylated STAT3 dimerizes and translocates in the nucleus and acts as a TF to induce target genes, including pluripotency and self-renewal genes[64]. Phosphorylated STAT3 can also act on chromatin accessibility at target genes[65]. This cascade is negatively regulated by a feedback loop as phosphorylated STAT3 induces the expression of Suppressor of cytokine signaling 3 (SOCS3) that blocks JAKs activity acting as a pseudo substrate for a short period[66]. However, it is also positively regulated by a feedback loop via the increase of GP130, LIFR and STAT3 expressions[61].

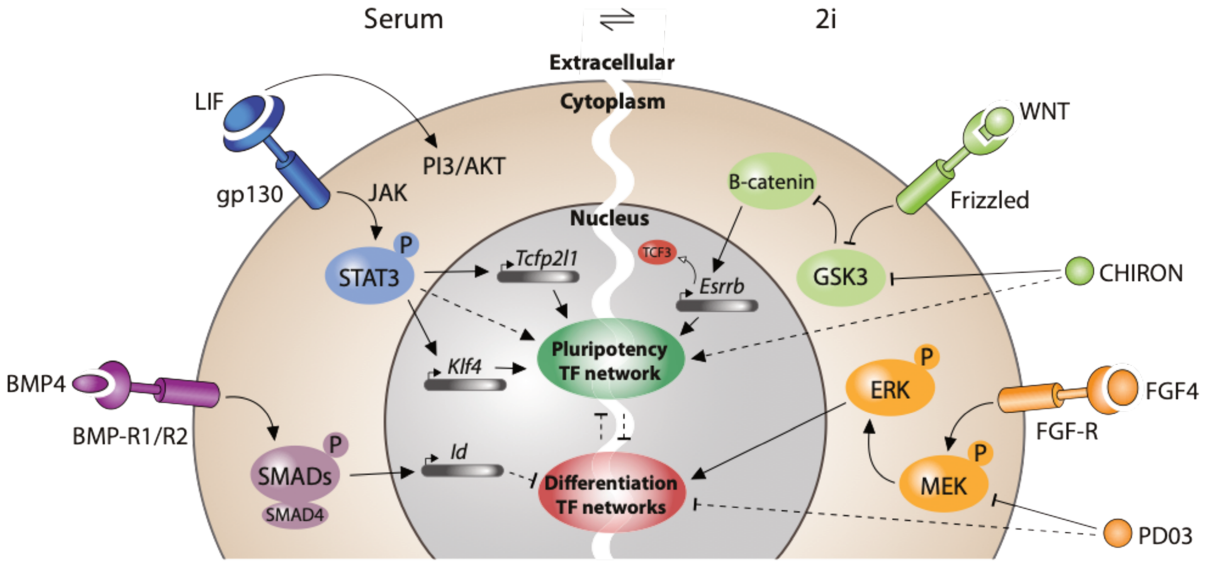


Figure 3: **Extrinsic signaling pathways that modulate the pluripotency.**
Schematic representation LIF, BMP4, WNT and FGF4 pathways that modulate the expression of pluripotency and differentiation TF network (from Hackett et al.[67]).

LIF signaling indirectly activates two other intracellular cascades: JAKs activates Pi3K that then activates Akt and mTOR pathways leading to regulation of pluripotency and proliferation-associated genes[68]. This axis plays an important role in the maintenance of self-renewal though the JAK pathway has been found to be sufficient for self-renewal[69]. Surprisingly, JAKs proteins also activate different Mitogen-Activated Protein Kinase (MAPK) signaling cascades that promote differentiation. Indeed, via Growth factor receptor-bound protein 2 (GRB2), Extracellular signal-Related Kinase 1/2 (ERK1/2) is activated promoting differentiation by several ways, including repression of pluripotent genes. The differentiation promoting Juno Amino-terminal kinase (JNK) signaling cascade is also activated by phosphorylated STAT3 [61,70]. STAT3 is also essential for neural differentiation[71]. This suggests that *Stat3* can have a cell type-specific role. Notably, JAKs-STAT3 and MAPK-ERK pathways converge on the TF couple Transcription Factor CP2-like Protein 1 (TFCP2L1)/ Transcription factor E2-alpha (TCF3)[42]. Indeed, TFCP2L1 is a TF that plays a role in pluripotency maintenance and it is repressed by TCF3[72]. Thus JAKs-STAT3 cascade increases TFCP2L1 expression, being involved in the pluripotency network regulation. By contrast, MAPK-ERK pathway activates TCF3 expression, leading to TFCP2L1 repression thus destabilizing the pluripotency network[73].

ESC plated on gelatin in serum and LIF (FCS/L), became the canonical culture conditions sustaining pluripotency and self-renewal. However, these culture conditions create an heterogeneous cell population with spontaneous differentiation[42]. Later on, other molecular pathways and external cues were identified actors of ESC maintenance. Serum was replaced by BMP4 molecule, not allowing *de novo* derivation but ESC maintenance and self-renewal[74] **(Fig. 3)**. BMP4 is normally provided by feeders. It belongs to the Transforming Growth Factor ß (TGFß) family that binds specifically to Type I receptor[75]. The receptor is composed by an intracellular tyrosine kinase domain that, upon ligand binding, interacts specifically with R-Smad proteins to phosphorylate them. Once R-Smad proteins are phosphorylated, they translocate in the nucleus and act as a co-TF to induce *Id* genes that play a role in differentiation inhibition[75]. Along the same lines, BMP4 also induces the inhibition of ERK and p38 MAPK pathways[76]. Interestingly, others TGFß family members are involved in pluripotency maintenance. ACTIVN, NODAL and LEFTY also bind to Type I receptors leading to R-Smad phosphorylation. NODAL and ACTIVIN are auto-produced by cells, promoting self-renewal[77]. Another external cue was found to be modulated by feeders as they absorb FGF4 through FGFR2 receptor, preventing ERK activation[78,79] and therefore differentiation through

TFCP2L1/TCF3. Later on, high cell density or LIF presence inhibits the FGF pathway, helping to prevent differentiation[80] **(Fig. 3)**. In addition, chemical ERK inhibition in addition to LIF or BMP4 allowed ESC derivation from new mice strains such as C57BL/6 and CBA[81]. Finally, the WNT pathway transducing through the Glycogen synthase kinase-3 (GSK3) was found to be important for self-renewal: GSK3 is inhibited by WNT pathway and then prevents ß-CATENIN phosphorylation and its degradation by the proteasome. Nuclear ß-CATENIN prevents TCF3 binding to DNA and its repression of the pluripotency network[82]. Interestingly, WNT signaling has a role via ß-CATENIN in embryo formation, essentially in axis formation and mesoderm differentiation[83] **(Fig. 3)**.

In order to increase derivation efficiency from various mice strains, a new culture medium was set up using findings on signaling pathways involved in ESC maintenance. ESCs were *de novo* derived from several mice strains with high efficiency using chemical inhibition of GSK3 (GSK3i – CHIRON), FGF4 (FGF4i) and MEK1/2 (Meki – PD03) (3i medium) in absence of serum and LIF[81]. Since FGF4i and Meki were redundant, GSK3i and Meki, or 2i **(Fig.3)**, were found to be sufficient enough for ESC maintenance. Culture in 2i in addition to LIF (2i/L) induces a ground pluripotent state: "robust self-renewal of a biologically homogeneous population of cells, each with unbiased potential to form all somatic cell lineage and germ cells"[42]. However, several studies suggest that long-term cultured 2i/L cells showed a high genomic and chromosomic instability, most probably due to the Meki[84]. Nevertheless, the most common culture media without serum remains the 2i/L and induces the "ground pluripotency state".

Two others physiological aspects were found to be relevant for ESC maintenance. The first one was the adhesion dynamic and the role of the extra cellular matrix. Cell adhesion is playing on lineage specification and self-renewal via the integrins that can modulate MAPK activity via src kinase regulation. Notably, cell culture in 2i/L presents a lower focal adhesion than in FCS/L, it might be one of the reasons of the enhanced self-renewal in 2i/L[85,86]. The second aspects is metabolism regulation. The metabolism of ESC and ICM cells metabolisms is glycolysis based which promotes proliferation. The acetyl-coA, the major actor of glycolysis is an acetyl donor for lysine of histones, which plays a major role in epigenetic regulation in pluripotency[87].

In summary, ESC are extracted from the ICM to be established in cell lines that are pluripotent and can self-renew in an indefinite number of time. ESC maintenance heavily depends on culture conditions. External cues are important to induce the proper signaling

cascades that will play at different levels: maintaining self-renewal, inhibiting initiation of differentiation and sustaining pluripotency. The major part of the signaling pathways are converging on a certain number of TFs, including the couple TFCP2L1/TCF3 that then modulate the core pluripotency network, mostly composed by transcription factors (called pluripotent transcription factors - pTF).

4) Pluripotent transcription factors and the spectrum of pluripotencies

Pluripotent transcription factors (pTF) constitute a complex and unique network that sustains pluripotency on different levels: maintenance of self-renewal, activation of pTF and repression of lineage specific genes.

a) Pluripotency: a unique and complex network of transcription factors

The first identified pTF was POU-domain TF OCT4, also named POU5F1 or OCT3[88]. OCT4 is expressed in a uniform and continuous way from the oocyte, throughout formation and maturation of the epiblast and even in PGCs[89]. It is also expressed in ESC and EpiSC. Over expression of *Oct4* leads to differentiation[90]. However, low levels of OCT4 impair differentiation with no impact on self-renewing[91]. Studies demonstrate that *Oct4* has to be tightly regulated to balance pluripotency and differentiation. A supporting evidence is that *Oct4* is still expressed in the post implantation epiblast which also suggests a role in the progress of lineage commitment *in vivo*[92]. It is now known that OCT4 has multiple partners.

Then the second pTF, identified as an OCT4 partner, is SRY-box TF 2 (SOX2)[93]. Indeed, OCT4 and SOX2 are forming a complex that binds at a specific DNA motif[95]. OCT4 and SOX2 motifs are close at pluripotency and developmentally related genes[86,96]. It was further shown that the juxtaposition of the two motifs was important for such elements to display their activator role. The overexpression of *Sox2* predisposes mESCs to differentiation in neuroectoderm, mesoderm, and trophectoderm but not endoderm, suggesting that *Sox2* expression levels should be tightly regulated for efficient self-renewal, as shown previously for OCT4 levels[97,98]. From the same perspective, *Sox2* knock out (KO) in ESCs phenocopies the loss of *Oct4*[93]. Nevertheless, *Sox2* KO in the mouse is lethal during the early post-implantation development[99]. A third factor, NANOG, was shown to share a great overlap of targets with OCT4/SOX2 throughout the genome.

Nanog was first identified in an ES-specific set of genes and was later described as a homeobox-containing gene preferentially expressed in pre-implantation embryos and mESCs[100,101]. Contrary to *Oct4* and *Sox2*, *Nanog* expression in embryo is restricted to naïve

pluripotent compartments, particularly to epiblast[102]. Indeed, loss of NANOG in embryo induces an absence of epiblast[103]. Interestingly, forced expression of *Nanog* impairs differentiation while KO in ESCs is tolerated, although it induces massive differentiation[104]. In addition, overexpression of *Nanog* confers LIF independency for self-renewal[104]. *Nanog* was shown to be regulated by OCT4 and SOX2 which directly bind to its promoter and activate its expression[105]. NANOG is also repressed by TCF3 and consequently upregulated by GSK3i[106]. Notably, it has been shown that NANOG has many interacting partners, in addition to OCT4 and SOX2[86].

In summary, OCT4, SOX2 and NANOG have been shown to cross-regulate each other and consequently are proposed to form a triumvirate that maintains pluripotent state through self-sustaining positive feedback. However, it is not sufficient to explain observed properties of ESCs. Indeed, overexpression of OCT4 or SOX2 leads to differentiation rather than the stabilization of pluripotent state. Furthermore, NANOG is dispensable for ESC self-renewal and pluripotency.

Later on, several other pTF have been functionally implicated in ESCs pluripotency and self-renewal. Estrogen-Related Receptor B (ESRRB)[107] has been identified through forced expression profiles. Several Krüppel-like factors (KLFs) *Klf2*, *Klf4* and *Klf5* have a role in the control of mESC pluripotency[108]. More precisely, *Esrrb* and *Klf4* deletions can be tolerated by ES cells, although self-renewal is compromised in the case of *Esrrb* KO[106]. *Esrrb* is upregulated by GSK3i through a repression by TCF3[106]. Moreover, ESRRB appears pivotal, probably because NANOG also acts in part through direct upregulation of *Esrrb*[109]. Forced expression of *Esrrb* is sufficient to render ESCs independent of GSK3i[110]. On the other hand, *Klf4* has been identified as a direct target of STAT3[109]. Nevertheless, its overexpression or loss of function only partially reproduce or abrogate the LIF effects. Finally, the pTF TFCP2L1 is another, very abundant, STAT3 direct targeted gene[72]. Interestingly, *Tfcp2l1* is also regulated by many pTF as OCT4, KLF4 and regulates other ones like *Esrrb*. Unlike *Klf4,* forced expression of *Tfcp2l1* at endogenous levels fully recapitulates LIF effects on the self-renewal. Conversely, *Tfcp2l1* downregulation impairs the ES cell self-renewal response to LIF, indicating that TFCP2L1 activity cannot be compensated by other STAT3 targets and plays a central role in the pluripotency network[72,107]. In addition, *Tfcp2l1* is also a target of TCF3, placing it at the convergence of the LIF and GSK self-renewal signaling pathways[72].

Another interesting feature of the pluripotency network is its ability to restore pluripotency. Indeed, in 2006, Yamanaka and Takahashi demonstrate that a specific cocktail of

TF – OCT4, SOX2, KLF4 and C-MYC – is able to revert fibroblast differentiation into induced pluripotent stem cells, the iPS, close to ESCs[111]. This first study established the paradigm that differentiation is not a dead end. iPS can then be directed towards a specific fate, by using small molecules and specific factors. Other reprogramming cocktails have been discovered, by mixing OCT4, SOX2 and ESRRB. Besides, efforts also focused to avoid reprogramming into pluripotency and generate multipotent progenitors. An example is the study by Ieda et al.[113] that reported a combination of three cardiac developmental TF – GATA4, MEF2C and TBX5. From dermal or cardiac fibroblasts, they were able to reprogram them in induced cardiomyocyte-like cells, progenitors of cardiac cells. Trans-differentiation has also been studied. The historical example is reprogramming of fibroblasts in myoblasts with MyoD forced expression[114]. Finally, in the last generation of studies, the research concentrated on the *in vivo* reprogramming, to convert in situ cells for tissue regeneration[115].

Altogether, a complex network of pTF is implicated in the pluripotency and the self-renewal. This network is centered on the OCT4 and SOX2 couple: they are necessary but not sufficient to confer full self-renewal. A second group of pTF composed partly by NANOG, ESRRB and KLF4 are individually dispensable and collectively sustain self-renewal. Finally, some pTF like TFCP2L1 constitute the bridge between the extrinsic signal for pluripotency – LIF and GSK3 pathways – and the intrinsic pluripotency network.

b) The in vitro *pluripotency: capture or creation?*

In vitro pluripotency constitutes a precarious balance between maintenance of pluripotency and differentiation. Several parameters as the pluripotency network level of expression and the developmental stage of derivation allow to determine different types of pluripotencies: naïve and primed. Moreover, *in vitro* pluripotencies use unique mechanisms to be maintained while the embryonic pluripotency is transient and highly dynamic along the development. Therefore, it is currently unclear whether these different pluripotencies represent any point of *in vivo* pluripotency.

Most of ESCs lines are derived from the naïve epiblast at E3.5-4.5 and attempts to derive from later embryos have been highly unsuccessful[42]. They need extracellular cues to maintain a naïve pluripotency and express the main actors of the pluripotency network : *Oct4*, *Sox2*, *Nanog*, *Klf4*,… However, LIF is not required for normal pre-implantation development while it is necessary for embryos in diapause[116]. ESCs maintained in FCS/L conditions are extremely heterogeneous with spontaneous differentiation. One postulates that this heterogeneity might be linked to a heterogeneous expression of pTF network. Indeed, cells expressing low or

undetectable levels of NANOG have a great tendency to differentiate, although they are not yet committed since they can also re-express NANOG and generate undifferentiated ESC colonies[68,102]. It has been speculated that variability in pTF expression may be an important parameter of pluripotency that allows opportunity for lineage commitment[117]. Nevertheless, when the cells are cultured in 2i/L medium, cell populations and pluripotent network expression are homogeneous, corresponding to a metastable pluripotency[68,102].

To challenge the pluripotency, different functional assays are traditionally used. Non-directed differentiation is a simple and rapid mean to test differential potential[118]. In FCS without LIF, ESCs form endoderm and mesoderm, while differentiation from FCS-free media, as the 2i/L base media called N2B27, leads to neural differentiation[74,119]. Several protocols for directed-differentiation have been optimized using cytokine and small molecule inhibitor cocktails to push cells uniformly toward a developmental direction. The result is generally a good approximation of the targeted cell type, at least at the transcriptional level, although, the cells are often not fully functional[118]. A second functional assay to test the pluripotency has been recently developed: the 3-dimensional (3D) differentiation protocols. The historical embryoid bodies-like structures (EBs) are still used and optimized[120]. EBs are three dimensional structures that are able to mimic features of early embryonic development. Indeed, most of EBs cultures tend to self-organize and initiate something that resembles early peri-implantation development with the three embryonic layers and ultimately a yolk sac-like structure. Other 3D structures have been developed as organoids or gastruloids[121]. Finally, a third functional test can be performed: ECS reintroduction into an embryo to assess their contribution to the different germ layers. This test has been done at different developmental stages. While the contribution is very efficient at a range of pre-implantation stages, it is highly inefficient into post-implantation embryos indicating an incompatibility[118].

Cells can also be derived from post-implantation epiblast, between E5.0 and E8.0, and are called EpiSCs[51]. They are associated with the “primed” pluripotency. They are very distinct from ESCs in behavior, culture conditions, morphology, transcriptional and epigenetic profiles and functional properties[118]. EpiSCs are maintained with FGF2 and ACTIVIN A with FCS; they exhibit spontaneous differentiation[51]. It is also possible to maintain them without FCS but with an additional inhibitor of the Wnt pathway to FGF2 and ACTIVIN[122]. From a transcriptional point of view, EpiSCs are often associated to E7.25-8.0 embryos. EpiSCs downregulate some genes specific of ESCs as *Klf4* and *Fgf4* while they maintain the pluripotency triumvirate *Oct4*, *Sox2* and *Nanog*. In the meanwhile, some EpiSCs specific markers are upregulated as *Otx2*, *Fgf5*, *Nodal* and some lineage specific genes are co-expressed

as *Brachury* and *Foxa2*. These results suggest that the EpiSCs are a merge of several cell populations[123]. From a functional point of view, EpiSCs can generate derivates of all germ layers *in vitro* and *in vivo*[124]. Similarly to ESCs, non-directed differentiation with serum containing media leads to mesodermal or endodermal differentiation while in N2B27 media, cells give rise to neuroectodermal differentiation[124]. However, it has been noticed that some subpopulations are biased toward a particular lineage. For instance, cells that are BRACHURY positive tend to differentiate into mesoderm[125]. When reintroduced in the embryo, cells have little to no contribution in pre-implantation embryos and fully participate after E8.5[51,124], suggesting that naïve and primed pluripotency are incompatible.

Finally, multiple states of pluripotency have been captured *in vitro*: naïve ESCs and primed EpiSCs. However, whether it is possible to capture an intermediate pluripotency when the naïve transcriptional program has been turned off is still unclear but the lineage-associated markers have not been upregulated. This "formative" pluripotency would then correspond to a post-implantation Epiblast, before the establishment of the germ layers. Several attempts have been performed and two are particularly promising. The first one was to culture ESCs in the EpiSCs conditions with FGF2 and ACTIVIN, leading to Epi-like Stem Cells (EpiLSC)[126]. As early as 2 days later, EpiLSCs display a downregulation of pluripotency markers and a low expression of lineage-associated genes. However, this state in transient and has not be fully functionally characterized. The second attempt to capture formative pluripotency was to culture ESCs in combined naïve and primed conditions. Indeed, Wnt signaling stimulates differentiation in primed pluripotency and stabilizes pluripotency in naïve conditions. Therefore, several studies cultured ESCs with FGF2, ACTIVIN and Wnt. The cells retain ESCs features as the morphology, naïve markers expression and X chromosome activation. They also develop EpiSC characteristics as the co-expression of pluripotency and germ layer markers[127]. In addition, they can contribute to pre-implantation embryo but the post-implantation has not been assessed yet.

In summary, the spectrum of pluripotency present in the embryo along with the development is difficult to capture *in vitro*. Naïve and primed pluripotencies might correspond to stable attractor states, similar but not fully identic to their *in vivo* counterpart. The formative pluripotency emerges *in vitro* from different culture conditions but still need to be characterized to assess its potential correspondence with a developmental stage *in vivo*.

II) Chromatin and epigenetics: a complex relationship

1) History and characterization of the chromatin

a) Discovery of the chromatin components and structure

In the 1880's, while Flemming was studying the nuclear division, he observed a colorable substance within nuclei that he named "chromatin"[128,129]. We now know that the chromatin is a complex assemblage of histone proteins wrapping with 146-147 pb of DNA and additional non-histone proteins. This forms the basic structure of chromatin: the nucleosome[130]. A canonical nucleosome is composed by two copies of each of the four core histones: H3, H4, H2A and H2B[130]. Studies demonstrate the presence of non-canonical histones. H1 is a linker histone that promote a higher order of chromatin organization. It has been shown that several histone variants such as H2A.B, H2A.X, H2A.Z, H3.3 and CENP-A are closely related to the chromatin conformation[131]. Indeed, nucleosomes are highly dynamic structures in time, composition and conformation. Particularly, nucleosomes can transition between different states of post-transcriptional modifications, subtly altering their structure and their interaction properties. It will be developed later in the part II-5.

In 1928, during development of chromatin staining methods, Emil Heitz noticed some regions that were not stained after the cellular division while other stayed colored, he termed them "euchromatin" (EuC) and "heterochromatin" (HC) respectively[132]. Going further in the observation, he noticed that some regions were only stained in certain cells while other were present in all cell types, these regions will be called later facultative and constitutive HC. In 1930, Muller experimentally displaced a gene responsible for eye color in Drosophilia and observed a pattern of variegated gene expression[133]. This was later directly associated to the repression of this gene expression when displaced close or within the HC. This phenomenon was termed Position effect variegation (PEV). In 1950, McClintock was studying transposable elements in plants and proposed the first link between the dynamics in the HC organization and the transcriptional status of genes[134]. Later on, improvements in both microscopy technics and chromatin staining as well as the emergence of the epigenetics, allowed to observe that the HC was definitively subdivided in two parts: the facultative and the constitutive HC (fHC and cHC respectively)[135].

b) Characterization of the different types of chromatin

Nowadays, the EuC is characterized by an open, gene-rich and easily transcribed chromatin. It is usually thought to form a 11 nm fiber and associated to H2A.Z and H3.3 histone

variants[131]. In contrast, the HC is defined as the molecular subtype of condensed and transcriptionally repressed chromatin that extends beyond a single gene or regulatory elements, speculated as a 30 nm fiber. Usually, H1 is a fifth histone characteristic of the HC. The fHC is a transcriptionally silent chromatin but it retains the possibility to interconvert between HC and EuC[136]. It forms at various chromosomal regions which usually contain developmental important genes that must be kept repressed. This feature confers to fHC the ability to control the cell-type specification by constituting a physical barrier to TF-mediated reprogramming and stabilizing the cell-type identity of differentiated cells[136]. Nevertheless, silent regions may decondense and allow transcription within the following contexts: (1) temporal cues during development, or even along cell cycle, (2) spatial localization with changes from the center to the periphery or vice versa and (3) parental characteristics, as monoallelic gene expression. Given its dynamic nature, fHC likely adopts a wide range of chromatin condensation states[136]. The transcriptionally competent 11 nm fiber is devoid of histone H1. Then, histone H1 can be incorporated in a localized manner without necessarily forming a potential 30 nm fiber but instead can participate in the formation of locally condensed chromatin structures that are repressive for transcription[137]. From this perspective, histone H1 might actually be ‘‘recruited’’ to a restricted region of the genome. One canonical example of fHC is its implication in the sexual chromosome dosage compensation in female cells. In mammals, one of the X chromosome is silent and reactivated in a dynamic way over the development[138].

The other type of HC is the constitutive HC (cHC). The cHC can be found at the same genomic regions in every cell-type with no or few genes[139]. Rather, it contains repetitive elements.

Repetitive elements correspond mainly to either simple shorts repeats (5 to 200 pb) or transposable elements (Te). Simple short repetitive elements are mainly concentrated at 3 different regions in the genome. The centromeres are the first region and are composed by minor satellites (~600 kb of 120 bp units)[140]. They correspond to regions where kinetochore are assembled. They are also characterized by a particular histone variant CENP-A[141]. The pericentric regions correspond to the second region that contains major satellite repeats (6 megabases of 234 bp units), with a wide variety of repeats that are not well conserved[142]. And finally, the third region corresponds to telomeres composed by short conserved DNA motifs at the end of chromosomes arms. cHC that form at those different regions from different chromosomes can aggregate to form chromocenters[143].

Te are other repetitive elements that are present all along the genome and more densely represented at some regions. Te can be divided into 2 classes: retrotransposons and DNA transposons[144]. DNA transposons class is ancient and rare, composed by two families: the cut-and-paste, using double stranded DNA and a transposase to mobilize and the rolling-circle that is copying itself into simple stranded DNA[144]. Retrotranposable elements are also using the "copy and paste" system with an intermediary step of reverse transcription. One family of retrotransposons is the Long Terminal Repeats (LTRs) that include the Endogenous Retroviruses (ERVs), and are autonomous for the transposition[145]. The other family of retrotransposons corresponds to the non-LTR. They englobe Long Interspersed Elements (LINEs) and Short Interspersed Elements (SINEs)[144]. LINEs encode for 1 to 3 proteins that are part of the enzymatic machinery used for retrotransposition. They are autonomous. SINE are usually Te derived from both rDNA repeats and LINEs. They parasite the LINEs enzymatic machinery. These repetitive sequences, by their mobility inside the genome, are a threat to its stability. Indeed, mutations produced by integration or excision of transposable elements and recombination between repeats can be deleterious and need to be repressed. This happens through cHC condensation and silencing. cHC has been proved to be crucial for proper function of all those repetitive sequences and regions.

In conclusion, the chromatin is a highly dynamic structure that can be present in different states: euchromatin, associated with transcriptionally active regions; facultative heterochromatin, associated with gene repression in a cell-type or cell-stage-specific manner and constitutive heterochromatin, usually enriched at permanently silent genomic regions such as pericentromeric regions and other repetitive elements. These HC are vital for genomic integrity as well as cell identity. Traditionally, these two types of heterochromatin have been characterized by distinct epigenetic signatures.

2) From the chromatin to epigenetics

a) The difficult definition of "Epigenetics"

The term epigenetics was coined by Conrad Weddington in 1942. In an attempt to unified genetics and embryology fields, he defined the epigenetics as "the branch of biology that studies the causal interactions between genes and their products which brings the phenotype into being"[146]. Today, his definition would correspond to the concept of phenotypic plasticity. In 1958, Nanney brought an evolution to definition of epigenetics, in addition to discussing the variability in expression pattern: he emphasized the fact that expression states could persist

through cell division[147]. In 1982, Lincoln and colleagues generalized the definition: "pertaining to the interaction of genetic factors and the developmental processes through which the genotype is expressed in the phenotype"[148,149]. This definition has the advantage to have a broadened meaning and to emphasize the importance of genetic and nongenetic factors in controlling gene expression. In 1994, Holliday redefined epigenetics in a more specific way to inheritance of expression states[150]. He proposed two definitions that need to be taken together to cover all the concepts: (1) "the study of changes in gene expression, which occur in organisms with differentiated cells, and mitotic inheritance of given patterns of gene expression" and (2) "nuclear inheritance, which is not based on differences in DNA sequences". Finally, it is only in 2001 that the two definitions were conciliated in one by Wu and Morris: "the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail change in DNA sequence"[151].

In summary, Waddington's definition described the interplay of genetic and cytoplasmic elements that produce emergent phenotypes and is mostly used to report the impact of the environment on phenotypes. This is particularly important in the field of the ecology physiology. In contrast, the definition of Holliday, more centered on how the pattern of expression persist across the mitosis and meiosis, is mainly used in the genetics field. While the two definition exist, they may not necessarily be related to each other and they settle down the ambiguity and the confusion across the field[149].

b) The fundamental components of the epigenetics

Despite foundational differences in definition, the use of "epigenetics" has been essentially used to refer to chemical modifications of histones and DNA, or more generally to chromatin modifications. The ensemble of such modifications has been referred to as the "epigenome". All the processes involved in the epigenetic modifications relies on the same mechanisms: post-translational modifications (PTMs) of the histones or methylation of the DNA[152]. The different histone PTMs and DNA methylation mentioned in this section II are summarized in **Figure 4**.

i) DNA methylation

The methylation of DNA cytosine residues (5mC) is an epigenetic mark present in all vertebrates[153]. The epigenetic mark often happens in the context of CpG dinucleotides but can also happens on non-CpG regions. Some regions are rich in CpG and called CpG islands (CGIs)[154]. Several recent studies demonstrate that sequences can be methylated *de novo*, inherited through mitosis and propagated. *De novo* methylation happens mostly for

nucleosomal DNA and is influenced by histone post-transcriptional modifications[155]. In addition, 5mC can be converted to thymine by spontaneous or enzymatic deamination[156]. CpG are densely present at promoters although they can also be present in gene bodies, regulatory elements, and even gene deserts[157]. Position of the methylation in the transcription unit influences the gene expression. For instance, methylation of immediate vicinity of transcription start sites (TSS) blocks the initiation. However, in somatic cells most TSS CpG-rich are non-methylated[154]. Nevertheless, when TSS are methylated, it is often associated to long-term stabilization of the repressed state for the associated gene[154]. It has been mainly observed on genes located in the inactivated X chromosome, called "imprinted genes". The level of methylation at non-CpG promoters is variable and has not clearly been linked to a fluctuation of gene expression. Methylation can additionally be present on gene bodies. They are mostly CpG-poor regions with multiple repeats and transposable elements, but extensively methylated. It is associated to transcription, more particularly to a stimulation of the elongation[158]. Moreover, it has also been linked with the splicing regulation and the regulation using of alternative promoters[159,160]. DNA methylation can also occur at repeated regions such as centromeric ones, conferring chromosome stability during the mitosis and more generally genomic stability[154]. Besides, It has been shown to be present regulatory elements as enhancers and insulators, influencing their functions.

ii) Histone methylation

Other canonical epigenetic marks are the PTMs of histones. The N- and C-terminal tails of histones, particularly H3 and H4, can be modified by processes such as acetylation, methylation or phosphorylation. They change the electronic charge and structure of these tails, which bind to the DNA, to alter the chromatin status and then gene expression[161].

The methylation generally occurs at the lysine (K) of H3 and H4, with the addition of one (me1), two (me2) or three (me3) methyl groups. Histone marks can be associated to either active or repressed transcription[161].

- Methylation of H3K4 is considered as an active mark, highly enriched at enhancer and promoter regions as well as TSS. H3K4me1 is mainly present at enhancers[162], but H3K4me2 and H3K4me3 are located around the promoter region. While H3K4me2 marks the 5' side of the gene, H3K4me3 is on the intergenic side. H3K4me3 is usually associated to actively transcribed genes and poised developmental genes[163,164]. In addition, H3K4me3 can be non-canonically arranged as broad domains at super-enhancers[165].

- H3K36 methylation is another histone mark commonly associated to active transcription. It is enriched at both intergenic and intragenic regions. When present in the gene body, H3K36 methylation gradually increases from me1 to me3 through the promoter towards the 3' end. H3K36me3, when located in gene body, interplays with other methylations such as DNA methylation to poise spurious transcription initiation[166]. H3K36me2 presence in gene bodies has been proved to regulate the establishment and the maintenance of DNA methylation.
- Finally, methylation of H3K79 is principally present at the coding regions. Although it has been reported to be implicated in the telomeres silencing, DNA damages responses and transcription elongation[161,167,168], it has been less studied that the previously mentioned ones.

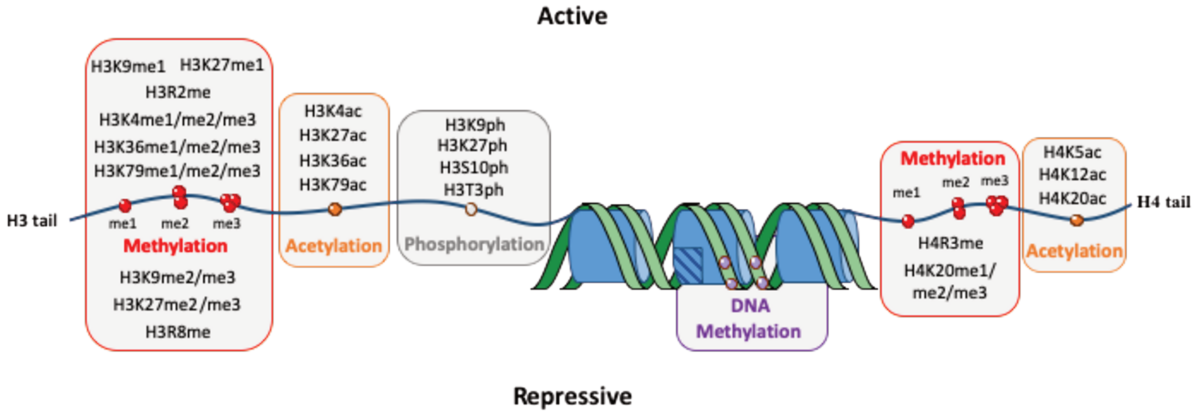


Figure 4: **Insight into different components of the epigenetic.**
Schematic representation of chromatin with nucleosomes in blue, DNA in green and H3 or H4 tails with the blue lines. The different chromatin modifications are depicted in red for histone methylation with the addition of one (me1), two (me2) or three (me3) methylation, in orange for acetylation, in grey for phosphorylation and in purple for DNA methylation. Above the line are indicated to marks that are associated with an active or below a repressed gene expression. The hatched represents an histone variant.

- The methylation of H3K9 is usually associated to the HC[169]. More precisely, whereas H3K9me1 is enriched at active genes TSS[170], H3K9me2 and H3K9me3 are associated to fHC and cHC[168], respectively. Interestingly, H3K9me1 and H3K9me2 can be found in the cytoplasm as well as in the nucleus while H3K9me3 is only nuclear, suggesting that H3K9 could be mono- and di-methylated co-translationally on the contrary to H3K9me3[171]. When H3K9 is mono-methylated directly in the nucleus, it is used as a substrate to reinforce H3K9me3 propagation and the HC.
- H3K27 methylation is another repressive mark. Similarly, to H3K9 methylation, H3K27me1 is distributed at the actively transcribed gene promoters,[172] then associated to active transcription, while H3K27me2 and H3K27me3 are correlated to poised promoters and enhancers[172,173]. This suggest that H3K27me2 and H3K27me3 are involved in the transcription repression, particularly for developmental genes. Additionally, H3K27me3 can form broad domains around promoters and is often associated with H3K4me3 and H3K4me1 at bivalent

promoters and enhancers, respectively[165]. These features are thought to prepare the chromatin to quickly answer to developmental cues.

- Repressive histone marks can also be present on H4. This is the case for the methylation of H4K20. H4K20me1 is enriched at lowly transcribed genes[174]. Mono- and di-methylation are involved in DNA replication and more generally[175] methylation of H4K20 has been proved to confer genomic stability and to help during the chromosome compaction during the mitosis and for the nucleosome turn over[161].

The methylation can also happen at arginine residues. The most common ones are histone 3 arginine 2 (H3R2) and H4R3 associated with active transcription and ribosomal biosynthesis, H3R8 is linked to transcriptional repression[176].

iii) Chromatin remodelers and other histones post-transcriptional modifications: acetylation and phosphorylation

Other PTMs are implicated in the genetic expression control. The acetylation (ac) can happen on spots similar to the methylation such as H3K4, H3K27, H3K36, H3K79 and H4K20, but also to different spots: for instance, H4K5 and H4K12[161]. The acetylation reduces the positive charges of the lysine residues which subsequently decreases the interactions between the histone tail and the DNA leaving the DNA exposed and accessible to different machineries[161]. Therefore, histone acetylation is usually considered as an active histone mark. Methylation and acetylation can be present on the same residue and act as antagonists to inhibit each other, leading to the crosstalk between different histone marks[161]. Another and one of the most common histone modifications is the phosphorylation. It can occur on lysine, tyrosine (T) and serine (S) residues. Phosphorylation (ph) of H3K9 and H3K27 can be simultaneous with the methylation and forming a “KS domain” and altering the histone mark modifications[177]. H3S10ph and H3T3ph are implicated in the mitosis proper progression[178]. The first one is mainly enriched at pericentromeric regions, especially with a maximal increase at the metaphase. The second one influences kinetochore-microtubule interactions. It has also been reported to be implicated in H3K9me2 regulation[179]. Finally, the chromatin remodelers are also able to modify interaction between the DNA and the histones to change nucleosomes position. They have been shown to be important for different biological processes such as replication, chromosome assembly and transcription[161].

3) Characterization of the main actors of the epigenome

Histone PTMs is a very dynamic process that needs to be tightly regulated to quickly respond to the environment stimuli, such as developmental cues. These modifications rely on the cooperation of different enzymatic actors: the "readers" which are able to recognize and bind to specific chromatin locus, the "writers" that can modify the nucleosomes and the "erasers" which can remove it **(Fig. 5)**. For each modification, the enzymatic families are different[180]. We will focus on the enzymes implicated in the DNA and histone methylation, and more particularly to two key histone marks of HC: H3K9 and H3K27 methylation. These different enzymes are summarized in **Table 1**.

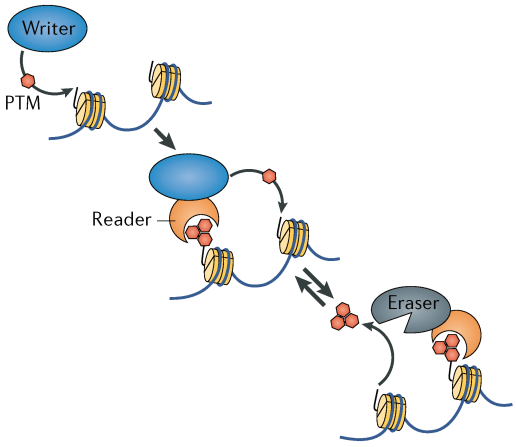


Figure 5: **The different actor of the epigenetics.**
Schematic representation of histone PTM establishment and erasure. Different « writer » enzymes catalyze the addition of a PTMs. « Reader » proteins recognize and bind PTM, they usually recruit writers to spread the PTMs or « eraser » enzymes that remove PTMs (from Allshire et al.[152]).

a) DNA methyltransferases and demethylases

The enzymes involved in the DNA methylation are the DNA MethylTransferases (DNMTs) that transfer a methyl group from S-adenyl methionine (SAM) to the fifth carbon of the Cytosine residue to form 5mC[155]. DNMTs is an enzyme family composed of four members. DNMT1 is the "maintenance DNMT": it preferentially methylates hemimethylated DNA[181], especially during the DNA replication[182], to precisely mimic the original methylation pattern. DNMT1 also has the ability to repair DNA methylation. *Dnmt1* KO in mice results in embryonic lethality at the end of the gastrulation and the beginning of organogenesis[183]. Nevertheless, mESCs KO for *dnmt1* remain viable[184]. DNMT3A and DNMT3B, although sharing a similar structure with DNMT1, have different functions. They are called "de novo DNMTs" because they can introduce methylation into naked DNA[185]. They have different pattern of expression, especially DNMT3B which is required during development. DNMT3B KO is lethal during mouse development while DNMT3A KO mouse survived 4 weeks after birth[185]. The mechanisms to target *de novo* DNMTs to specific regions are not fully elucidated yet[186]. Different paths are possible and one strong hypothesis is that the implication of the TF that can recruit the DNMTs on their binding sequences[187]. Another member of DNMTs enzymes is *dnmt3l*. DNMT3L has no catalytic functions of its own, however, it associates with

DNMT3A and DNMT3B to stimulate their activity[188]. DNMT3L is mainly expressed in early development[188], and is restricted to germ cells and thymus during adulthood. It is required for establishing both maternal and paternal genomic imprinting, methylation of retrotransposons and the compaction of X chromosome. Finally, a novel DNMT has been discovered recently. Thought first as a pseudogene, the DNMT3C is a *de novo* methyltransferase that protects the male germ cell from genomic instability by repressing retrotransposons[189].

DNA demethylation can be an active or a passive process. Passive DNA demethylation occurs during cellular division, naturally diluting the methylation[154,176]. Active DNA demethylation happens in both dividing and non-dividing cells, with specific enzymatic reactions to revert the 5mC[190]. This demethylation occurs through a series of chemical reactions that modify 5mC to form a final product recognized by the Base Excision Repair pathway that replace the modified base by a naked cytosine[186]. Several families of enzymes have been proved to be involved in those chemically reactions: the Activation-Induced Cytidine Deaminase (AID), the Ten-Eleven Translocation (TET) methylcytosine dioxygenase[191,192] and the Thymine DNA glycosylase (TDG)[193]. Importantly, the TET enzyme can add a hydroxyl group to the 5mC to form 5hmC. It is thought to be also involved on its own in the regulation of gene expression. Finally, DNA methylation can be recognized by different families of proteins: the MBD proteins, the UHRF proteins and the zinc-finger proteins[154]. The MBD proteins contain a methyl-CpG-binding domain and has a high affinity for single methylated CpG sites[194]. They can act as a platform to recruit transcriptional machinery elements. The UHFR, for Ubiquitin-like, containing domain that can directly bind methylated cytosine DNA[195]. However, they first bind to DNMT1 and then direct it to hemimethylated DNA. The last family binds to the DNA by zinc-finger domain. It is composed by Kaiso, ZBTB4 and ZBTB38 proteins. It can recognize two consecutives methylated CpGs and can repress transcription in a methylated DNA-dependent manner[196]. Interestingly, the readers of DNA methylation are a strong link between DNA methylation and histone modifications.

b) Histone methyltransferases and demethylases

For lysine methylation, writers are Histone MethylTransferases (HMT and in our case, KMT) that are able to add one, two or three methyl groups on a specific histone's lysine. They usually contain a Su(var)3-9, Enhancer-of-zeste and Trithorax domain (SET): a 130 amino acid catalytic domain that use the SAM as a methtyl-group donor[161], similarly to the DNMTs. KMT have a high degree of enzymatic specificity for lysine within the substrate and for the degree of methylation. Erasers are histone demethylase (HDM or KDM), they use Jumonji C (JmjC)

domain to catalyze demethylation through oxidation of methyl groups[180]. As for KMT, they are specific for both site and degree of methylation. Finally, readers are proteins that contains methyl-lysine-binding motif such as Plant HomeoDomain (PHD) and chromo domains, among others[180]. As for DNA methylation, they can recognize specific sites based on their methylation status and the surrounding amino-acid sequences. They consequently recruit KMT and KDM and stimulate their activities.

PTM	Writers	Erasers
DNA Methylation	DNMT1 DNMT3A, 3B, 3C DNMT3L *(no catalytic functions)*	AID TET TDG
H3K9 Methylation	SUV39H1/KTM1A SUV39H2/KTM1B SETDB1/KTM1E G9a/KTM1D GLP/KTM1C PRDM3, PRDM16	JHDM2/KDM3 JHDM3/KDM4 *(also H3K36)* JHDM1/KDM7 *(also H4K20)*
H3K27 Methylation	EZH1/KTM6B EZH2/KTM6A	JHDM1/KDM7 *(also H4K20)* UTX/KDM6A *(also H3K4)* UTY/KDM6B JMJD3/KDM6B

Table 1: **Insights into different writers and erasers of DNA, H3K9 and H3K27 methylation.** Presentation of the different enzymes involved in the PTM biogenesis. Between parenthesis and in italics is indicated the other PTM target of the enzyme for erasers, and a comment for the writers.

In mammalian cells, several KMT specific for H3K9 have been identified. The first family is the SUVAR family, with SUV39H1/KMT1A and SUV39H2/KMT1B that are able to di- and tri- methylation, especially at the HC[180]. This family of HMT will be developed later. Another HMT is the SET Domain Bifurcated 1 (SETDB1/KMT1E) that can mono-, di- and trimethylates H3K9, both at the EC and HC[197]. G9a/KMT1D and G9a-like Protein (GLP/KMT1C) are HMT that forms heterodimers to catalyze mono- and di-methylation at the EuC[198]. They interact independently from their SET-domain. *In vitro*, they have been shown to form homodimers that can form H3K9me1, H3K9me2 and H3K9me3. The last identified family are the Positive Regulatory Domain containing (PRDM) proteins, composed of 17 members[199]. They contain a PR domain, similar to the SET, but only few of the family members have a catalytic activity such as PRDM3 and PRDM16 that are able to deposit H3K9me1. The others are thought to recruit other HMT such as SUV39H1/H2 or G9a to influence the gene expression.

Three classes of mammalian proteins have H3K9 demethylation activity through a JmjC domain: (1) Jumonji domain-containing histone demethylase 2 (JHDM2/KDM3) family can remove me1 and me2 groups[200], (2) JHDM3/KDM4 proteins revert me2 and me3 on H3K9 as well as H3K36[201] and (3) JHDM1/KDM7 family, including PHD Finger Protein 8 (PHF8), contains a PHD domain and are capable of H3K27me2/1, H3K9me2/1 and H4K20me1

demethylations[202]. PHF8 needs the presence of H3K4me3 in the surrounding environment, explaining the mutual exclusion of H3K9 and H3K4 methylations.

Enhancer-of-zest homolog 1 and 2 (EZH1/KTM6B and EZH2/KTM6A) are the HMT responsible for the mono-, di- and trimethylation of H3K27[203]. They are part of the Polycomb Repressive Complex 2 (PRC2). Their interaction with other PRC2 complex subunits as Suppressor Of Zeste 12 (SUZ12) and Embryonic Ectoderm Development (EED) that enhance their methylation activity of H3K27. In addition, SUZ12 can recognize the H3 N-terminal tail and EED can bind to H3K27me3. This allows a feedback loop that spreads H3K27me3 marks. The demethylation of H3K27 is mainly catalyzed by the KDM7 family[202], similarly to H3K9 demethylation and by the KDM6 family that can remove di- and trimethylation. The family is composed by Ubiquitously-Transcribed X/Y (UTX/KDM6A – UTY/KDM6C) and JmjC Domain containing protein 3 (JMJD3/KDM6B) proteins. Interestingly, UTX is also part of MLL4 complex that methylated H3K4, mediating the cross-talk between H3K27 and H3K4 methylations[204].

c) Focus on Su(var) and Suv39h families

Among the numerous HMT important for H3K9 methylation, *Suv39h1* and *Suv39h1* have a critical role in the HC biology. They have been the first HMT family described, opening the field of the epigenomic actors. We will now focus on their discovery and characterization.

As mentioned before, the PEV was described for the first time in 1930 by Muller in drosophilia[133]. PEV corresponds to chromosomal rearrangements with one breakpoint in cHC and a second in EuC, juxtaposing these two alternate states of chromatin. At the breakpoint, the condensed heterochromatic conformation spreads into EuC section. This phenomenon appeared to be a useful tool to dissect important actors of the chromatin structure. Molecular genetic manipulation to inactivate genes allowed to identify genes that suppress or enhance PEV, suggesting they encode chromatin proteins. Two family of genes have been discovered[205]. The first one is the enhancer of variegation *E(var)* genes family. For example, E(var)3-93D is a chromatin protein that establishes and maintains an open chromatin conformation[206], implicated in the homeotic genes' regulation. The second family is the Suppressor of Variegation *Su(var)* family. Many genes have been identified, such as the *Histones* genes, *Su(var)2-5* that is now known as Heterochromatin Protein 1 (HP1)[207] and *Su(var)3-7* that contains a zinc-finger domain to interact with DNA[208]. A particular gene of this family received particular attention: *Su(var)3-9*. Tschiersch et al.[209] showed in 1994 that *Su(var)3-9* contains a chromodomain similar to the ones of *Su(var)2-5* and *Polycomb* genes at the N-terminal end. It is able to

recognize methylated lysine. They also observed that the C-terminal contained a domain comparable to the Polycomb chromatin regulator Enhancer of Zester (E(z)) and Trithorax proteins. This domain was consequently named SET. Then, in 1998, Jenuwein et al.[210] established that the SET domain, conserved from yeast to mammals, was mechanistically responsible from the chromatin modulation.

Shortly after, *Su(var)3-9* homologs were identified in the yeast (*clr4*)[211], in mouse (*Suv39h1)* and in human (*SUV39H1*)[212], exhibiting the same structure with a chromodomain to interact with chromatin and a SET-domain to catalyze the methylation process. In parallel, O'Carroll et al. isolated and characterized a second HMT *Suv39h2* displaying 59% identity with *Suv39h1*[213]. In the yeast, Clr4 is responsible for the mono-, di- and trimethylation whereas man and mouse SUV39H1/H2 are only able to di- and trimethylate *in vivo*. SUV39H1 can also be phosphorylated, mainly during the transition between the G1 and the S phases, with an accumulation at the mitotic chromosome centromeres[214]. SUV39H1 is mostly present at the HC such as telomeres[215] and major satellites[216], with consequently an enrichment at chromocenters. The functional impact of Suv39h1 in the HC regulation will be developed later in this chapter. For now, we will focus on SUV39H1 mechanism of action.

Muller et al. published a model for SUV39H1 with a two-stage activation mechanism[217] **(Fig. 6A)**. In few words, in its free form, SUV39H1 is inactive and samples chromatin through its chromodomain. When the chromodomain recognize H3K9me3, it allosterically activates the enzyme that anchor itself at the chromatin. It secondly stimulates H3K9 methylation by the SET domain. The N terminus has probably a role in this transition, either by an additional chromatin-binding activity or mediation of cooperative interaction involving the chromo- and the SET-domain. Although this model has been built with *in vitro* studies, it demonstrates the important for SUV39H1 to be anchored at the chromatin. Other studies demonstrated that the chromodomain is critical for the catalysis process, and vice versa[218,219]. Furthermore, being capable of binding its own production and then being stimulated suggest a positive feedback loop that contributes in the HC spreading. To function in a proper way, SUV39H1 is interacting and stabilized by different proteins: HP1 for the non-phosphorylated form[220] **(Fig. 6B)** and the antiphosphatase SET Binding Factor 1 (SBF1) for the phosphorylated one[221]. More precisely, HP1, through its chromodomain, has been proved to recruit and then stabilize SUV39H1 at the HC by a direct protein-protein interaction. Interestingly, the Retinoblastoma protein (Rb) can recruit SUV39H1 at the promoter of active genes to repress them, emphasizing its role in the facultative HC[222]. These recruitment and stabilization can also occur through other interactions. It has been shown that RNA chromatin-associated are important for SUV39H1 stabilization at

the cHC, such as the major satellites, as well as at the centromere during the mitosis[223,224]. In addition, SUV39H1 level of expression is modulated by DDB1 And CUL4 Associated Factor 13 (DCAF13) that induces its degradation via ubiquitinylation[225].

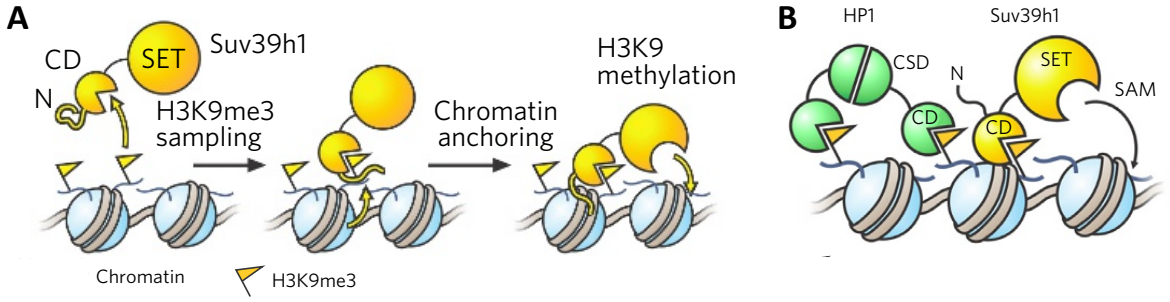


Figure 6: **SUV39H1 mechanisms of action.**
Schematic representation of SUV39H1 with its Chromodomain (CD), its SET domain (SET) and N tail (N) and dimerized HP1 with its Chromo-shadow domain (CSD) and CD. **(A)** Müller et al. model for a two-state activation of Suv39h1. **(B)** Depiction of H3K9me3-modified nucleosomes that serve to recruit HP1 and SUV39H1 through their CD. SUV39H1 catalyzes the SAM-dependent methylation of H3K9 via its SET domain (adapted from Müller et al.[217]).

SUV39H1 also operates in cooperation with the epigenome and the epigenetic machinery. Indeed, SUV39H1 interferes with H3S10 phosphorylation[226], most probably having a role in the chromatin condensation during the mitosis. It also interacts with the DNMTs to coordinate histone and DNA methylation[227]. SUV39H1/H2 and DNMT3A/B interact which each other and can be recruited together via HP1. It is participating in the reinforcement of the condensed chromatin structure. From the same perspective, H3K9me3 by SUV39H1 directs DNA methylation at the major satellites repeats. Furthermore, MBD1 interacts with the complex SUV39H1/HP1 to induce gene repression[228]. SUV39H1 is also able to induce SUMOylation of HP1, in an independent way of its HMT activity[229]. SUV39H1 acts in cooperation with the histone deacetylase (HDA) of H3K16 SIRTUIN 1 (SIRT1), at the HC, coordinating the deacetylation and the methylation on a given locus[230]. In addition, SIRT1 stabilizes and activates SUV39H1 activity[231]. This complex can be disrupted by different proteins such as SIRT7 and Deleted in breast cancer 1 (DBC1)[232,233], adding a new layer of complexity in the regulation of SUV39H1 activity. Finally, SUV39H1 can be part of a multimeric complex composed by different HMT: G9a, GLP and SETDB1[234]. These different HMT cooperate functionally to establish and maintain the repression at major satellites repeats as well as G9a target genes. A recent study suggested that the complex HP1-SUV39H1-TRIM28 undergoes a phase-separation process at the HC[235]. This phenomenon seems to be important for the HC compartmentalization, excluding the binding of transcriptional proteins such as TF. Nevertheless, a contradictory work suggested that HC is more consistent with collapsed polymer globules that undergo ‘digital’ switching between silent and active states[236].

SUV39H1 and SUV39H2 are very similar on a structural and mechanistical point of view[237]. In addition, they have an overlapping profile of expression during mouse embryogenesis[213], suggesting a functional redundancy. However, several specificities arose for SUV39H2. First, it has a specific expression in the adult testes, suggesting a particular role compared to SUV39H1[237]. Then, from a mechanistical point of view, it has been proved that the human SUV39H1 prefers H3K9me0 as substrate and proceeds directly to the two methylation and sometimes, to the third one in a slower way[237]. Surprisingly, SUV39H2 can also automethylated itself to modulate its own activity[238]. Finally, despite the fact that a KO for *Suv39h1* or a KO for *Suv39h2* are separately viable for mouse and mESC, the double KO for *Suv39h1* and *Suv39h2* leads to a loss of H3K9me3 at the pericentromeric HC specifically, suggesting SUV39H1 and SUV39H2 specific activity at those regions[239].

In conclusion, the actors of the epigenomic are enzymes able to modify chromatin, and more precisely, the interaction between DNA and histones. These interactions are modulating the accessibility of the chromatin to the transcriptional machinery and therefore impacting the gene expression. These enzymes need to be tightly regulated to initiate, maintain and spread the HC in coordination to the cellular physiology. More particularly, we will investigate the role of HC in mESCs and during the early development.

4) Heterochromatin biogenesis and function

Globally, HC is characterized by hypoacetylation and specific methylation marks on histones. More precisely, cHC exhibits H3K9me3, H3K64me3 and H4K20me3 whereas H3K9me2 and H3K27me3 are features of the fHC[152]. HC histone marks lead to fiber compaction and lack of transcription. To ensure its biological function, the HC needs to be initiated at the right locus, controlled and inherited to the future generations. To this purpose, the activities of the chromatin-modifying enzymes are tightly regulated.

a) Heterochromatin biogenesis: a complex orchestration

Histone PTMs play a crucial role in the control of gene expression. Indeed, modifications of histones tails affect inter-nucleosomal interactions and in turn the overall chromatin structure. Some histone modifications alter both the charge and the structural interaction within the nucleosome. It is the case of acetylation that neutralizes the lysine's positive charge, and it potentially loosen the histones–DNA interaction, favoring chromatin opening and gene accessibility to the transcription machinery[240]. Unlike acetylation, histone

methylation is a relatively inert modification[241] and it can be associated to both transcription activation and repression. The positions of the methylated histone residue suggest that the modified residue may function as a steric hindrance and as a nucleation site for 'effector' proteins[241]. Regarding HC, chromatin remodelers may be recruited to move nucleosomes position and compact the chromatin. H1 is characteristic of HC. DNA wrapping by H1 includes an additional 20 bp of DNA with respect to other histones, leading to a more compacted chromatin[242].

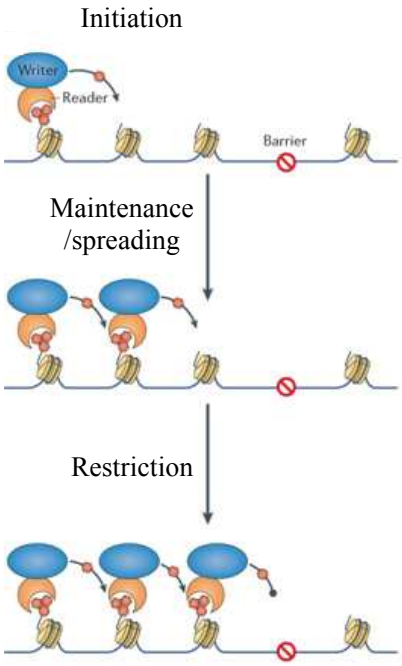


Figure 7: **Model for hetero-chromatin biogenesis.** Schematic representation hete-rochromatin initiation with the cooperation of a writer and a reader, followed by the maintenance/spreading of the heterochromatin enhanced by the already PTM chromatin and finally spreading restriction (adapted from Allshire et al.[152]).

The biogenesis of HC is also a complex orchestration **(Fig. 7)**. The mechanisms used to initiate the HC are not clearly deciphered yet. In contrast to telomeres, the repeated sequences at centromeric and pericentromeric regions are poorly conserved without notable genic elements as promoters, suggesting that their functions do not depend on specific DNA sequences but rather on epigenetic features. The repetitive nature of the sequences is nevertheless thought to trigger HC formation[139]. Some HMT have been shown to methylate histones prior to HC assembly. This is the case for PRDM3, PRDM16 and SETDB1, which are able to deposit H3K9me1[243a], the obligate template of SUV39H1 and SUV39H2 for di- and trimethylation. Interestingly, some chromodomains have been shown to interact DNA, giving a potential explanation to a *de novo* H3K9me1 deposition which then triggers the formation of HC[243b]. In addition, it has been shown that the HC is not transcriptionally inert. Satellites sequences are transcribed, from both strands, at low level – with respect to their amount – and despite the presence of high levels of H3K9me3 and H4K20me3[224]. Transcription occurs at different moment during the cell cycle, mainly during DNA replication, when DNA becomes accessible, but also in a cell-type dependent fashion, during the development for example. Transcription has also been suggested to participate in HC initiation by recruiting specific factors[139], but the exact recruitment mechanisms are still unclear. In addition, it has been suggested that clustered DNA TFs binding motifs outside of a genic context are critical for HC initiation[244]. Concerning the fHC, its initiation is thought to mainly occur through the recruitment of HMT and HDA by specific proteins, such as Rb for

the *Cyclin-E* gene[222]. Those enzymes will then induce H3K9me2 and H3K27me3 at the gene promoter, shutting down transcription.

Once the HC has been initiated, it needs to be maintained and spread across appropriate regions. This happens through the intrinsic properties of HMT that usually anchor to histones at the very modifications they catalyze and therefore expand, as already documented for SUV39H1[218,219]. Other factors are however important to recruit HMT. HP1 is one of the main examples, acting as coordinator of multiple actors and activities. For instance, HP1 can be a recruiting platform for multiple HMT including SUV39H1 and DNMT3B, thereby linking H3K9 and DNA methylation[227]. Indeed, some epigenetic marks are linked. The *Suv39h1/h2* double KO alters the DNA methylation profile at satellites sequences[216] and reciprocally, the *Dnmt1/3A/3B* triple KO results in a loss of H3K9me3[245]. Another example is SUV420H1 and SUV420H2 that depend on H3K9me3 to catalyze H4K20me3. SUV420H also interacts with HP1[246]. Their role in transcriptional silencing is still elusive, but they are required for a proper sister chromatid cohesion and further segregation[247]. The cross talks between EZH2 and DNMTs are not yet fully understood. Due to their physical interaction, EZH2 has been suggested to be required for DNA methylation[248], nevertheless, H3K27-methylated regions are largely devoid of DNA methylation[249]. Other mechanisms are important for the HC maintenance: (1) the relatively low turnover of nucleosome, most probably due to the condensed state of the chromatin[250] and (2) recruitment of HDA such as SIRTUINS proteins[251]. Indeed, SIRT1 KO in mESC leads to an increase of acetylated marks as H3K9ac and a decrease of H3K9me3 and H4K20me1.

Nevertheless, the cHC needs to be tightly restricted to not spread over deleterious regions. Different spreading control processes have been studied. One of the main examples are the transfer RNA (tRNA) genes[252]. They constitute a class of HC-spreading-barrier conserved from yeast to human. They have large nucleosome-free regions that form a “gap” that some reader-writer machineries cannot cross. There is also the recruitment of antagonist factors that are able to remove the heterochromatic marks with KDM and add other ones as histone acetylation. In yeast, it has been proved that the reader *swi6* recruit *epeI*, an histone demethylase, at HC[152]. Finally, the EuC itself can antagonize to HC spreading. The presence of active marks such as H3K4me3 and the histone variant H2A.Z at the boundaries of nucleosome depleted regions acts as a natural barrier[152].

Another important concept is the inheritability of HC. The inheritance is fundamental during strong cellular changes upon differentiation for example, or during the development.

Both cHC and fHC are maintained throughout the cell-cycle to the daughter cells. Indeed, during DNA replication, the parental nucleosome is randomly distributed to sister chromatids. The histone modifications are restored on the newly synthesis strand by copying from the parental nucleosome. This is the case for H3K9, H3K27 and DNA methylation. This suggests that HC self-propagates in a DNA sequence-independent manner[152]. This is called "cis inheritance of a chromatin state". Interestingly, H3K9me3-mediated repression is thought to be reversible whereas DNA methylation is thought to be more permanent, with the silencing for many cell divisions[253].

b) Functions of the heterochromatin in cellular physiology

The HC is known to have three main functions: (1) locking a cell-type specific identity upon differentiation (2) ensuring a proper centromere function during the cell division and (3) protecting the genome by repressing transposable elements from producing genomic instability. The studies of HMT played a pivotal role in these functions understanding.

Gene silencing by HC allows controlling cell-type specification. Large regions of DNA with H3K9me3 form barrier to TF-mediated cell-type reprogramming[254]. *Suv39h1* is also important to lock up the cell identity. The overexpression of the human version of SUV39H1 in mouse impaired erythroid differentiation[255].

Another main function of *Suv39h1* and *Suv39h2* is their role in centromere functions. Although both *Suv39h1* or *Suv39h2* single KO are viable, most probably due to their functional redundancy, the double KO (dn) is inducing chromosome mis-segregation during both mitosis and meiosis in male germ cells[239]. In addition, Chu et al.[256] suggest that SUV39H1 generates a gradient of methylation marks at the kinetochore that provides spatiotemporal information essential for accurate chromosome segregation in mitosis. Whereas SUV39H1 level appears to be constant during the cell cycle, 2 phosphorylated forms are increased specifically during the mitosis phase. Others HMT are important during the cell cycle, G9A and GLP interact with DNMTs to coordinate DNA and histone methylation during the replication[257].

Finally, SETDB1 as well as SUV39H1 are playing a central role in the genome stability. Indeed, both SUV39H1 overexpression and the dn are inducing an increased risk of tumorigenesis via genomic instability and deregulation of several tumor suppressor proteins, respectively[239,255]. The transcriptional repression of transposable elements is one of the main barriers to genomic instability. SETDB1 and SUV39H1/H2 have all been shown to repress repeated sequences[258]. SETDB1 is able to propagate H3K9me3 at some ERVs elements[259], adding an additional layer of epigenetic silencing. From the same perspective, Lehnertz et al[216],

reported that the dn in mESC failed to propagate H3K9me3 and DNA methylation at pericentric satellite repeats, underlying the importance of SUV39H in repeats silencing. Another mechanism to insure genomic integrity is the epigenetic regulation of telomeres. SUV39H family play a key role in telomere length, the dn results in a decrease in H3K9me2/me3 level at the telomere concomitant with an increase of H3K9me1, inducing an increase length of telomeres[215].

5) Heterochromatin and mouse development

The HC, via its function in genomic silencing, plays a pivotal role in maintenance of pluripotency and during early development. It provides a barrier to reprogramming by locking the cell identity and alternative differentiation pathways but it also contributes to the chromatin plasticity to allow cells to adapt quickly to environmental cues as it can happen during development.

a) Heterochromatin and pluripotency

From the same perspective, mESCs have a more open and plastic chromatin compared to differentiated cells, with a global lower density of nucleosome. Tosolini et al.[260] performed a comparative study of the different pluripotencies, comparing 2i/L and FCS/L culture conditions. The two pluripotent states are different but completely interconvertible, including at the genetic and epigenomic levels. mESCs cultured in 2i/L display a global DNA hypomethylation and a low level of H3k27me3 at gene promoters with respect to FCS/L condition. In addition, the chromocenters are rearranged, with a local diminution of H3K9me3, DNA methylation and an increase of H3K27me3, suggesting a redistribution from promoters to pericentric HC. Furthermore, there is also a clear decrease of H3K9me2. Finally, the inhibition of the ERK pathway, constitutive of the 2i/L media, induces the upregulation of JHDM3C, a KDM of H3K9, reinforcing the decrease of H3K9 methylation and the redistribution of H3K27me3. In FCS/L, cells exhibit higher level of H3K9me3 and 5mC at major satellites sequences. Interestingly, *Suv39h* KO or *Dnmts* KO induce a loss of H3K9me3 and 5mC respectively at the pericentric HC[216,245], replaced by H3K27me3 locally. Those two phenomena illustrate the plasticity between the different components of the HC, H3K9me3 and H3K27me3. Along the same line, differentiating cells show accumulation of H3K9me3 and DNA methylation.

Another study by has been performed to compare more specifically the transposable elements behavior between FCS/L and 2i/L[261]. Upon culture in 2i/L, H3K9me3 remains stable

while H3K9me2 decreases and H3K27me3 accumulates at transposons, illustrating again HC plasticity. Of note, H3K9me3 and H3K27me3 are present at different transposon families or different territories within the same family.

Despite the derivation of mESCs directly from the embryo and the fact that they constitute a powerful system to study pluripotency, the chromatin in ESCs differs significantly from the embryo. The two main different characteristics are nuclear organization and prevalence of cHC.

b) Heterochromatin during the mouse early development

One of the main difficulties to study chromatin in the embryo is that many biochemical approaches cannot be applied owing to the small amount of material available from each embryo. Nevertheless, in the recent years, technological advances allow to study more precisely the nuclear organization, mainly through the genome accessibility and the 3D structures. Indeed, at large scales, chromosomes segregate into regions of two mutually exclusive types of chromatin. The compartment "A" is gene- and active chromatin-rich while the compartment "B" is mostly composed by repressive chromatin. At a smaller scale, chromosomes fold into domains with preferential intradomain interactions (as compared to interdomain interactions), these contact domains are called "topologically associated domains" (TADs)[262]. Another important concept in the study of early development is the sexual-specific imprinting. It is an epigenetic mechanism that induces parental-specific gene expression and affects ~1% of genes. It is mainly characterized by a differentiated DNA methylation on one of the two copies. The imprinted genes expression is controlled by an imprinted control region (ICR)[263]. To date, 23 ICRs have been identified that control 130 imprinted genes[264]. This feature constitutes a parental-inherited long-term silencing mechanism in the progeny, possibly even until adulthood.

<u>i) Heterochromatin fate during the germ line differentiation</u>

The embryonic chromatin is very particular and dynamic during the early developmental events. Starting from highly epigenomically different germ cells, the embryo undergoes a wave of epigenetic reprogramming to enable them to generate the multiple cell types in the organism and to support extra-embryonic tissues. Maternal and paternal gametes are very asymmetric. During the gametogenesis, they experience two waves of reprogramming to erase somatic epigenetic marks and establish sex-specific epigenomic profile, including the X chromosome reactivation for the female gamete[264].

The first epigenetic reprogramming occurs around the sexual specification step during the migration to the genital ridge, between E7.25 and E9.5. There is a loss of H3K9me2 via the diminution of GLP expression followed by a gradual loss of H3K27me3[265]. The 5mC is also loss around E8.0, concomitantly with a decrease of DNMT3A/3B expression. At E11, there is another and wider reprogramming with a complete loss of DNA methylation and histone modifications via a genome wide histone replacement[266]. Demethylation occurs at many repetitive elements and imprinted genes.

In the male germ cell, the DNA methylation is re-established from E13.5 to birth. The paternal genome experiences dramatic changes in addition to DNA methylation. Nucleosomes are replaced by protamines, smaller structural proteins that lead to more condensed chromatin[264]. However, in mice 2-4% of the genome still contains nucleosome[267]. Interestingly, H2A.B is a testis-specific variant histone that plays a role in meiotic germ cells and has also a post-fertilization role[268]. Protamines positioning happens just after a wave of chromatin hyperacetylation. In addition, the paternal genome also harbors H3K4me2/me3 at genes important for spermatogenesis and basic cellular processes. Despite a clear erasure of HC marks, some genes important for the development as *Sox2* and *Gata6* present H3K27me3 at promoters[269]. Furthermore, some promoters are bivalent, with the co-existence of H3K27me3 and H3K4me3, poising the future transcription[270]. There is also low levels of H3K9me3, ensuring sex-specific marks and gene imprints[263].

The epigenomic profile of the oocyte is close to somatic cells. Oocyte undergoes an acute transcriptional activity during the growth phase of development, allowing the formation of RNAs and proteins stocks necessary for the early steps of the embryonic development[42]. In parallel, the acetylation level increases during the transcription followed by a sharp decrease as the meiosis starts[271]. The maternal genome exhibits active marks as H3K4me2/me3, H3K36me3 and H3K79me2/me3 as well as repressive ones as H3K9me2/me3 and H3K27me2/me3[264]. Alternative oocyte-specific promoters are commonly located upstream of somatically used promoters. These downstream promoters frequently become methylated in the oocyte and not in the sperm, such as genes encoding members of the DNMT family, *Dnmt1*, *Dnmt3b*, and *Dnmt3L*[272]. Oocyte chromatin presents also non-canonical histone modifications[165]. Non-canonical H3K4me3 (ncH3K4me3) corresponds to the propagation of H3K4me3 to broad domains around promoters and at intergenic regions. This correlates with a lack of expression. ncH3K27me3 is also present at intergenic regions and gene-deserts. In the maternal progenitor, *de novo* DNA methylation also occurs but after birth during follicular growth. This methylation in dependent on transcription, inducing the presence of partially

methylated domains (PMD). ncH3K4me3 and ncH3K27me3 overlap with partially methylated domains but within different subregions[155]. In addition, oocytes present a very particular HC nuclear organization with the formation, at the end of its maturation, of a ring of HC around the nucleoli, forming SN GV[273]. The maternal gamete displays with very few compartmentalization and TADs, however the presence of H3K27 methylation induces the formation of "polycomb-associated domain" (PADs), specific of the oocytes[274]. The somatic subtypes of linker histone H1 are not present during oogenesis. Instead, an oocyte-specific H1 subtype H1oo is present[275]. The role of this H1oo in the development is unclear, but in mESCs, it binds to chromatin and hinders the differentiation by maintaining expression of pluripotent genes through regulation of the chromatin structure. For the development, it has been suggested to be important for transcriptional activation following fertilization.

This epigenetic establishment results in two highly asymmetric gametes that will be merged together at fecundation, inducing a novel and last wave of reprogramming to erase the sex-specific marks to a certain extent and allow the constitution of a totipotent zygote. We will now present the dynamic of the embryonic chromatin along the first steps of the development.

ii) Heterochromatin dynamic during the pre-implantatory development

Right after the fecundation, the two parental genomes remain physically separated as PN, and although they both have access to the same maternal factors, they go through very distinct chromatin remodeling events. Indeed, even after physical reunion, the paternal and maternal genomes remain compartmentalized until the 8 cell-stage, when the embryonic cells start to be determined.

The paternal PN displays TADs and nuclear compartment A and B as soon as the G1-phase[274]. However, it is not clear if it is an inherited or very quickly established structure. Upon the fertilization, the paternal PN undergoes several decondensation cycles that allow a more permissive chromatin. Meanwhile, protamines are replaced with histones that are hypomethylated and hyperacetylated, with specific inclusions of H3.3 and phosphorylated H2A.X[276]. These histones are free of methylated marks such as H3K9me2/me3, H4K20me2/me3, H3K4me2/me3, H3K27me2/me3. This absence of cHC marks induces a replacement by H3K27 methylation at cHC. Especially, the loss of H3K9m2/me3 is supposed to facilitate the subsequently DNA demethylation[277]. Indeed, the paternal genome experiences an active, quick and global DNA demethylation before the beginning of the DNA replication. Some TET-independent pathways are involved, such as the replacement of protamine by hypomethylated histone[276]. Moreover, recent studies showed that some of 5mC are not removed

but in fact hydroxylated into 5hmC by TET enzymes. However, some regions remain methylated: some imprinted genes, repeats such as IAP and centromeric regions[277]. Interestingly, some H3K9me1 have been detected[278] which are hypothesized to protect loci from DNA demethylation but also to be a marker for later H3K9 methylation. In addition, the hyperacetylation associated with DNA hypomethylation allows the start of transcriptional activity at the S-phase and followed by a progressive increase in next developmental steps.

The maternal PN develops very differently. Soon after the fertilization, meiosis resumption generates the second polar globule and the formation of the PN. Then, the genome undergoes a wave of DNA demethylation much slower than the paternal genome[273]. It is thought to be a passive process associated to cell division and also due to the limited available amount of DNMTs. The DNA methylation colocalizes with H3K9me2 and not H3K9me3, suggesting a role for H3K9me2 in the DNA protection from demethylation[279]. Meanwhile, HC features such as H4K20me3, are removed. H3K9me2/me3 is also lost in a passive manner, mainly by dilution and prevention of *de novo* methylation[278,280,281]. This loss of cHC is followed by an acquisition of H3K27 methylation at those loci. H3K27 methylation also plays a role in genomic imprinting at hypomethylated promoters[283]. This H3K27 imprinting is lost at the epiblast specification but it is an important step for proper DNA imprinting and later development.

Both paternal and maternal PN are exhibiting a very particular nuclear organization with the localization of major satellites in rings around the nucleolar-like bodies[273]. This structure persists until the late 2-cell stage and is reorganized in chromocenters at the 4-cell stage. Along the cleavages, the interactions inside TADs are increasing in both PN, mainly dependent on the DNA replication rather than the ZGA[284]. They also undergo a wave of histone re-mono-methylation, followed in a delayed manner by the di- and tri-methylation, especially for H3K9. This delay was first thought to be triggered by maternal inhibitory factors, maintaining a differential level of H3K9me2/me3 between maternal and paternal genome until the late 2-cells stage[280]. A recent study however demonstrated that SUV39H2 catalyzes H3K9me3 at the paternal PN in the late zygote, and this *de novo* deposition does not repress transcription[285]. The previously established ncH3K4me3 represent 22% of maternal genome and are replaced during the ZGA. This is one of the reasons why KDM5A/B are essential for the ZGA[286]. Unexpectedly, repetitive sequences, including Te, are reactivated around the 2-cell stage[287] with the loss of H3K9 and DNA methylation and a deposit of H3K4me3. MERVL are reactivated and drives ZGA transcription and pluripotency acquisition[288]. LINE elements are also expressed along with a low level of retrotransposition events. Transposable elements activation plays a role in

the ribosomal DNA (rDNA) expression control, the exit of the 2-cell stage and the chromatin remodeling. It has also been shown that LINE1 RNAs degradation is necessary for the development[289]. The chromatin remodeling and accessibility, synchronized between male and female genome, is essential for a proper ZGA and the binding of the pioneering factors. Oct4 is playing a central role to this purpose[284].

From the morula and at the blastocyst stage, specific cell-lineages start to be determined. H3K9 methylation is progressively re-established, especially at LTR and other repeated sequences. The deposition of H3K9me3 at promoters does not happen until the post-implantation stages. H3K9me3 is thought to be a barrier to reprogramming to totipotency[290]. More specifically, a loss of H3K9me3 induces abnormal 3D chromatin structure, suggesting that H3K9me3 also constitute a barrier to chromatin structure reprogramming [284]. In parallel, DNA methylation is also re-propagated, in a lineage-specific manner. The presence of bivalent promoters is reduced in pre-implantation embryos compared to mESC[291]. Nevertheless, their numbers increase from the blastocyst stage with some “super bivalency” regions in epiblast around E6.5 that are involved in the regulation of developmental genes[284].

In conclusion, the development period immediately after fertilization is characterized by a unique organization of the DNA in the nuclear space. The atypical epigenetic reprogramming is an exquisitely controlled process, involving both global re-establishment of most epigenetic marks and locus-specific regulation. It is leading to a particular chromatin, particularly open and plastic that will support the huge changes that will occur during the development.

III) Long Non-Coding RNAs

1) History and classification of non-coding RNAs

a) Discovery of non-coding RNAs

The central dogma of biology states that genetic information stored into DNA is transcribed into RNA and subsequently translated into proteins. However, for several decades, scientists have known the existence of RNAs that are not translated into proteins, formally known as non-coding RNAs (ncRNAs). Indeed, tRNA and ribosomal RNA (rRNA), with structural and regulatory roles in translational machinery, are abundant ncRNAs and represent notable and accepted exceptions to the central dogma[292]. With the new generation of whole-genome technologies, it is known that 70 to 90% of genome is transcribed at some point during development[293]. However, since coding sequences represent near 1% of eukaryotic genome, it was proposed that thousands of genes are potentially transcribed as ncRNAs[293]. In the past, this pervasive transcription was first thought to represent junk, debris or by-products of coding gene transcription. Nevertheless, several lines of evidences established that these ncRNAs could play a pivotal role in gene regulation. Already in the 60's, Jacob and Monod suggested that RNA might be involved in transcriptional and/or post-transcriptional regulation of gene expression[294]. In 1984, the first ncRNA, a transcript of *micF* gene[295] was identified in bacteria Escherichia Coli. This ncRNA is operating through antisense base pairing to regulate the expression of a coding gene. Then, in mammals, the first ncRNA identified was the abundant hepatic fetal-specific transcript of *H19* gene[296]. This gene, supposed to be involved in liver development, is transcribed by RNA Pol II, spliced and polyadenylated but does not encode a protein. In their study, the authors suggested that the gene might act through its cytoplasmic RNA molecule rather than a protein. Two years after, the X inactivation specific transcript (*Xist*) ncRNA was identified as playing a role in sexual chromosome dose-compensation in mammals[297].

b) Non-coding RNA classification

It was difficult to identify a simple and unique biological function of ncRNAs because they display a very diverse pattern of expression, processing and localization. Then, it was necessary to classify them into different categories. There is a class of "small" ncRNAs, usually involved in gene regulation, such as small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), microRNAs (miRNAs), short interfering RNAs (siRNAs) and Piwi-interacting RNAs

(piRNAs)[298]. Among them, miRNAs, piRNAs and siRNAs are the most thoroughly investigated classes. miRNAs directly interact with partially complementary target sites located in the 3′ untranslated region (UTR) of mRNAs and repress their expression. More than 60 % of mRNAs miRNA target sites, suggesting a tight regulation by miRNAs[299] . In addition, it has been shown that many miRNAs are able on their own to target several hundred mRNAs, suggesting a complex and combinatorial mode of miRNA action in mRNA regulation. miRNAs are widely expressed in most mammalian cells and tissues where they play essential roles during differentiation and development[299]. They are produced from dsRNA, originated from extended hairpin structures or from base-pairing of sense and antisense transcripts, through cleavages by RNAses[299]. Another class of small ncRNA are the piRNAs class. They also have a wide diversity of biogenesis among mammals, flies and nematodes. In mammals, they appear to mainly function in the germline cells where they target and repress expression of transposable and repetitive elements to maintain genomic stability[299]. Another group of small regulatory RNAs, called endogenous siRNAs (endo-siRNAs) are involved in the silencing of repetitive transposable elements in the female germline cells of mammals. They share miRNAs maturation pathway, however, they have been shown to be specifically important for a proper meiotic progression during ovogenesis[300a].

Besides the biological relevance of the small ncRNAs, we will now focus on another class of ncRNAs called the long ncRNAs (lncRNAs). They are defined as non-coding transcripts of more than 200 nt in length[299] but their functional prediction and classification method for lncRNAs is still lacking. However, based on their relative position to neighboring coding-genes, they can be grouped into several broad but mutually nonexclusive categories[298] **(Fig. 8)**. In respect to their location to a nearby coding-gene, lncRNAs can be named antisense when they are transcribed from the opposite direction to their coding counterpart or named divergent when they originate from bivalent promoters which also control protein-coding genes. Intronic lncRNAs results from intron transcriptions within protein-coding genes and are transcribed to the same direction transcripts with intronic and/or exonic overlaps. Finally, many lncRNAs are transcribed from loci devoid of protein-coding genes, at distance at least 5kb, and are named long intergenic noncoding RNAs, lincRNAs (or large intervening). Recently, the enhancer RNAs (eRNAs)[300b] have been described as a new class of ncRNAs. They are produced by activity-dependent RNA Pol II binding at specific enhancers and they contribute to enhancer functions.

c) Long non-coding RNAs vs coding mRNAs

LncRNAs share substantial similarities with mRNAs. First, they are transcribed by the RNAPol II and most of them are spliced, capped and polyadenylated[293], with the exception of eRNAs that are unspliced and non-polyadenylated[300]. LncRNAs are commonly expressed at lower levels than mRNAs, are slightly shorter and exhibit more tissue- or cell-type specific patterns of expression[301]. Indeed, they are more tightly regulated than protein-coding genes transcripts with promoters as conserved as the mRNAs' ones. Second, they also exhibit a similar chromatin state to mRNAs, with H3K4me3, H3K9ac and H3K27ac marks at the promoters[302]. Notably, the particular class of eRNAs are marked by histone modifications such as H3K4me1/2 and H3K27ac[300b].

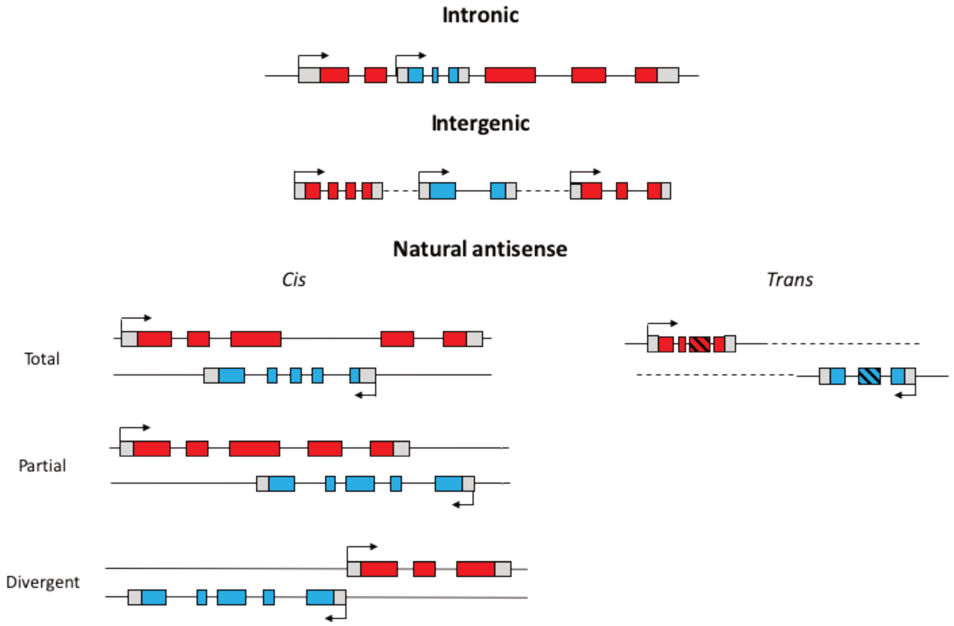


Figure 8: **Classification of lncRNAs according their relative position to neighboring coding genes.** Schematic representation of the different classes of lncRNA (in blue) according their neighboring coding-gene (in red). The boxes represent the exons, with grey ones for untranslated regions. The line within genes represents the intron, and outside DNA. Dashed lines indicate a faraway genomic location between two genes. Hatched boxes indicate a partial sequence complementarity.

However, lncRNAs and mRNAs also display different characteristics. Indeed, lncRNAs might also have a specific circuitry of regulation. For example, the KO of *Dicer1* in mESC, an enzyme intervening in the biogenesis of miRNA, results in the lower expression of hundreds of lncRNAs, particularly divergent transcripts[302]. It was also recently showed that the asymmetric distribution of polyadenylation and splicing signal sequences in the sense and antisense directions from a promoter dictates the strong difference in mRNA and divergent lncNA elongation and stability[304]. The enrichment of polyadenylation signals in the antisense side and splicing signal in the sense's one enhances an early and efficient polyadenylation on the antisense lncRNA and productive splicing and elongation of the mRNA. In the same

perspective, another recent study demonstrates that a component of transcriptional machinery WDR82 is also involved in the early termination of antisense lncRNAs[305].

Sequence conservation is on average much lower for lncRNAs than for their coding counterparts. The evolutionary conservation of lncRNAs suggests that they have a functional relevance. LncRNAs have a function independent of their sequence as the formation of secondary structures. Indeed, some mutations modify the primary sequence of an RNA but still preserve their base pairing[301]. LncRNAs often fold into complex, thermodynamically stable secondary and tertiary structures, which can be of crucial importance for their function. The protein coding potential of lncRNA is still under debate. Several studies have shown that RNAs that had been previously described as lncRNA are in fact associated with the ribosomes, in the cytoplasm, suggesting that they may be translated, in particular into small peptides[306]. However, ribosome occupancy alone is not sufficient to distinguish between coding and non-coding transcripts that may control translation[307]. Whether the production of small peptides from certain lncRNA represent "translational noise" or a new layer of regulation, remains still controversial.

2) Long non-coding RNAs mechanisms of action: from transcription to translation

a) High diversity of mechanisms

LncRNAs are thought to serve as molecular signals. Their cell type-specific expression and control by diverse stimuli suggest that their expression is under considerable transcriptional control[308]. For instance, many lncRNAs were linked to the regulatory network of TFs supporting pluripotency and lineage determination. A large subset of them represent the potential target of Oct4, Nanog and Sox2 in both mouse[309] and human[310]. LncRNAs transcription at a very specific time and place allow them to integrate developmental cues, interpret cellular context or respond to diverse stimuli. Then, categorizing lncRNAs on the basis of their molecular function – as generally done with proteins – is difficult. lncRNAs can potentially control every step of gene expression, from nuclear and chromatin organization to the efficiency of translation. Moreover, individual lncRNAs can act at different and independent stages of gene regulation. lncRNAs mentioned in this chapter III are summarized according to their mechanism of action in **Table 2**.

Some of nuclear lncRNAs act at the chromatin level to activate or repress gene expression. They can for instance interact with chromatin regulators to enable targeting to a DNA site, potentially explaining how protein complexes without intrinsic-sequence specific

DNA-binding ability find their DNA targets, in cis or in trans[308]. This is the case for the HOTTIP lncRNA that recruits the adaptor WDR5 and directs WDR5/MLL complex across its target genes, inducing H3K4me3 deposition[311]. From the same perspective, RNAs produced from the promoters of ribosomal DNA recruit DNMTs to induce their own silencing[312]. As the last example, this is also the case of *Xist*. X-chromosome inactivation is the epigenetic process that equalizes the dosage of X-linked genes between female (XX) and male (XY) in mammals. This operates through the lncRNA *Xist*. It is up-regulated from one of the two X chromosomes, and its RNAs spread in cis along the entire X chromosome and triggers the inactivation of more than 1,000 genes by PRC2 recruitment to induce H3K27me3 and subsequent chromosome-wide repression of gene expression[138].

LncRNAs can also act as a scaffold to enable interactions between different actors as proteins, RNA and DNA. The class of eRNAs are particularly suitable to illustrate this concept[300b]. Indeed, they can establish local interactions with proteins that either enhance or inhibit the transcriptional activity of the enhancer. They can regulate the chromatin accessibility to the transcriptional machinery or they can favor the elongation by interacting with the RNA Pol II at the gene locus. The transcription from pericentric major satellites sequences recruit HP1, SUV39H1 and other chromatin modifiers to induce its own silencing[224]. They can also interact only with proteins by modulating their enzymatic activity. This is the case of the *lnc-DC* which was found to regulate the gene expression involved in dendritic cell differentiation by interacting with STAT3 to prevent its dephosphorylation by tyrosine phosphatase SHP1[313]. In another case, LncRNAs can also replace DNA to sequester proteins. Indeed, Telomeres are transcribed into the lncRNA telomeric repeat-containing RNA (*TERRA*)[314], which physically interact with the telomerase to sequester it while it inhibits its function. Another example of lncRNA as a molecular decoy is the Growth Arrest-Specific 5 lncRNA (*Gas5*). The lncRNA imitates the DNA motif of an hormone response elements found in the promoter regions of responsive genes. Then, it competes for binding with the hormone and effectively precludes its interaction with the DNA[315].

LncRNAs are involved in the 3D nuclear organization. As the more classical example, lncRNAs *NEAT1*[316] and *MALAT1*[317] are involved in the nucleation of paraspeckles and nuclear speckles respectively. *NEAT1* is an abundant mono-exonic lncRNA, which by its own transcription and processing, starts the nucleation of the paraspeckle, a subnuclear structures characterized by high local concentrations of specific proteins and RNAs implicated in regulation of gene expression. It mainly occurs through binding of paraspeckle protein with the lncRNA. *MALAT1* is able to sequester several splicing factors to initiate the formation of

nuclear speckles. It is therefore acting as a molecular decoy. Finally, eRNAs are also participating to 3D structure of the genome by promoting enhancers-promoters loops[300b].

LncRNAs can also act at the post-transcriptional level both in the nucleus by influencing alternative splicing and in the cytoplasm. The lncRNA produced from the human FGFR2 locus promotes tissue-specific alternative splicing of the corresponding coding gene. Through the recruitment of Polycomb-group protein and KDM2A, it creates a chromatin environment that prevents binding of a repressive chromatin-splicing adaptor complex important for alternative splicing[318]. In addition, lncRNAs can influence the stability and translation efficiency of mRNAs, either as being a source of production of miRNAs, or by acting as a "sponge" that captures miRNAs inhibiting specific mRNAs. For example, the lncRNA Regulator of Reprogramming (*linc-RoR*) acts as a microRNA sponge, modulating the concentration of specific miRNA. It shares miRNA-response elements with the pluripotency TFs NANOG, OCT4 and SOX2 and prevents them from miRNA-mediated suppression in self- renewing human ES cells[319].

b) A particular class of long non-coding RNAs: Natural Antisense Transcripts

Natural Antisense Transcripts (NATs) form a particular but very frequent category of lncRNAs. They are endogenous RNA molecules containing sequences that are complementary to other transcripts. NATs can be divided into two categories: cis-NATs, which are transcribed from opposing DNA strands at the same genomic locus, and trans-NATs, which are transcribed from separate loci **(Fig. 8)**. Cis-NAT pairs display perfect sequence complementarity whereas trans-NAT pairs display imperfect complementarity and can therefore target many sense targets to form complex regulation networks[320]. This phenomenon is conserved across all species, from yeast to animals including mouse and human. The proportion of cis-NAT transcription is variable, with 29% in mouse genome[321] and 22% in human[322]. There are different criteria to classify NATs, but the degree of overlap is the most important one **(Fig. 8)**. The overlap can correspond only to the promoter in case of bidirectional divergent promoters. It can be partial, this is mainly the case for NATs with independent promoters but which are head to head and with a convergent transcription. It also can be total when the NAT has a cryptic promoter inside the sense gene and is internal. Several genome wide studies reported that the head to head are the most prevalent category[320].

Because of their position, this class of lncRNA presents a particularly high number of potential mechanisms to control the linked sense gene. They have been divided into four

categories: transcription interference, chromatin remodeling, RNA-RNA interactions in the cytoplasm and/or the nucleus and RNA-DNA interactions.

The transcriptional interference mechanism can only revolve on cis-acting NATs. It has been reported in budding yeast for the gene pair *gal10* and *gal7*[323]. Their competitive transcription elongation step leads to the collision of the two RNA Polymerase II. Such interference induces mutual repression, and alternatively results in the shutdown of both transcripts. Both cis- and trans-NATs can modify chromatin environment. *ANRIL* is a NAT that overlaps with *INK4b/ARF/INK4a* locus. *ANRIL* induces silencing of the locus through the recruitment of DNMTs and EZH2 methyltransferases[324]. LncRNAs can also act to activate gene expression. This is the case for the NATs named *Evx1as*, which is transcribed from the opposite DNA strand of *Evx1*. It recruits MLL HMT that propagates H3K4me3 on the locus[325]. Finally, epigenetic regulation can occur through DNA-RNA-Protein interactions. *HOTAIR* is a lincRNA encoded in antisense orientation in the HOX-C cluster. It interacts with PRC2 in initiation silencing and with specific DNA binding motif to target PRC2 onto a precise locus, but not necessarily the one of its transcription[326]. This is a trans-NAT.

Mechanisms	Chromatin modifications		3D organisation		Protein Activity modulation		Molecular Decoy		Transcriptional Interference	Alternative Splicing	
Interactions	RNA-Protein	Prot-RNA-DNA	RNA-Protein	RNA-DNA	RNA-Protein	RNA-RNA	RNA-Protein	RNA-RNA	Protein-Protein	RNA-RNA	RNA-Protein
lncRNAs	Xist, **Tsix**, HOTTIP, Major Satellites RNAs, rRNA, eRNA, FGFR2, **ANRIL**, **Evx1as**, **Kcnq1ot1**, **Airn**, lincGET, Oct4P4, Evf2, TUNA	**HOTAIR**	NEAT1, MALAT1	eRNA	In-DC, TERRA, lincU	**Uchl1-AS**	TERRA, Gas5, MALAT1	linc-RoR	**Gal7-Gal10**, Evf2	**ErbAa**	FGFR2

Table 2: **An insight into the high diversity of long non-coding RNAs mechanisms of action.** Presentation of the different lncRNAs mentioned in the section III according their mechanism of action and the different necessary interactions for their activity. LncRNAs in bold are NATs.

NATs can also act by RNA-RNA interactions. In the nucleus, it can induce the masking of specific sequences, as illustrated by NAT of the thyroid hormone receptor gene *ErbAα*[327]. Indeed, transcribed NAT masks a splicing site shifting the balance between two alternative variants. Moreover, NATs can also regulate their targets in the cytoplasm through RNA-RNA interaction. One suggested mechanism is that NATs produced from the 3'-UTR of a coding gene can help to stabilize or destabilize the mRNA by recruiting different factors or competing for the translational initiation complex recruitment. One illustration is the *Uchl1-AS*, which regulates the translation of its antisense mRNA by interacting with *Uchl1* mRNA, recruiting ribosomes and promoting translation[328].

Finally, one of the most famous NAT is *Tsix*, the antisense of the *Xist* gene, which illustrates how one lncRNAs can act at multiple levels on its mRNA target[329]. *Tsix* is a full

overlapping antisense of *Xist*. Interestingly, *Tsix* operates via its own act of transcription by keeping the chromatin open within *Xist*. It induces efficient H3K4 methylation over the entire *Xist/Tsix* unit. Nevertheless, the RNA molecule also recruits repressive HMT complexes to enrich *Xist* promoter with repressive marks as DNA and H3K9 methylation.

3) lncRNAs impact pluripotency and early development

At present time, several lncRNAs have been involved in a wide range of biological phenomena such as developmental processes, disease and cancer biology. Increasing efforts have been made to dissect functional relevance of lncRNAs in pluripotency and during development. As already noted above, lncRNA can act as environmental signal integrator, and particularly they are linked to the pluripotency network. Moreover, it was shown that loss-of-function of several lncRNAs is similar to knock-down of well-established ES cell regulators[309]. It is, at present, known that lncRNA operates on different aspects of development and pluripotency.

LncRNAs are involved in a part of mESC pluripotency network. Several of them have been characterized. One main example is the *linc-ROR*, which enhances reprogramming efficiency of iPS cells by promoting the activity of pTF as OCT4, NANOG and SOX2[319]. Since *linc-ROR* expression is controlled by pTF, it creates a self-sustaining feedback loop. The lncRNA *TUNA* (for Tcl1 upstream neuron-associated lncRNA) is also involved the regulation of pTFs. Indeed, with other RNA-binding protein (PTBP1, hnRNP-K and NCL), it binds to NANOG and SOX2 promoters to modulate their expression[330]. Finally, a recent identified lncRNA *lincU* was reported to be regulated by Nanog. It stabilizes DUSP9 protein, an ERK-specific phosphatase, preventing its ubiquitination and degradation. This effect results in the repression of the ERK1/2 signalling pathway activity[331].

LncRNAs also play a role in the cell fate specification. *Evf2* lncRNA has been shown to be involved in the brain development[332]. DLX5 and DLX6 genes are transcribed near *Evf2* locus and have a role in development of specific neurons. *Evf2* operates through transcriptional interference in cis to repress *Dlx6* expression while it act in trans via chromatin remodeling to shutdown *Dlx5* expression. The antisense lncRNA *Oct4P4*, which is transcribed from a pseudogene *Oct4,* have been reported to regulate *Oct4* expression[333]. In that case, Oct4-lncRNA is thought to be associated with HP1 and SUV39H1 and targets *Oct4* promoter among others to induce heterochromatinization, impacting global pluripotency network and the cell fate specification. Recently, the lncRNA *lincGET* have been reported to bind the Histone arginine

transferase CARM1 to promote the nuclear localization and favor the H3R26 methylation at ICM-specific genes[28]. This subsequently increases chromatin accessibility. This phenomenon happens as soon as 2-cell state.

Gene imprinting is a phenomenon primordial for the development that also includes participation of lncRNAs. Interestingly, even though not all imprinted lncRNAs may have a functional role in imprinting, it is intriguing that most of known imprinted domains contain at least one lncRNA with anti-correlated expression[334]. The two canonical examples of lncRNA involved in the imprinting silencing of a specific allele include the NATs *kcnq1ot1* and *Airn*. They map into *Kcnq1* and *Igfr2* imprinted gene clusters respectively. *Kcnq1ot1* interacts with the HMT G9a and PRC2, inducing the repression in cis of the neighboring genes specifically of the paternal allele[335]. Similarly, *Airn* is imprinted and expressed only from the paternal allele, and its transcription is required for repression of several imprinted genes on paternal chromosome, specifically in the placenta. It is thought to recruit chromatin modifiers enzymes to induce the silencing[336].

Finally, the paradigm that involved lncRNA, epigenetic silencing and functional relevance during development is X chromosome inactivation by *Xist*[329]. As previously explained, sexual dimorphism necessitates to equalize sex-chromosome dosage between male and female. This is occurring through the inactivation of whole X chromosome. This inactivation can be imprinted, leading to specific paternal X chromosome inactivation in mouse extra embryonic tissues as placenta[329]. Inactivation can also be random, more particularly in embryonic tissues. Inactivation in reverted upon pluripotency acquisition, in ICM for example or in mESCs, leading to two active X chromosome[138]. Moreover, the two X chromosomes are also active during oocyte maturation[329]. *Xist* is originated from Xi center (Xic) and initiate X chromosome inactivation by heterochromatinization of the targeted X chromosome through the recruitment of HMT complexes[138]. Then, this silencing needs to be maintained. Interestingly, the lncRNA *Xist* is itself regulated by a NAT lncRNA *Tsix*[337]. Navarro et al[338,339] have shown that both lncRNAs are the target of pTFs. OCT4, NANOG and SOX2 bind in the first intron of *Xist* to repress its expression. REX1, KLF4 and C-MYC are located on *Tsix* promoter to active it. Moreover, REX1 is required for a proper *Tsix* elongation. Therefore, distinct complexes of pTFs couple X-inactivation reprogramming and pluripotency. In addition, a recent study[340] detailed the *Tsix* induced chromatin regulation of *Xist*. Firstly, *Tsix* transcription creates a transient open chromatin conformation with an increase of H3K4me3. Secondly, *Tsix* establishes a reversible silencing state with an increase of H3K36me3 and H3K9me3 and a decrease of H3K4me3. Finally, DNA methylation achieved the silencing in an irreversible state.

Thesis objectives

Pluripotency is a precarious equilibrium based on the expression of a pTFs network which activates pluripotency supporting genes and represses differentiation promoting genes. This balance, tightly regulated during development, is artificially maintained by culture conditions *in vitro*. Nevertheless, according culture conditions and developmental stage of derivation, there is a wide variety of mESCs with different characteristics. Therefore, it is unclear whether the spectruum of *in vitro* pluripotencies has any functional relevance for *in vivo* pluripotency.

Then, when a cell commit into a specific fate, during the development for example, genes important for cell fate specifically are expressed while genes of alternative fates are shut down. This repression is mainly occurring through their local heterochromatinization, mainly through H3K9 methylation. Interestingly, even between different *in vitro* pluripotent states, there is a global reorganization of the heterochromatin associated with an increase of H3K9 methylation. Moreover, there is several waves of epigenetic reprogramming during gametogenesis and first steps of embryogenesis, establishing a very low and particular embryonic heterochromatin.

Therefore, when a cell differentiates, pTF network and heterochromatin need to be tightly regulated. Nevertheless, molecular link between pluripotency and heterochromatin control has never been formally studied. Nevertheless, lncRNA have been reported to be important for pluripotency as well as heterochromatin. Important examples are lncRNAs *Xist* and *Tsix*. They are involved in the chromosome X inactivation, an essential feature for development.

The aim of my thesis was to identify a H3K9 specific HMT regulated by pTFs that would synchronize pluripotency dynamic and heterochromatin spreading. Thus, we found a pluripotency-associated lncRNA transcribed in antisense to *Suv39h1*. In a first step, we assessed if and how the lncRNA regulates *Suv39h1* and its functional impact in mESCs. In a second step, we also tried to evaluate its role *in vivo* during oogenesis and early embryogenesis.

Materials and Methods

The following materials and methods are presented in the chapter IV.

• **Cell culture conditions**

• **Generation and validation of Suv39h1as KO cell lines by CRISPR/Cas9 technology** (CRISPR for Clustered Regularly Interspaced Short Palindromic Repeats – for gRNA sequences, see **Table 3**)

• **Gene expression analysis** with RNA extraction, Reverse Transcription (RT) and quantitative real-time PCR, and genome-wide sequencing (for primers sequences, see **Table 3**)

• **Chromatin Immunoprecipitation** (ChIP) with the chromatin preparation and the Immunoprecipitation (for antibodies, see **Table 6**)

• **siRNA assays** (for sequences and pooling, see **Table 4**)

We will now present the materials and methods specific for the chapters V and VI.

• Generation of SunTag and Suv39h1as cDNA overexpressing cell lines

- Plasmids cloning

Several guide RNAs (gRNAs) were designed with the help of online design tools (sam.genome-engineering.org/database/, crispr.mit.edu/, chopchop.cbu.uib.no/) providing a region from -50bp to -500bp upstream target gene TSS as input. Highly ranked gRNAs for off-targets as well as on-target predictions score were selected. gRNAs sequences were inserted into plasmids containing a U6-promoter driven expression cassette along with a puromycin selection cassette (see guides sequences in **Table 3**). Suv39h1as cDNA was ordered as a block at IDT. It was cloned under the control of a CAGG promoter in a plasmid containing a puromycin resistance cassette.

- Lipofection for the SunTag system gRNA insertion or transient cDNA expression

E14Tg2a WT or SunTag cells[341] were plated in a 6-well plate at day 0 (D0). Meanwhile, either 1 µg of transposase coding plasmid (Piggybac) and 1 µg of plasmid containing the gRNA with a Puromycin cassette or a Suv39h1as cDNA containing plasmid with a Puromycin cassette were pre-mixed in 250 µL of DMEM Medium without serum. 5 µL of Lipofectamin 2000 (Invitrogen, 11668-019) was added to 250 µL of DMEM Medium without serum in a separated tube. After 5 min, both tubes were mixed (final volume 500µL) for 30 min at room temperature to allow complexes formation. Finally, complexes were added to the culture medium to allow cell delivery. For the SunTag cell lines, puromycin (1 µg/mL – P9620-10mL, Sigma) selection

was performed from D2 to D5 to select cells that have integrated sgRNA plasmids. For the cDNA transient overexpression, cells were selected for 3 days and harvested for RNA extraction.

- Electroporation for stable cDNA insertion

10 µg of the plasmid was first linearized by enzymatic restriction. Then the linearized plasmid was precipitated and resuspended in 100 µL of PBS. In parallel, 10 million of cells were harvested and resuspended with the linearized plasmid in a total of 800 µL of PBS. The solution is then placed in a 4 mm cuve and placed in the electroporator (ECM 830, BTX Harvard apparatus). The run settings were: LV and unipolar mode, 240V voltage and Plenght 002 ms. When the run is done, cells were placed in a pre-warm media, incubated for 5 to 10 min to recover and distributed at a clonal density. The day after, the media was renewed and an antibiotic selection can be added if necessary.

- Cell line establishment by cell clone picking

For the SunTag and the stable cDNA insertion cell lines, after 10 days of growth at clonal density, colonies were picked to establish homogeneous cell lines. Then cells were harvested for RNA extraction and gene expression measure to validate the cell lines.

• Topo cloning for mouse genotyping or isoforms

For mouse genotyping, the genomic DNA was obtained from the lysis of a piece of tail with a lysis tail buffer (50 mM TrisHCl pH 8.5; 100 mM NaCl; 0.5% Tween20 and 100 mg/ml proteinase K (GEXPRK01-B5, Eurobio scientific) and incubated overnight at 56°C then 10 min at 96°C. For the identifications of Suv39h1 isoforms, the cDNA was obtained from a oligo dT RT reaction (First Strand cDNA Kit, Roche, 04379012001) following manufactory's instructions. The PCRs were performed on 1 µL of mouse genomic DNA or 2 to 5 µL of cDNA, with primers indicated on **Table 3** and with LongAmp Taq PCR kit (BioLabs, E5200S) or Q5 High Fidelity (NEB, M0491S) following manufactory's instructions. The PCR product was run on an agarose gel for validation and purified thanks to NucleoSpin Gel and PCR Clean-Up (Macherey-Nagel, 740609.50). 2 to 4 µL of the purified product was added to pCR™-Blunt II-TOPO® solution (ThermoFisher, cat n°450245) as indicated by the manufactory's instructions. The solution was incubating for 30 min at room temperature and transformed in competent bacteria and incubated overnight at 37°C. 10 colonies were picked, replated on a bacterial plate and exploded in 100 µL of water and incubate at 95°C for 5 min, in the purpose to use their DNA for subsequent colony-PCR assays. PCRs were performed on the bacterial DNA with

LongAmp Taq PCR kit (BioLabs, E5200S) and run on agarose gel for validation. The correctly amplificated products were sent for sequencing (by Eurofin – GATC).

• Strand specific RT reactions

RT reactions were performed with locus specific primers and a Tbp-R primer (see primers **Table 3**) on 0,05 µg of total RNA according to manufactory's instructions (First Strand cDNA Kit, Roche, 04379012001). In parallel, the same reaction was performed with only Tbp-R primer. The cDNA was then diluted 6 times and used to proceed with qPCR.

• Embryoid bodies differentiation

- Not size controlled

Cells were seeded at 0,04 million cells/cm^2 and cultures for 3 days. After, cells were washed once with PBS 1X and treated with pre-warmed trypsin. Partial dissociation of ES cells colonies was evaluated under the microscope and inactivation with a large volume of FCS medium was done 1 or 2 min after trypsinization to allow small clumps of cells to be maintained. Cells were carefully recovered with a 10 mL pipette and transferred in a 50 mL falcon tube to avoid further dissociation. After few minutes, when the clumps progressively reached the bottom of the tube, as much as possible supernatant was gently removed without perturbing the accumulated clumps of cells. 10 mL of FCS medium was gently added to the tube and the cells clumps were precociously resuspended and transferred into bacterial Petri dishes thus precluding cell adhesion (Day 0). Dishes were incubated at 37°C in 7% CO2. Medium was changed every day by carefully collecting the clumps of cells with a 10 mL pipette and replacing them into bacterial Petri Dishes in FCS medium for 6 additional days and by splitting them, if necessary, into several dishes. At Day 6, the biggest embryoid bodies were collected by allowing the clumps to decant for a short time followed by the quick aspiration of the supernatant. They were subsequently replated on gelatinized surfaces at low density to allow for bodies adhesion and cell differentiation. One day later, adhesion of the embryoid bodies was checked under the microscope. Medium was changed every day and differentiating samples were collected at Day 6 and Day 10 for RNA extraction and gene expression analysis.

- Size controlled: hanging-drop protocol

A suspension of 0.1M cells/mL was distributed in 20 µl drops (2000 cells/drop) onto inverted cover plates of several dishes filled with PBS 1X to avoid evaporation. The day after, the aggregates formed in each drop were pooled and seeded on non-cell culture treated dishes and

cultured in FCS media without LIF for 8 days. EB were harvested for RNA extraction at day 6 and day 8.

• EpiLSC differentiation[126]

Cells cultured in FCS/L medium were adapted to 2i/L medium for 3 passages (9 days in total) before starting the differentiation protocol. The EpiLSCs differentiation was induced by plating 0.23 million ES cells on a well of a 6-well plate coated with human plasma fibronectin (16.7 mg/ml – F2006, Sigma) in N2B27 medium containing activin A (20 ng/ml – 338-AC-010, Biocompare), rhFGF (12 ng/ml – 233-FB, R&Dsystems), and KSR (1% – 10-828-010, Gibco). Medium was changed every day until day 3 of differentiation. Cells were harvested along the assay (2i/L 3 passages, Day 1, 2 and 3 EpiLCs) for RNA extraction and gene expression analysis.

• Single molecule Fluorescence In Situ Hybridization (sm-FISH)

- Probes design

Using the online tool "Stellaris RNA FISH probes designer" from LGB biosearch laboratories, two probes have been designed. The first one is for the Suv39h1 gene with 47 oligos, 30 exonic and 17 intronic fused to the Quasar670 dye (equivalent to Cy5). The second one is for Suv39h1as with 35 probes only exonic fused to Quasar570 dye (equivalent to Cy3). Two other probes were design in Suv39h1as introns and fused to Quasar670 dye, AS1 with 47 oligos and AS2 with 41 oligos (sequences presented on **Table 5**). Each oligo is strand-specific. They were resuspended in TE Buffer (10 mM Tris, 1 mM EDTA, pH8) at a concentration of 12,5 µmol/L.

- Slides preparation

Cells harvested and counted. Cells were fixed in PFA 4% for 10 min at room temperature and then quenched with 150 mM Glycine/PBS. After a wash in cold PBS, cells were concentrated at 10 million cells/mL. 10 µL of the solution was dropped on a slide, spread to form a spot. Slides were let settle down in a humid chamber for at least 15 min. Then liquid was aspirated with a paper towel and the slide was air dried before being kept in 70% ethanol at least one day, or until use.

- Probes hybridization for sm-FISH

Slides were washed in 100% ethanol for 2 min and then air dried quickly. Then slides were placed in a wet chamber at 37°C overnight with 10 µL of hybridization cocktail per spot of cells

with a coverslip. The hybridization cocktail is composed of: SSC (2X – S6639-1L, Sigma), Dextran (Life technologies, 5%), Formamide (Sigma F9037, 2,5 M (10%)), E. Coli tRNAs (Sigma 10109541001, 2 µg/µL), Ribonucleoside Vanadyl Complex (VRC - NEB S1402S, 5 mmol/L), BSA (0,5 µg/µL – B9001S, NEB) and FISH probes for Suv39h1 and Suv39h1as (LGC Biosearch laboratories, both at 0,6 µmol/L). After removing the coverslip, slides were washed in a fresh SSC 2X/Formamide 10% for 30 min at 37°C. Finally, single spots were mounted using a small coverslip with 8 µL of Vectashield Antifade Mounting medium with DAPI (Vector Laboratories, H-1200-10) and sealed with nail polish. Sm-FISH images were acquired with a wide-field fluorescence inverted microscope Nikon Ti-Eclipse with X63 oil immersion objective (N.A1.4).

- Chromatic aberration assessment with tetracolored beads

To assess to chromatic aberration present between the Cy3 and Cy5, we used Tetraspeck microsphere beads (0,1 µm diameter - T7279 Invitrogene) that are colored with Blue 360/430 - DAPI, Green 505/515 - FITC, Orange 560/580 – Cy3 and Dark Red 660/680 – Cy5. They were attached to a super frost slide and image on a wide-field fluorescence inverted microscope Nikon Ti-Eclipse with X63 oil immersion objective (N.A1.4).

• Suv39h1as knock-out mouse line generation

All experiments were conducted at the Institut Pasteur according to the French and European regulations on care and protection of laboratory animals (EC Directive 86/609, French Law 2001-486 issued on June 6, 2001) and were approved by Ministère de l'enseignement supérieur, de la recherche et de l'innovation (N°089)

- CRISPR/Cas9 system electroporation in mouse embryos

The following steps were performed by the Mouse Genetics Engineering Center directed by Francina Langa Vives. 48 C57BL/6 J embryos were electroporated with 50 or 100 ng/µL of sgRNA ordered at IDT (sequences on **Table 3**) and with 100 or 200 ng/µL of Cas9 protein. They were then re-implanted in a surrogate mouse. Once born, the pup were biopsied for subsequent genotyping.

- Genotyping strategy

Genotyping was performed by PCR after tissue lysis in 50 mM Tris, pH 8.5, 100 mM NaCl, 0.5% Tween 20, and 100 mg/ml proteinase K at 56 °C overnight, followed by a 10-min incubation at 96 °C. PCR was performed with LongAmp Taq PCR kit (BioLabs, E5200S) with

primers indicated on **Table 3**. PCR products were separated by electrophoresis on 2,5% agarose gels (16500-500, Life Technologies). The PCR product was also cloned in a topo vector for subsequent sequencing (see section Topo cloning for mouse genotyping or isoforms)

- Generation of homozygous KO females

3 mice founders were chosen to established mouse lines, females #13 and #26 to obtain homozygous KO females and a male #44 to obtain WT females. The three founders were crossed with males and females WT C57BL/6 J to generate the F1 generation. For the line 13 and 26, the expected proportions and genotype were obtained with 50% of heterozygous females and 50% of hemizygous males. For the 44 line, 50% of heterozygous females and 50% of WT males were obtained. Then, littermate was crossed together as followed to obtain the F2 generation. For the 13 and 26-line, heterozygous females were crossed with hemizygous male and gave birth to 25% of heterozygous females, 25% of homozygous females, 25% of hemizygous males and 25% of WT males. For the 44-line, heterozygous females were crossed with WT male and gave birth to 25% of heterozygous females, 25% of WT females, 25% of hemizygous males and 25% of WT males. From this F2 generation, crosses were established to proceed to fertility tests, as explained in the results section.

• Germinal Vesicles oocytes and zygotes immunofluorescence

- GV and zygotes harvest

3 to 4 weeks-old females were intraperitoneally injected with 2,5 IU of PMSG (Stock at 6000 UI Chronogest, PMS009 CENTRAVET). 42h to 48h after, either the every was extracted and GV isolated or 5 IU of hCG (Chorulon 1500, CHO004 CENTRAVET) was injected. In case of hCG injection, the female was placed right away with breeders' males. The next day, females are dissected to get the zygotes. GV and zygotes are kept in EmbryoMax M2 Medium with Phenol Red (MR-015-D SIGMA).

- GV and zygotes immunostaining

GV and zygotes were fixed for 10 min at room temperature with 4% paraformaldehyde and then permeabilized for 10 min at room temperature with 0.25% Triton X-100. Incubations with primary and secondary antibodies were performed in phosphate-buffered saline (PBS), 0.1% Tween-20, and 10% FCS. See **Table 6** for the list of used antibodies. GV and zygotes were incubated overnight at 4°C with primary antibodies and for 2 h or overnight with secondary antibodies. Nuclei were counterstained with Hoechst 33342 (10150888, FISHER

SCIENTIFIC) stain, and embryos were placed in different wells of microfabricated device[342]. Fluorescent images were obtained using a confocal laser scanning microscope (LSM510, Zeiss).

• **Microscopy and image analysis**

Acquisitions were performed using wide-field fluorescence inverted microscope Nikon Ti Eclipse or a confocal laser scanning microscope (LSM510, Zeiss). Light exposure time and LED intensity were identical between the compared conditions. The images were generated by multi-position acquisition to reduce microscopy bias. Then images were analyzed manually with ImageJ/Fiji (version 2.1.0/1.53c) to count single molecules RNA-FISH dots or to quantify a sum of pixel intensities.

• Statistical Analysis

Statistical analysis were performed with Rstudio software (Version 1.1.463 – © 2009-2018 RStudio, Inc.). Replicates distributions were submitted to normality test with Shapiro test. When they were normally distributed, a t-test was applicated, alternatively when the distribution was not normal, Wilcoxon test was used. For **Figure 12E**, a z-test was performed. When the p value was inferior to 0,05, results were considered as significant and indicated with *** if inferior to 0,005, ** if comprised between 0,005 and 0,025, * if between 0,025 and 0,05 and n.s if p value is superior to 0,05.

Results

IV) Oct4 activates a Suv39h1-repressive antisense lncRNA to couple histone H3 Lysine 9 methylation to pluripotency

In this first part of the results, some experiments and analyses were performed by different persons. ChIP-qPCR were performed by Agnès Dubois and RNA-seq analyses and assemblies were done by Noa Gil, Nick Owens, Almira Chervova and Igor Ulitsky.

During development, establishment and maintenance of distinct gene expression patterns supporting identity of each cell type are closely linked to the regulation of chromatin states. The chromatin is a highly dynamic structure composed by: EuC, associated with transcriptionally active regions[152] and HC, associated with gene repression in a cell-type-specific manner, the fHC[136], and at permanently silent regions such as transposable elements and pericentromeric regions, the cHC[139]. These two types of HC have been thought to be distinguishable by distinct molecular signatures, with facultative heterochromatin being characterized by H3K27me3[173] and constitutive heterochromatin by H3K9me3[216].

Methylation of H3K9 has been demonstrated to be important for genome stability[244] and during development[239]. Indeed, mouse KO models of H3K9 HMTs display penetrant phenotypes, particularly during gastrulation when pluripotency is lost and major differentiation events take place[198]. In mESCs, H3K9 methylation is controlled at the level of its abundance: during differentiation the global levels of H3K9me2 and H3K9me3 increase drastically[260]. Indeed, H3K9 methylation has been shown to act as a major epigenetic barrier for reprogramming back to pluripotency [290].

How the global levels of H3K9 methylation are seemingly coupled to the acquisition and loss of pluripotency, and what consequences this coupling has, remain open questions.

We found Suv39h1 to be the only HMT tightly connected to the network of TFs supporting pluripotency, particularly to its main player OCT4. The analysis of the mechanisms of Suv39h1 repression by OCT4 led us to the identification of an antisense long non-coding (lnc) RNA (Suv39h1as) to the Suv39h1 gene. Thanks to CRISPR/Cas9 technology, we suppressed Suv39h1as expression, validating its role in Suv39h1 repression and the consequent impact on global H3K9me2 and H3K9me2 levels. Therefore, we identified a simple genetic network based on the activation of the Suv39h1 antisense by OCT4 only which, in turn, represses Suv39h1 expression thereby coupling H3K9 methylation to pluripotency. This link

highlights the role of pluripotent TFs on chromatin modifications, similarly to Nanog that controls H3K27me3 during early differentiation[341].

Then, we investigated also the molecular mechanism of Suv39h1 regulation by Suv39h1. We found that Suv39h1as has no impact on Suv39h1 mRNA stability but we observed by sm-FISH that it controls the frequency of Suv39h1 transcription. This control of transcription frequency is possibly due to the establishment of a complex local chromatin by Suv39h1as. Indeed, by ChIP-qPCR, we noted that Suv39h1as triggers euchromatin through the gene body but not at the promoter. This kind of regulation reminds the one of Tsix, an antisense lncRNA that regulates Xist[340].

Finally, we also study the functional impact of Suv39h1 upregulation in mESCs. In line with previous studies of Suv39h1 and other H3K9 methyltransferases[225,239,343], we found that H3K9me2 and H3K9me3 increase has very limited impact on mESCs. They are able to self-renew and differentiate properly. Nevertheless, we demonstrated that H3K9 methylation contributes to time the irreversible commitment into differentiation. This is underscoring the role of H3K9 methylation in the formation of an epigenetic barrier providing robustness to differentiation.

In the following chapters, we will first study more precisely Suv39h1 mechanism of regulation by Suv39h1as, questioning the presence of additional mechanisms of regulation. We will also discuss other few functional assays to study the role of Suv39h1as in mESCs biology.

In second, we will investigate the functional impact of Suv39h1as *in vivo* with the generation and the characterization of a mouse line KO for Suv39h1as. Indeed, several studies indicated the importance of Suv39h1 tight regulation during the early development. Moreover, since Suv39h1as is mainly expressed in the late oocyte, we will examine the impact of Suv39h1as expression suppression on GV oocytes and zygotes.

The activation of a *Suv39h1*-repressive antisense lncRNA by OCT4 couples the control of H3K9 methylation to pluripotency

Laure D. Bernard[1,2], Agnès Dubois[1], Victor Heurtier[1,2], Almira Chervova[1], Alexandra Tachtsidi[1,2], Noa Gil[3,4], Nick Owens[1,5], Sandrine Vandormael-Pournin[1], Igor Ulitsky[3], Michel Cohen-Tannoudji[1], and Pablo Navarro[1,*]

[1]Epigenomics, Proliferation, and the Identity of Cells, Department of Developmental and Stem Cell Biology, Institut Pasteur, CNRS UMR3738, 75015 Paris, France
[2]Sorbonne Université, Collège doctoral, F-75005 Paris, France
[3]Department of Biological Regulation, Weizmann Institute of Science, Rehovot, Israel
[4]Current address: Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland
[5]Current address: Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter, UK
[*]**Corresponding author: pnavarro@pasteur.fr**

Histone H3 Lysine 9 (H3K9) methylation, a characteristic mark of heterochromatin, is progressively implemented during development to contribute to cell fate restriction as differentiation proceeds. For instance, in pluripotent mouse Embryonic Stem (ES) cells the global levels of H3K9 methylation are rather low and increase only upon differentiation. Conversely, H3K9 methylation represents an epigenetic barrier for reprogramming somatic cells back to pluripotency. How global H3K9 methylation levels are coupled with the acquisition and loss of pluripotency remains largely unknown. Here, we identify SUV39H1, a major H3K9 di- and tri-methylase, as an indirect target of the pluripotency network of Transcription Factors (TFs). We find that pluripotency TFs, principally OCT4, activate the expression of an uncharacterized antisense long non-coding RNA to *Suv39h1*, which we name *Suv39h1as*. In turn, *Suv39h1as* downregulates *Suv39h1* transcription in cis via a mechanism involving the modulation of the chromatin status of the locus. The targeted deletion of the *Suv39h1as* promoter region triggers increased SUV39H1 expression and H3K9me2 and H3K9me3 levels, leading to accelerated and more efficient commitment into differentiation. We report, therefore, a simple genetic circuitry coupling the global levels of H3K9 methylation to pluripotency in mouse ES cells.

Introduction

During development, the establishment and maintenance of distinct gene expression patterns supporting the identity of each cell type are closely linked to the regulation of chromatin states[1]. Two broad states have been clearly and unambiguously identified: euchromatin, associated with transcriptionally active regions, and heterochromatin, associated with gene repression[2-5]. Two major states of heterochromatin have been traditionally considered. Facultative heterochromatin refers to a repressive chromatin environment displaying high variability across developmental stages, cell types and cell states. Indeed, silent developmental genes are usually embedded in facultative heterochromatin[3,5]. In contrast, ubiquitously silent elements such as retrotransposons and pericentromeric regions are locked by constitutive heterochromatin[4,5]. These two types of heterochromatin have been thought to be distinguishable by distinct molecular signatures, with facultative heterochromatin being characterized by trimethylation of histone H3 lysine 27 (H3K27me3) and constitutive heterochromatin by H3K9me3, among other chromatin features[2-5]. Nevertheless, recent data has challenged these strict definitions[3]. On the one hand, constitutive heterochromatin can under some circumstances be transcribed or decorated by marks previously associated with facultative heterochromatin[6-8]. On the other, while H3K27me3 and H3K9me2 were considered as major repressive mark for developmental genes, an increasing body of evidence points to H3K9me3 as an additional mean to silence developmental regulators as their expression is definitely shut down in particular lineages[9]. Hence, even though the role of H3K9 methylation in genome stability is unquestionable[10], its importance in gene regulatory mechanisms during development appears to be equally important. Indeed, mouse knock-out (KO) models of H3K9 histone methyltransferases (HMTs) display penetrant phenotypes, particularly during gastrulation when pluripotency is lost and major differentiation events take place[11,12]. Conversely, before reaching pluripotency during early mouse embryogenesis the levels of H3K9 methylation are strictly controlled; promoting their increase, for instance by overexpressing the HMT SUV39H1, leads to developmental defects at the compaction stage[13,14].

While extensive research has contributed to our understanding of how the establishment and maintenance of H3K27me3 regulates developmental transitions, how the levels of H3K9 methylation are developmentally regulated is less clear. Yet, a major distinction has been identified, particularly using pluripotent cells such as mouse Embryonic Stem (ES) cells. Indeed, H3K27me3 characterizes developmental genes even before differentiation, when they are embedded in the so-called bivalent chromatin, which is simultaneously enriched for H3K27me3 and for marks of activity[15]. Upon differentiation, H3K27me3 is either consolidated or erased in a cell-type-dependent manner[16]. On the contrary, H3K9 methylation is more largely controlled at the level of its abundance: during differentiation the global levels of H3K9me2 and H3K9me3 increase drastically[17,18]. Conversely, during

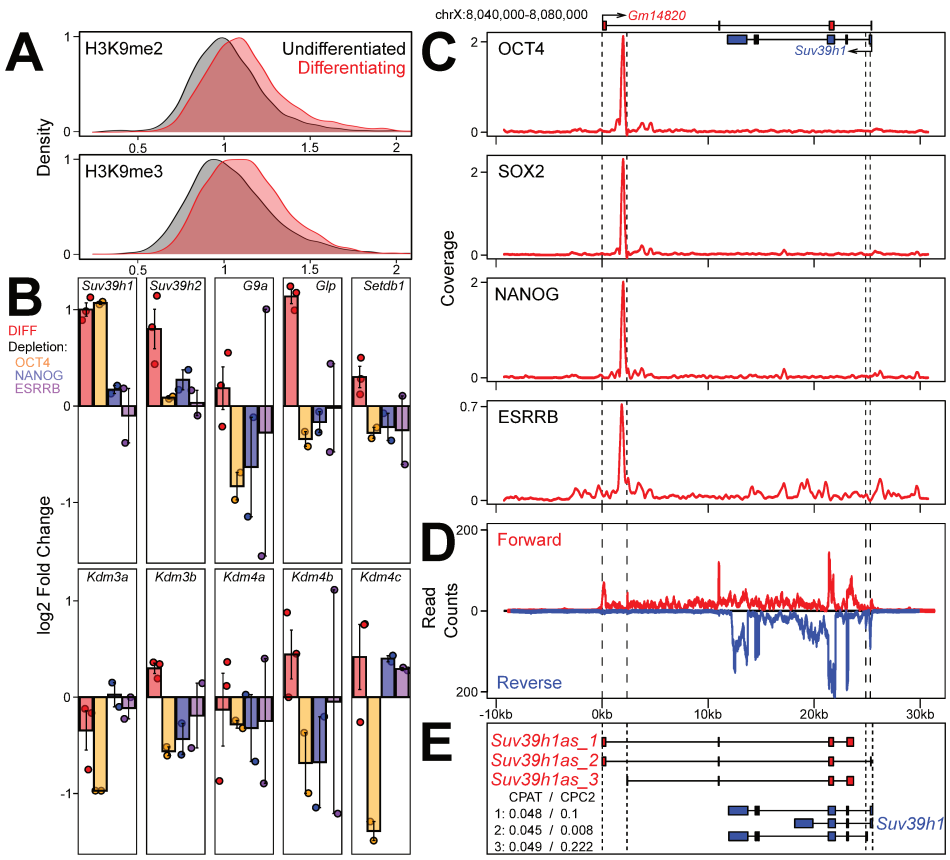


Fig. 1. ***Suv39h1* is downregulated by OCT4, which binds to the promoter of a *Suv39h1* antisense lncRNA. (A)** Distribution of H3K9me2 (top) and H3K9me3 (bottom) in undifferentiated and differentiating ES cell populations assessed by immunofluorescence (Y-axis, density; X-axis, relative mean intensity per cell). Undifferentiated cells (black) are E14Tg2a cells cultured in FCS/LIF (black; n= 4503 cells for H3K9me2 and 3150 for H3K9me3); differentiating cells were obtained by LIF withdrawal (red; n= 6231 cells for H3K9me2 and 3755 cells for H3K9me3). **(B)** Log2 fold change of the indicated gene after differentiating ES cells as in (A) or 24h after inducing the depletion of individual TFs (OCT4, NANOG or ESRRB, as indicated) using Dox-inducible knock-out cells. Each dot represents an independent replicate and the bar and error bars the corresponding means and standard errors. **(C)** Average binding profile of OCT4, SOX2, NANOG and ESRRB (reads per million) across the *Suv39h1*/*Gm14820* locus (mm10, chrX:8,040,000-8,080,000 – 40 kb). *Suv39h1* and *Gm14820* are schematically represented on top. **(D)** RNA-seq profile across the *Suv39h1*/*Gm14820* locus, with forward and reverse fragment counts expressed with positive and negative values. **(E)** Schematic representation of *Gm14820*/*Suv39h1as* (red) and *Suv39h1* (blue) isoforms. The coding probabilities calculated with CPAT and CPC2 algorithms are shown for the three isoforms of *Suv39h1as*. The vertical dashed lines in (C), (D) and (E) mark the position of *Suv39h1as* or *Suv39h1* promoters.

the induction of pluripotency *in vitro* through reprogramming processes, H3K9 methylation has been shown to act as a major epigenetic barrier that is in part overcome by globally reducing its levels[18,19]. Therefore, while H3K27 methylation is mainly controlled by altering its genomic distribution, the global levels of H3K9 methylation display correlated changes to the differentiation status. Beyond the role of H3K9 methylation to stabilise somatic cell identities[20], how its global levels are seemingly coupled to the acquisition and loss of pluripotency, and what consequences this coupling has, remain open questions.

In this study, we aimed at understanding the molecular basis of the link between H3K9 methylation and pluripotency. We find *Suv39h1* to be the only HMT tightly connected to the network of transcription factors (TFs) supporting pluripotency, particularly to its main player *Oct4*. The analysis of the mechanisms of *Suv39h1* repression by OCT4 led us to the identification of a repressive antisense long noncoding RNA (lncRNA[20]) to the *Suv39h1* gene. Our work identifies a simple genetic network based on the activation of the *Suv39h1* antisense by OCT4 which, in turn, represses *Suv39h1* transcription thereby coupling H3K9 methylation to pluripotency. Using CRISPR-Cas9 mediated deletion of the antisense promoters, we show that its activity controls the timing and efficiency of ES cell commitment into differentiation. Our results therefore provide a mechanistic perspective into how the global levels of H3K9 methylation are regulated at the onset of differentiation to irreversibly exit pluripotency.

Results

Suv39h1 expression is under the control of OCT4 in ES cells.

Using immunofluorescence, we first confirmed that differentiation of ES cells by LIF withdrawal leads to an increase of both H3K9me2 and H3K9me3 (Fig.1A; $p<10^{-15}$), as expected. We hypothesized that one or several histone methyltransferases or lysine demethylases (HMTs and KDMs, respectively)[2,4] could be differentially expressed upon differentiation, linking the loss of pluripotency to increased H3K9 methylation. To assess this, we monitored mRNA levels of HMTs and KDMs using published RNA-seq datasets[22] (Fig.1B; Table S1). We found three HMTs to be upregulated upon differentiation: *Suv39h1* (p=0.004), *Suv39h2* (p=0.044) and *Glp* (p=0.004). In contrast, all tested KDM displayed minor changes below 2-fold. We reasoned that the increase of *Suv39h1*, *Suv39h2* and *Glp* expression could either be due to a direct control of their transcription by pluripotency TFs or to alternative mechanisms. To address this, we assessed the impact of the loss of individual pluripotency TFs (*Oct4*, *Nanog* and *Esrrb*) using available dox-inducible knock-outs[23-25] (Fig.S1A). Only one, *Suv39h1*, was found upregulated 24h after inducing the loss of pluripotency TFs, particularly of OCT4, which depletion leads to a 2-fold increase in *Suv39h1* mRNA levels (Fig.1B; Table S1). Hence, after confirming *Suv39h1* expression changes by RT-qPCR (Fig.S1B), we hypothesized that OCT4 may act as a repressor of *Suv39h1* expression to maintain low levels of H3K9 methylation until the onset of differentiation. Exploration of

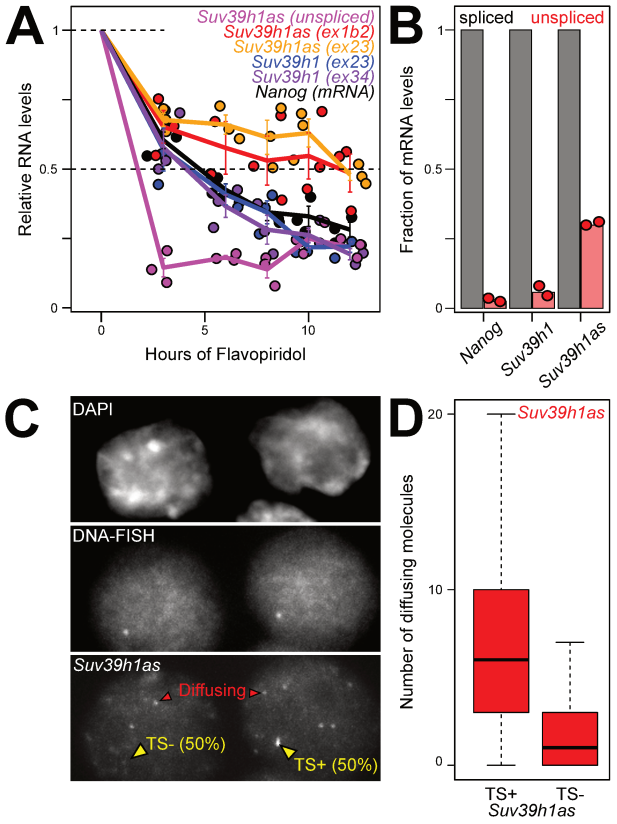


Fig. 2. ***Suv39h1as*** **is a nuclear, stable and lowly expressed lncRNA. (A)** RT-qPCR analysis of the half-life of several RNA species during a transcription inhibition assay with Flavopiridol (X-axis, hours of treatment): *Suv39h1* mRNA, using two trans-exonic primer pairs between exons 2 and 3 (blue – ex23) or exons 3 and 4 (purple – ex34); *Suv39h1as*, using two trans-exonic primer pairs between exons 1b and 2 (red – ex1b2) or exons 2 and 3 (orange – ex23) or primer pairs amplifying the unspliced RNA (magenta – pre-mRNA); *Nanog* mRNA (black). Each dot represents an independent replicate and the line the corresponding mean and standard error. Ribosomal RNA (*28S*) was used for normalization. **(B)** Histogram representing unspliced RNA levels (red) relative to corresponding spliced RNAs (black) for *Nanog*, *Suv39h1* and *Suv39h1as*, as measured by RNA-seq. Each dot represents an independent replicate and the bar the corresponding mean. **(C)** Representative sm-FISH followed by DNA-FISH visualizing *Suv39h1as* RNA molecules and the *Suv39h1*/*Suv39h1as* locus, respectively, in undifferentiated E14Tg2a cells. Red arrowheads indicate RNAs diffusing away from the locus, which is indicated by a yellow arrow. The proportion of actively transcribing cells is indicated (n=358 cells). **(D)** Boxplots (median; 25-75% percentiles; error bars) showing the number of *Suv39h1as* diffusible molecules counted in cells presenting an active (TS+) or inactive (TS-) *Suv39h1as* gene (n=358).

available ChIP-seq datasets[26] (Fig.1C) and direct validation by ChIP-qPCR (Fig.S1C) identified a hotspot of pluripotency TFs, including OCT4, in the vicinity of *Suv39h1*. However, this TF binding hotspot was found located 3' to *Suv39h1*, at around 27 kb of its promoter region. Notably, we noticed that this region coincides with the promoter region of an uncharacterized gene, *Gm14820* (*AK010638*), antisense to and largely overlapping *Suv39h1* (top of Fig.1C).

Identification of a *Suv39h1* antisense lncRNA.

Stranded RNA-seq confirmed *Gm14820* to be expressed in ES cells, at levels comparable to *Suv39h1* (Fig.1D). Using *de novo* transcript assembly with our RNA-seq datasets, together with direct cDNA cloning, sequencing and RT-qPCR, we identified several isoforms expressed in ES cells (Fig.1E). All isoforms initiate from two distinct promoters, located in proximity to the region bound by pluripotency TFs, exhibit overlapping exons with *Suv39h1* and terminate within *Suv39h1* or in the vicinity of its 5' end. Notably, *Gm14820* is annotated as a lncRNA[21]. Accordingly, using two different algorithms (CPAT[27] and CPC2[28]), the nearly absent coding potential of all *Gm14820* isoforms was confirmed (Fig.1E). Thereafter, we refer to *Gm14820* as *Suv39h1as*. To further characterize *Suv39h1as*, we assessed the stability of its RNA products and found the half-life of its spliced and unspliced forms to be around 12h and 1h30, respectively (Fig.2A). However, *Suv39h1as* splicing is relatively inefficient compared to *Suv39h1* or another protein coding gene, *Nanog* (Fig.2B), as is generally the case for lncRNAs[21]. Moreover, *Suv39h1as* was efficiently captured in poly-A selected RNA-seq, suggesting it is normally poly-adenylated (Table S1). Next, we aimed at visualizing *Suv39h1as* RNA molecules in single cells. For this, we designed oligonucleotides targeting *Suv39h1as* exons and performed strand-specific single molecule RNA-FISH (smFISH) coupled to DNA-FISH to identify the *Suv39h1as*/*Suv39h1* locus, using a fosmid covering the whole region (Fig.2C). We observed that *Suv39h1as* is mainly detected as a bright point in the nucleus, likely representing actively transcribed loci as it coincides with the DNA-FISH signal. A small number of single *Suv39h1as* RNA molecules could also be detected diffusing in the nucleus and, more rarely, in the cytoplasm. Quantification of the smFISH/DNA-FISH suggested a transcriptional frequency of around 50% in the population, with a median of 6 freely diffusing RNAs in cells presenting a transcriptionally active locus (Fig.2D). Hence, we conclude that the pluripotency TFs bind close to the two promoters of a *Suv39h1* antisense lncRNA, which is mostly localised at its site of transcription, poly-adenylated and poorly spliced even though the spliced isoforms are relatively stable.

Anticorrelated expression patterns of *Suv39h1* and *Suv39h1as*.

To investigate *Suv39h1as* and *Suv39h1* expression patterns we differentiated ES cells using three independent protocols based on LIF withdrawal, N2B27 or EpiLC-directed differentiation. These three assays showed a strong reduction of *Suv39h1as* expression after 3 days of differentiation, when *Suv39h1* expression increases (Fig.3A; $p<0.05$ (*Suv39h1*) and $p<0.01$ (*Suv39h1as*) for all differentiation assays). Moreover, the depletion of OCT4 showed a rapid downregulation of *Suv39h1as* ($p<10^{-3}$ at all time-points), reaching minimal levels of expression within 12h and accompanied by a marked increase of *Suv39h1* expression ($p<0.05$ at all time-points) that reached maximal levels after 24h (Fig.3B). Thus, *Suv39h1* and *Suv39h1as* display anticorrelated expression levels. To explore this observation at the single cell level, we designed oligonucleotides across *Suv39h1* exons and introns to monitor *Suv39h1*/*Suv39h1as* expression by smFISH in parallel to DNA-FISH (Fig.3C,D). We found around 20% of cells actively transcribing both sense/antisense genes and around 30% transcribing either one or the other. Moreover, cells actively transcribing *Suv39h1as* displayed significantly ($p<10^{-11}$) fewer *Suv39h1* mRNA molecules (Fig.3D). In

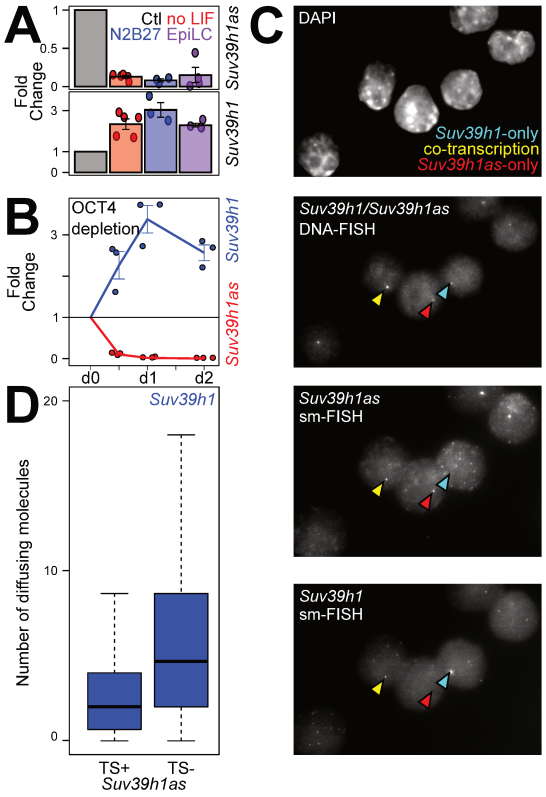


Fig. 3. Anticorrelated expression of *Suv39h1* and *Suv39h1as*. **(A)** Fold change expression of *Suv39h1as* (top) or *Suv39h1* (bottom) measured by RT-qPCR in differentiating E14Tg2a cells versus undifferentiated controls (ctl; gray). Differentiation was triggered for three days with three independent protocols: LIF withdrawal from FCS/LIF cultures (no LIF; red), 2i and LIF withdrawal from 2i+LIF cultures (N2B27; blue) or EpiLC differentiation from 2i+LIF cultures (EpiLC; purple). Values were normalized to *Tbp* and fold changes calculated to their respective control culture. Each dot represents an independent replicate and the bar and error bars the corresponding means and standard errors. **(B)** RT-qPCR analysis of *Suv39h1as* (red) and *Suv39h1* (blue) expression upon OCT4 depletion for the indicated time (X-axis). Values were normalized to *Tbp*. Each dot represents an independent replicate and the line the corresponding mean with standard errors. **(C)** Representative sm-FISH of *Suv39h1as* and *Suv39h1* RNA molecules, followed by DNA-FISH visualising the *Suv39h1/Suv39h1as* locus in E14Tg2a cells (n= 358). Selected loci transcribing either *Suv39h1*, *Suv39h1as* or both genes are indicated with arrow heads: blue, *Suv39h1*-only (30%); red, *Suv39h1as*-only (30%); yellow for cells transcribing both (20%). **(D)** Boxplots (median; 25-75% percentiles; error bars) showing the number of *Suv39h1* diffusible molecules counted in cells presenting an active (TS+) or inactive (TS-) *Suv39h1as* gene, counted in 358 E14Tg2a cells.

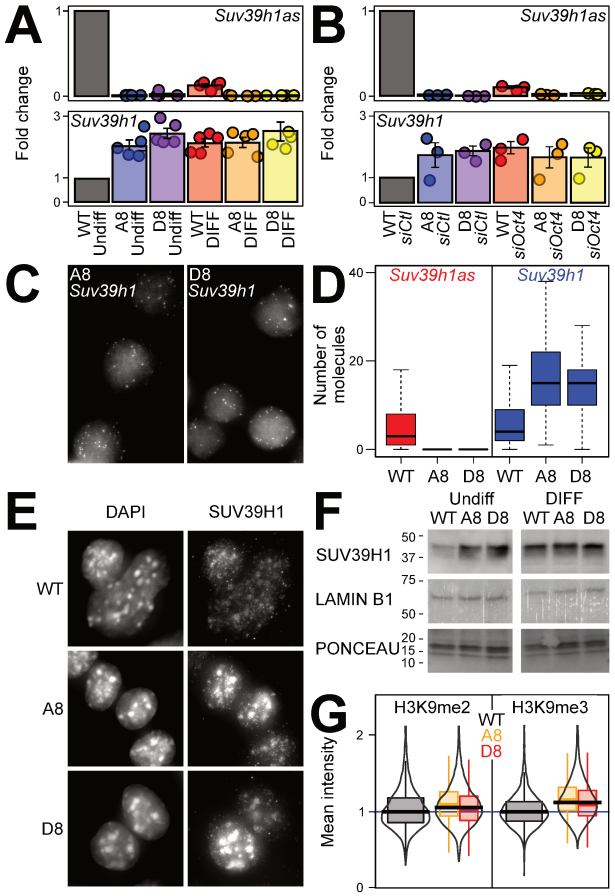


Fig. 4. OCT4 represses *Suv39h1* via *Suv39h1as* to reduce H3K9 methylation in ES cells. **(A)** Expression fold change of *Suv39h1as* (top) and *Suv39h1* (bottom) in WT (E14Tg2a) and *Suv39h1as*-mutant cells (A8 and D8) cultured in undifferentiated (FCS/LIF) or differentiating conditions (3 days without LIF). Values were normalized to *Tbp*. Each dot represents an independent replicate and the bar and error bars the corresponding means and standard errors. **(B)** Expression fold change of *Suv39h1as* (top) and *Suv39h1* (bottom) in wild-type (WT – E14Tg2a) and *Suv39h1as*-mutant cells (A8 and D8) knocked-down with either control or *Oct4*-targetted siRNAs. Values were normalized to *Tbp*. Each dot represents an independent replicate and the bar and error bars the corresponding means and standard errors. **(C)** Representative sm-FISH images of *Suv39h1* in A8 and D8 cells (wild-type cells presented in Fig.3C). **(D)** Boxplots (median; 25-75% percentiles; error bars) showing the number of *Suv39h1as* (red) or *Suv39h1* (blue) diffusible molecules counted in wild-type (WT; n= 358) or mutant cells (A8, n= 289; D8, n= 270). **(E)** Representative SUV39H1 immunofluorescence of wild-type (E14Tg2a) and *Suv39h1*-mutant ES cells (A8 and D8). **(F)** Representative Western-Blot of SUV39H1, LAMIN B1 and corresponding Ponceau for wild-type (E14Tg2) and mutant cells (A8 and D8) in undifferentiated (FCS/L) and differentiating (3 days without LIF) conditions. On the left of each image is indicated protein scale in kDa. **(G)** Violin and boxplots (median; 25-75% percentiles; error bars) presenting relative H3K9me2 (left) and H3K9me3 (right) mean intensity values of E14Tg2a (WT – black; n=12881 for H3K9me2 and n=12053 for H3K9me3) or mutant cells (A8, yellow, n=3553 cells for H3K9me2 and 2081 for H3K9me3; D8, red, n=5050 cells for H3K9me2 and 2641 for H3K9me3), assessed by immunofluorescence.

summary, both smFISH and RT-qPCR experiments showed that *Suv39h1* and *Suv39h1as* display anticorrelated expression patterns. This anticorrelation stems from single cell dynamics where the transcription of the antisense leads to a slight reduction of the transcriptional frequency of *Suv39h1* and, more significantly, to reduced expression of diffusing mRNAs. These observations indicate that *Suv39h1as* acts as a repressor of *Suv39h1*. Since *Suv39h1as* is downregulated upon differentiation and upon the loss of OCT4, our data suggests that pluripotent TFs activate *Suv39h1as* transcription which, in turn, downregulates *Suv39h1* expression.

A simple OCT4-*Suv39h1as*-*Suv39h1* circuitry couples H3K9 methylation to pluripotency.

To functionally establish the relationships between OCT4, *Suv39h1as*, *Suv39h1* and H3K9 methylation, we designed two gRNAs to delete 5,5kb encompassing the two promoters of *Suv39h1as*. Two independent KO clones, A8 and D8, were generated. PCR genotyping confirmed the deletion and analysis of their karyotypes did not reveal any overt defect. Moreover, RT-qPCR showed a complete loss of *Suv39h1as* expression (Fig.4A). We then addressed the impact of *Suv39h1as* depletion on *Suv39h1* expression, both before and during differentiation. In the two mutant clones we observed an increase of *Suv39h1* expression in undifferentiated cells (p=0.011 and 0.0016 for A8 and D8, respectively), reaching the levels observed upon differentiation in WT cells

(Fig.4A). In differentiating cells, when *Suv39h1as* is naturally silenced, the deletion had no impact (Fig.4A), as expected, regardless of the differentiation protocol (Fig.S1D). Therefore, these results indicate that *Suv39h1as* acts as a pluripotency-associated repressor of *Suv39h1* expression. However, this does not imply that OCT4 exclusively represses *Suv39h1* expression via *Suv39h1as*. To address this, we used siRNAs targeting *Oct4* to test whether in the absence of *Suv39h1as*, the loss of OCT4 would lead to any modification of *Suv39h1* expression. Whereas in wild-type cells the knock-down of *Oct4* (above 80% efficiency, Fig.S1E), led to higher (p=0.027) *Suv39h1* expression (Fig.4B), as expected, in mutant cells it was fully inconsequent (Fig.4B). This indicates that the OCT4-dependent repression of *Suv39h1* is exclusively mediated by *Suv39h1as*. Next, we performed sm-FISH to study *Suv39h1* upregulation with single cell resolution (Fig.4C). We observed a marked increase in the transcriptional frequency of *Suv39h1*, rising from 47.6% in WT to 76.3% and 74.1% in A8 and D8, respectively ($p<10^{-10}$ for both clones). Moreover, the number of *Suv39h1* mRNAs per cell also increased substantially ($p<10^{-15}$), with virtually no cell displaying an absence of *Suv39h1* mRNAs (Fig.4D). This increase in *Suv39h1* expression was accompanied by higher levels of SUV39H1 protein expression and accumulation at heterochromatic regions such as chromocenters, as established by immuno-staining and western-blot (Fig.4E,F; Fig.S1F). Consequently, in both mutant clones we observed higher levels of H3K9me2 and H3K9me3 ($p<10^{-8}$ for both marks and cell clones), establishing a direct link between *Suv39h1as* and the global levels of H3K9 methylation in ES cells (Fig.4G).

Suv39h1as modifies the chromatin of the *Suv39h1as/Suv39h1* locus.

We have observed that in mutant ES cells lacking *Suv39h1as* expression, the transcriptional frequency of *Suv39h1* increases from 50 to 75%. Moreover, the absence of *Suv39h1as* is not accompanied by increased stability of *Suv39h1* mRNAs (Fig.S1G). Therefore, *Suv39h1as* is likely to act as a transcriptional repressor of *Suv39h1*. To explore this, and given that other antisense transcription units have been shown to modify the chromatin of their corresponding sense gene[29-31], we used a ChIP approach to establish the histone modification profiles of the locus (Fig.5). First, we monitored H3K4 methylation profiles. We found H3K4me1 and me2, which usually mark transcriptionally competent regions[32], to globally decorate the locus with minimal focal accumulation at the promoters. Conversely, H3K4me3, a mark of activity[32], was focally enriched at the *Suv39h1* promoter and displayed low levels over the antisense promoter. We then profiled the active histone acetylation marks, H3K9ac and H3K27ac. Similarly to H3K4me3, we found H3K9ac to preferentially mark the *Suv39h1* promoter. In contrast, both sense and antisense gene promoters where enriched for H3K27ac. In mutant cells we observed a global decrease of H3K4me1/me2 over the region transcribed by *Suv39h1as* ($p<10^{-3}$), particularly before it reaches the *Suv39h1* gene body (Fig.5), indicating its transcription promotes the establishment of these marks. The lack of H3K4me1/me2 reduction within the region transcribed by both genes suggests that the increased transcription of *Suv39h1* may have a compensatory role. Moreover, H3K4me2, H3K9ac and H3K27ac, all marks of gene activity, showed a slight but statistically significant increase ($p<0.05$) at *Suv39h1* promoter in the absence of *Suv39h1as* (Fig.5). Altogether, this analysis suggests that the loss of *Suv39h1as* leads to increased *Suv39h1* transcription at least in part mediated by increased euchromatinisation of the Suvar39h1 promoter. These results suggest a simple mechanism where *Suv39h1as* modulates the local chromatin environment of *Suv39h1* to decrease the probability of its transcription, leading to a global lower level of *Suv39h1* mRNA and, subsequently, protein.

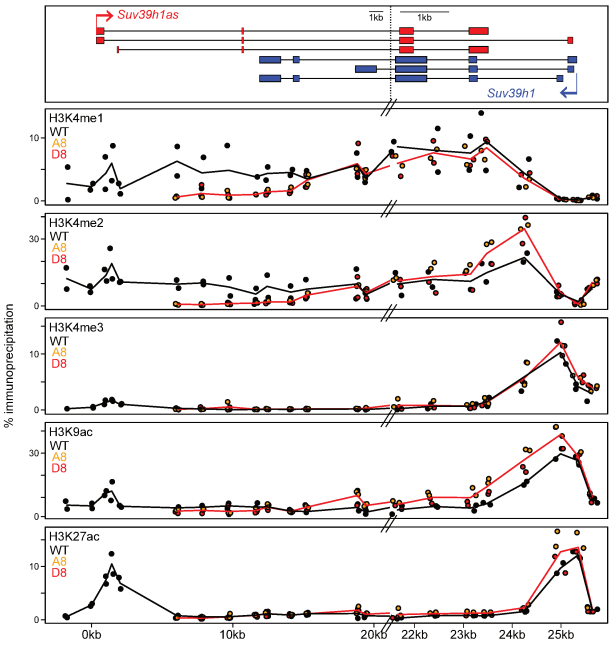


Fig. 5. *Suv39h1as* triggers complex chromatin changes across the locus. Chromatin immunoprecipitation profile (Y-axis, percentage of immunoprecipitation) of H3K4me1, H3K4me2, H3K4me3, H3K9ac and H3K27ac, as indicated, across *Suv39h1*/*Suv39h1as* locus in wild-type (E14Tg2a, black) and *Suv39h1as*-mutant cells (A8, yellow dots; D8, red dots; the red line represents the average of all data points for mutant clones). The X-axis represents genomic distances in kb with respect to the *Suv39h1as* transcription start site, as schematized on top. Note a break on the scale of the genomic coordinates at around X=21kb.

The lack of *Suv39h1as* leads to accelerated differentiation commitment.

We finally wondered whether the increase of H3K9 methylation taking place in *Suv39h1as* mutant cells had any physiological impact. First, we used clonal assays to assess self-renewal and differentiation efficiency (Fig.6A,B). Either in conditions of reinforced self-renewal (2i/LIF), in traditional serum-containing culture medium (FCS/LIF) or in the absence of LIF (FCS), the number of alkaline-phosphatase colonies, a marker of pluripotent cells, was identical between wild-type and mutant clones. Hence, the presence of increased H3K9 methylation is largely inconsequential for self-renewal and for the loss of pluripotency. In agreement, both wild-type and mutant cells proliferate and differenti-

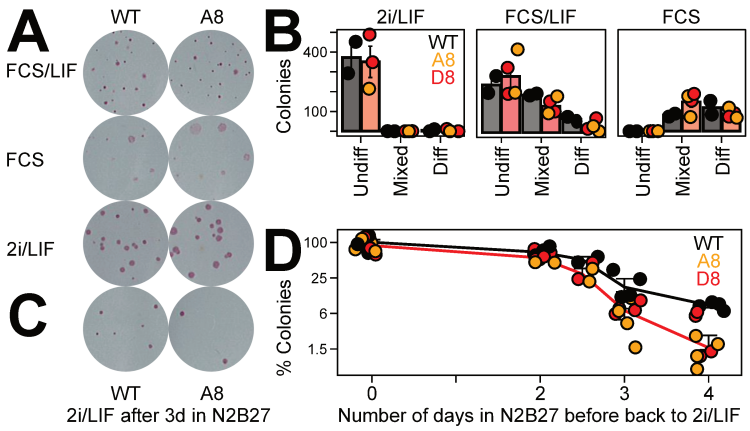


Fig. 6. Accelerated commitment into differentiation in the absence of *Suv39h1as*. **(A)** Representative alkaline-phosphatase staining of ES cell colonies cultured as indicated. **(B)** Number of wild-type (E14Tg2a, black) or *Suv39h1as*-mutant (A8, orange points; D8, red points) colonies characterized as undifferentiated, mixed or differentiated after culturing them as indicated. Each dot represents an independent replicate and the bar and error bars the corresponding means (black for wild-type and red for the mean of all data points for mutant clones) and standard errors. **(C)** Alkaline-phosphatase staining of ES cell colonies cultured in 2i/LIF after 3 days in N2B27 for wild-type (E14Tg2a) and *Suv39h1as*-mutant cells (A8). **(D)** Percentage of alkaline-phosphatase positive colonies cultured in 2i/LIF after differentiating them for the indicated number of days (X-axis) in N2B27. D0, undifferentiated cells were set as 100%. Each dot represents an independent replicate (wild-type, black; *Suv39h1as*-mutant clones in orange, A8, and red, D8) and the line the corresponding mean and standard error (all mutant datapoints were averaged to obtain the red line).

ate normally, as evaluated morphologically (Fig.S1H) and by marker expression (Fig.S1I). However, during differentiation, the role of H3K9 methylation is to restrict cell fate and developmental competence[18,33], more than to elicit differentiation. Therefore, we reasoned that the loss of *Suv39h1as* could modulate the timing of commitment into differentiation. To test this, we used an established assay[34] whereby wild-type and mutant clones were differentiated in parallel and, every day, the cells were harvested and reseeded clonally in 2i/LIF: only those cells that were not yet committed into irreversible differentiation can self-renew and form undifferentiated colonies (Fig.6C,D). As previously shown, we observed that commitment took place between days 2 and 3 in wild-type cells, with a reduction in clonogenicity of nearly 90% (Fig.6D). In mutant cells, however, the reduction in the number of undifferentiated colonies was more marked from day 2.5 onwards (Fig.6D; p=0.0666, 0.0020 and 0.0041 for days 2.5, 3 and 4, respectively). Therefore, the premature establishment of higher levels of H3K9me2/me3 in ES cells facilitates the irreversible commitment into differentiation, in line with the role of these repressive marks in locking cell fate changes[20].

Discussion

In this study, we have identified a genetic network linking the control of the global levels of H3K9 methylation to pluripotency. The pluripotency network, mainly through OCT4, activates an antisense lncRNA to the *Suv39h1* gene (*Suv39h1as*), repressing *Suv39h1* expression. Consequently, the level of H3K9 methylation is reduced. Upon differentiation, the collapse of the pluripotency network leads to the silencing on *Suv39h1as*, enabling increased SUV39H1 expression and H3K9 methylation, which controls the timing and efficiency of the irreversible commitment into differentiation. This genetic setup may act as a time-delay generator, enabling ES cells to filter out transient fluctuations in the activity of the pluripotency network[35]: only a long decrease in activity of the pluripotency network may be sufficient to elicit the increase in SUV39H1 expression that will follow the extinction of *Suv39h1as*.

Antisense lncRNAs are frequent in mammals, with 29% of canonical protein coding genes displaying antisense transcription[36]. Given their antisense orientation and the resulting complementarity, antisense lncRNAs can theoretically regulate their cis-linked sense gene through a wide variety of mechanisms. By deleting the *Suv39h1as* promoter region, we found that *Suv39h1as* controls the frequency of *Suv39h1* transcription and has no impact on *Suv39h1* mRNA stability. However, the exact mechanisms by which *Suv39h1as* represses *Suv39h1* transcription remain open, even though we show it has complex influences on the chromatin. On the one hand, *Suv39h1as* triggers H3K4me1 and me2 throughout the locus; on the other, it slightly reduces H3K4me2, H3K4me3, H3K9ac and H3K27ac at the *Suv39h1* promoter. Notably, these effects are reminiscent of those associated with the regulation of *Xist* by its antisense *Tsix*[29-31], possibly revealing a general property of antisense transcription. While it is possible that the slight increase of euchromatic marks at the *Suv39h1* promoter are sufficient to increase the frequency of *Suv39h1* transcription, other potential mechanisms cannot be excluded. For instance, the antisense/sense polymerases could enter into physical collision and transcriptional abortion[37], modulating *Suv39h1* elongation. Moreover, *Suv39h1as* presents two exons nearly fully overlapping *Suv39h1* exons, a gene structure that is conserved at the human *Suv39h1as*/*Suv39h1* locus (Fig.S1J). Therefore, while the transcriptional effects are clear, *Suv39h1as* may deploy additional post-transcriptional mechanisms to downregulate *Suv39h1* mRNA expression, such as masking splice sites as it has been reported for *ErbAa*[38]. Finally, other mechanisms fully independent of *Suv39h1as* transcription or of the produced lncRNA cannot be excluded, such as direct competition between the sense and antisense promoters[39].

The deletion of *Suv39h1as* promoter and the ensuing increase in H3K9 methylation, appears to be largely inconsequent for ES cells: they self-renew and differentiate efficiently. This observation is in line with others, where histone modifiers have been either invalidated or ectopically expressed in ES cells with minor consequences for self-renewal[40,41]. However, despite the fact that *Suv39h1as* mutant cells self-renew and differentiate normally, we asked whether the timing of commitment into differentiation is altered. Our results showed a faster and more efficient com-

mitment into differentiation, suggesting that the global levels of H3K9 methylation contribute to irreversibly lock the loss of pluripotency. This observation adds to the notion of H3K9 methylation acting as an epigenetic barrier providing robustness to cell fate changes[33]. Moreover, our results also underscore the dominance of pluripotency TFs over chromatin modifications[41]. We had already shown that NANOG, another key pluripotency TF, controls H3K27me3 levels, particularly during early differentiation[22]. Here, we complement this notion with OCT4 controlling H3K9me3 via the *Suv39h1as*/*Suv39h1* tandem. Together, these results place the control of global levels of heterochromatin marks under the activity of the pluripotency network, extending the concept of the genetic dominance of pluripotency. Whether our observations and conclusions can be extended to early mouse embryogenesis and to the acquisition and loss of pluripotency *in vivo* is now a question of primary importance. Notably, H3K9 methylation levels are exquisitely regulated during early embryogenesis[42]. It is noteworthy that SUV39H1 is absent in oocytes and its expression starts at the 2-4 cell transition stage[43], when the reconfiguration of constitutive heterochromatin as chromocenters is initiated. Moreover, the overexpression of SUV39H1 during the early stages of embryogenesis leads to developmental defects. Indeed, using different strategies, Zhang et al.[13] and Burton et al.[14] overexpressed SUV39H1 in the zygote and observed an increase of H3K9me3 levels leading to early developmental arrest at the time of compaction. Therefore, an important hypothesis emerges from this work: *Suv39h1as* could be a key regulator of *Suv39h1* during early embryogenesis, holding its expression until the appropriate time to enable the timely establishing of the first heterochromatic structures in the embryo.

Supplementary information One supplementary figure accompanies this manuscript, it can be found at the end of this document. Two Supplementary Tables and Methods are available online.

Acknowledgements. This study was conceived by P.N with inputs from L.B. and V.H.. Experiments were designed and executed by L.B., with help from A.D., V.H., A.T., S.V-P. and M.C-T. RNA-seq analyses and assembly were done by N.G., N.O., A.C. and I.U. The paper was written by L.B. and P.N. L.B. acknowledges the Ecole Normale Supérieure and Sorbonne Université for funding. P.N. and M.C-T. acknowledge the Labex Revive (Investissement d'Avenir; ANR-10-LABX-73), the Institut Pasteur and the CNRS for funding.

Declaration of interests. The authors declare no competing interests.

References

1. Li E. Chromatin modifications and epigenetic reprogramming in mammalian development. Nat Rev Genet 3, 662-673 (2002).

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

3. Trojer, P. Reinberg, D. Facultative Heterochromatin: Is There a Distinctive Molecular Signature? Molecular Cell 28, 1–13 (2007).

4. Saksouk, N., Simboeck, E. Déjardin, J. Constitutive heterochromatin formation and transcription in mammals. Epigenetics Chromatin 8, 3 (2015).

5. Liu, J., Ali, M. Zhou, Q. Establishment and evolution of heterochromatin. Annals of the New York Academy of Sciences 1476, 59–77 (2020).

6. Tosolini, M. et al. Contrasting epigenetic states of heterochromatin in the different types of mouse pluripotent stem cells. Scientific Reports 8, 5776 (2018).

7. Walter, M., Teissandier, A., Pérez-Palacios, R. Bourc'his, D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. Elife 5, (2016).

8. Johnson, W. L. et al. RNA-dependent stabilization of SUV39H1 at constitutive heterochromatin. eLife 6, e25299 (2017).

9. Nicetto, D. and Zaret, K. Role of H3K9me3 heterochromatin in cell identity establishment and maintenance. Curr Opin Genet Dev, 55, 1-10 (2019).

10. Janssen, A., Colmenares, S. U. Karpen, G. H. Heterochromatin: Guardian of the Genome. Annu. Rev. Cell Dev. Biol. 34, 265–288 (2018).

11. Dodge, J. E., Kang, Y.-K., Beppu, H., Lei, H. Li, E. Histone H3-K9 Methyltransferase ESET Is Essential for Early Development. MCB 24, 2478–2486 (2004).

12. Tachibana, M. et al. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes Dev 16, 1779–1791 (2002).

13. Zhang, Y.-L. et al. DCAF13 promotes pluripotency by negatively regulating SUV39H1 stability during early embryonic development. EMBO J. 37, (2018).

14. Burton, A. et al. Heterochromatin establishment during early mammalian development is regulated by pericentromeric RNA and characterized by non-repressive H3K9me3. Nat Cell Biol 22, 767–778 (2020).

15. Bernstein, B. E. et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125, 315–326 (2006).

16. Mikkelsen, T. S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553–560 (2007).

17. Wen, B., Wu, H., Shinkai, Y., Irizarry, R. A. Feinberg, A. P. Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. Nat Genet 41, 246–250 (2009).

18. Sridharan, R. et al. Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1γ in reprogramming to pluripotency. Nat Cell Biol 15, 872–882 (2013).

19. Wei, J. et al. KDM4B-mediated reduction of H3K9me3 and H3K36me3 levels improves somatic cell reprogramming into pluripotency. Sci Rep 7, 7514 (2017).

20. Becker, J. S., Nicetto, D. Zaret, K. S. H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. Trends in Genetics 32, 29–41 (2016).

21. Staněk, D. Long non-coding RNAs and splicing.

Essays Biochem EBC20200087 (2021).

22. Heurtier, V. et al. The molecular logic of Nanog-induced self-renewal in mouse embryonic stem cells. Nat Commun 10, 1109 (2019).

23. Festuccia, N. et al. Esrrb Is a Direct Nanog Target Gene that Can Substitute for Nanog Function in Pluripotent Cells. Cell Stem Cell 11, 477–490 (2012).

24. Niwa, H., Miyazaki, J. Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat. Genet. 24, 372–376 (2000).

25. Navarro, P. et al. OCT4/SOX2-independent Nanog autorepression modulates heterogeneous Nanog gene expression in mouse ES cells. EMBO Journal 31, 4547–4562 (2012).

26. Festuccia, N. et al. Transcription factor activity and nucleosome organization in mitosis. Genome Research 29, 250-260 (2019).

27. Wang, L. et al. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res 41, e74 (2013).

28. Kang, Y.-J. et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic Acids Res 45, W12–W16 (2017).

29. Ohhata, T. et al. Dynamics of transcription-mediated conversion from euchromatin to facultative heterochromatin at the Xist promoter by Tsix. Cell Reports 34, 108912 (2021).

30. Navarro, P., Pichard, S., Ciaudo, C., Avner, P. Rougeulle, C. Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: implications for X-chromosome inactivation. Genes Dev. 19, 1474–1484 (2005).

31. Navarro, P., Page, D. R., Avner, P. Rougeulle, C. Tsix-mediated epigenetic switch of a CTCF-flanked region of the Xist promoter determines the Xist transcription program. Genes Development 20, 2787–2792 (2006).

32. Santos-Rosa, H. et al. Active genes are tri-methylated at K4 of histone H3. Nature 419, 407–411 (2002).

33. Wang, C. et al. Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. Nat Cell Biol 20, 620–631 (2018).

34. Kalkan, T. et al. Tracking the embryonic stem cell transition from ground state pluripotency. Development 144, 1221–1234 (2017).

35. Torres-Padilla, M.-E. Chambers, I. Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage. Development 141, 2173–2181 (2014).

36. Katayama, S. et al. Antisense transcription in the mammalian transcriptome. Science 309, 1564–1566 (2005).

37. Prescott, E. M. Proudfoot, N. J. Transcriptional collision between convergent genes in budding yeast. Proc Natl Acad Sci U S A 99, 8796–8801 (2002).

38. Hastings, M. L., Milcarek, C., Martincic, K., Peterson, M. L. Munroe, S. H. Expression of the thyroid hormone receptor gene, erbAalpha, in B lymphocytes: alternative mRNA processing is independent of differentiation but correlates with antisense RNA levels. Nucleic Acids Res 25, 4296–4300 (1997).

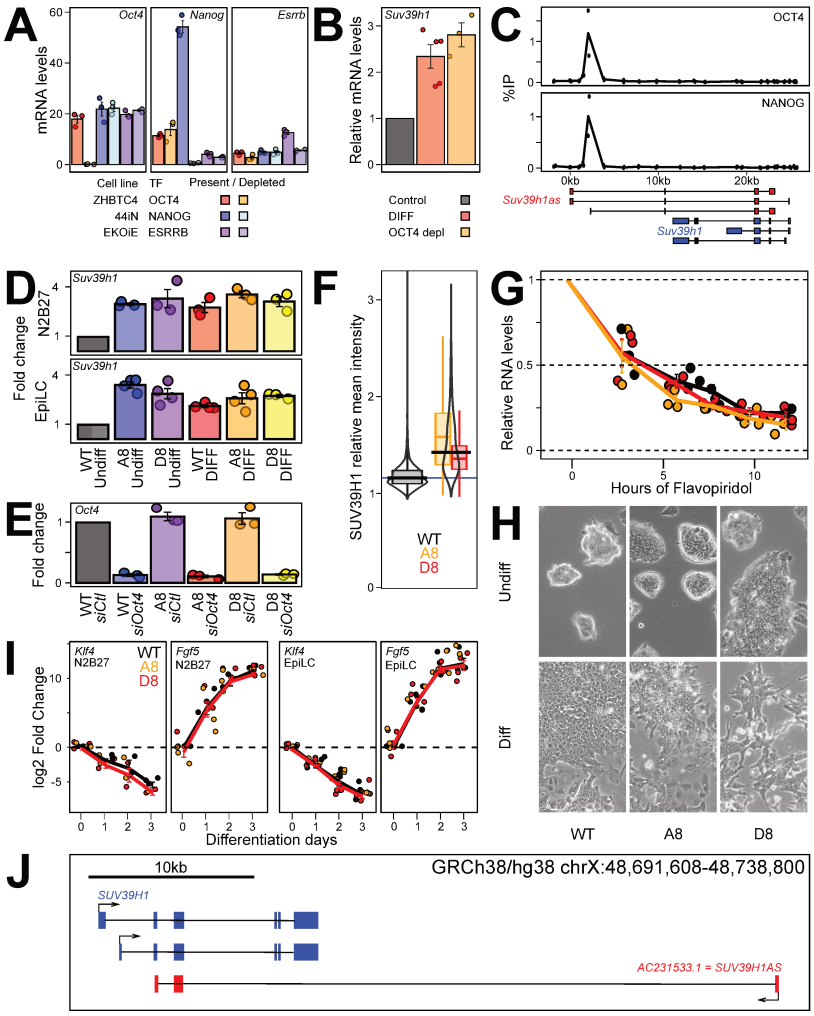
39. Cho, S.W. et al. Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. Cell. 173(6):1398-1412.e22 (2018).

40. Peters, A. H. F. M. et al. Loss of the Suv39h Histone Methyltransferases Impairs Mammalian Heterochromatin and Genome Stability. Cell 107, 323–337 (2001).

41. Festuccia, N., Gonzalez, I. Navarro, P. The Epigenetic Paradox of Pluripotent ES Cells. Journal of Molecular Biology 429, 1476–1503 (2017).

42. Liu, H., Kim, J.-M. Aoki, F. Regulation of histone H3 lysine 9 methylation in oocytes and early pre-implantation embryos. Development 131, 2269–2280 (2004).

43. Deng, Q., Ramskold, D., Reinius, B. Sandberg, R. Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells. Science 343, 193–196 (2014).



Supplementary Information, Fig. S 1. Additional information on *Suv39h1* regulation by OCT4 and *Suv39h1as*. **(A)** Expression of *Oct4*, *Nanog* or *Esrrb* upon inducing their depletion in specific dox-inducible knock-out lines. Note in EKOiE the remnant expression of *Esrrb* produces a truncated, non-functional protein[23]. Each dot represents an independent replicate and the bar the corresponding mean and standard error. **(B)** RT-qPCR validation of *Suv39h1* overexpression upon differentiation (DIFF, 3 days without LIF, red) or upon OCT4 depletion (24h, orange). Each dot represents an independent replicate and the histogram the corresponding mean and standard error. **(C)** ChIP-qPCR validation of OCT4 and NANOG binding at the promoter region of *Suv39h1as* in E14Tg2a (X-axis represents genomic distances in kb with respect to the *Suv39h1as* transcription start site). Each black dot represents the percentage of immunoprecipitation (%IP; Y-axis) of an independent replicate and the line the corresponding mean. A schematic representation of the locus with different isoforms is shown below. **(D)** RT-qPCR confirmation of lack of increased *Suv39h1* upregulation in differentiated *Suv39h1as*-mutant cells obtained by N2B27 or EpiLC assays as compared to wild-type cells. Each dot represents an independent replicate and the histogram the corresponding mean and standard error. **(E)** RT-qPCR confirmation of *Oct4* knock-down in wild-type and *Suv39h1as*-mutant cells. Each dot represents an independent replicate and the histogram the corresponding mean and standard error. **(F)** Violin and box-plots (median; 25-75% percentiles; error bars) showing immunofluorescence quantification of SUV39H1 mean intensity in wild-type (black; n= 4048 cells) and mutant cells (A8 – yellow; n= 4949 cells and D8 – red; n= 4448 cells). **(G)** Analysis of *Suv39h1* mRNA half-life in wild-type and *Suv39h1as*-mutant cells, performed and presented as in Fig. 2A. **(H)** Representative bright-field photomicrographs of wild-type (E14Tg2a) and mutant (A8 and D8) cells cultured in undifferentiating (FCS/LIF) and differentiating (3 days of LIF withdrawal) conditions. **(I)** RT-qPCR analysis of ES (*Klf4*) and differentiation (*Fgf5*) markers during differentiation in N2B27 or in EpiLC conditions, as indicated, in wild-type (E14Tg2A, black) and in *Suv39h1as*-mutant cells (A8, orange, and D8, red). Each dot represents an independent replicate and the lines corresponding means (Black for WT and red for both mutant) with error bars. **(J)** Schematic representation of the *SUV39H1/SUV39H1as* locus in the human genome (Gencode V36 assembly).

V) Complementary studies of Suv39h1as in the mESCs

In this second part of the results, some experiments and analyses were performed by different persons. Victor Heurtier created the SunTag cell line and introduced the gRNA specific for Suv39h1as. Almira Chervova identified and quantified different isoforms by analysing RNA-seq data. And, Agnès Dubois performed one of EBs differentiation presented.

Until now, we have been using a loss-of-function method by deleting both Suv39h1as promoters with the CRISPR-Cas9 system. To assess the impact of Suv39h1as on Suv39h1 expression, we also tried a gain-of-function approach by overexpressing Suv39h1as specifically with the SunTag system, without altering the DNA sequence.

1) Suv39h1as enhanced expression by the SunTag system

a) The SunTag cell line: an elegant system to enhance the transcription of a specific gene

The SunTag system relies on the use of gRNAs and an enzymatically inert Cas9 (dCas9)[341]. In this method, dCas9 is fused to 10 GCN4 epitopes acting as a recruitment platform for diffusible antibodies fused to the potent VP64 transactivator. As a result, whenever a gRNA targets a promoter, up to 10 copies of VP64 are recruited, inducing strong levels of transcription. In the lab, this system was further engineered such that both dCas9 and VP64 are cloned in an established cell line and are expressed only upon the addition of Dox. Therefore, 2 clones, A1 and E1, can introduce the desired gRNA expression vector. Initially, 6 gRNAs were designed and cloned in a gRNA expressing vector to target the 5' extremity of the isoform 3 of Suv39h1as within a DNase accessibility region and between the two promoters **(Fig. 9A)**.

After lipofection in A1 cells and puromycin selection, the bulk of cells was treated with Dox for 3 days. We observed that 3 gRNAs increased transcriptional activity of Suv39h1as **(Fig. 9B)**. Subsequently, these 3 gRNA expression vectors were individually lipofected in A1 and E1 cells and selected in puromycin for 6 days. The resistant cells were then plated at clonal density and 6 independent clones representing all combinations between A1/E1 and the 3 gRNAs were selected and expanded for further analyses.

b) The Suntag System enhance Suv39h1 transcription to a moderated extent

In order to assess whether the up regulation of Suv39h1as can induce a variation in Suv39h1 expression, cells were cultured for 3 days in either FCS/L or FCS with and without Dox to activate the SunTag system. We observed that Suv39h1as was systematically

upregulated in the presence of Dox, albeit at very low levels, in particular in differentiating cells **(Fig. 9C)**.

Unfortunately, the moderate increase of Suv39h1as had a very small impact of Suv39h1 expression **(Fig. 9C)**. In conclusion, the SunTag system was efficient to some extent to activate Suv39h1as but the increase of Suv39h1as expression only led to a minor Suv39h1 downregulation. Moreover, in a differentiation context, Suv39h1as that is normally lowly expressed, was still overexpressed in small proportions by the SunTag system. It suggests that the system can be shut down upon differentiation, or that it cannot override other regulatory impositions on the locus.

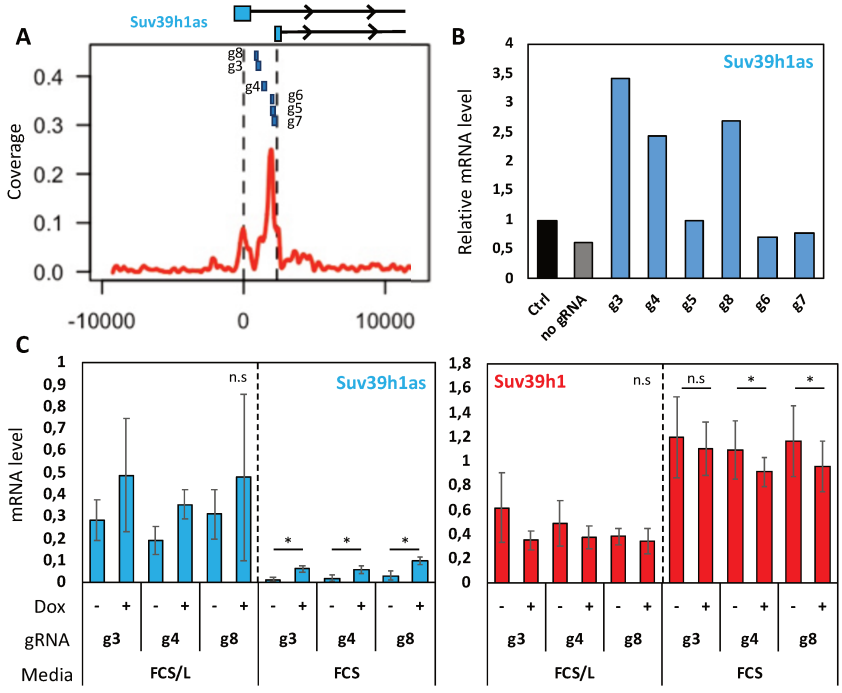


Figure 9: **Suv39h1as overexpression by SunTag system.** **(A)** Representation of potential ST guides positions relative to the opening of the region (shown by DNAseI sensitivity) centered on Suv39h1as transcription start sites. **(B)** Histogram showing the rela-tive expression level of Suv39h1as in cells transfected by the different guides (in blue) or not (in grey) and cultured with Dox for 3 days or not (Ctrl in black). **(C)** Histograms showing expression of Suv39h1as (blue) and Suv39h1 (red) measured by RT-qPCR and normalized to Tbp. Each gRNA value (g3, g4 and g8) is the mean of two clones (A1 and E1) and two replicates cultured with (+) or without (-) dox in FCS and FCS/L for 48h.

Finally, with the SunTag system, it was not possible to overexpress Suv39h1as enough to be conclusive. We chose this approach to avoid the DNA sequence alteration, but, as it is not efficient, an alternative would be to replace Suv39h1as' promoters with CAG promoters in order to strongly overexpress it. Nevertheless, as the introduction of the CAG sequence will introduce a huge change on the locus, it would be nice to have a Flox CAG to be able to remove it and assess its effect specifically.

2) Potential existence of an additional mechanism of regulation by Suv39h1as

In the previous section, we elaborated a mechanism of Suv39h1 regulation by Suv39h1as. Suv39h1as acts as a transcriptional repressor of Suv39h1 via the control of Suv39h1 frequency of transcription and has no impact on Suv39h1 mRNA stability. Suv39h1as

has complex influences on the chromatin with a slight local euchromatinazation. Nonetheless, these mild effects may not fully explain the 2 to 3-fold mRNA and protein increase. We hypothesize that other mechanisms might be involved.

To support this idea, the quantification of Suv39h1 pre-mRNA level by RNA-seq **(Fig. 10A)** showed a very minor upregulation in the absence of Suv39h1as. To confirm it, we performed a strand-specific RT-qPCR, priming the RT in Suv39h1 exon 3 **(Fig. 10B)**. Trans-exonic amplification between exon 2 and exon 3 allows to measure the level of mature mRNA (**Fig. 10B** – red) while intronic amplification of intron 2 measures the level of pre-mRNA ((**Fig. 10B** – yellow). We also performed a control strand-specific RT-qPCR in parallel with Tbp only priming to measure the background (**Fig. 10B** – dark and light greys). First, we confirmed the mature mRNA upregulation then, pre-mRNA displays a very low level of expression, close to background noise. This is in line with the RNA-seq results. Furthermore, there is no upregulation of pre-mRNA in the mutated cells, confirming the RNA-seq results.

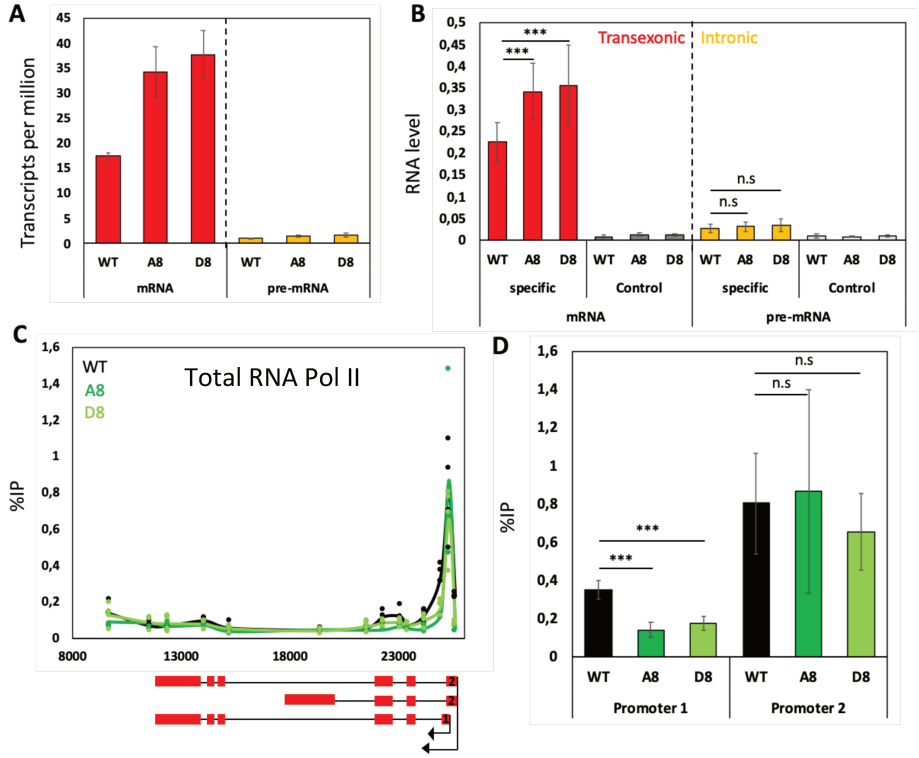


Figure 10: **Suv39h1as impacts in a very mild way Suv39h1 transcription intensity.**
(A) Histogram presenting count of transcripts per million of Suv39h1 mature mRNA and pre-mRNA in WT and mutated cells (A8 and D8). Each value is the mean of 2 replicates. **(B)** Histogram showing expression level of mature Suv39h1 mRNA (in red) and pre-mRNA (in yellow) measured normalized to Tbp. The strand-specific RT-qPCR was primed either in Tbp gene and Suv39h1 exon3 (Specific) or in Tbp only (Control – dark and light grey). Each value is the mean of 4 replicates. **(C)** Profile of total RNA Pol II across Suv39h1 locus (X-axis represents genomic distances from Suv39h1as promoter 1 in bp) in WT (in black) and mutated clones A8 (green) and D8 (light green). Each dot represents the % of immunoprecipitation (%IP in Y-axis) of independent replicates and the line represent the average of the 4 replicates at each genomic location. Under the graph is presented a schematic representation of Suv39h1 isoforms, with promoter 1 and promoter 2 indicated. **(D)** Histogram representing the % of IP at Suv39h1 promoter 1 and 2 for each cell line indicated. Each value is the average of 4 independent replicates.

We also performed a total Pol II ChIP-qPCR on the locus which showed no particular difference between WT and mutated clones **(Fig. 10C)**. When looking more carefully at Suv39h1 promoters, we observed no increase of transcription initiation **(Fig. 10D)** and, even more, a decrease at promoter 1. This might be explained by the loss of Suv39h1as transcription which is compensated at Promoter 2. Therefore, there might be a small upregulation of Suv39h1 transcription initiation, largely not present at the pre-mRNA. Moreover, it would not explain the 2 to 3-fold upregulation of the mRNA.

Overall, these results suggest that the transcriptional effect only partially explains regulation. Moreover, it is known that a single lncRNA can operate at different mechanisms[329]. Thus, the next sections will develop different attempts to elucidate a potential additional mechanism.

3) Study of the impact of Suv39h1as RNA molecule in the mechanism of regulation

a) Transitory and stable Suv39h1as overexpression

To understand whether Suv39h1as can act from outside of its locus of origin, we used two approaches. First, we constructed a plasmid containing the cDNA of isoform 2 of Suv39h1as, which was then cloned under the control of a strong CAG promoter with a puromycin resistance cassette, and finally lipofected in WT and Suv39h1as KO cell lines A8 and D8. After that, the cells were selected for 3 days and harvested to extract RNAs. In parallel, the same plasmid without cDNA was also lipofected as a control (**Fig. 11A**). The measure of Suv39h1as expression (in blue) confirmed its huge overexpression – by a factor of 50. In this context, Suv39h1 expression (in red) doesn't vary in the WT cell line. However, it is slightly downregulated in A8 and D8, suggesting that the RNA molecule has an effect. To assess the cDNA overexpression impact at a single cell level, a sm-FISH for Suv39h1as and Suv39h1 was performed on the lipofected and selected cells (**Fig. 11B**). We first observed a very variable Suv39h1as overexpression at the single cell level. Cells were also presenting foci of Suv39h1as RNA (yellow arrow head), probably because of a local accumulation of the plasmid.

Second, to overpass this single cell variability of overexpression, we decided to generate a cell line with a stable Suv39h1as overexpression by electroporation. In that way, 4 independent stable clones were generated in WT and A8 cells with the same cDNA construct (**Fig. 11C**). The clones were chosen to exhibit different level of Suv39h1as overexpression (in blue). However, whatever the level of Suv39h1as expression, no effect on Suv39h1 mRNA was

observed (in red). This suggested that the RNA molecule was not acting on Suv39h1 expression and the previous result was actually an artifact, most probably due to the plasmid presence.

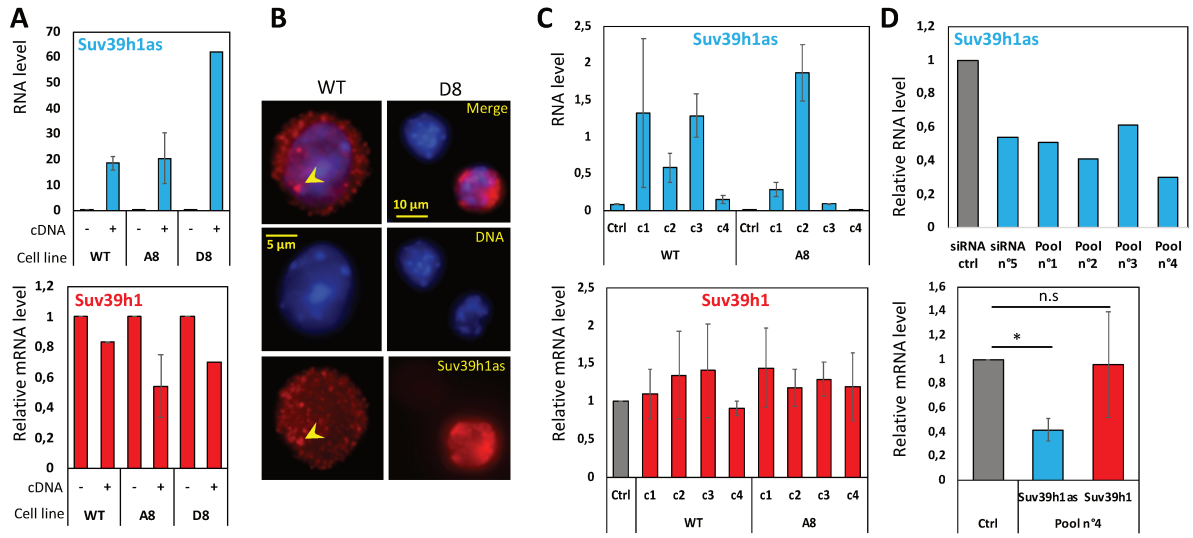


Figure 11: **Suv39h1as is not acting via a trans-mechanism.**
(A) Histograms showing the RNA level of Suv39h1as (in blue) or the relative RNA level of Suv39h1 (in red) normalized to Tbp in WT, A8 and D8 cell lines transfected with Suv39h1as cDNA containing plasmid (+) or an empty plasmid (-). For Suv39h1, the values were also normalized to corresponding empty plasmid condition. Each value in the average of 2 independent replicates except for D8 transfected with a cDNA. **(B)** Representative pictures of WT or D8 cells transfected with Suv39h1as cDNA and hybridized with Suv39h1as sm-FISH probe. The yellow arrow head indicated the presence of local accumulation of Suv39h1as transcript. **(C)** Histograms presenting the level of expression for Suv39h1as (in blue) and Suv39h1 (in red) in WT and A8 parental lines (Ctrl) or in 4 independent cell lines transfected with Suv39h1as cDNA (from c1 to c4). Values were normalized to Tbp and for Suv39h1, also to the parental cell line values. Each value is the average of 3 independent replicates. **(D)** Histograms showing the level of Suv39h1as expression (in blue) or Suv39h1 (in red) in cells nucleofected with individual or pooled Suv39h1as siRNA or non-binding siRNA as control (Ctrl). The values are either a single replicate value or the average of 4 values and were normalized to Tbp and siRNA control value.

b) Suv39h1as knock-down by siRNAs

To confirm the RNA molecule lack of impact on Suv39h1as expression, we decided to use a third strategy. We produced 11 siRNAs to knock-down (KD) Suv39h1as RNA. To design those siRNAs, we used Dharmacon and Blockit websites to generate their sequences and estimate their efficiency. Among the most likely to be effective, we scored them according to criteria suggested by Fakhr et al.[344] choosing the most relevant again. From those filters, 11 siRNA were chosen. They are spread along the cDNA sequence. The siRNAs were pooled (see **Table 4**) and tested on WT cells, which were then harvested 24h after nucleofection for expression analysis. A control pool of non-binding siRNA was also nucleofected in parallel. The different combinations of pooled siRNAs were not efficient enough, except for the pool of the 6 most efficient siRNAs **(Fig. 11D)**. Despite a 50% to 70% reduction in Suv39h1as expression, we did not observe any change in Suv39h1 mRNA levels **(Fig. 11D)**. These results confirmed that, indeed, the RNA molecule was not involved in the mechanism of regulation.

However, the KD of Suv39h1as is only partial, reducing probably mainly the number of diffusible molecules and not at the locus.

Finally, these different assays indicate that Suv39h1as RNA itself is not important for Suv39h1 regulation outside of the locus. As the transcriptional effect is moderate, it suggests an additional mechanism impacting the locus. It could be transcriptional with for example the induction of early termination during Suv39h1 transcription. The hypothetical other mechanism can also be post-transcriptional with the modulation of Suv39h1 splicing.

4) Study of Suv39h1as/Suv39h1 acts of transcription at single cell level by sm-FISH

As discussed above, we hypothesized a transcriptional additional mechanism via a mutual exclusion of Suv39h1as and Suv39h1 RNA Pol II on the locus. This hypothesis is linked to an unexpected observation: when we performed sm-FISH on Suv39h1 and Suv39h1as RNAs, using WT cells, 25% of cells were exhibiting co-existence of Suv39h1 and Suv39h1as pin points at the DNA locus, indicating a possible "co-transcription" of the genes. However, the two dots were ***NEVER*** overlapping (nOv) as illustrated on **Fig. 12B**.

The microscopic resolution was not high enough to distinguish the production of the two RNAs on the locus. We first thought that this phenomenon was caused by a physical artefact: the chromatic aberration due to the use of different wavelengths for the two genes. To test this idea, we examined tetra-colored beads with sm-FISH probes colors (**Fig. 12C**). However, we saw the different colors overlapping in the middle. They slightly shift on the side but not in the extent of the sm-FISH signals, arguing that the chromatic aberration was not sufficient to explain such phenomenon.

After, we realized that Suv39h1as probe was only exonic (called thereafter ASe – **Fig. 12A**), detecting mostly mature RNA and probably not, or very badly, the pre-mRNA on the locus. We then decided to design two new sets of intronic probes for Suv39h1as schemastically presented in **Fig. 12A**:

(1) AS1: a set of probes was hybridizing on the 5' of Suv39h1as RNA, outside of the Suv39h1 overlap.

(2) AS2: a set of probes was on the 3' part of the RNA, on the section overlapping with Suv39h1.

We then performed sm-FISH mixing AS1 or AS2 with the canonical Suv39h1 that binds in Suv39h1 introns and exons **(Fig. 12A)**. We first observed for both AS1 and AS2 that the number of diffusible molecules decreased drastically, as expected for intronic probes **(Fig. 12D)**. We

also noted one or two bright dots corresponding most probably to the transcription site (TS). Furthermore, we observed the co-existence of Suv39h1 and Suv39h1as dots at the same location, with, the two signals sometimes overlapping **(Fig. 12D** – Ov**)**.

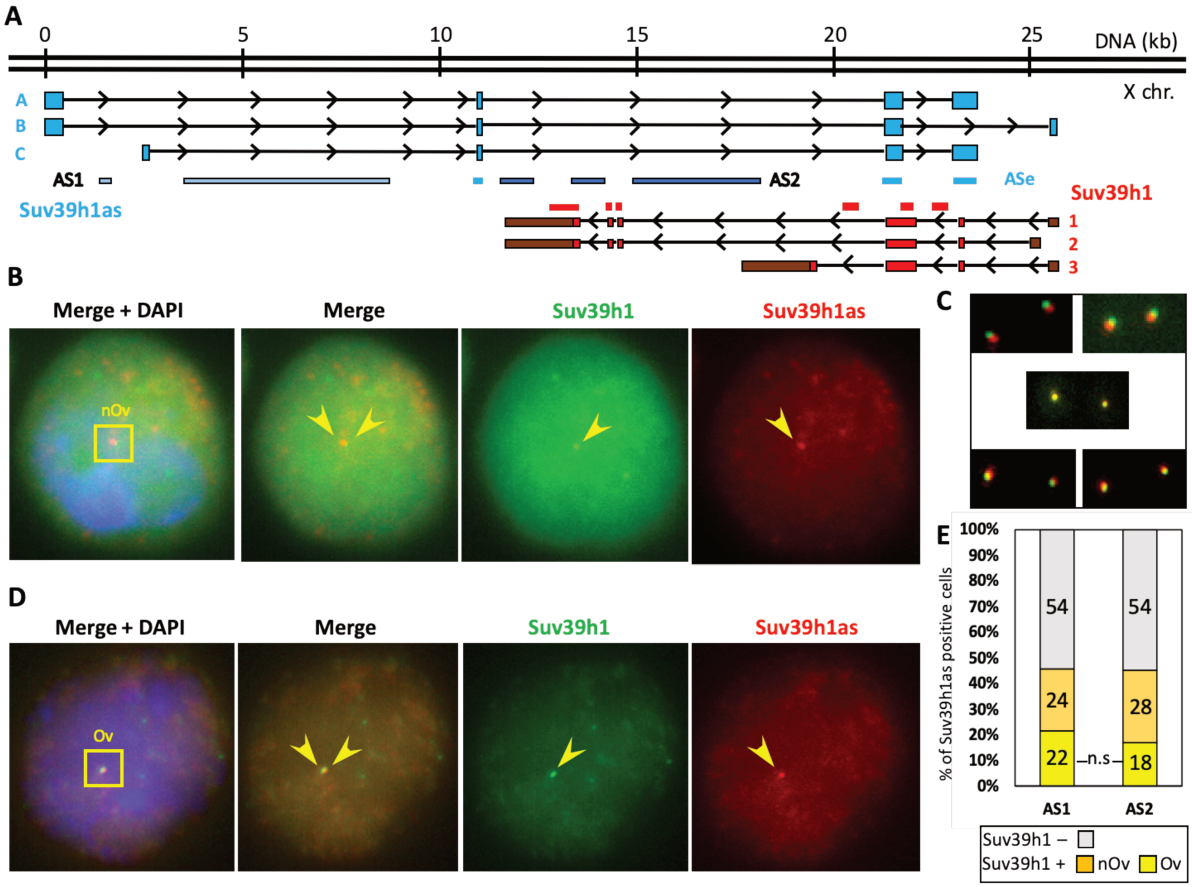


Figure 12: **Suv39h1as pre-RNA overlaps with Suv39h1 transcript at the locus.**
(A) Schematic representation of Suv39h1as/Suv39h1 locus with position of Suv39h1as intronic (AS1 and AS2) and exonic (ASe) sm-FISH probes as well as Suv39h1 intronic and exonic probes. Boxes represent exons and black line the introns. Suv39h1as isoforms are depicted in blue (A, B and C) and Suv39h1 in red (isoforms 1, 2 and 3). The genomic approximative location is represented with the double line in kb. **(B)** Representative picture of a cell displaying non-overlapping (nOv) but close Suv39h1 and Suv39h1as signals. Cells have been hybridized with ASe and Suv39h1 probes. Yellow arrow head are pointing Suv39h1as and Suv39h1 signals. **(C)** Representation of a field with beads presenting a signal in Suv39h1 (green) and Suv39h1as (red) wave length. The images are placed according their position in the microscope field. **(D)** Representative picture of a cell displaying overlapping (Ov) but close Suv39h1 and Suv39h1as signals. Yellow arrow head are pointing Suv39h1as and Suv39h1 signals. Cells have been hybridized with AS1, AS2 and Suv39h1 probes. Yellow arrow head are pointing Suv39h1as and Suv39h1 signals. **(E)** Stacked bar chart displaying the proportion of cells with Suv39h1as signal only (in grey) or with Suv39h1 and Suv39h1as signals overlapping (Ov - yellow) or not (nOv – orange) in cells hybridized with AS1 or AS2. Proportions are calculated on 305 cells for AS1 and 291 cells for AS2.

This phenomenon was present for both AS1 and AS2. This result suggested that the first probe, only exonic, might have detected the presence of mature RNA still on the locus instead of the active transcription. In addition, it highlights the necessity to detect Suv39h1as pre-mRNA to actually visualize the active transcription at the TS.

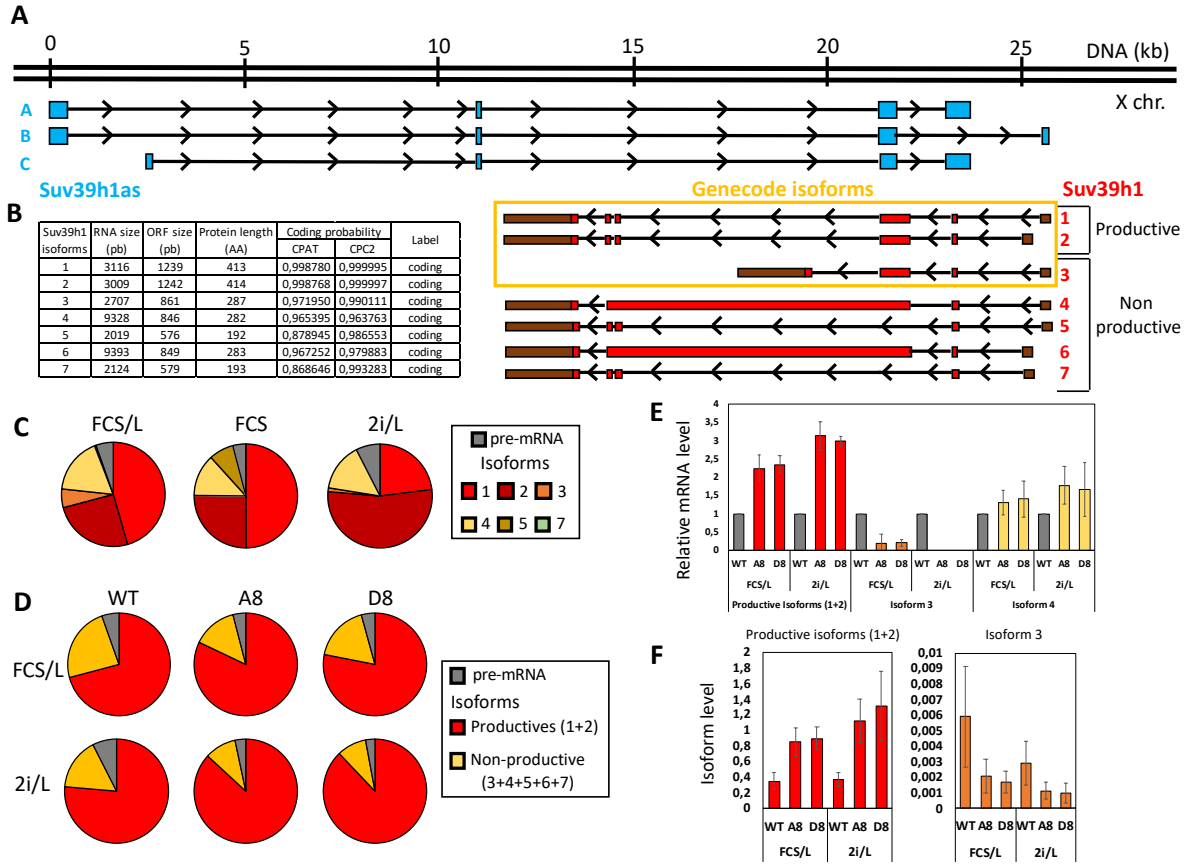
To go further in the hypothesis of the mutual exclusion of the RNAs Pol II at the locus, we counted the number of occurrences of "overlapping" (Ov – **Fig. 12D**) and "non-overlapping" (nOv **– Fig. 12B**) signals at the transcription site in the different setups (**Fig. 12E**). We were expecting a higher proportion of nOv signals for the AS2 setting compared to AS1. Indeed, the production of AS2 part of Suv39h1as implies that Suv39h1as RNA Pol II is going through Suv39h1 overlapping section, possibly preventing the transcription of Suv39h1 on the other strand. At first, we first noticed that both genes were expressed at the TS (gOv = Ov + nOV) among 45% of the cells expressing Suv39h1as, for both AS1 and AS2. Hence, detection of Suv39h1as appears to be similar in both cases. But for AS1, we observe an equal repartition of the Ov and nOv among the gOv (24% and 22% meaning 50/50 among gOv) whereas for AS2, there is a slight increase of nOv signals (28% and 18% meaning 40/60 among gOv). These results are in line with our hypothesis of a mutual exclusion of RNA Pol II, however, the increase of nOv signal in AS2 experiment is not significative. Therefore, if this event is happening, it is most probably only a tendency linked to the antisense orientation of Suv39h1as and not a general mechanism of action. Nevertheless, Suv39h1 probe is a mix of intronic and exonic oligos, detecting both mature mRNAs and pre-mRNAs. Therefore, it is possible that we don't detect properly Suv39h1 active transcription, and reduce the differences between AS1 and AS2. It would be valuable to design new probes for Suv39h1, one intronic (S1) only and the other exonic only (S2), in order to perform combinatory sm-FISH experiment by mixing AS1, AS2 or ASe with S1 or S2, respectively.

5) Study of Suv39h1 isoforms expression in mouse embryonic stem cells

a) Suv39h1 isoforms identification

In our quest to identify an additional mechanism of regulation, we noticed that several isoforms of Suv39h1 RNAs were annotated on Genecode. Therefore, we used *de novo* transcript assembly with our RNA-seq datasets, together with direct cDNA cloning, sequencing and RT-qPCR to identify and validate several isoforms expressed in mESCs (**Fig.13A**). In agreement with annotated isoforms, we found three protein-coding mRNA isoforms, among which a shorter one that produced a truncated and not functional protein. We also identified several other isoforms that had never been annotated. One type of isoform was missing the full exon 3, resulting from an alternative splicing. The other type of isoform was the consequence of introns 3 and 4 retention. Interestingly, these alternative processes always occur between exon 2 and exon 3, the two exons that overlap with Suv39h1as exons. In addition, those two types of

isoform initiate from both promoters, suggesting a general phenomenon, independent of the promoter used. By processing all isoforms cDNA through CPAT[345] and CPC2[346] to assess their coding potential (**Fig. 13B**), we can observe that they all keep their coding potential despite a slight decrease of their coding probability. Nevertheless, given the intron inclusion or loss of exon, they probably encode for non-functional proteins although we didn't address this specific question.



Suv39h1 isoforms	RNA size (pb)	ORF size (pb)	Protein length (AA)	Coding probability		Label
				CPAT	CPC2	
1	3116	1239	413	0,998780	0,999995	coding
2	3009	1242	414	0,998768	0,999997	coding
3	2707	861	287	0,971950	0,990111	coding
4	9328	846	282	0,965395	0,963763	coding
5	2019	576	192	0,878945	0,986553	coding
6	9393	849	283	0,967252	0,979883	coding
7	2124	579	193	0,868646	0,993283	coding

Figure 13: **Characterization and quantification of Suv39h1 isoforms.**
(A) Schematic representation of Suv39h1as (in blue)/Suv39h1 (in red) locus with presentation of Suv39h1 already annotated (in yellow box) in Genecode or *de novo* isoforms. Isoforms can give a productive protein (1 & 2) or most probably a non-productive one (3 to 7). **(B)** Table recapitulating the different characteristics of Suv39h1 RNAs isoforms and their coding probabilities calculated by CPAT and CPC2 online tools. **(C)** Pie charts of Suv39h1 pre-mRNA and isoforms proportions for WT cells cultured in FCS/L, 2i/L and FCS. **(D)** Pie charts of Suv39h1 pre-mRNA, productive (in red) and non-productive (in yellow) isoforms proportions for WT, A8 and D8 cells cultured in FCS/L and 2i/L. **(E)** Histogram of expression level of isoform 3, 4 or productive ones in WT, A8 and D8 cells cultured in FCS/L and 2i/L. The values are normalized to corresponding WT condition. The values of **(C)**, **(D)** and **(E)** are extracted from RNA-seq analysis and are the average of 2 replicates. **(F)** Histograms of isoform 3 (in orange) and productive ones (in red) level of expression measured by RT-qPCR in WT, A8 and D8 cells cultured in FCS/L and 2i/L. The values are the mean of 41 replicates for WT FCS/L, 25 replicates for A8 and D8 in FCS/L and 12 replicates for 2i/L conditions.

b) Quantification of Suv39h1 isoforms

To investigate a potential regulation of those different isoforms by Suv39h1as, we used our RNA-seq to quantify each isoform level in different pluripotent conditions (**Fig. 13C**). In general, we observe that the major part of the expression matches the isoform annotated as coding for a productive protein (Isof 1 & 2 – productive isoforms), with 70% in FCS/L, 75% in differentiation and 76% in 2i/L. Among the non-productive isoforms, Isoform 4 is the major one, with 17% in FCS/L, 7% in FCS and 15% in 2i/L. Isoform 3, encoding for a truncated protein, is lowly expressed and specifically in FCS/L media. Those two isoforms are in line with the RNA-seq reads coverage on the locus, depicting an accumulation right after the exon 3 in the WT and not in the mutant cell lines (**see figure 1D and 1E, chapter IV**). Interestingly, in FCS/L, the proportion of non-productive isoform decreases from 23% to 15% in the A8 and D8, respectively, with a subsequent increase of productive isoforms from 70% to 80% (**Fig. 13C**). Isoform 3 is specifically decreasing in the mutant cell line while Isoform 4 is staying at the same level (**Fig. 13E**). This might be explained by an increase of the elongation concomitantly with a reduction of the splicing probability that leads to the retention of intron 3 and 4. Interestingly, in 2i/L where the absence of Suv39h1as induces the highest Suv39h1 upregulation, the percentage shifts are the most important from 16% to 9% (**Fig. 13D**). Furthermore, the decrease of Isoform 3 in the mutant cell has been confirmed by RT-qPCR (**Fig. 13F**). Nonetheless, repeating the same procedure for the isoform 4 is impossible by classic RT-qPCR because it is impossible to design a couple of primers specific to this isoform.

In summary, the non-productive isoforms are representing a modest part of the transcripts. They are also slightly decreasing in the mutant cell line. However, all these results are coming from RNA-seq analysis data and would need confirmation by RT-qPCR. To go further, we may postulate that Isoform 2 is a *de novo* assembly artifact, reflecting an early termination of transcription inside the intron 3 instead of a full intron retention. Indeed, in addition to the profile depicted by the RNA-seq, it would be in line with the previously exposed possible RNA Pol II mutual exclusion suggested by sm-FISH . These elements allow us to build a model where Suv39h1as transcription reduces the capacity of Suv39h1 transcribing polymerases to elongate properly, leading to short and non-productive RNAs. This model will be discussed in the section VII.

6) Other differentiation assays in Suv39h1as knock-out cell lines

In our search for a phenotype produced by the absence of Suv39h1as, we conjecture that Suv39h1 overexpression may alter the differentiation process. The increase of H3K9 methylation may have induced deleterious gene silencing. Therefore, we tried several protocols to differentiate the cells. The differentiation by the withdrawal of 2i and LIF factors (2iOFF) have already been exposed in the previous chapter. We also tested differentiation in EpiLSCs and embryoid bodies (EBs).

a) Differentiation in EBs

A classical way to functionally test the pluripotency is to differentiate cells into EBs. During this assay, cells are supposed to self-organize and initiate the three embryonic layers[120]. The advantage of EBs is to be a non-directed differentiation, allowing cell to "naturally" differentiate. We tested two different protocols to generate EBs, with the EBs size controlled or not. We monitored gene expression at day 6 and day 8 after the beginning of the differentiation. We first observed that the differentiation was efficient, with the downregulation of the pluripotent TF Nanog and Klf4 (**Fig. 14A**) and the upregulation of markers of the differentiation such as Gata4, Gata6 and Sox7 (**Fig. 14B**). Suv39h1as is also completely disappears by day 6 while Suv39h1 is mildly upregulated at day 6 and keeps increasing at day 8 (**Fig. 14C**). By examining the gene expression in the mutant cell lines, we observed no upregulation of Suv39h1 at day 0. This phenomenon might be due to the fact that the cells were cultured at very high density just before processing them to form EBs. In the WT cell, Suv39h1as level of expression was already very low. Therefore, despite a small reduction observed in the level of differentiation markers, it is difficult to link Suv39h1 dynamic with the absence of Suv39h1as. Furthermore, even if there is only one replicate for each protocol, the result suggests that this assay was not appropriate in the study of Suv39h1 and Suv39h1as expression.

b) Differentiation in EpiLSCs

By exploring RNA-seq data previously made in the lab, we noticed an anticorrelated pattern of expression between Suv39h1as and Suv39h1 during the differentiation in EpiLSCs. Suv39h1as is shut down while Suv39h1 is upregulated. To investigate this, we applied the protocol established by Hayashi et al.[126] to WT and mutant cell lines, and monitored the genes expression during the differentiation process. First, we confirmed the differentiation was efficient by observing a downregulation of pluripotent factors as Nanog and Klf4 (**Fig. 14D**)

and an upregulation of markers of the differentiation such as Fgf5 and Wnt3 **(Fig. 14E)**. We also validated the anticorrelated pattern of expression between Suv39h1as and Suv39h1 after 1 day of differentiation. At 3 days, Suv39h1as is completely lost while Suv39h1 is maximal **(Fig. 14F)**. In the mutant cell lines, Suv39h1 level is already maximal at day 0. Nevertheless, the expression profiles of pluripotent and differentiation markers are not altered in the mutant cell lines, suggesting that the differentiation is happening correctly. Additionally, the cells are exhibiting normal morphologic changes during the differentiation. In conclusion, despite anticipating overexpression of Suv39h1, we cannot observe a clear phenotype.

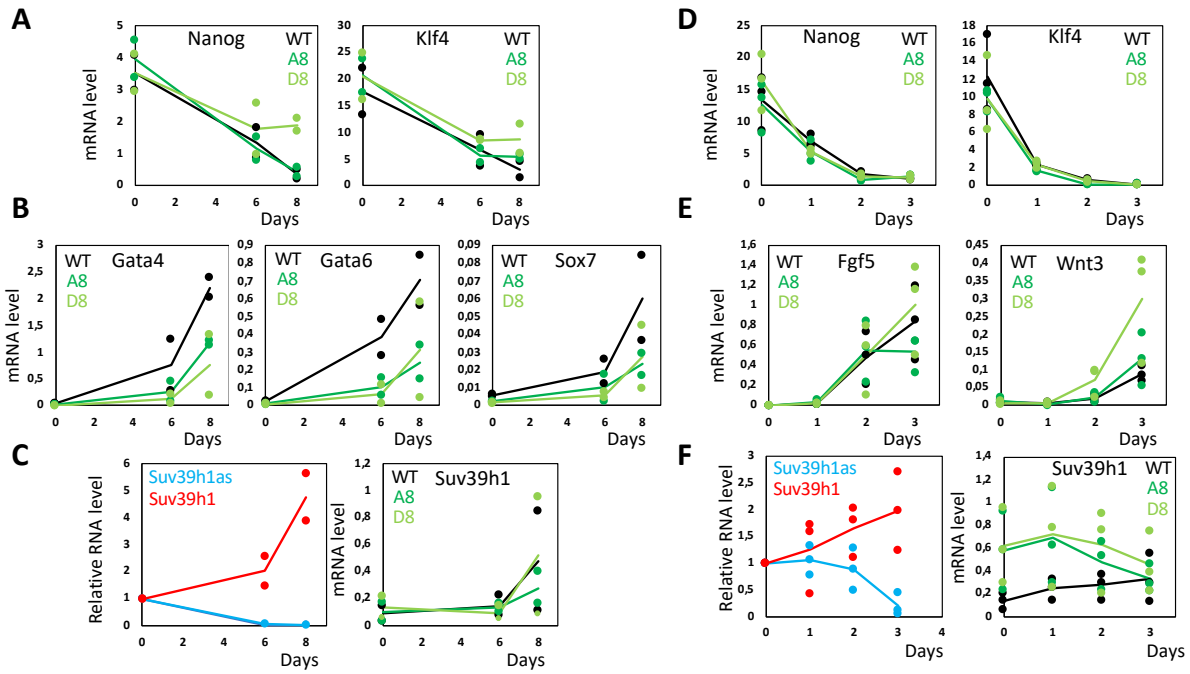


Figure 14: **EBs and EpiLSC differentiation in WT and mutated cells.**
RT-qPCR analysis of markers of pluripotency (**A** and **D**), differentiation (**B** and **E**) or Suv39h1 and Suv39h1as expression (in WT cells specifically – **C** and **F**) after 0, 6 and 8 days of EBs differentiation (**A**, **B** and **C)** or 0, 1, 2 and 3 days of EpiLSC differentiation (**D**, **E** and **F**) in WT (black), A8 (green) and D8 (light green) cells. Tbp was used for normalization, and in some case, values were also normalized to D0. Each dot represents an independante replicate and lines are the mean values. There is 2 replicates for **(A)**, **(B)** and **(C)** and 3 replicates for **(D)**, **(E)** and **(F).**

VI) Study of Suv39h1as knock-out mouse model

In this last part of the results, some experiments were performed by different persons. Sandrine Vandormael-Pournin and Michel Cohen-Tannoudji helped me a lot in the experimental design of mouse experiments and performed the different immunostainings on oocytes and zygotes.

1) Rationale to study Suv39h1as expression during the early development

The heterochromatin is highly dynamic during the early development. More precisely, the level of H3K9 methylation is kept low during the very first steps, especially the di- and tri-methylations[280]. They are re-established only after few cleavage events. It suggests that the HMT specific for H3K9 needs to be tightly regulated. Moreover, several studies in literature showed that SUV39H1 is important for the development. Despite the fact that Suv39h1 KO in the mouse is totally viable, the double KO for SUV39H1 and SUV39H2 leads to growth retarded mice at birth and during the adulthood[239]. Besides, the overexpression of SUV39H1 in the early stages of the embryo causes developmental defects. As shown by Burton et al.[285] and Zhang et al.[225], when they overexpressed SUV39H1 by different means in the zygote, they observed a subsequent increase in H3K9me3 level and then default of compaction leading to early developmental arrest.

With the re-analysis of a single embryonic cell RNA-seq data[347], we observed that Suv39h1 was not expressed before the 2-cell stage when the loss of Suv39h1as was complete **(Fig. 15A)**. Furthermore, Suv39h1as expression was maximal at the end of the oocyte maturation and decreased progressively after the fecundation[348] **(Fig. 15B)**. This suggests that Suv39h1as is only expressed during the oocyte development repressing the expression of Suv39h1 in the oocyte and possibly until the 2-cell stage. To assess the *in vivo* function of Suv39h1as, we generated mouse KO for Suv39h1as by mimicking the mESCs deletion of the two promoters with CRISPR-Cas9 system.

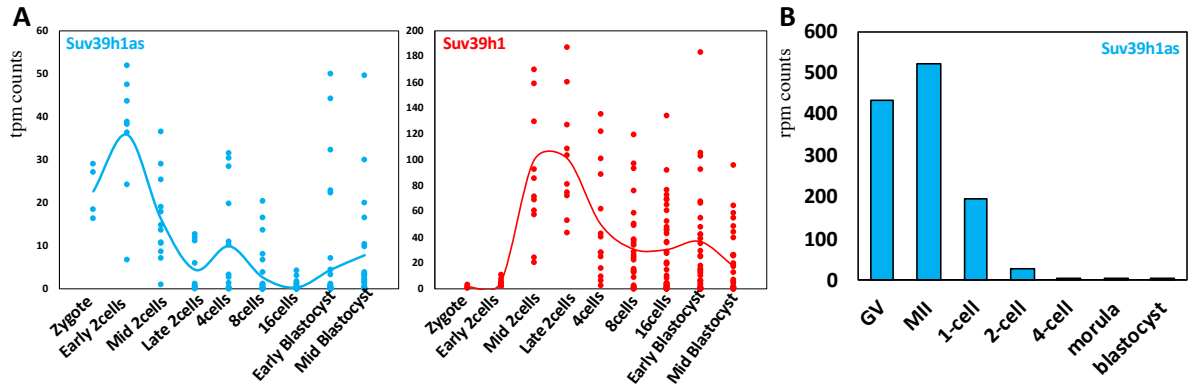


Figure 15: **Suv39h1as expression during early mouse development.**
(A) Re-analysis of single cell RNAseq published datasets (Deng et al.[347]) showing that Suv39h1 (red) expression initiates at the mid-2cell stage, coinciding with a downregulation of Suv39h1as (blue). Each dot represents a single cell quantification. **(B)** RNA-seq published dataset (Karlic et al.[348]) showing Suv39h1as in the late oocyte (GV), ovulated oocytes (MII) and during the first steps of the development.

2) Generation of mouse line knock-out for Suv39h1as

To study the functional impact of Suv39h1as expression *in vivo*, we decided to create a mouse line KO for Suv39h1as. With the help of the Institut Pasteur platform Mouse Genetic Engineering Center (CIGM) under the direction of Francina Langa-Vives, C56BL/6 J mouse embryos were microinjected or electroporated with the CRISPR/Cas9 system and two gRNAs. We used the same gRNA as for the generation of the mESCs in the hope to mimic as much as possible the mESCs deletion (**Fig. 16A**). The deletion was efficient: among the 48 transfected embryos, 14 were presenting an alteration at the locus. Among the 14 mice, 2 males were presenting a hemizygous deletion and 5 females were homozygous. The genotype of those 7 mice was assessed by PCR assay and sequence by PCR cloning.

We chose to generate 2 mouse lines to obtain homozygous and heterozygous females for the deletion and one to have WT females as control. The purpose to control mouse line is to have WT mice in the same experimental context: same animal facility environment and same potential off-targets. The mice 13 and 26 were presenting the same deletion (**Fig. 16B**), very similar to the A8's one with the exception of 1 addition base pair deletion. Then, there were crossed WT males with the same background. 44 males were crossed with WT female to generate WT mice. The deletion was present in the germ cells and transmitted to offspring. F2 mice for the three different lines were used for further studies.

The RNAs were extracted from total ovaries from 3 mice per condition to quantify Suv39h1as expression **(Fig. 16C)**. We observed an absence of Suv39h1as expression in KO/KO females and similar level of expression in KO/+ and WT mice, confirming that Suv39h1as had indeed been invalidated by the promoters' deletion. In addition, we also observed by RT-qPCR

with trans-exonic primers spanning different exons that only isoforms A and/or B (**Fig. 13A**) are expressed in WT oocytes. The second promoter is not expressed. Besides, the expression of Suv39h1 was also quantified with no difference between homozygous, heterozygous and WT mice. This was expected since Suv39h1 is ubiquitously expressed in the ovary and the potential upregulation in the oocyte might not be detectable. Once the KO model was validated, we decided to test the females' fertility.

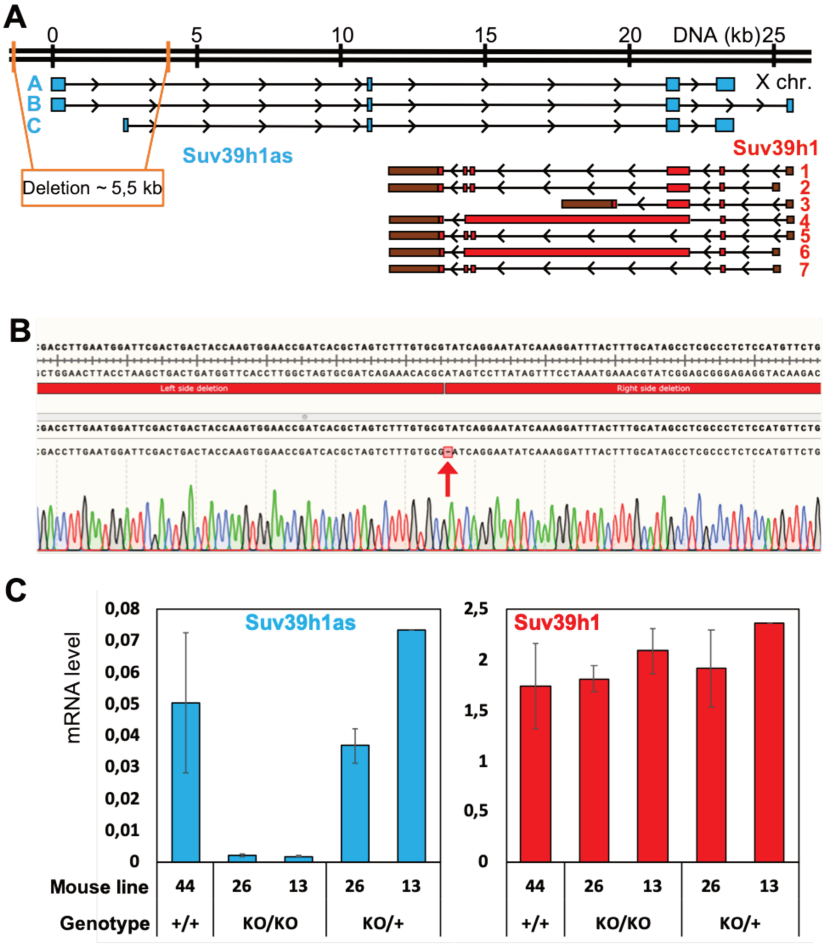


Figure 16: **Validation of Suv39h1as KO mouse line.** **(A)** Schematic representation of Suv39h1/Suv39h1as locus with the localization of Suv39h1as promoters deletion as it is present in mESCs. **(B)** Snapgene capture of mouse line deletion sequence (bottom sequence) aligned against A8 deletion centered on break point (top sequence), with the chromatograph profile. The red arrow points the additional bp deletion. **(C)** Histograms showing Suv39h1as (blue) and Suv39h1 expression levels in 44, 26 and 13 mouse lines ovaries, with WT, homozygous (KO/KO) and hetero-zygous (KO/+) mouse. The values are the mean of 3 inde-pendent replicates except for heterozygous of 13 line, with only one replicate.

3) Fertility tests for Suv39h1as knock-out mouse line

As we postulate that Suv39h1as is important to regulate Suv39h1 during the oocyte maturation, we established different crosses for the mouse lines 13 and 26 to test whether females were fertile. For each mouse line, we implemented:

- homozygous (KO) crosses with 4 females KO/KO and 2 males KO/0, named 13KO and 26KO

- heterozygous crosses with 4 females KO/+ and 2 males KO/0, named 13H and 26H.

In parallel, we also set up a control cross with 4 females and 2 males WT from the line 44, named 44WT. For each cross, we monitored the litter with the date and the number of pups, dead or alive, over 100 days. The fertility graphs represent the cumulative number of pups over the number of days since the beginning of the cross for each litter, with the cumulation of 3 to 4 mice for each cross.

The graph indicates no difference between the heterozygous and the WT crosses in term of number of pups and interval between two litters (**Fig. 17B** & **17D**). For the KO crosses, there might have been a slight decrease in the number of pups per litter and a small increase in the interval between two litters (**Fig. 17A**). The number of days between the first three litters was displayed for each cross, and confirmed the slight increase for the KO crosses (**Fig. 17C**). Nevertheless, the differences were very small and might just be the result of a difficult animal facility environment.

Therefore, we can conclude that the absence of Suv39h1as expression doesn't impact fertility. As a next step, we tried to ascertain whether the absence of impact on fertility comes from Suv39h1as expression having no functional role or whether it was rather due to an absence of Suv39h1 mRNA and protein upregulations. In order to do so, we decided to evaluate the level of Suv39h1 protein in the oocyte.

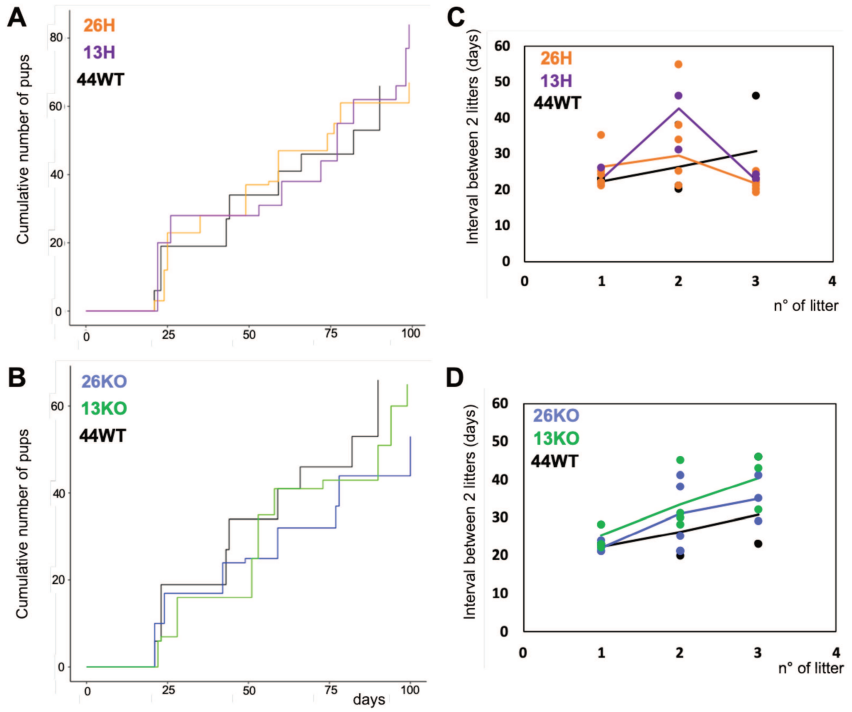


Figure 17: **Suv39h1as KO mouse line are fertile.** **(A)** Fertility plots showing the cumulative number of pups (dead or alive) in function of the number of days between the beginning of the cross (0) and the date of the birth, over 100 days of crossing. **(B)** Number of days between two litter for the 3 first litters. Those plots are ploting each cross: 26KO in blue, 13KO in green, 26H in orange, 13H in purple and 44WT in black.

4) Study of Suv39h1 expression in the oocyte

To assess the effects of Suv39h1as absence on Suv39h1 expression, it was technically difficult to evaluate it at the mRNA level, so we decided to measure directly the protein level by immunofluorescence (IF). To concentrate the potential signal, we performed the experiment on the latest stage of the oocyte maturation with still a nuclear envelop, the GV stage. GVs were harvested from hormonally stimulated WT and homozygous KO mice from the different lines. The GV were then stained for SUV39H1, H3K9me2 and H3K9me3.

The IF for Suv39h1 was performed on 16 to 34 GV depending on the mouse line. Representative immune-stained GV pictures are presented on **Figure 18**. We can observe a clear SUV39H1 expression in both 26 and 13 homozygous GV and no protein in the 44 WT GV **(Fig. 18A)**. SUV39H1 staining have been quantified, which confirmed this increase in the KO mouse lines compared to the control **(Fig. 18B)**. That being said, we also observed nucleoli staining in the 26 line and not in the 13 line, which explains the huge difference in the quantification. However, we suspect the nucleolus staining to be nonspecific. Indeed, the IF of 26 GV has been performed separately from the 13 and 44 GV with slightly different experimental conditions. This might be nonspecific signal from the secondary antibody. Alternatively, it is possibly an additional effect coming from off-target mutations. In any case, we need to perform the staining again, all three mouse lines at the same time. Nevertheless, we can clearly conclude an unusual expression of SUV39H1 protein in the GV of Suv39h1as homozygous KO mice. In the light of this finding, we wondered if the SUV39H1 expression had an impact on H3K9me2 and H3K9me3 levels (**Fig. 18C**). By IF realized on 12 to 42 GVs, we could see a clear increase of H3K9me3 level in both KO lines compared to control. For H3K9me2, no major differences could be observed. By quantifying the different staining, we confirmed H3K9me3 increase and no change in H3K9me2 level (**Fig. 18D**). Nevertheless, it should be stressed that the different quantifications have been normalized to Hoechst staining. Although it was confirmed to be similar inside the same experiment (between 13 and WT GVs), it is different for 26 GV and it is staining the heterochromatin which is our subject of study. Upon repeating the experiment, an additional H3 total staining would be a good control staining for normalization. In summary, the absence of Suv39h1as expression leads to the expression of a functional SUV39H1 protein that consequently propagates H3K9me3 but not H3K9me2 at the heterochromatin. Knowing that, it is quite surprising to have no functional impact on fertility. Therefore, we wondered if these different upregulations were conserved upon fertilization in the zygote.

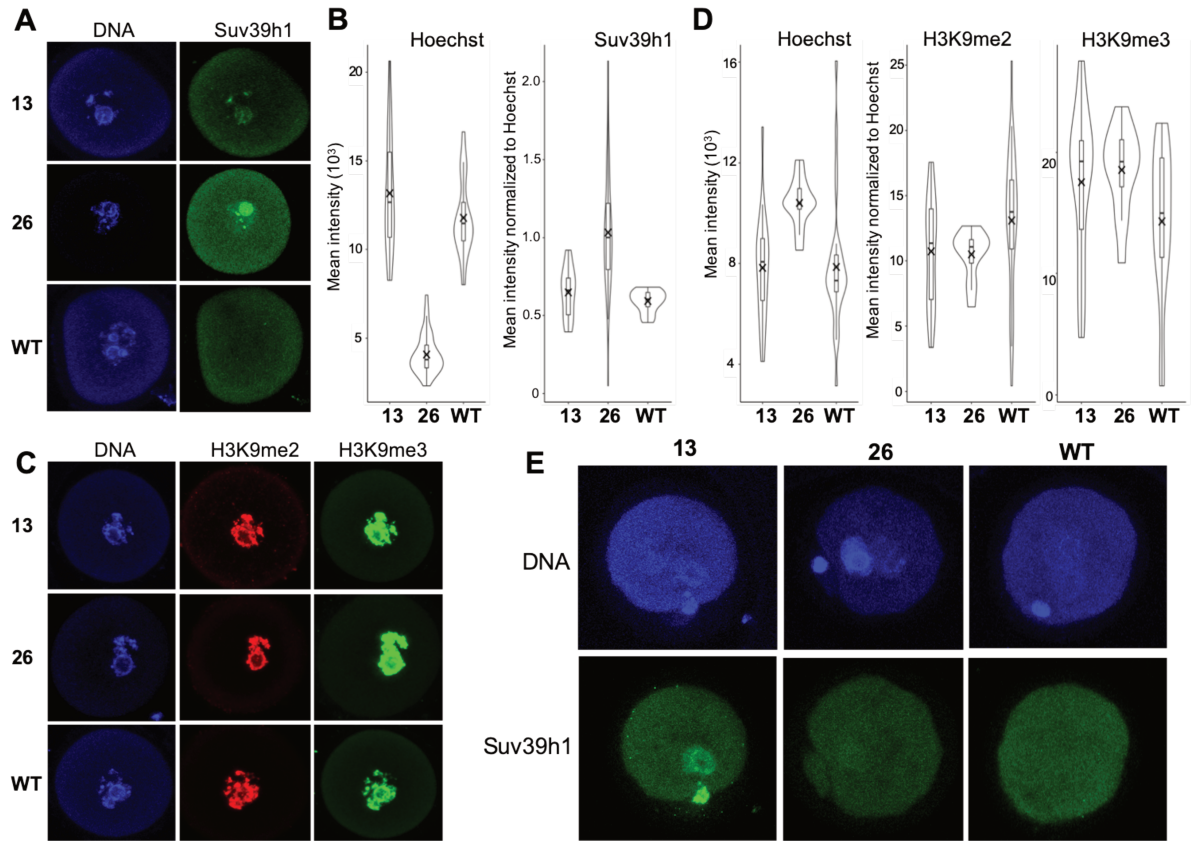


Figure 18: **SUV39H1 and H3K9me3 are upregulated in Suv39h1as KO GVs.**
(A) and **(C)** Representative pictures of SUV39H1, H3K9me2 or H3K9me3 IF on WT, 13 and 26 homozygous KO GVs. DNA is stained with Hoechst. GVs are at SN stage. SUV39H1 IF was performed in 17 GVs for WT, 27 for 13 and 34 for 26 GVs. H3K9me2 and H3K9me3 IFs were performed on 42 for WT, 32 for 13 and 12 for 26 GVs. **(B)** and **(D)** Violin plots showing the mean intensity, normalized or not, for Hoechst, SUV39H1, H3K9me2 and H3K9me3 staining in WT, 13 and 26 homozygous KO GVs. The cross indicates the mean value of the samples. **(E)** Representative pictures of SUV39H1 immunostaining on WT, 13 and 26 homozygous KO zygotes. DNA is stained with Hoechst. Zygotes are at PN5 stage. SUV39H1 IF was performed in 4 for WT, 8 for 13 and 2 for 26 zygotes.

To test that, we performed IF on zygotes obtained from the crossing of hormonally stimulated homozygous females 13 and 26 with a WT male. In parallel, a WT female was also crossed with a WT male. Zygotes were harvested twelve hours after mating and stained for SUV39H1 **(Fig. 18E)**. In zygotes at the PN5 stages, before the first mitosis, we can see a clear SUV39H1 upregulation in the female PN compared to control. However, this upregulation is not present in the 26-mouse line. This variability in SUV39H1 expression after the fecundation is surprising and may suggest additional off-target effects. Nonetheless, the fecundation rate was very low for 26 and 44 mouse line compared to 13 line, resulting in a small number of actual zygotes at the PN stage: 8 for the line 13, 2 for the line 26 and 4 for the WT. This experiment needs to be performed again, with additional immunostaining for H3K9me2 and H3K9me3.

Discussion

VI) Suv39h1as links H3K9 methylation to pluripotency through an unclear mechanism

1) Suv39h1as coordinates H3K9 methylation with pluripotency

We unraveled a genetic network linking pluripotency and H3K9 methylation through the regulation of an antisense lncRNA. This regulation is controlled by one of the core pTF OCT4 through the transcription of a lncRNA that will, in turn, regulates the coding gene expression Suv39h1. This is not the first model to link pTF with heterochromatin control. Heurtier et al. showed that Nanog blocks differentiation by sustaining H3K27me3 at developmental regulators[340]. The indirect regulation of SUV39H1 by OCT4 might have different advantages. First, the antisense transcription allows to maintain a minimal level of SUV39H1 expression without closing the locus, as illustrated by the loss of H3K4me1 and H3K4me2 along the gene body when Suv39h1as is not transcribed. The lncRNA establishes a local chromatin environment that will favor future Suv39h1 expression, upon differentiation for instance. Second, the indirect mode of regulation can also be seen as a buffering mechanism. Indeed, OCT4 is a core pTF whose level of expression is very tightly regulated to maintain pluripotency[93]. Therefore, any change in the quantity of protein would lead to cell differentiation. Nevertheless, intrinsic variations are unavoidable and it is important to buffer them. Regulating Suv39h1 through Suv39h1as transcription would then allow to really impact Suv39h1 expression only when OCT4 is downregulated definitively. Finally, the fact that Suv39h1 indirect regulation is operated through the transcription of a lncRNA is a quicker mechanism than going via the full process of protein production. It might be an important parameter to synchronize OCT4 loss with H3K9 methylation levels.

From the same perspective, the timing between OCT4 fluctuation and SUV39H1 expression is important to coordinate the H3K9 methylation level with the commitment to differentiation. Indeed, the cells lacking Suv39h1as expression are committing earlier resulting in an irreversible differentiated status. However, this acceleration has no major effect on the differentiation themselves. As already mentioned, the EBs assay is not appropriate since the high density of cell erase WT and A8/D8 differences in Suv39h1 expression levels. EpiLSCs differentiation, as a directed differentiation, might not be relevant for our case. However, these results are in line with the general very mild phenotype of the mutant cells and more widely, of the alteration of H3K9 methylation in mESCs[139,225,343,349]. Moreover, Montavon et al.[350] published very recently the study of immortalized mouse embryonic fibroblasts KO for 6

HMTs: Suv39h1, Suv39h1, Setdb1, Setdb2, G9a and GLP. They studied the impact of individual, paired and total KO of those genes on the heterochromatin status. They showed that any distinct gene pair of either Suv39h1/Suv39h2, SetDB1/SetDB2 or G9a/GLP maintain heterochromatin structure, confirming the functional redundancy of these different HMT. Nevertheless, it is important to notice that even when the 6 genes are KO, the collapse of heterochromatin is not present in 100% of cells, indicating most probably other compensation mechanism, such as DNA or H3K27 methylations. It might explain the mild phenotype we observed in our Suv39h1as KO cell lines.

2) A partially-elucidated mechanism of regulation

In the first section of results, we demonstrated that Suv39h1as regulated Suv39h1 through a cis-mechanism. We established a clear transcriptional component of the regulation, with a complex modulation of local chromatin. These effects are reminiscent of those associated with the regulation of Xist by its antisense Tsix[339]. That being said, several elements indicate that other mechanisms could be involved in the regulation. The presence of many Suv39h1 non-productive protein coding isoforms and the two overlapping Suv39h1as/Suv39h1 exons are a first indication. A second hint is the absence of a clear increase in pre-mRNA level and transcription initiation in absence of Suv39h1as expression. Finally, a third point is the intriguing observation in sm-FISH that Suv39h1as exonic signal never overlap with Suv39h1's one. Those facts raise two questions: (1) Is there additional mechanisms involved in the regulation? (2) Is there a transcriptional interference at the elongation step? In addition, it is known that a single lncRNA can act through several mechanisms[351].

First, we investigated potential trans-mechanism with ectopic Suv39h1as overexpression and siRNA assays. The different assays didn't modulate Suv39h1 expression. In addition, Suv39h1as didn't alter Suv39h1 mRNA stability. Consequently, we concluded that there is no trans-acting regulation of Suv39h1 expression. However, sm-FISH experiment proved that something was happening at the locus. Indeed, only intronic signal of Suv39h1as can overlap with Suv39h1 signal, which is surprising given the microscopic resolution. It is suggesting that both mature RNAs are processed after their transcription in an unknown way.

Then, we supposed that physical interaction between the sense and antisense RNA Pol II might interfere in their respective elongation. With distinct 5'end and 3'end of intronic probes, AS1 and AS2 respectively, we observed a slightly higher proportion of non-overlapping signals for AS2. This hypothesis can also be supported by the fact that Suv39h1 exhibits many

isoforms that are always modified on exon 3, which overlaps with Suv39h1as's one. By looking carefully at the RNA-seq reads spreading along Suv39h1 gene body, we might also postulate that isoforms 4 and 6 are a *de novo* assembly artefact. **Figure 19** is a schema presenting an hypothetical mechanism. An alteration of Suv39h1 RNA Pol II processivity by Suv39h1as transcription **(A)** or RNA **(B)** should lead to a variety of shorter versions of Suv39h1 pre-mRNA that would either be degraded via non-sense mediated decay for instance, or lead to non-productive isoforms as the isoforms 3, 4 and 6. Possibly, Suv39h1as might modulate Suv39h1 splicing **(C)**. Accordingly, the quantification by RT-qPCR of Isoform 3 shows a decrease in mutant cells line. However, how this alteration of RNA Pol II processivity would occur is still a fully open question. We can envisage 3 hypothesis:

(1) by a physical hinder between the two RNA Pol II **(A)** as it has been described in the yeast[323]?

(2) by a specific binding of the lncRNA by the common exons with Suv39h1 via base pairing (**B** and **C**) similarly to linc-ROR that acts as a sponge of miRNAs[319]?

(3) by the recruitment of an additional protein that would induce an early termination **(B)**, like WDR82[305]?

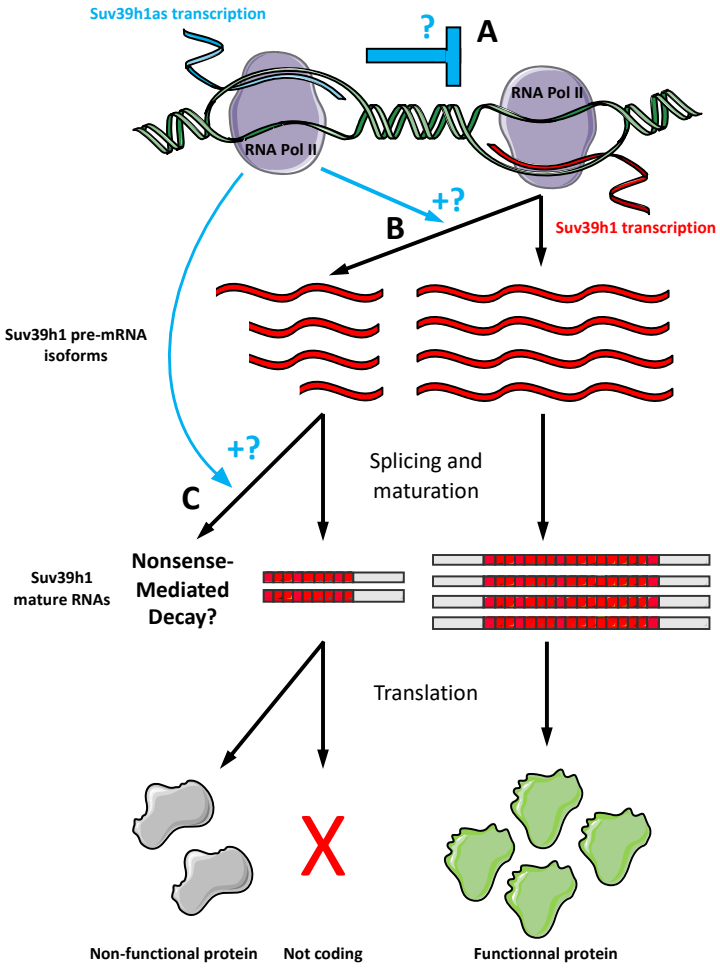


Figure 19: **Hypothetical Suv39h1 mechanism of regulation by Suv39h1as.**
In our hypothesis, Suv39h1as transcription impacts the production of different versions of Suv39h1 pre-mRNA. It can happen either by a transcriptional interference between the two RNA Pol II **(A)** or through the common exons between Suv39h1as and Suv39h1 **(B)**. The different versions the pre-mRNA would be either degraded through Nonsense-Mediated Decay or spliced. Possibly, Suv39h1as RNA might also influence Suv39h1 splicing **(C)**. Then the different isoforms would give either a functional protein (Isoform 1 and 2) or a non-functional protein or are not coding (Isoforms 3, 4, 5, 6 and 7).

Finally, a last mechanism can be proposed. The presence of isoforms missing in the complete exon 3 suggests a possible alternative splicing. This can occur similarly to erbAa lncRNA[327], where the RNA is masking splicing sites to modulate the production of isoforms. It could also occur through the establishment of a splicing site-specific chromatin signature, as was reported for *FGFR2* locus[318] .

Nevertheless, the different elements that are supporting this model are very subtly different between mutant and control cells, suggesting that the hypothetical mechanism is a "side effect" of the antisense transcription rather than a fully established mechanism. Still, it would be interesting to perform additional experiments to actually work out this question. First, it would be important to assess the coding potential of Suv39h1 isoforms and more precisely, the productivity of the resulting proteins. *In vitro* translation of the different isoforms[352], combined to *in vitro* methylation assays would allow to test the coding potential and the efficiency of the produced proteins[217]. After that, it would be interesting to perform a strand specific Transient-Transcription RNA-seq (TT-seq) on control and mutant cells[344] in order to sequence nascent RNAs and assess the production of shorter versions of Suv39h1 pre-mRNAs. In this scenario, we would expect a gradient of pre-mRNA in the WT from the 5' end to the 3' end that would decrease in the mutant cells.

VII) Suv39h1as regulates Suv39h1 in the oocyte but doesn't impact fertility

1) Suv39h1 expression in the oocyte and the zygote

Because of the re-analyzed RNA-seq and several studies demonstrating the importance of SUV39H1 in the early embryo, we postulate that Suv39h1as is regulating Suv39h1 in the mature oocyte and the early embryo. Therefore, to test its functional relevance *in vivo*, we created mice lines KO for Suv39h1as. We generated a deletion similar to that of mESC, with the deletion of both Suv39h1as promoters. After confirming Suv39h1as loss of expression, we examined the expression of SUV39H1 in the GV.

We did confirm an absence of SUV39H1 in WT GV as expected[285]. Both KO mouse lines were expressing SUV39H1, which proved that the observed upregulation in mESCs was also present in GV. This also validated our hypothesis that Suv39h1as was repressing Suv39h1 expression in the oocyte. Nevertheless, although the two KO mouse lines were exhibiting heterochromatin specific signal, the line 26 was also presenting a marking in the nucleolus. Except for technical consideration, it is interesting to notice that Zhang et al.[225] did observe a nucleolus marking of SUV39H1. Indeed, they overexpressed SUV39H1 in a strong way in HeLa cells and described a marking in the nucleolus, specifically of the overexpressed protein. In addition, they described an alteration of rDNA transcription in the embryo. From this perspective, if the nucleolus staining is specific, why is it not present in the line 13? Despite the exact same deletion, it might be due to a different level of overexpression. In any case, we need to perform the experiment again, with the 3 mouse lines together to confirm specificity of the nucleolus staining.

In a subsequent experiment, we also investigated the expression of SUV39H1 in the zygotes. By comparing zygotes at the PN stage, we observed SUV39H1 at the heterochromatin of the maternal PN with nothing at the paternal one. This staining is present specifically in the line 13, not in the WT and not in the line 26. A difference in Suv39h1 expression could be observed once more, even though this is strikingly opposite from the GV staining. Besides, the experiment was inefficient for the 26 and the WT lines. Although this experiment clearly needs to be replicated, we can nonetheless observe that Suv39h1 was still expressed in the zygotes of the line 13. That suggests that the overexpression is maintained despite the fecundation. It might be either *de novo* expressed, since some transcription has been reported to start as early as the first cell stage[353] or, most probably, maternally inherited as it is present in maternal PN only. It would have been interesting to compare the level of SUV39H1 expression to the one of

overexpressed SUV39H1 in Burton et al.[285] and Zhang et al.[225] studies, however, the quantifications were done on 2-cells embryos or later, which makes the comparison irrelevant. In addition, the ectopic expression is introduced in the zygote stage, in the maternal and paternal PN, another difference with our biological set up.

As for how Suv39h1 regulation by Suv39h1as happens *in vitro*, several mechanisms *in vivo* can be suggested. Suv39h1as transcription might establish a local chromatin context that would program later Suv39h1 transcription, similarly to what has been reported for Zdbf2 expression[354]. Alternatively, if we postulate that Suv39h1as is modulating Suv39h1 isoforms production by inducing the transcription of non-coding or defective proteins coding isoforms, we can hypothesize that mostly non-productive isoforms are transcribed and maternally inherited by the zygote. This would allow the locus to be ready to be actively and efficiently transcribed as soon as Suv39h1as is turned off.

Nevertheless, the comparison between *in vitro* and *in vivo* models has its limitation. In mESCs, we observed a 2 to 3 fold increase of Suv39h1 at the RNA and protein levels, while in oocytes, the increase is almost "infinite" since there is no detectable expression in control conditions. Moreover, if SUV39H1 is really present in the zygote, it would be interesting to know if it is maintained upon cleavages until the "real" expression at the 8-cells stage. Then, even at the stage, it would be evaluating the expression in comparison with the control. If SUV39H1 is not expressed in the zygote, it would mean that other mechanisms are present to insure the absence of SUV39H1 during the first developmental stages. It could be a role attributed at DCAF13 that starts to be expressed at 4-cells stage[225], after the ZGA, and could induce SUV39H1 degradation by ubiquitylation. It stresses the importance of heterochromatin regulation during the development with several mechanisms insuring the absence of SUV39H1 until the 8-cell stage[285]. Those different aspects also underlie differences between *in vitro* and *in vivo* pluripotencies, suggesting the implication of different mechanisms in HMTs regulation.

2) H3K9me2 and H3K9me3 levels in the oocytes and females fertility

Once we established the presence SUV39H1 in the oocytes and most certainly in the zygote, we wondered if this protein was functional and if it would lead to an increased level of H3K9 methylation. We did observe an increase of H3K9me3 level but not of H3K9me2 in both mutant mouse lines, which confirmed that overexpressed SUV39H1 was functional and active in late oocytes. The absence of H3K9me2 level modifications was surprising, and not totally in line with the results in mESCs, where both H3K9 marks are enhanced. A saturation effect might

explain it. Other HMTs are active in the oocyte, such as G9a[198], which has been proven to be important for H3K9me2 methylation. Therefore, H3K9me2 might already be maximal. Nevertheless, we did observe a marked increase of H3K9me3 level at the heterochromatin specifically, and it introduces the question of its impact on the development. Indeed, one would expect that the high level of H3K9me3 would make the oocyte incompetent for fertilization and subsequent development, as it is thought to be a barrier to reprogramming into pluripotency[290]. Thus, we need to perform H3K9me2 and H3K9me3 staining on early zygotes to assess whether the high level of H3K9me3 is maintained. It would also be interesting to investigate later stages of zygote to estimate how long SUV39H1 overexpression is persisting, as well as its impact on H3K9me3 level.

Furthermore, fertility tests demonstrated that Suv39h1as loss of expression doesn't impact the global development and females' fertility. Given the increase of H3K9me3 level, it suggests that either higher levels of H3K9me3 have no impact on fertility and development, being in complete opposition to Zhang et al.[225] and Burton et al.[285] studies, or, most probably, that a normal level is re-established after fertilization. If we suppose that H3K9me3 increase is maintained upon fertilization, it would be interesting to investigate where the deposition is happening. It might be at the pericentromeric and centromeric regions, since this is where SUV39H1 is reported to be the most active normally[216]. However, it is now known that their expression in important for a proper development[355], and it would suggest that it happens at other regions as genes promoters, already shut down. Otherwise, if we speculate that other mechanism are re-establishing after fertilization, it might be due to a low level of SUV39H1 quantity. Indeed, if SUV39H1 is maternally inherited, it is diluted after the fertilization and even further after the first division. A low level of SUV39H1 might not be enough to enhance H3K9me3 level at the maternal genome and even more at the paternal genome. Alternatively, compensation mechanisms can also be present, with active H3K9 demethylases that would erase the excess of H3K9me3. It has been proven that the H3K9 demethylase KDM4A is important for the oocyte development and the maternal-to-zygotic transition, mainly through the specific demethylation of H3K9me3[356]. In any case, it illustrates differences between *in vivo* and *in vitro* heterochromatin impact on pluripotency. In mESCs, heterochromatin is more plastic, allowing fluctuations in some extent, while in at the early stages of development, the heterochromatin in more tightly regulated with alternative mechanisms to regulate SUV39H1 expression and H3K9me3 levels.

Besides, our hypothesis that a lncRNA expressed in the oocyte might be important for oocyte and/or embryo development is not a totally innovative idea. Indeed, a very recent

study[357] identified a lncRNA, called Rose, important for oocyte and embryo development. Some of lncRNA Rose isoforms are specifically expressed in the growing oocyte. It is a fully overlapped antisense lncRNA by the coding gene Transmembrane protein 108 (Tmem108). Therefore, it presents similarities to Suv39h1as which has oocyte-specific isoforms and is an antisense too. Nevertheless, in contrast to Suv39h1as, Rose has a clear role in oocyte cytokinesis and early embryo development.

In conclusion, we propose a simple genetic network to link the pluripotency dynamic to the regulation of the heterochromatin. The antisense lncRNA Suv39h1as regulates Suv39h1 expression *in vitro* and *in vivo*. This regulation is OCT4-dependent in mESCs. OCT4 is abundant in the mature oocyte and only reactivated prior the 8-cells stage in the embryo, correlating well with Suv39h1as *in vivo* and supporting the idea that it is also associated with the pluripotency. Its functional role in mESCs is very mild, mostly consisting in coordinating heterochromatin status with commitment into differentiation. *In vivo*, it seems to have no functional importance for oocyte maturation and early development, joining the long list of lncRNA that have no clear function. Nevertheless, the mechanism of regulation seems to be complex and unusual even though it has not been fully elucidated.

Appendices

Index of abbreviations

(p)TF: pluripotent Transcription Factor
3D: 3 dimensional
cHC: constitutive heterochromatin
ChIP: Chromatin ImmunoPrecipitation
CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats
DAPI: 4',6-Diamidino-2-Phenylindole
dn: *Suv39h1*/*Suv39h2* double knock-out
DNMTs: DNA MethylTransferases
E(var): enhancer of variegation
EBs: embryoid bodies-like structures
EPI: Epiblast
EpiLC: Epiblast-like Stem Cells
EpiSC: Epiblast Stem Cell
eRNA: enhancer RNA
ERV: Endogenous Retroviruse
EuC: Euchromatin
FCS/L: ESC culture media containing Fetal Calf Serum and LIF
fHC: facultative heterochromatin
gRNA: guide RNA
GV: Germinal Vesicle
HC: heterochromatin
HDM/KDM: Histone/Lysine DeMethylase
HMT/KMT: Histone/Lysine MehtylTransferase
HP1: Heterochromatin Protein 1
HxKyac: acetylation of lysine y on Histone x
HxKymeZ: Z-methylation of lysine y on Histone x
HxRymeZ: Z-methylation of argenine y on Histone x
HxSyph: phosphorylation of serine y on Histone x
HxTyph: phosphorylation of tyrosine y on Histone x
ICM: Inner Cell Mass
IF: Immunofluorescence
KO/KD: Knock-Out/Konck-Down
LIF: Leukemia Inhibitory Factor
LINE: Long Interspersed Element
lncRNA: long non-coding RNAs
LTRs: Long Terminal Repeats
mESC: mouse Embryonic Stem Cells
miRNA: micro RNAs
ncRNA: non-coding RNA
PEV: Position effect variegation
PGC: Primordial Germ Cell
piRNA: piwi RNA
PMF: Primordial Follicles
PN: Pro-Nuclei
PRC2: Polycomb Repressive Complex 2
PrE: Primitive Endoderm
PTM: Post-Transcriptional Modification
rDNA/RNA: ribosomal DNA/RNA
RT: Reverse Transcription
SAM: S-adenyl methionine
SINE: Short Interspersed Element
siRNA: small interference RNA
smFISH: single molecule Fluorescence In Situ Hybridization
Su(var): Suppressor of variegation
TAD: Topologically Associated Domains
Te: transposable elements
TE: Trophectoderm
TET: Ten-Eleven Translocation methylcytosine dioxygenase
tpm/rpm: Transcript/Read per million
Transfer RNA: tRNA
T(S)S: transcription (start) sites
UTR: untranslated region
WT: Wild Type
Xist: X inactivation specific transcript
ZGA: Zygotic Genome Activation

Supplementary methods of IV)

OCT4 activates a *Suv39h1*-repressive antisense lncRNA to couple histone H3 Lysine 9 methylation to pluripotency

Laure D. Bernard et al.

Cell lines and generation of A8 and D8 Suv39h1as mutant cells.

WT cells in this study are E14Tg2a ES cells, from which all mutant cells were derived. Dox-inducible KO cells have been described before (*Esrrb*: EKOiE[23], *Oct4*: Zhbtc4[24] and *Nanog*: 44iN[25]). To generate ES cells deleted for the Suv39h1as promoter, several gRNAs were designed using several online resources (sam.genome-engineering.org, crispr.mit.edu, chopchop.cbu.uib.no). Highly ranked gRNAs for off-targets as well as on-target predictions score were selected. The gRNAs were cloned into plasmids containing a U6-promoter driven expression cassette along with a puromycin selection cassette or Cas9 enzyme linked to a mCherry reporter (addgene 51133 and 64324 respectively, see gRNA sequences on Table S2). E14Tg2a cells were lipofected (Lipofectamin 2000; Invitrogen, 11668-019) with 1 μg of the plasmid containing the right deletion site guide and the Cas9 cassette and 3 μg of the plasmid containing the left deletion site guide and a Puromycin cassette. Puromycin (1 μg/mL – Sigma, P9620-10ml) selection was performed for 3 days and mCherry fluorescence checked by microscopy. Resistant cells were plated at clonal density and colonies picked and expanded. Genomic DNA was isolated with NucleoSpin Tissue DNA extraction Kit (Macherey-Nagel, 740952.50), and screened by qPCR and PCR. PCR was performed with LongAmp Taq PCR kit (BioLabs, E5200S) following manufactory's instructions and sequenced. qPCR and PCR primers are available in Table S2. Two KO clones, A8 and D8, were selected for further analyses of *Suv39h1* expression. The selected clones were checked by PCR on genomic DNA with primers indicated on Table S2 and with LongAmp Taq PCR kit (BioLabs, E5200S) or Q5 High Fidelity (NEB, M0491S) following manufactory's instructions. PCR product was validated by agarose gel migration and purified (NucleoSpin Gel and PCR Clean-Up; Macherey-Nagel, 740609.50). PCR products were purified and cloned (ThermoFisher, cat n°450245) as indicated by the manufacturer. Ten bacterial colonies were picked, expanded and checked by PCR and sequencing. The precisely expected deletion was observed for A8 (5.5kb); D8 showed a 5.7kb deletion exhibiting a shift of 700bp compared to A8 but encompassing both *Suv39h1as* promoters. Both clones were karyotyped, with A8 exhibiting a normal karyotype and D8 presenting 50% of cells with an extra chromosome.

Regular cell culture.

Cells were cultured at 37°C, 7% CO2 on 0.1% gelatin-coated plates (SIGMA, G1890-100G) in FCS base [DMEM+GlutaMax-I (Gibco, 31966-021), 10% FCS (Sigma F7524), 100µM 2-mercaptoethanol (Gibco, 31350-010), 1X MEM non-essential amino acids (Gibco, 1140-035)], supplemented with 10 ng/mL recombinant Leukemia Inhibitory Factor (MILTENYI BIOTEC, 130-099-895), named then FCS/LIF. In 2i/LIF medium, cells were grown in N2B27 base [0.5X DMEM/F-12 (Life Technologies, 31331-093), 0.5X Neurobasal (Life Technologies, 21103-049), 0.5X N2 (Life Technologies, 17502-048), 0.5X B27 (Life Technologies, 17504-044), Insulin 10 µg/mL (Sigma, I-1882), 2mM L-Glutamin (Life Technologies, 25030-024), 37.5 µg/mL BSA (Sigma, A3311-10G) and 0.1mM 2-mercaptoethanol (Life Technologies, 31350010)], supplemented with 10ng/mL recombinant LIF (MILTENY BIOTEC, 130-099-895), 1µM PD0325901 (Axon 1408) and 3µM CHIRON99021 (Axon 1386). Doxycycline (1 µg/mL – D3072, Sigma), Puromycin (1 µg/mL – Sigma, P9620-10ml) or Flavopiridol (400 nM – Selleckchem, S2679), were extemporaneously added, as indicated. Cells were passaged every 2-3 days, when they reached 70-80% confluency.

Differentiation of ES cells.

Cells cultured in FCS/LIF were differentiated by seeding 300000 cells per well of gelatin-coated 6-wells plate and withdrawing LIF. N2B27 and EpiLC differentiation assays were performed with cells cultured in 2i/LIF for a minimum of 3 passages (9 days). For N2B27 differentiation, 50000 cells per well were seeded on wells of 6-wells plates coated overnight with poly-L-ornithine 0.01% (Sigma, Cat# P4957) at 37°C and 2h with 1X laminin (Sigma, Cat# L2020) and LIF, PD0325901 and CHIRON99021 were withdrawn. For EpiLC differentiation, 230000 cells were seeded per well of 6 wells plates coated with human plasma fibronectin (10 µg/well – F2006, Sigma) and cultured in N2B27 medium containing activin A (20 ng/mL – 338-AC-010, R&D Systems), rhFGF (12 ng/mL – 233-FB, R&Dsystems), and KSR (1% – 10-828-010, Gibco).

Clonal and differentiation commitment assays.

For comparing self-renewal and differentiation capacity of WT and *Suv39h1as* clones, 600 cells were plated in gelatin-coated wells of a six well plate. Cells were cultured for 7 days in the indicated media and stained for alkaline phosphatase activity (Sigma, cat. 86R-1KT), following the manufacturer's instructions. Colonies were scored as undifferentiated, mixed and

differentiated using a stereo-microscope (NIKON-SMZ1500). For commitment assays, 600 cells obtained every day of differentiation in N2B27 were plated in poly-L-ornithine/laminin-coated wells of a 6 well plate, cultured for 7 days in 2i/LIF and stained for alkaline phosphatase activity.

Assessment of RNA half-lives.

One million cells were plated in a single well of a 6-well plate and treated the next day with Flavopiridol (400 nM – S2679, Selleckchem) for the indicated times. All samples were harvested at the end of the assay for RNA extraction and RT-qPCR analysis.

Oct4 knock-down.

Oct4 knock-downs were performed with siRNAs (Dharmacon, ON-TARGETplus Mouse Pou5f1 siRNA – L-046256-00-0005) and compared to untergated control siRNAs (On Target plus control pool D-001810-10-05). Cells were nucleofected with 200 pmol of siRNA using a Mouse ES Cell Nucleofector kit (Lonza, VPH-1001, program A30) and cultured for 24h.

Reverse transcription and real time polymerase chain reaction (RT-qPCR).

Total RNAs were isolated with Trizol (Invitrogen, 15596026) according to the manufacture's protocol and absence of DNA contamination was ensured by additional DNase I digestion (Qiagen, 79254). Reverse transcription (RT) reactions were performed with random hexamers on 0,5 – 2 µg of total RNA according to manufacturer's instructions (First Strand cDNA Kit, Roche, 04379012001). Real-time PCR reactions were performed in duplicate in 384-well plates with a 480 Light Cycler (Roche) using Light Cycler 480 SYBR Green I Master Mix (Roche, 04707516001) and qPCR primers at 0.4 µM final concentration. qPCR primers sequences are listed in Table S2. Standard and melting curves were generated to verify the amplification efficiencies (> 85%) and the production of single amplicons. Relative DNA amount was obtained from Cp (Crossing point) calculated from the second derivative of the DNA amplification signal over time. For RT-qPCR experiments, values for gene expression were normalized to the levels of housekeeping gene *Tbp* mRNA. Assays to measure RNA half-lives were normalized to the levels of *28s* mRNA.

RNA-seq analysis and annotation of Suv39h1as.

Poly-A selected RNAs were sequenced by Novogene Ltd (stranded, PE150) and reads aligned to the mm10 genome using STAR[44], quantified by RSEM[45] and counts transformed into

Transcripts per million (TPM, Table S1). To annotate *Suv39h1as*, reads were mapped to the genome using Hisat2[46]. Bam files were used to build new transcript models with Stringtie[47] based on the Gencode (vM12) gtf (default parameters except -m 300). The resulting gtf files were merged (default parameters, except -m 300 -c 0.5 -F 0.5 -f 0.05), and all *Suv39h1as* isoforms annotated. Only those experimentally validated by Topo-cloning of poly-A selected cDNAs and RT-qPCR quantifications were retained. *Suv39h1as* isoforms 1 and 2, as reported in Fig.1, were validated by Topo-cloning and sequenced with primers Suv-as_ex1a-1-F/Suv-as_ex4-3-R and Suv-as_ex4s-1-R, respectively. Isoform 3 was validated with primers Suv-as_ex1b2-F/Suv-as_ex4-3-R. Splicing events were also validated by RT-qPCR: splicing from promoter 1a (Isoforms 1 and 2) and promoter 1b (Isoform 3) with exon 2, with primers Suv-as_ex1b2-F/Suv-as_ex1b2-R and Suv-as_ex1a2-F/Suv-as_ex1a2-R, respectively, and splicing from exon 2 and 3 with primers Suv-as_ex23-F/Suv-as_ex23-R. All primer sequences are available in Table S2.

Chromatin immunoprecipitation (ChIP).

Ten million cells were crosslinked either for 45 min with DSG 1X plus 10 min in Formaldehyde (FA) 1% in PBS 1X for TF binding analysis or only for 10 min in FA 1% for histone modification analysis. Formaldehyde was quenched with 125 mM glycine for 5 min at room temperature. Nuclei were prepared in 1mL of ice-cold swelling buffer (25 mM Hepes pH7.95, 10 mM KCl, 10 mM EDTA) freshly supplemented with 1X protease inhibitor cocktail (PIC - Roche, Cat# 04 693 116 001) and 0.5% IGEPAL (Sigma, Cat#I8896) for 20 min on ice and 50 passes in a dounce homogenizer. Nuclei were then resuspended in ice-cold D3 buffer (0.1% SDS, 15 mM Tris pH 7.6, 1 mM EDTA), freshly supplemented with 1X PIC, and sonicated using a Covaris M220-Setpoint at 6°C and 10 (FA only) to 15 (DSG/FA) cycles with the following parameters for each cycle: 60 sec duration, peak power of 67W, duty factor of 15% and cycles/burst of 500. A delay of 45 sec is added at the end of each cycle. Result of the average power is 10W. After centrifugation (15 min, 14000 rpm, 4°C), the supernatant was stored at −80 °C until use. 20 µL were used to quantify the chromatin concentration and check DNA size (typically 200-600 bp for Covaris). 15 to 20 µg of chromatin were used for each ChIP after pre-clearing it for 1.5 hours rotating on-wheel at 4 °C in 1 mL of TSE150 (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH8, 150 mM NaCl) buffer containing 50 µL of protein G Sepharose beads (Active Motif, Cat#37499) 50% slurry, previously blocked with BSA (0.5 mg/mL; Roche, Cat# 10711454001) and yeast tRNA (1 µg/mL; Roche Cat# 10109495001). Immunoprecipitations were performed overnight rotating on-wheel at 4°C in

500 µL of TSE150. 20 µL were set apart for input DNA extraction and precipitation. 50 µL of blocked protein G beads 50% slurry was added for 2h rotating on-wheel at 4°C. Beads were pelleted and washed for 5 min rotating on-wheel at room temperature with 1 mL of buffer in the following order: 2X TSE150, 1X TSE500 (as TSE150 but 500 mM NaCl), 1X washing buffer (10 mM Tris-HCl pH8, 0.25M LiCl, 0.5% IGEPAL, 0.5% Na-deoxycholate, 1 mM EDTA), and 2X TE (10 mM Tris-HCl pH8, 1 mM EDTA). Elution was performed in 100 µL of elution buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8) for 15 min at 65°C after vigorous vortexing. Eluates were collected after centrifugation and beads rinsed in 150 µL of TE-1%SDS. After centrifugation, the supernatant was pooled with the corresponding first eluate. For both immunoprecipitated and input chromatin, the crosslinking was reversed overnight at 65°C, followed by proteinase K treatment, phenol/chloroform extraction and ethanol precipitation. Input and IP samples were analysed by qPCR using primers in Table S2. The 2dCt method was used. All values were corrected to the input. The antibodies used and their working dilution are indicated in Table S2.

RNA/DNA Fluorescent In Situ Hybridization (FISH).

Using the online tool "Stellaris RNA FISH probes designer" from LGB biosearch laboratories, single-strand probes for *Suv39h1* (47 oligos, 30 in exons, see sequences Table S2) and *Suv39h1as* (35 exonic oligos, see sequences Table S2) were designed for single-molecule FISH (smFISH). DNA probes for DNA-FISH were generated by nick translation (Vysis Nick Translation Kit; Abbott, cat. 32-801300) using a fosmid clone (WIBR1-2188H11 – from Children's Hospital Oakland Research Institute, bacpac.chori.org) covering the entire locus. Cells were fixed in Formaldehyde 4%, quenched with Glycine (1M) and cytospun (Cytospin3, Shandon, at 400 rpm for 5 min with a low acceleration) onto slides that were kept in Ethanol 70% at 4°C until use. Slides were washed in 100% ethanol for 2 min and air dried. For each spot, 10 µL of hybridization cocktail (SSC2X – S6639-1L, Sigma; Dextran – Life technologies, 5%; Formamide 10% – Sigma F9037; 2 µg/µL E. Coli tRNAs – Sigma 10109541001; 5 mmol/L Ribonucleoside Vanadyl Complex – NEB S1402S; 0,5 µg/µL BSA – NEB B9001S) and smFISH probes for *Suv39h1* and *Suv39h1as* (each at 0,6 µmol/L) were used for overnight incubation into a humid chamber at 37°C. Slides were washed in fresh SSC 2X/Formamide 10% for 30 min at 37°C, mounted and counterstained with Vectashield Antifade Mounting medium with DAPI (Vector Laboratories, H-1200-10). Sm-FISH images were acquired with an inverted Nikon Eclipse X microscope equipped with: X63 oil immersion objective (N.A1.4); LUMENCOR excitation diodes; Hamamatsu ORCA-Flash 4.0LT camera; NIS Elements 4.3

software. The position of each image was recorded on the microscope. Subsequently, the coverslips were removed and the slides washed 3 times in washing medium (4X SSC, 0,2% Tween-20) at 37°C, and treated with RNaseA 10U/mL (Invitrogen, cat. EN0531) in 2XSSC at 37°C for 1h. DNA denaturation was performed in 50% formamide/2XSSC at 80°C for 30 min. Slides were dehydrated in cold ethanol and hybridized overnight in a 50% Formamide/ 2X Hybridization cocktail (4X SSC – Sigma, S6639-1L; 20% Dextran sulfate – Life technologies; 2 mg/mL Bovine Serum Albumine –NEB B9001S; and 40 mM Ribonucleoside Vanadyl Complex – NEB S1402S) at 37°C with 0.3 ng of DNA-Fish probe, 3 µL of mouse Cot1 DNA (Invitrogen, cat 18440016) and salmon sperm DNA (Invitrogen, cat.15632011), previously denatured in Formamide (7 min at 75°C). After overnight hybridization of the probes, the slides were washed 3 times in 50% Formamide/2X SSC buffer at 37°C for 5 min and 3 times in 2XSSC buffer at 37°C for 5 min, mounted with Vectashield containing DAPI and imaging on the previous recorded positions.

Western-blot.

Cell were lysed in in Laemmli buffer (1 000 000 cells per µL ; #1610747, BioRad) at 95°C for 5 min and samples run in a mini-PROTEAN® TGX Stain-Free Precast Gel (Bio-Rad, 456-8086) in 25 mM Tris, 0,21 M Glycine, 50% SDS at 120V with constant voltage and transferred onto nitrocellulose membranes (Life Science, 10600003) in a 25mM Tris, 0.21 M glycine, 20% ethanol solution. The membrane was blocked in phosphate-buffered saline (PBS; 0.8 g/L NaCl, 0,02 g/L KCl, 0.144 g/L Na2HPO4, 0.024 g/L KH2PO4, pH = 7.2) with 0.1% Tween (PBST), 5% Bovin Serum Albumin (BSA) for 1 h and incubated in 3 ml of PBST 5% BSA with different antibodies listed Table S2, overnight at 4°C. Membranes were washed 3 times for 5 min in PBST and incubated with Pierce® goat anti-rabbit IgG-HRP conjugated secondary antibody (Thermo Scientific #314666, 0,1 µg/mL or 50 ng/mL). Membranes were washed in PBST and developed using a Pierce® ECL Western Blotting substrate kit (Thermo Scientific, #32109) or Pierce® ECL plus Western Blotting substrate kit (Thermo Scientific, #32134) for 5 min or 1 min at RT and luminescence detected using a ChemiDoc MP Imaging Systems with Image LabTMTouch Software Version 2.2.0.08.

Immunostainings.

Cells were trypsinized, counted and resuspended at 1 million/mL in FCS free medium (DMEM-Glutamax/100 mM 2-mercaptoethanol/NEAA 1X) into sterile 1.5 mL Eppendorf tubes. To ensure direct comparison of WT and mutant cells, they were then individually incubated either

with 10 µM Rhodamine Red dye (InVitrogen, Cat#CMTPX C34552) or 1µM Deep Red dye (InVitrogen , Cat#C34565) for 20 to 40 min at 37°C. The labeled cells were then collected by centrifugation, washed with PBS1X, resuspended in DMEM/10%FCS/LIF medium and mixed at a 1:1 ratio for Rhodamine and Deep Red labelled cells (usually ~0.4M each). 800 000 mixed cells were seeded onto Poly-L-Ornithine/Laminin coated single wells of a µ-slide 4 well Ph+ibiTreat (Ibidi GmbH Ref#80446) and incubated for 6H at 37°C and 7% CO2. Cells were then fixed directly into the well with freshly prepared PFA 4% (Fisher Scientific, Cat#16431755) for 10 min at room temperature in the dark and washed twice in PBS1X for 10 min. Cells were permeabilized with PBS1X/0.1% Triton X-100 (Sigma, Cat#T8787) for 10 min at room temperature. After three washes with PBS 1X, cells were blocked with PBS 1X/3% Donkey Serum (Sigma, Cat#D9663) for 30 min in the dark and incubated overnight with primary antibodies (diluted in PBS 1X/10% DS). Following three washes with PBS, 1h incubation with secondary antibodies at room temperature in the dark and 3 washes with PBS 1X, nuclei were counterstained with DAPI (Sigma, Cat#D9542), washed in PBS1X and imaged with an inverted Nikon Eclipse X microscope equipped with: X20/0.45 (WD 8.2-6.9) objective; LUMENCOR excitation diodes; Hamamatsu ORCA-Flash 4.0LT camera; NIS Elements 4.3 software. Quantifications were performed using Cell Profiler[48]. For each experiment, Rhodamine/Deep Red labelled cells were attributed using the FlowJo software. For each experiment, the fluorescence intensity of mutant cells was normalised to the median of the corresponding wild-type cells intensities imaged on the same spot. The data was plotted using the ggplot2 package[49] in R.

Statistical analyses

Statistical significance of gene expression and chromatin immunoprecipitation differences measured by qPCR were assessed with Student's t-test. For the analysis of the locus-wide effects of the loss of *Suv39h1as* on H3K4me1 and me2, all values obtained with primer pairs located between coordinates +6 and +15 were considered; for the effects measured at the promoter region, the position showing the highest difference for each histone mark was used. Single molecule FISH data on RNA counts per cell was evaluated with Mann-Whitney-Wilcoxon tests and differences in transcription frequency with a Chi-squared test. Population differences observed with immunofluorescence were assessed with Kolmogorov-Smirnov tests. Differences in clonal assays were evaluated with Mann-Whitney-Wilcoxon tests using raw colony counts.

Additional references

44. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
45. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323 (2011).
46. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37, 907–915 (2019).
47. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 33, 290–295 (2015).
48. Carpenter, A. E. et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol 7, R100 (2006).
49. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. (Springer-Verlag, 2009). doi:10.1007/978-0-387-98141-3.

Materials and Methods tables

Table 3: **List of primers and gRNAs.**

Primer type	Experiment	oligo name	Sequence
ChIP primers	ChIP-primers	c3-F	AACTACGGGCTGAACATGTG
	ChIP-primers	c3-R	TGCCCAATCCCTGTGTTTAT
	ChIP-primers	c5-F	GGGTTCATTGGCAGAGTGAT
	ChIP-primers	c5-R	CGCTGGGAGTTCTTAGCTGT
	ChIP-primers	c6-F	TGTTAGTGGGGACATTCTTCG
	ChIP-primers	c6-R	AACTGGGCAAGAGCTGATGT
	ChIP-primers	c7-F	TGTTGTGGGGTTTTGTTTCA
	ChIP-primers	c7-R	CAGAGCAAAAGTGGGGAAGA
	ChIP-primers	c7a-F	CACGTGCTCTAGCCCATTTT
	ChIP-primers	c7a-R	CAATGAGAGAGCAGCCATGA
	ChIP-primers	c8-F	CATTCCTGAGACTGGAGATGC
	ChIP-primers	c8-R	GACCTCAAAACTTGCCCCTAC
	ChIP-primers	c8b-F	TGTAGGGGCAAGTTTTGAGG
	ChIP-primers	c8b-R	GAGGAGCTGAGGGGAGAGTT
	ChIP-primers	c8a-F	TCCTTGTACCTTGCCTTTGG
	ChIP-primers	c8a-R	CTCAGGAGGTGGAACTCTGC
	ChIP-primers	c10-F	CTCCAACTCAAGGAACCTTC
	ChIP-primers	c10-R	GCTTGGGTCTGTTTTTCAGC
	ChIP-primers	c12-F	TCCTGGTAGGGGAGGTCTCT
	ChIP-primers	c12-R	GCCCATAGACCCACAAAGAA
	ChIP-primers	c13-F	ATCCCAAGACACAGGAGTGC
	ChIP-primers	c13-R	TTCTGCATTTTCCAGCACAG
	ChIP-primers	c14-F	TCTTCTGACCCGGAGAAATG
	ChIP-primers	c14-R	GGCCAGGAATACCAGAAACA
	ChIP-primers	c27-F	GCCCAGATTTGAGGTTGAAA
	ChIP-primers	c27-R	CACAGCCCAGAGCTTAGACC
	ChIP-primers	c28-F	GGAAGTATGGGCAGGAAACA
	ChIP-primers	c28-R	GGCAATGTCTAGGCCAAAAA
	ChIP-primers	c29-F	GCCCACTATCTGCCAAGTGT
	ChIP-primers	c29-R	TGCCCCTCCTGACTGTAATC
	ChIP-primers	c30-F	CAAGGAGAAACTCGCTGACC
	ChIP-primers	c30-R	CTGAGTCCCCTTTCCATTCA
	ChIP-primers	c32-F	CAGCAATGGTGGCACATATC
	ChIP-primers	c32-R	CCATGTGGCCTCTGAAAAGT
	ChIP-primers	c15-F	AGAGTGGAGGAGCAAGGTCA
	ChIP-primers	c15-R	TGAATGATTTTGGGCAGTCA
	ChIP-primers	c16-F	ACTACACGGTTTGGGCAGTC
	ChIP-primers	c16-R	CAGGTAGCTGTTGGCTGTGA
	ChIP-primers	c25-F	GGTCCCCTGCCAGTGTAGTA
	ChIP-primers	c25-R	CCTGCCCTTGGTGTTTCTAA
	ChIP-primers	c33-F	GGTCCCCTGCCAGTGTAGTA
	ChIP-primers	c33-R	CCTGCCCTTGGTGTTTCTAA
	ChIP-primers	c34-F	CAATTCTCAGGGCAACACCT
	ChIP-primers	c34-R	TTGCCATTCCAGAGTTAGGG
	ChIP-primers	c24-F	CTCGGGGGTGAAAAGTACAA
	ChIP-primers	c24-R	GAAGGGATGTGCTCTGAAGC
	ChIP-primers	c21-F	GGTCCGTTCCACTTCTTCCT
	ChIP-primers	c21-R	GCTCTCAGTGACCGTTACCC
	ChIP-primers	c22-F	GCCAGTAGCTCGCGTCTCTA
	ChIP-primers	c22-R	GCCCGAATGTTGTTAGCTGT
	ChIP-primers	c23-F	CGCCCCTTGTCAGTCATATT
	ChIP-primers	c23-R	AGAATTGGGGTTGCACAAAA

Primer type	Experiment	oligo name	Sequence
House keeping genes	ChIP-primers	Actin Promoter-f	CCGTTCCGAAAGTTGCCTT
	ChIP-primers	Actin Promoter-r	CGCCGCCGGGTTTTATA
	RT-primers	Tbp-f	GGGGAGCTGTGATGTGAAGT
	RT-primers	Tbp-r	CCAGGAAATAATTCTGGCTCA
	RT-primers	28S-f	GAAGGCAAGATGGGTCACCA
	RT-primers	28S-r	GAACTTCCGTGGGTGACTCC
Locus gene expression	RT-primers	Suv_int1-F	CCTCCCATCATACGACTCAAA
	RT-primers	Suv_int1-R	GTATTTGCAGGCCACTTCTTG
	RT-primers	Suv_ex23-F	GGCTAGGTTTGGGTCCAGAT
	RT-primers	Suv_ex23-R	CCTGCCCTTGGTGTTTCTAA
	RT-primers	Suv_ex34-F	CCACGTAGTCCAGGTCAAAGA
	RT-primers	Suv_ex34-R	GACTGCCCAAACCGTGTAGT
	RT-primers	Suv-as_int1-F	TCTTCTGACCCGGAGAAATG
	RT-primers	Suv-as_int1-R	GGCCAGGAATACCAGAAACA
	RT-primers	Suv-as_ex23-F	AGGAGACTTTGGGATCTGGAA
	RT-primers	Suv-as_ex23-R	CGCTACGATCTCTGCATCTTC
	RT-primers	Suv-as_ex1b2-F	AGTACCACATCTGCCAAACG
	RT-primers	Suv-as_ex1*2-R1	TTCCAGATCCCAAAGTCTCC
	RT-primers	Suv-as_ex1a2-F	TTTCCGACAACACAGATGGA
	RT-primers	Suv-as_ex1*2-R	TTCCAGATCCCAAAGTCTCC
	RT-primers	Suv-as_int2-1-F	GTGCTTTACCCAGGTTCCAA
	RT-primers	Suv-as_int2-1-R	CCAAATCCCAGGACACAGAT
Pluripotent markers	RT-primers	Nanog-f	AGGATGAAGTGCAAGCGGTG
	RT-primers	Nanog-r	TGCTGAGCCCTTCTGAATCAG
	RT-primers	Oct4-f	CCCCAATGCCGTGAAGTTG
	RT-primers	Oct4-r	TCAGCAGCTTGGCAAACTGTT
	RT-primers	Sox2-f	CACAGATGCAACCGATGCA
	RT-primers	Sox2-r	GGTGCCCTGCTGCGAGTA
	RT-primers	Klf4-f	CAAGTCCCCTCTCTCCATTATCAAGAG
	RT-primers	Klf4-r	CCACTACGTGGGATTTAAAAGTGCCTC
	RT-primers	Esrrb_1 F	CCAACCAGAAGTGGGTCTTGTTCCT
	RT-primers	Esrrb_1 R	TGTGGAAGGATCCTGGGACACAGAT
Differentiation markers	RT-primers	Rex1-f	CAGCTCCTGCACACAGAAGA
	RT-primers	Rex1-r	ACTGATCCGCAAACACCTG
	RT-primers	Tbra-F	GTGACTGCCTACCAGAATGA
	RT-primers	Tbra-R	ATTGTCCGCATAGGTTGGAG
	RT-primers	Prdm14_F	ACAGCCAAGCAATTTGCACTAC
	RT-primers	Prdm14_R	TTACCTGGCATTTTCATTGCTC
	RT-primers	Dnmt3b_F	GGCAAGGACGACGTTTTGTG
	RT-primers	Dnmt3b_R	GTTGGACACGTCCGTGTAGTGAG
	RT-primers	Otx2_F	AATCAACTTGCCAGAATCCAGGG
	RT-primers	Otx2_R	GCTGTTGGCGGCACTTAGC
	RT-primers	Blimp1_F	AGCATGACCTGACATTGACACC
	RT-primers	Blimp1_R	CTCAACACTCTCATGTAAGAGGC
	RT-primers	Gata4_F	TTCCTCTCCCAGGAACATCAAA
	RT-primers	Gata4_R	GCTGCACAACTGGGCTCTACTT
	RT-primers	Gata6_F	TGCAAGATTGCATCATGACAGA
	RT-primers	Gata6_R	TGACCTCAGATCAGCCACGTTA
	RT-primers	Wnt3_F	CAAGCACAACAATGAAGCAGGC
	RT-primers	Wnt3_R	TCGGGACTCACGGTGTTTCTC
	RT-primers	Fgf5-F	GTTTCCAGTGGAGCCCTTC
	RT-primers	Fgf5-R	GAGACACAGCAAATATTTCCAAAA
	RT-primers	Sox7_F	GGATGAGAGGAAACGTCTGG
	RT-primers	Sox7_R	TCTGCCTCATCCACATAGGG
	RT-primers	Eomes_F	AACATGCAGGGCAATAAGATG
	RT-primers	Eomes_R	GAAATCTCCTGCCTCATCCA

Primer type	Experiment	oligo name	Sequence
RT-strand specific	RT-primers	RT2_suv_ex23-F	AAACTTGTGCAGGGATGCTC
	RT-primers	RT2_suv_ex23-R	ACTTCAGCGTTGGGAACAAG
	RT-primers	RTS_Suv_ex23-F	GCTCCCAGGTGTTTTCTGAG
	RT-primers	RTS_Suv_ex23-ex-R	GGTTGCAGTGTGTGCTGTAAA
	RT-primers	RTS_Suv_ex23-F	GCTCCCAGGTGTTTTCTGAG
	RT-primers	RTS_Suv_ex23-int-R2	CTACAGGAGTTTGGGCCTGA
	RT-primers	RT2_suv_ex34-F	CCATAATAAGCGGCATCCAC
	RT-primers	RT2_suv_ex34-R	TACCTCAGAGGAGGCAGAGC
	RT-primers	RTS_Suv_ex34-F2	CCACGTAGTCCAGGTCAAAGA
	RT-primers	RTS_Suv_ex34-ex-R1	GACTGCCCAAACCGTGTAGT
	RT-primers	RTS_Suv_ex34-F2	CCACGTAGTCCAGGTCAAAGA
	RT-primers	RTS_Suv_ex34-int-R1	CTGTACCTTGCCACCTCTCC
Isoform identfication	PCR primers	Suv-as__ex1a-1-F	CTAAGAACTCCCAGCGTTCC
	PCR primers	Suv-as__ex4s-1-R	GGTTAGGCCCGAATGTTGTT
	PCR primers	Suv-as__ex1a-1-F	CTAAGAACTCCCAGCGTTCC
	PCR primers	Suv-as__ex4-3-R	TTGCCATTCCAGAGTTAGGG
	PCR primers	Suv-as_ex1b2-F	AGTACCACATCTGCCAAACG
	PCR primers	Suv-as__ex4-3-R	TTGCCATTCCAGAGTTAGGG
	PCR primers	Suv39h1_ex1a-R	GGAGAAAGATGGCGGAAAA
	PCR primers	Suv39h1_ex6-1-F	GGAAGTATGGGCAGGAAACA
	PCR primers	Suv39h1_ex1b2_ex-R	AGAAACACCAAGGGCAGGAC
	PCR primers	Suv39h1_ex6-1-F	GGAAGTATGGGCAGGAAACA
	PCR primers	Suv39h1_ex1b2_ex-R	AGAAACACCAAGGGCAGGAC
	PCR primers	Suv39h1_ex3f-1-F	CAGGGTTAGGGGCACACATA
	PCR primers	Suv39h1_ex1a-R	GGAGAAAGATGGCGGAAAA
	PCR primers	Suv39h1_ex4f-2-F	GAGACCACGTCTTTCCTGCT
Mouse and ES genotyping	PCR primers	D8-seq-F	ACCACAACAAATGCCGAAAT
	PCR primers	v-as_PrDel_PCR_extKO	AGGAACCTTCCCTGCAGAAC
	PCR primers	v-as_PrDel_PCR_extgD	GCTTGGGTCTGTTTTTCAGC
	PCR primers	v-as_PrDel_PCR_extKO	ACGGTCTGATCCGCAACTAC
gRNAs	gRNA_ST	g3	GACTAGGGAGTCTAGGCTCT
	gRNA_ST	g4	GTTATTAATCAATACGAGCTG
	gRNA_ST	g5	CAAGATGCCACACCTATATT
	gRNA_ST	g6	GATGCCACACCTATATTGGG
	gRNA_ST	g7	GCCAAGATGCCACACCTATAT
	gRNA_ST	g8	GGACTAGGGAGTCTAGGCTC
	gRNA_CRISPR	gD	TTGCTCTGAGGAGGTCTATC
	gRNA_CRISPR	gG	ACGCTAGTCTTTGTGCGCAA

Table 4: **List of individual and pooled siRNAs.**

individual siRNA	siSuv-as-1	CTTCCTGCCATATGGACTA
	siSuv-as-2	CCAAAGAGCCATTTCTAAA
	siSuv-as-3	GTACTGGACTGCTGGTGTA
	siSuv-as-4	GGCCATCATTAGTGCGGAA
	siSuv-as-5	CCAATTCCTTGAACCCTAATGCTCA
	siSuv-as-6	TGGCAAGGATGAAGTGGCGCTTGTA
	siSuv-as-7	GGGCAACAAACAGAGGTGTTGCTAT
	siSuv-as-8	CGGAAGATGCAGAGATCGT
	siSuv-as-9	GATCTGGAACCTAGGAAGT
	siSuv-as-10	GAAACTGTGTGAAATATCA
	siSuv-as-11	GGACTAAACTCTTGAAACT
Pool n°1	siSuv-as-1	CTTCCTGCCATATGGACTA
	siSuv-as-4	GGCCATCATTAGTGCGGAA
	siSuv-as-5	CCAATTCCTTGAACCCTAATGCTCA
Pool n°2	siSuv-as-1	CTTCCTGCCATATGGACTA
	siSuv-as-5	CCAATTCCTTGAACCCTAATGCTCA
	siSuv-as-8	CGGAAGATGCAGAGATCGT
Pool n°3	siSuv-as-4	GGCCATCATTAGTGCGGAA
	siSuv-as-5	CCAATTCCTTGAACCCTAATGCTCA
	siSuv-as-8	CGGAAGATGCAGAGATCGT
Pool n°4	siSuv-as-1	CTTCCTGCCATATGGACTA
	siSuv-as-4	GGCCATCATTAGTGCGGAA
	siSuv-as-5	CCAATTCCTTGAACCCTAATGCTCA
	siSuv-as-8	CGGAAGATGCAGAGATCGT
	siSuv-as-10	GAAACTGTGTGAAATATCA
	siSuv-as-11	GGACTAAACTCTTGAAACT

Table 5: **List of smFISH probes.**

Probe name	Oligo name	Sequence	Probe name	Oligo name	Sequence
ASe	Suv-as_1	tcctcatgagagtctgatat	Suv39h1	Suv_1	tgctcacggatcttcttata
	Suv-as_2	ctaggttccagatcccaaag		Suv_2	gccacttaaccaggtaatac
	Suv-as_3	gtgtgataccggctcaaaac		Suv_3	tgttttctgagtcgggatac
	Suv-as_4	actaatgatggccgaggctg		Suv_4	acattttagattctgccgtg
	Suv-as_5	gctacgatctctgcatcttc		Suv_5	ccttgtggaactgcttaaga
	Suv-as_6	aaccgtgtagtccagaaagg		Suv_6	tcggacaagctctctttcta
	Suv-as_7	tgttgctgtggctatgactg		Suv_7	taattggctaggtttgggtc
	Suv-as_8	catctacgagtgcaactccc		Suv_8	ggtaataatctctcccacat
	Suv-as_9	caggtgcgactgaaagctgg		Suv_9	gtatatacgtcttccacgta
	Suv-as_10	agtttgcctacaatgaccaa		Suv_10	ttgccataataagcggcatc
	Suv-as_11	ctcaaccaggtagctgttgg		Suv_11	gggatcacaactatggttga
	Suv-as_12	tgagtatcgagttggtgagg		Suv_12	gaatacgttgtacacctgca
	Suv-as_13	caaggtcctttgtctatatc		Suv_13	tagtcgctcatcaaggttgt
	Suv-as_14	gcggatcaccgtggagaatg		Suv_14	ccacttgcatgttgtaatca
	Suv-as_15	cttcagcgttgggaacaaga		Suv_15	caaagttggagtccattcgg
	Suv-as_16	gttgaatacctggtgcagaa		Suv_16	tgcattcaatacggactcgt
	Suv-as_17	cccttggtgtttctaagaag		Suv_17	ctcagggctagaagaggtat
	Suv-as_18	cgactagccaagctttcttg		Suv_18	tcatccttgacaagagggta
	Suv-as_19	ttggaatcaactgcaggacc		Suv_19	gcagacaaggctctgattga
	Suv-as_20	tgcagtgtgtgctgtaaatc		Suv_20	tagggtaggtgaggacaagt
	Suv-as_21	ctttggccttctcattctag		Suv_21	ggagttagtcctcacaacag
	Suv-as_22	ccgtcgattacatttttgtg		Suv_22	ggagttttaagtgcaatgcc
	Suv-as_23	ttgcagaggatcttagcttt		Suv_23	ccaactctagggaatcatga
	Suv-as_24	ttcaaggaattggcacttcc		Suv_24	tggacctccattatatgaca
	Suv-as_25	ccctgagaattgagcattag		Suv_25	tctgttttacagccgagttc
	Suv-as_26	ctggtgtgtaggcttatttc		Suv_26	caacaggtcacacacaagct
	Suv-as_27	gggaaagacaggttgatgct		Suv_27	cttcctcataccattagagt
	Suv-as_28	ataggagaactggacccttc		Suv_28	tatctcagctctgatgtctg
	Suv-as_29	agatgagggcataggtcctc		Suv_29	ctgatccagcttgatgttag
	Suv-as_30	gccagagtttggagtagaag		Suv_30	ctctgaagattgccttttct
	Suv-as_31	cttgccattccagagttagg		Suv_31	taaggttaataccagccagc
	Suv-as_32	attgctacaagcgccacttc		Suv_32	aactgttgagaccctatagg
	Suv-as_33	agtggatccctttatggacg		Suv_33	gcatttgggcaactactttt
	Suv-as_34	ttaaggccttggatgttctg		Suv_34	gttctagtggccatcaaatg
	Suv-as_35	cgaaggtaagtctgcaggca		Suv_35	aatacctagatcttggttcc
				Suv_36	cattggaggattcattctgc
				Suv_37	ccactcttaggcacaaagta
				Suv_38	gtaactgtttatggttgggc
				Suv_39	ctctaactatcctgctacta
				Suv_40	aagcatgctactttcagtgg
				Suv_41	gcctgagtcactatatcaac
				Suv_42	caaggggagcaggtgtaata
				Suv_43	gtcccctactataaaatgct
				Suv_44	ctgactcattagcaacttcc
				Suv_45	atacttccagccagtagata
				Suv_46	catacacctgctagttactc
				Suv_47	gtcttaccttacaatgtggt
				Suv_48	agatttgtctatatgcccag

Probe name	Oligo name	Sequence	Probe name	Oligo name	Sequence
AS1	AS1-1	gtccccacatcttgataaaa	AS2	AS2-1	agaatgcggtcagaaaggct
	AS1-2	ccccccaaatttaacaaact		AS2-2	tagaagctggggttgagttg
	AS1-3	caaggcagagctctgaagac		AS2-3	ccacttgtctcagttaactg
	AS1-4	ctcctcagagcaaagacttc		AS2-4	acagaaggcaatgtctaggc
	AS1-5	tgtggtatctagagctcttt		AS2-5	tcagggaagccttgaaaacc
	AS1-6	tcagtaccagaagttctgga		AS2-6	aatgccatcagaatggggat
	AS1-7	agaagccctgaaacatcagg		AS2-7	tctttgtgcttttgatgctg
	AS1-8	aaatgggactcattgggctg		AS2-8	gtggggttagtgtttatact
	AS1-9	ctgacaagagagcgttcagg		AS2-9	cactcagcaggaatcacttt
	AS1-10	tatgggacactggcatactg		AS2-10	gacaacggcttgccgaaaat
	AS1-11	accttacccccttgataaatg		AS2-11	catgcaggccttctaacaaa
	AS1-12	tgcaagctagggcaattgat		AS2-12	gtgttgtgctctagttctaa
	AS1-13	ttagagccacacaagcagat		AS2-13	agacacttggcagatagtgg
	AS1-14	ttcatttcctacttctgcaa		AS2-14	tccatcttaaatcacctacc
	AS1-15	ctaccagtcatgtagacaca		AS2-15	aacttacacttacaccacct
	AS1-16	tagcttcaccactggaaaca		AS2-16	gttttaagatgttcttcccg
	AS1-17	cattcaaaaccctctcagtc		AS2-17	tgatatgggatgtgaggacc
	AS1-18	acaaattctagggtacccat		AS2-18	tttgccacaagaaccatctg
	AS1-19	actgtgatatgtgcatacca		AS2-19	gaaccttcgtgaccaaatct
	AS1-20	taccctcacacacataactc		AS2-20	cctttcacaaactgtacctt
	AS1-21	caatttggatataatccccc		AS2-21	ttctggttctcaggatttga
	AS1-22	tcctagatctgacatgttgt		AS2-22	atttcctaggataggagcag
	AS1-23	taccaggaaaacgccagaga		AS2-23	gaagaaaccctggagtgagg
	AS1-24	cttagtcaagtcagagacct		AS2-24	aggaaggtctgtgtctgatg
	AS1-25	cacaaatgcaacaggtctcc		AS2-25	gcacagaggagggtatattt
	AS1-26	agctgtatagcacctgaatc		AS2-26	taatatagccttccttgcta
	AS1-27	cacatccatggattcatgtt		AS2-27	gtggtctggagagaatgtga
	AS1-28	aggccctgagaaatagatca		AS2-28	aggacctgagtgaatggaga
	AS1-29	ggagctcaagtcctagaata		AS2-29	agcctccagtgggaaaataa
	AS1-30	atactagttcatatcctggc		AS2-30	gatgtttccacgaactacgt
	AS1-31	atcaagaacctcagaggctg		AS2-31	agcatactttgccttacatg
	AS1-32	gatttactcataccctgaga		AS2-32	gtggacattccttaaggttt
	AS1-33	atctgcagatggatccaaga		AS2-33	tttcctagtctcattcacat
	AS1-34	ctaatgtatggacccagaca		AS2-34	gtcatgtaagaagtggtcca
	AS1-35	cttaatcagggaccacactg		AS2-35	actgtctagatgctcctaag
	AS1-36	ccaagttgtttagccttatt		AS2-36	tggttgtttaggttctttct
	AS1-37	ttgggatccacatcaacaga		AS2-37	gtacagtacagtaccaaggt
	AS1-38	catccaacatccctaagatt		AS2-38	ctagaggtcattccaagcaa
	AS1-39	ctcaccagtgaatacatcgg		AS2-39	ggcagtcaggtatatagaca
	AS1-40	gaatgccctcaacacagata		AS2-40	gagttctccatactttcaga
	AS1-41	tgggacatgtgggatcttag		AS2-41	atctgcagtaagagatggct
	AS1-42	caaactggggctatgacact		AS2-42	cgtggtctcaaggtagaaca
	AS1-43	agttcagggaacatccatga		AS2-43	atgactaggagaggctttca
	AS1-44	ctcagagagctgacactacg			
	AS1-45	gtacttaaagatgcactccc			
	AS1-46	accggcatgtttatcaacta			
	AS1-47	ccacaaagaccctaagtgac			
	AS1-48	ccaagtctcagaatggtcag			

Table 6: **List of antibodies.**

Antibody	Supplier	Reference	IF		Western Blot	ChIP
			mESC	GV/Zygotes		
Suv39h1	Cell Signalling	CST 8729	1/200	1/500	1/1000	-
H3K9me2	abcam	ab8898	1/200	1/500	1/10000	5 µL
H3K9me3	abcam	ab1220	1/500	/500 or 1/30	1/10000	3 µL
LaminB1	abcam	ab16048	-	-	1/10000	-
H3K9ac	Cell Signaling	CST 9649S	-	-	-	10 µL
H3K27ac	Active Motif	39133	-	-	-	10 µL
H3K4me1	Active Motif	39297	-	-	-	10 µL
H3K4me2	Abcam	ab32356	-	-	-	10 µL
H3K4me3	Active Motif	39159	-	-	-	5 µL
H3K36me3	Active Motif	61101	-	-	-	10 µL
Pol II H224	Santa Cruz	sc-9001	-	-	-	12 µL
anti-rabbit IgG HRP	Thermo Scientific	RB230254	-	-	1/1000	-
anti-mouse IgG HRP	Thermo Scientific	QB213868	-	-	1/1000	-
anti-mouse IgG AlexaFluor 488	JIR	715-545-150	1/500	-	-	-
anti-rabbit IgG AlexaFluor 488	JIR	711-545-152	1/500	-	-	-
anti-mouse IgG AlexaFluor 647	Invitrogen	A31571	-	1/300	-	-
anti-rabbit IgG AlexaFluor 488	Invitrogen	A21206	-	1/300	-	-

References

1. Nichols, J. & Smith, A. Pluripotency in the Embryo and in Culture. *Cold Spring Harbor Perspectives in Biology* **4**, a008128–a008128 (2012).
2. Piliszek, A., Grabarek, J. B., Frankenberg, S. R. & Plusa, B. Cell fate in animal and human blastocysts and the determination of viability. *Mol. Hum. Reprod.* **22**, 681–690 (2016).
3. Santos, F., Hendrich, B., Reik, W. & Dean, W. Dynamic reprogramming of DNA methylation in the early mouse embryo. Dev Biol 241, 172–182 (2002).
4. Adenot, P. G., Mercier, Y., Renard, J. P. & Thompson, E. M. Differential H4 acetylation of paternal and maternal chromatin precedes DNA replication and differential transcriptional activity in pronuclei of 1-cell mouse embryos. Development 124, 4615–4625 (1997).
5. Wang, Q. T. *et al.* A genome-wide study of gene activity reveals developmental signaling pathways in the preimplantation mouse embryo. *Dev. Cell* **6**, 133–144 (2004).
6. Li, L., Zheng, P. & Dean, J. Maternal control of early mouse development. *Development* **137**, 859–870 (2010).
7. Handyside, A. H. Distribution of antibody- and lectin-binding sites on dissociated blastomeres from mouse morulae: evidence for polarization at compaction. *J Embryol Exp Morphol* **60**, 99–116 (1980).
8. Tabansky, I. *et al.* Developmental bias in cleavage-stage mouse blastomeres. *Curr. Biol.* **23**, 21–31 (2013).
9. Tarkowski, A. K. & Wróblewska, J. Development of blastomeres of mouse eggs isolated at the 4- and 8-cell stage. *Development* **18**, 155–180 (1967).
10. Yao, C., Zhang, W. & Shuai, L. The first cell fate decision in pre-implantation mouse embryos. *Cell Regeneration* **8**, 51–57 (2019).
11. Smith, R. & McLaren, A. Factors affecting the time of formation of the mouse blastocoele. *J Embryol Exp Morphol* **41**, 79–92 (1977).
12. Chazaud, C., Yamanaka, Y., Pawson, T. & Rossant, J. Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway. *Developmental Cell* **10**, 615–624 (2006).
13. Leitch, H. G., Tang, W. W. C. & Surani, M. A. Primordial Germ-Cell Development and Epigenetic Reprogramming in Mammals. in *Current Topics in Developmental Biology* vol. 104 149–187 (Elsevier, 2013).

14. Ohinata, Y. *et al.* Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* **436**, 207–213 (2005).
15. Molecular Mechanisms of Cell Differentiation in Gonad Development. vol. 58 (Springer International Publishing, 2016).
16. Sasaki, H. & Matsui, Y. Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat. Rev. Genet.* **9**, 129-140 (2008).
17. Lawson, K. A. *et al.* Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev.* **13**, 424–436 (1999).
18. Saitou, M., Barton, S. C. & Surani, M. A. A molecular programme for the specification of germ cell fate in mice. *Nature* **418**, 293–300 (2002).
19. Hajkova, P. *et al.* Epigenetic reprogramming in mouse primordial germ cells. *Mech. Dev.* **117**, 15–23 (2002).
20. Hajkova, P. *et al.* Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452**, 877–881 (2008).
21. Yu, F.-X. & Guan, K.-L. The Hippo pathway: regulators and regulations. *Genes Dev.* **27**, 355–371 (2013).
22. Nishioka, N. *et al.* The Hippo Signaling Pathway Components Lats and Yap Pattern Tead4 Activity to Distinguish Mouse Trophectoderm from Inner Cell Mass. *Developmental Cell* **16**, 398–410 (2009).
23. Rayon, T. *et al.* Notch and hippo converge on Cdx2 to specify the trophectoderm lineage in the mouse blastocyst. *Dev. Cell* **30**, 410–422 (2014).
24. Niwa, H. *et al.* Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123**, 917–929 (2005).
25. Ng, R. K. *et al.* Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. *Nat. Cell Biol.* **10**, 1280–1290 (2008).
26. Lorthongpanich, C. *et al.* Temporal reduction of LATS kinases in the early preimplantation embryo prevents ICM lineage differentiation. *Genes Dev.* **27**, 1441–1446 (2013).
27. Zhang, J. *et al.* Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. *Nat. Cell Biol.* **8**, 1114–1123 (2006).
28. Wang, J. et al. Asymmetric Expression of LincGET Biases Cell Fate in Two-Cell Mouse Embryos. Cell 175, 1887-1901.e18 (2018).
29. Ohnishi, Y. *et al.* Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. *Nat. Cell Biol.* **16**, 27–37 (2014).

30. Yang, D.-H. *et al.* Disabled-2 is essential for endodermal cell positioning and structure formation during mouse embryogenesis. *Dev. Biol.* **251**, 27–44 (2002).
31. Azami, T. et al. Regulation of the ERK signalling pathway in the developing mouse blastocyst. Development 146, (2019).
32. Pepling, M. E. & Spradling, A. C. Mouse Ovarian Germ Cell Cysts Undergo Programmed Breakdown to Form Primordial Follicles. Developmental Biology 234, 339–351 (2001).
33. Pedersen, T. & Peters, H. Proposal for a classification of oocytes and follicules in the mouse ovary. Reproduction 17, 555–557 (1968).
34. Reddy, P. et al. Oocyte-Specific Deletion of Pten Causes Premature Activation of the Primordial Follicle Pool. Science 319, 611–613 (2008).
35. Zhang, H. et al. Somatic Cells Initiate Primordial Follicle Activation and Govern the Development of Dormant Oocytes in Mice. Current Biology 24, 2501–2508 (2014).
36. Magoffin, D. A. Ovarian theca cell. The International Journal of Biochemistry & Cell Biology 37, 1344–1349 (2005).
37. Tang, S.-B. et al. Multiple superovulations alter histone modifications in mouse early embryos. Reproduction 157, 511–523 (2019).
38. MacLennan, M., Crichton, J. H., Playfoot, C. J. & Adams, I. R. Oocyte development, meiosis and aneuploidy. Seminars in Cell & Developmental Biology 45, 68–76 (2015).
39. Tsafriri, A. & Dekel, N. Intra- and Intercellular Molecular Mechanisms in Regulation of Meiosis in Murid Rodents. Chapter 2, Oocyte Maturation and Fertilization: A Long History for a Short Event, 38-63 (2011)
40. Robker, R. L., Hennebold, J. D. & Russell, D. L. Coordination of Ovulation and Oocyte Maturation: A Good Egg at the Right Time. 10 (2018).
41. Oocytes: Maternal Information and Functions. vol. 63 (Springer International Publishing, 2017).
42. Martello, G. & Smith, A. The Nature of Embryonic Stem Cells. Annu. Rev. Cell Dev. Biol. 30, 647–675 (2014).
43. Bochner, S. S. & Martin, W. T. Spontaneous testicular teratomas in an inbred strain of mice. Leroy C. Stevens, JR., & C. C. Little. Proc. N. A. S. 8 (1954).
44. Martin, G. R. & Evans, M. J. Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. Proc. Natl. Acad. Sci. U.S.A. 72, 1441–1445 (1975).
45. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. Nature 292, 154–156 (1981).

46. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Natl. Acad. Sci. U.S.A. 78, 7634–7638 (1981).
47. Bradley, A., Evans, M., Kaufman, M. H. & Robertson, E. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature 309, 255–256 (1984).
48. Brook, F. A. & Gardner, R. L. The origin and efficient derivation of embryonic stem cells in the mouse. Proc. Natl. Acad. Sci. U.S.A. 94, 5709–5712 (1997).
49. Boroviak, T., Loos, R., Bertone, P., Smith, A. & Nichols, J. The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. Nat. Cell Biol. 16, 516–528 (2014).
50. Nichols, J. & Smith, A. The origin and identity of embryonic stem cells. Development 138, 3–8 (2011).
51. Brons, I. G. M. et al. Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature 448, 191–195 (2007).
52. Kojima, Y. et al. The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. Cell Stem Cell 14, 107–120 (2014).
53. Matsui, Y., Zsebo, K. & Hogan, B. L. Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. Cell 70, 841–847 (1992).
54. Rastan, S. & Robertson, E. J. X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation. Development 90, 379–388 (1985).
55. Schulz, E. G. et al. The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. Cell Stem Cell 14, 203–216 (2014).
56. Koopman, P. & Cotton, R. G. A factor produced by feeder cells which inhibits embryonal carcinoma cell differentiation. Characterization and partial purification. Exp. Cell Res. 154, 233–242 (1984).
57. Smith, T. A. & Hooper, M. L. Medium conditioned by feeder cells inhibits the differentiation of embryonal carcinoma cultures. Exp. Cell Res. 145, 458–462 (1983).
58. Smith, A. G. et al. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature 336, 688–690 (1988).
59. Williams, R. L. et al. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature 336, 684–687 (1988).
60. Rathjen, P. D. et al. Developmentally programmed induction of differentiation inhibiting activity and the control of stem cell populations. Genes Dev. 4, 2308–2318 (1990).

61. Ohtsuka, S., Nakai-Futatsugi, Y. & Niwa, H. LIF signal in mouse embryonic stem cells. JAK-STAT 4, 1–9 (2015).
62. Yasukawa, H. et al. The JAK-binding protein JAB inhibits Janus tyrosine kinase activity through binding in the activation loop. EMBO J 18, 1309–1320 (1999).
63. Stahl, N. et al. Choice of STATs and other substrates specified by modular tyrosine-based motifs in cytokine receptors. Science 267, 1349–1353 (1995).
64. Wegenka, U. M., Buschmann, J., Lütticken, C., Heinrich, P. C. & Horn, F. Acute-phase response factor, a nuclear factor binding to acute-phase response elements, is rapidly activated by interleukin-6 at the posttranslational level. Mol Cell Biol 13, 276–288 (1993).
65. Ho, L. et al. An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. PNAS 106, 5187–5191 (2009).
66. Naka, T. et al. Structure and function of a new STAT-induced STAT inhibitor. Nature 387, 924–929 (1997).
67. Hackett, J. A. & Surani, M. A. Regulatory Principles of Pluripotency: From the Ground State Up. Cell Stem Cell 15, 416–430 (2014).
68. Niwa, H., Ogawa, K., Shimosato, D. & Adachi, K. A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. Nature 460, 118–122 (2009).
69. Niwa, H., Burdon, T., Chambers, I. & Smith, A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes Dev. 12, 2048–2060 (1998).
70. Cheng, A. M. et al. Mammalian Grb2 regulates multiple steps in embryonic development and malignant transformation. Cell 95, 793–803 (1998).
71. Huang, G., Yan, H., Ye, S., Tong, C. & Ying, Q.-L. STAT3 phosphorylation at tyrosine 705 and serine 727 differentially regulates mouse ESC fates. Stem Cells 32, 1149–1160 (2014).
72. Martello, G., Bertone, P. & Smith, A. Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. EMBO J. 32, 2561–2574 (2013).
73. Ogawa, K., Nishinakamura, R., Iwamatsu, Y., Shimosato, D. & Niwa, H. Synergistic action of Wnt and LIF in maintaining pluripotency of mouse ES cells. Biochem. Biophys. Res. Commun. 343, 159–166 (2006).
74. Ying, Q. L., Nichols, J., Chambers, I. & Smith, A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell 115, 281–292 (2003).
75. Park, K.-S. Tgf-Beta family signaling in embryonic stem cells. Int J Stem Cells 4, 18–23 (2011).

76. Qi, X. et al. BMP4 supports self-renewal of embryonic stem cells by inhibiting mitogen-activated protein kinase pathways. Proc Natl Acad Sci U S A 101, 6027–6032 (2004).
77. James, D., Levine, A. J., Besser, D. & Hemmati-Brivanlou, A. TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. Development 132, 1273–1282 (2005).
78. Kunath, T. et al. FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. Development 134, 2895–2902 (2007).
79. Burdon, T., Stracey, C., Chambers, I., Nichols, J. & Smith, A. Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. Dev Biol 210, 30–43 (1999).
80. Soszyńska, A., Klimczewska, K. & Suwińska, A. FGF/ERK signaling pathway: how it operates in mammalian preimplantation embryos and embryo-derived stem cells. Int J Dev Biol 63, 171–186 (2019).
81. Ying, Q.-L. et al. The ground state of embryonic stem cell self-renewal. Nature 453, 519–523 (2008).
82. Wray, J. & Hartmann, C. WNTing embryonic stem cells. Trends Cell Biol 22, 159–168 (2012).
83. ten Berge, D. et al. Wnt signaling mediates self-organization and axis formation in embryoid bodies. Cell Stem Cell 3, 508–518 (2008).
84. Choi, J. et al. Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. Nature 548, 219–223 (2017).
85. Taleahmad, S. et al. Low Focal Adhesion Signaling Promotes Ground State Pluripotency of Mouse Embryonic Stem Cells. J. Proteome Res. 16, 3585–3595 (2017).
86. Hassani, S.-N., Moradi, S., Taleahmad, S., Braun, T. & Baharvand, H. Transition of inner cell mass to embryonic stem cells: mechanisms, facts, and hypotheses. Cell. Mol. Life Sci. 76, 873–892 (2019).
87. Taleahmad, S., Hassani, S. N., Hosseini Salekdeh, G. & Baharvand, H. Metabolic Signature of Pluripotent Stem Cells. Cell J 20, 388–395 (2018).
88. Okamoto, K. et al. A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. Cell 60, 461–472 (1990).
89. Schöler, H. R., Dressler, G. R., Balling, R., Rohdewohld, H. & Gruss, P. Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. EMBO J 9, 2185–2195 (1990).

90. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat. Genet. 24, 372–376 (2000).
91. Karwacki-Neisius, V. et al. Reduced Oct4 expression directs a robust pluripotent state with distinct signaling activity and increased enhancer occupancy by Oct4 and Nanog. Cell Stem Cell 12, 531–545 (2013).
92. DeVeale, B. et al. Oct4 is required ~E7.5 for proliferation in the primitive streak. PLoS Genet. 9, e1003957 (2013).
93. Masui, S. et al. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat. Cell Biol. 9, 625–635 (2007).
94. Ambrosetti, D. C., Basilico, C. & Dailey, L. Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. Mol. Cell. Biol. 17, 6321–6329 (1997).
95. Chen, X. et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell 133, 1106–1117 (2008).
96. Cole, M. F., Johnstone, S. E., Newman, J. J., Kagey, M. H. & Young, R. A. Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. Genes Dev 22, 746–755 (2008).
97. Kopp, J. L., Ormsbee, B. D., Desler, M. & Rizzino, A. Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. Stem Cells 26, 903–911 (2008).
98. Zhao, S., Nichols, J., Smith, A. G. & Li, M. SoxB transcription factors specify neuroectodermal lineage choice in ES cells. Mol. Cell. Neurosci. 27, 332–342 (2004).
99. Avilion, A. A. et al. Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev 17, 126–140 (2003).
100. Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R. C. & Melton, D. A. 'Stemness': transcriptional profiling of embryonic and adult stem cells. Science 298, 597–600 (2002).
101. Silva, J. et al. Nanog is the gateway to the pluripotent ground state. Cell 138, 722–737 (2009).
102. Chambers, I. et al. Nanog safeguards pluripotency and mediates germline development. Nature 450, 1230–1234 (2007).
103. Mitsui, K. et al. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell 113, 631–642 (2003).

104.Chambers, I. et al. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell 113, 643–655 (2003).

105.Boyer, L. A. et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122, 947–956 (2005).

106.Festuccia, N. et al. Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. Cell Stem Cell 11, 477–490 (2012).

107.Ivanova, N. et al. Dissecting self-renewal in stem cells with RNA interference. Nature 442, 533–538 (2006).

108.Bourillot, P.-Y. & Savatier, P. Krüppel-like transcription factors and control of pluripotency. BMC Biology 8, 125 (2010).

109.Bourillot, P.-Y. et al. Novel STAT3 target genes exert distinct roles in the inhibition of mesoderm and endoderm differentiation in cooperation with Nanog. Stem Cells 27, 1760–1771 (2009).

110.Martello, G. et al. Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. Cell Stem Cell 11, 491–504 (2012).

111.Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676 (2006).

112.Feng, B. et al. Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. Nat Cell Biol 11, 197–203 (2009).

113.Ieda, M. et al. Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell 142, 375–386 (2010).

114.Srivastava, D. & DeWitt, N. In Vivo Cellular Reprogramming: The Next Generation. Cell 166, 1386–1396 (2016).

115.Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell 51, 987–1000 (1987).

116.Nichols, J., Chambers, I., Taga, T. & Smith, A. Physiological rationale for responsiveness of mouse embryonic stem cells to gp130 cytokines. Development 128, 2333–2339 (2001).

117.Silva, J. & Smith, A. Capturing pluripotency. Cell 132, 532–536 (2008).

118.Morgani, S., Nichols, J. & Hadjantonakis, A.-K. The many faces of Pluripotency: in vitro adaptations of a continuum of in vivo states. BMC Developmental Biology 17, 7 (2017).

119.Ying, Q.-L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat Biotechnol 21, 183–186 (2003).

120.Brickman, J. M. & Serup, P. Properties of embryoid bodies. Wiley Interdiscip Rev Dev Biol 6, (2017).

121.Cermola, F. et al. Gastruloid Development Competence Discriminates Different States of Pluripotency. Stem Cell Reports 16, 354–369 (2021).

122.Sumi, T., Oki, S., Kitajima, K. & Meno, C. Epiblast Ground State Is Controlled by Canonical Wnt/β-Catenin Signaling in the Postimplantation Mouse Embryo and Epiblast Stem Cells. PLOS ONE 8, e63378 (2013).

123.Kojima, Y. et al. The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. Cell Stem Cell 14, 107–120 (2014).

124.Tesar, P. J. et al. New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature 448, 196–199 (2007).

125.Tsakiridis, A. et al. Distinct Wnt-driven primitive streak-like populations reflect in vivo lineage precursors. Development 141, 1209–1221 (2014).

126.Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. Cell 146, 519–532 (2011).

127.Tsukiyama, T. & Ohinata, Y. A Modified EpiSC Culture Condition Containing a GSK3 Inhibitor Can Support Germline-Competent Pluripotency in Mice. PLOS ONE 9, e95329 (2014).

128.Olins, D. E. & Olins, A. L. Chromatin history: our view from the bridge. Nat Rev Mol Cell Biol 4, 809–814 (2003).

129.Flemming, W. Zellsubstanz, Kern und Zelltheilung. 1–472 (F.C.W. Vogel, 1882). doi:10.5962/bhl.title.168645.

130. Zhou, K., Gaullier, G. & Luger, K. Nucleosome structure and dynamics are coming of age. Nat Struct Mol Biol 26, 3–13 (2019).

131.Biterge, B. & Schneider, R. Histone variants: key players of chromatin. Cell Tissue Res 356, 457–466 (2014).

132.Passarge, E. Emil Heitz and the Concept of Heterochromatin: Longitudinal Chromosome Differentiation was Recognized Fifty Years Ago. 10.

133.Muller, H. J. Types of visible variations induced by X-rays inDrosophila. Journ. of Gen. 22, 299–334 (1930).

134.McClintock, B. The origin and behavior of mutable loci in maize. PNAS 36, 344–355 (1950).

135.Brown, S. W. Heterochromatin. Science 151, 417–425 (1966).

136. Trojer, P. & Reinberg, D. Facultative Heterochromatin: Is There a Distinctive Molecular Signature? Molecular Cell 28, 1–13 (2007).
137. Vaquero, A. et al. Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin. Mol Cell 16, 93–105 (2004).
138. Loda, A. & Heard, E. Xist RNA in action: Past, present, and future. PLoS Genet 15, e1008333 (2019).
139. Saksouk, N., Simboeck, E. & Déjardin, J. Constitutive heterochromatin formation and transcription in mammals. Epigenetics & Chromatin 8, 3 (2015).
140. Wong, A. K. & Rattner, J. B. Sequence organization and cytological localization of the minor satellite of mouse. Nucleic Acids Res 16, 11645–11661 (1988).
141. Guenatri, M., Bailly, D., Maison, C. & Almouzni, G. Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. Journal of Cell Biology 166, 493–505 (2004).
142. Joseph, A., Mitchell, A. R. & Miller, O. J. The organization of the mouse satellite DNA at centromeres. Experimental Cell Research 183, 494–500 (1989).
143. Hsu, T. C., Cooper, J. E., Mace, M. L. & Brinkley, B. R. Arrangement of centromeres in mouse cells. Chromosoma 34, 73–87 (1971).
144. Platt, R. N., Vandewege, M. W. & Ray, D. A. Mammalian transposable elements and their impacts on genome evolution. Chromosome Res 26, 25–43 (2018).
145. Lowy, D. R., Rowe, W. P., Teich, N. & Hartley, J. W. Murine leukemia virus: high-frequency activation in vitro by 5-iododeoxyuridine and 5-bromodeoxyuridine. Science 174, 155–156 (1971).
146. Waddington, C. H. The epigenotype. 1942. Int J Epidemiol 41, 10–13 (2012).
147. Nanney, D. L. Epigenetic control systems. Proc Natl Acad Sci U S A 44, 712–717 (1958).
148. R. J. Lincoln, G. A. Boxshall & P. F. Clark, compilers. A Dictionary of Ecology, Evolution and Systematics. viii, 298 pp. Cambridge University Press, 1982. (Paperback edition 1983.) Price £9.95. Journal of the Marine Biological Association of the United Kingdom 64, 970–970 (1984).
149. Deans, C. & Maggert, K. A. What do you mean, 'epigenetic'? Genetics 199, 887–896 (2015).
150. Holliday, R. Epigenetics: an overview. Dev Genet 15, 453–457 (1994).
151. Wu Ct, null & Morris, J. R. Genes, genetics, and epigenetics: a correspondence. Science 293, 1103–1105 (2001).

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

153.Hotchkiss, R. D. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. J Biol Chem 175, 315–332 (1948).

154.Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 13, 484–492 (2012).

155.Ooi, S. K. T. et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448, 714–717 (2007).

156.Cortellino, S. et al. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell 146, 67–79 (2011).

157.Takai, D. & Jones, P. A. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci U S A 99, 3740–3745 (2002).

158.Nguyen, C. T., Gonzales, F. A. & Jones, P. A. Altered chromatin structure associated with methylation-induced gene silencing in cancer cells: correlation of accessibility, methylation, MeCP2 binding and acetylation. Nucleic Acids Res 29, 4598–4606 (2001).

159.Maunakea, A. K. et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 466, 253–257 (2010).

160.Shukla, S. et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature 479, 74–79 (2011).

161.Histone Mutations and Cancer. (Springer Singapore, 2021). doi:10.1007/978-981-15-8104-5.

162.Heintzman, N. D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 39, 311–318 (2007).

163.Santos-Rosa, H. et al. Active genes are tri-methylated at K4 of histone H3. Nature 419, 407–411 (2002).

164.Mikkelsen, T. S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553–560 (2007).

165.Xu, Q. & Xie, W. Epigenome in Early Mammalian Development: Inheritance, Reprogramming and Establishment. Trends Cell Biol 28, 237–253 (2018).

166. Kizer, K. O. et al. A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. Mol. Cell Biol. 25, 3305–3316 (2005).

167.Huyen, Y. et al. Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. Nature 432, 406–411 (2004).

168. Mueller, D. et al. A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification. Blood 110, 4445–4454 (2007).
169. Bannister, A. J. et al. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature 410, 120–124 (2001).
170. Vavouri, T. & Lehner, B. Human genes with CpG island promoters have a distinct transcription-associated chromatin organization. Genome Biology 13, R110 (2012).
171. Towbin, B. D. et al. Step-wise methylation of histone H3tos positions heterochromatin at the nuclear periphery. Cell 150, 934–947 (2012).
172. Barski, A. et al. High-Resolution Profiling of Histone Methylations in the Human Genome. Cell 129, 823–837 (2007).
173. Rada-Iglesias, A. et al. A unique chromatin signature uncovers early developmental enhancers in humans. Nature 470, 279–283 (2011).
174. Sato, Y. et al. A Genetically Encoded Probe for Live-Cell Imaging of H4K20 Monomethylation. J Mol Biol 428, 3885–3902 (2016).
175. Kuo, A. J. et al. The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. Nature 484, 115–119 (2012).
176. Zhang, J., Jing, L., Li, M., He, L. & Guo, Z. Regulation of histone arginine methylation/demethylation by methylase and demethylase (Review). Mol Med Rep 19, 3963–3971 (2019).
177. Fischle, W. et al. Regulation of HP1–chromatin binding by histone H3 methylation and phosphorylation. Nature 438, 1116–1122 (2005).
178. Wang, F. et al. Histone H3 Thr-3 phosphorylation by Haspin positions Aurora B at centromeres in mitosis. Science 330, 231–235 (2010).
179. Suzuki, T., Watanabe, H., Kita, K., Honma, T. & Ochi, T. Arsenite-induced histone H3 modification and its effects on EGR1 and FOS expression in HeLa cells. Journal of Applied Toxicology 38, 734–743 (2018).
180. Hyun, K., Jeon, J., Park, K. & Kim, J. Writing, erasing and reading histone lysine methylations. Exp Mol Med 49, e324–e324 (2017).
181. Pradhan, S., Bacolla, A., Wells, R. D. & Roberts, R. J. Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. J Biol Chem 274, 33002–33010 (1999).
182. Leonhardt, H., Page, A. W., Weier, H. U. & Bestor, T. H. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell 71, 865–873 (1992).

183.Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69, 915–926 (1992).

184.Chen, R. Z., Pettersson, U., Beard, C., Jackson-Grusby, L. & Jaenisch, R. DNA hypomethylation leads to elevated mutation rates. Nature 395, 89–93 (1998).

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

186.Moore, L. D., Le, T. & Fan, G. DNA Methylation and Its Basic Function. Neuropsychopharmacol 38, 23–38 (2013).

187.Brenner, C. et al. Myc represses transcription through recruitment of DNA methyltransferase corepressor. EMBO J 24, 336–346 (2005).

188.Aapola, U. et al. Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. Genomics 65, 293–298 (2000).

189.Barau, J. et al. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science 354, 909–912 (2016).

190.Oswald, J. et al. Active demethylation of the paternal genome in the mouse zygote. Curr Biol 10, 475–478 (2000).

191.Bhutani, N. et al. Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature 463, 1042–1047 (2010).

192.Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324, 930–935 (2009).

193.Cortázar, D. et al. Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. Nature 470, 419–423 (2011).

194.Nan, X., Meehan, R. R. & Bird, A. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nucleic Acids Res 21, 4886–4892 (1993).

195.Hashimoto, H., Horton, J. R., Zhang, X. & Cheng, X. UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications. Epigenetics 4, 8–14 (2009).

196.Filion, G. J. P. et al. A family of human zinc finger proteins that bind methylated DNA and repress transcription. Mol Cell Biol 26, 169–181 (2006).

197.Tchasovnikarova, I. A. et al. GENE SILENCING. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. Science 348, 1481–1485 (2015).

198. Tachibana, M. et al. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes Dev 16, 1779–1791 (2002).
199. Fumasoni, I. et al. Family expansion and gene rearrangements contributed to the functional specialization of PRDM genes in vertebrates. BMC Evol Biol 7, 187 (2007).
200. Yamane, K. et al. JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. Cell 125, 483–495 (2006).
201. Klose, R. J. et al. The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. Nature 442, 312–316 (2006).
202. Horton, J. R. et al. Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. Nat Struct Mol Biol 17, 38–43 (2010).
203. Czermin, B. et al. Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell 111, 185–196 (2002).
204. Issaeva, I. et al. Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. Mol Cell Biol 27, 1889–1903 (2007).
205. Reute, G. & Spierer, P. Position effect variegation and chromatin proteins. Bioessays 14, 605–612 (1992).
206. Dorn, R., Krauss, V., Reuter, G. & Saumweber, H. The enhancer of position-effect variegation of Drosophila, E(var)3-93D, codes for a chromatin protein containing a conserved domain common to several transcriptional regulators. PNAS 90, 11376–11380 (1993).
207. Eissenberg, J.C., T.C. James, D.M. Foster-Hartnett, T. Hartnett, V. Ngan & S.C.R. Elgin, 1990. Mutation in a heterochromatin- specific chromosomal protein is associated with suppression of position-effect variegation in Drosophila melanogaster. Proc. Nat. Acad. Sci. USA 87: 9923–9927.
208. Cléard, F. & Spierer, P. Position-effect variegation in Drosophila: the modifier Su(var)3-7 is a modular DNA-binding protein. EMBO Rep 2, 1095–1100 (2001).
209. Tschiersch, B. et al. The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. The EMBO Journal 13, 3822–3831 (1994).
210. Jenuwein, T., Laible, G., Dorn, R. & Reuter, G. SET domain proteins modulate chromatin domains in eu- and heterochromatin. Cellular and Molecular Life Sciences (CMLS) 54, 80–93 (1998).

211. Ivanova, A. V., Bonaduce, M. J., Ivanov, S. V. & Klar, A. J. The chromo and SET domains of the Clr4 protein are essential for silencing in fission yeast. Nat Genet 19, 192–195 (1998).
212. Aagaard, L. et al. Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31. EMBO J. 18, 1923–1938 (1999).
213. O'Carroll, D. et al. Isolation and characterization of Suv39h2, a second histone H3 methyltransferase gene that displays testis-specific expression. Mol. Cell. Biol. 20, 9423–9433 (2000).
214. Aagaard, L., Schmid, M., Warburton, P. & Jenuwein, T. Mitotic phosphorylation of SUV39H1, a novel component of active centromeres, coincides with transient accumulation at mammalian centromeres. J. Cell. Sci. 113 (Pt 5), 817–829 (2000).
215. García-Cao, M., O'Sullivan, R., Peters, A. H. F. M., Jenuwein, T. & Blasco, M. A. Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. Nat. Genet. 36, 94–99 (2004).
216. Lehnertz, B. et al. Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr. Biol. 13, 1192–1200 (2003).
217. Müller, M. M., Fierz, B., Bittova, L., Liszczak, G. & Muir, T. W. A two-state activation mechanism controls the histone methyltransferase Suv39h1. Nat Chem Biol 12, 188–193 (2016).
218. Krouwels, I. M. et al. A glue for heterochromatin maintenance: stable SUV39H1 binding to heterochromatin is reinforced by the SET domain. J. Cell Biol. 170, 537–549 (2005).
219. Chin, H. G., Patnaik, D., Estève, P.-O., Jacobsen, S. E. & Pradhan, S. Catalytic properties and kinetic mechanism of human recombinant Lys-9 histone H3 methyltransferase SUV39H1: participation of the chromodomain in enzymatic catalysis. Biochemistry 45, 3272–3284 (2006).
220. Yamamoto, K. & Sonoda, M. Self-interaction of heterochromatin protein 1 is required for direct binding to histone methyltransferase, SUV39H1. Biochem. Biophys. Res. Commun. 301, 287–292 (2003).
221. Firestein, R., Cui, X., Huie, P. & Cleary, M. L. Set domain-dependent regulation of transcriptional silencing and growth control by SUV39H1, a mammalian ortholog of Drosophila Su(var)3-9. Mol. Cell. Biol. 20, 4900–4909 (2000).
222. Nielsen, S. J. et al. Rb targets histone H3 methylation and HP1 to promoters. Nature 412, 561–565 (2001).

223. Johnson, W. L. et al. RNA-dependent stabilization of SUV39H1 at constitutive heterochromatin. eLife 6, e25299 (2017).
224. Velazquez Camacho, O. et al. Major satellite repeat RNA stabilize heterochromatin retention of Suv39h enzymes by RNA-nucleosome association and RNA:DNA hybrid formation. Elife 6, (2017).
225. Zhang, Y.-L. et al. DCAF13 promotes pluripotency by negatively regulating SUV39H1 stability during early embryonic development. EMBO J. 37, (2018).
226. Rea, S. et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature 406, 593–599 (2000).
227. Fuks, F., Hurd, P. J., Deplus, R. & Kouzarides, T. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Res. 31, 2305–2312 (2003).
228. Fujita, N. et al. Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression. J. Biol. Chem. 278, 24132–24138 (2003).
229. Maison, C., Quivy, J.-P. & Almouzni, G. Suv39h1 links the SUMO pathway to constitutive heterochromatin. Mol Cell Oncol 3, e1225546 (2016).
230. Vaquero, A. et al. SIRT1 regulates the histone methyl-transferase SUV39H1 during heterochromatin formation. Nature 450, 440–444 (2007).
231. Bosch-Presegué, L. et al. Stabilization of Suv39H1 by SirT1 is part of oxidative stress response and ensures genome protection. Mol. Cell 42, 210–223 (2011).
232. Li, Z. et al. Inhibition of SUV39H1 methyltransferase activity by DBC1. J. Biol. Chem. 284, 10361–10366 (2009).
233. Kumari, P. et al. Sirt7 inhibits Sirt1-mediated activation of Suv39h1. Cell Cycle 17, 1403–1412 (2018).
234. Fritsch, L. et al. A subset of the histone H3 lysine 9 methyltransferases Suv39h1, G9a, GLP, and SETDB1 participate in a multimeric complex. Mol. Cell 37, 46–56 (2010).
235. Wang, L. et al. Histone Modifications Regulate Chromatin Compartmentalization by Contributing to a Phase Separation Mechanism. Mol Cell 76, 646-659.e6 (2019).
236. Erdel, F. et al. Mouse Heterochromatin Adopts Digital Compaction States without Showing Hallmarks of HP1-Driven Liquid-Liquid Phase Separation. Molecular Cell 78, 236-249.e7 (2020).

237.Schuhmacher, M. K., Kudithipudi, S., Kusevic, D., Weirich, S. & Jeltsch, A. Activity and specificity of the human SUV39H2 protein lysine methyltransferase. Biochim Biophys Acta 1849, 55–63 (2015).

238.Piao, L. et al. Automethylation of SUV39H2, an oncogenic histone lysine methyltransferase, regulates its binding affinity to substrate proteins. Oncotarget 7, 22846–22856 (2016).

239.Peters, A. H. F. M. et al. Loss of the Suv39h Histone Methyltransferases Impairs Mammalian Heterochromatin and Genome Stability. Cell 107, 323–337 (2001).

240.Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. Cell Res 21, 381–395 (2011).

241.Ng, S. S., Yue, W. W., Oppermann, U. & Klose, R. J. Dynamic protein methylation in chromatin biology. Cell. Mol. Life Sci. 66, 407–422 (2009).

242.Zhou, Y. B., Gerchman, S. E., Ramakrishnan, V., Travers, A. & Muyldermans, S. Position and orientation of the globular domain of linker histone H5 on the nucleosome. Nature 395, 402–405 (1998).

243a.Pinheiro, I. et al. Prdm3 and Prdm16 are H3K9me1 methyltransferases required for mammalian heterochromatin integrity. Cell 150, 948–960 (2012).

243b.Eissenberg, J. C. Structural biology of the chromodomain: form and function. Gene 496, 69–78 (2012).

244.Bulut-Karslioglu, A. et al. A transcription factor-based mechanism for mouse heterochromatin formation. Nat Struct Mol Biol 19, 1023–1030 (2012).

245.Saksouk, N. et al. Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. Mol Cell 56, 580–594 (2014).

246.Sakaguchi, A., Karachentsev, D., Seth-Pasricha, M., Druzhinina, M. & Steward, R. Functional characterization of the Drosophila Hmt4-20/Suv4-20 histone methyltransferase. Genetics 179, 317–322 (2008).

247.Hahn, M. et al. Suv4-20h2 mediates chromatin compaction and is important for cohesin recruitment to heterochromatin. Genes Dev. 27, 859–872 (2013).

248.Viré, E. et al. The Polycomb group protein EZH2 directly controls DNA methylation. Nature 439, 871–874 (2006).

249.Schlesinger, Y. et al. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat Genet 39, 232–236 (2007).

250.Aygün, O., Mehta, S. & Grewal, S. I. S. HDAC-mediated suppression of histone turnover promotes epigenetic stability of heterochromatin. Nat Struct Mol Biol 20, 547–554 (2013).

251. Vaquero, A., Sternglanz, R. & Reinberg, D. NAD + -dependent deacetylation of H4 lysine 16 by class III HDACs. Oncogene 26, 5505–5520 (2007).
252. Raab, J. R. et al. Human tRNA genes function as chromatin insulators. EMBO J 31, 330–350 (2012).
253. Bintu, L. et al. Dynamics of epigenetic regulation at the single-cell level. Science 351, 720–724 (2016).
254. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell 151, 994–1004 (2012).
255. Czvitkovich, S. et al. Over-expression of the SUV39H1 histone methyltransferase induces altered proliferation and differentiation in transgenic mice. Mech. Dev. 107, 141–153 (2001).
256. Chu, L. et al. SUV39H1 orchestrates temporal dynamics of centromeric methylation essential for faithful chromosome segregation in mitosis. J Mol Cell Biol 4, 331–340 (2012).
257. Rothbart, S. B. et al. Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. Nat Struct Mol Biol 19, 1155–1160 (2012).
258. Loyola, A. et al. The HP1alpha-CAF1-SetDB1-containing complex provides H3K9me1 for Suv39-mediated K9me3 in pericentric heterochromatin. EMBO Rep 10, 769–775 (2009).
259. Kato, M., Takemoto, K. & Shinkai, Y. A somatic role for the histone methyltransferase Setdb1 in endogenous retrovirus silencing. Nat Commun 9, 1683 (2018).
260. Tosolini, M. et al. Contrasting epigenetic states of heterochromatin in the different types of mouse pluripotent stem cells. Scientific Reports 8, 5776 (2018).
261. Walter, M., Teissandier, A., Pérez-Palacios, R. & Bourc'his, D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. Elife 5, (2016).
262. Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. Science Advances 5, eaaw1668 (2019).
263. John, R. M. & Lefebvre, L. Developmental regulation of somatic imprints. Differentiation 81, 270–280 (2011).
264. Duffié, R. & Bourc'his, D. Parental Epigenetic Asymmetry in Mammals. in Current Topics in Developmental Biology vol. 104 293–328 (Elsevier, 2013).
265. Seki, Y. et al. Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. Development 134, 2627–2638 (2007).

266. Hajkova, P. et al. Chromatin dynamics during epigenetic reprogramming in the mouse germ line. Nature 452, 877–881 (2008).
267. Carone, B. R. et al. High-Resolution Mapping of Chromatin Packaging in Mouse Embryonic Stem Cells and Sperm. Developmental Cell 30, 11–22 (2014).
268. Molaro, A. et al. Biparental contributions of the H2A.B histone variant control embryonic development in mice. PLOS Biology 18, e3001001 (2020).
269. Brykczynska, U. et al. Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. Nat Struct Mol Biol 17, 679–687 (2010).
270. Bernstein, B. E. et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125, 315–326 (2006).
271. Kim, J.-M., Liu, H., Tazaki, M., Nagata, M. & Aoki, F. Changes in histone acetylation during mouse oocyte meiosis. Journal of Cell Biology 162, 37–46 (2003).
272. Smallwood, S. A. et al. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet 43, 811–814 (2011).
273. Burton, A. & Torres-Padilla, M.-E. Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. Nat Rev Mol Cell Biol 15, 723–735 (2014).
274. Ladstätter, S. & Tachibana, K. Genomic insights into chromatin reprogramming to totipotency in embryos. Journal of Cell Biology 218, 70–82 (2018).
275. Tanaka, M., Hennebold, J. D., Macfarlane, J. & Adashi, E. Y. A mammalian oocyte-specific linker histone gene H1oo: Homology with the genes for the oocyte-specific cleavage stage histone (cs-H1) of sea urchin and the B4/H1M histone of the frog. Development 128, 655–664 (2001).
276. Burton, A. & Torres-Padilla, M.-E. Epigenetic reprogramming and development: a unique heterochromatin organization in the preimplantation mouse embryo. Briefings in Functional Genomics 9, 444–454 (2010).
277. Lane, N. et al. Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. Genesis 35, 88–93 (2003).
278. Santos, F., Peters, A. H., Otte, A. P., Reik, W. & Dean, W. Dynamic chromatin modifications characterise the first cell cycle in mouse embryos. Dev Biol 280, 225–236 (2005).
279. Nakamura, T. et al. PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. Nature 486, 415–419 (2012).
280. Liu, H., Kim, J.-M. & Aoki, F. Regulation of histone H3 lysine 9 methylation in oocytes and early pre-implantation embryos. Development 131, 2269–2280 (2004).

282.Puschendorf, M. et al. PRC1 and Suv39h specify parental asymmetry at constitutive heterochromatin in early mouse embryos. Nat Genet 40, 411–420 (2008).

283.Inoue, A., Jiang, L., Lu, F., Suzuki, T. & Zhang, Y. Maternal H3K27me3 controls DNA methylation-independent imprinting. Nature 547, 419–424 (2017).

284.Xu, R., Li, C., Liu, X. & Gao, S. Insights into epigenetic patterns in mammalian early embryos. Protein Cell 12, 7–28 (2021).

285.Burton, A. et al. Heterochromatin establishment during early mammalian development is regulated by pericentromeric RNA and characterized by non-repressive H3K9me3. Nat Cell Biol 22, 767–778 (2020).

286.Dahl, J. A. et al. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. Nature 537, 548–552 (2016).

287.Evsikov, A. V. et al. Systems biology of the 2-cell mouse embryo. Cytogenet Genome Res 105, 240–250 (2004).

288.Kigami, D., Minami, N., Takayama, H. & Imai, H. MuERV-L Is One of the Earliest Transcribed Genes in Mouse One-Cell Embryos1. Biology of Reproduction 68, 651–654 (2003).

289.Wu, Y. et al. Nuclear Exosome Targeting Complex Core Factor Zcchc8 Regulates the Degradation of LINE1 RNA in Early Embryos and Embryonic Stem Cells. Cell Rep 29, 2461-2472.e6 (2019).

290.Wang, C. et al. Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. Nat Cell Biol 20, 620–631 (2018).

291.Liu, X. et al. Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. Nature 537, 558–562 (2016).

292.Pauli, A., Rinn, J. L. & Schier, A. F. Non-coding RNAs as regulators of embryogenesis. Nat Rev Genet 12, 136–149 (2011).

293.Kung, J. T. Y., Colognori, D. & Lee, J. T. Long noncoding RNAs: past, present, and future. Genetics 193, 651–669 (2013).

294.Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 3, 318–356 (1961).

295.Mizuno, T., Chou, M. Y. & Inouye, M. A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). Proc Natl Acad Sci U S A 81, 1966–1970 (1984).

296.Brannan, C. I., Dees, E. C., Ingram, R. S. & Tilghman, S. M. The product of the H19 gene may function as an RNA. Mol Cell Biol 10, 28–36 (1990).

297.Brockdorff, N. et al. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell 71, 515–526 (1992).

298.Mattick, J. S. & Rinn, J. L. Discovery and annotation of long noncoding RNAs. Nat Struct Mol Biol 22, 5–7 (2015).

299.Non-coding RNAs in Colorectal Cancer. vol. 937 (Springer International Publishing, 2016).

300a.Stein, P. et al. Essential Role for endogenous siRNAs during meiosis in mouse oocytes. PLoS Genet 11, e1005013 (2015).

300b.Kim, T.-K., Hemberg, M. & Gray, J. M. Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers. Cold Spring Harb Perspect Biol 7, a018622 (2015).

301.Quinn, J. J. & Chang, H. Y. Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet 17, 47–62 (2016).

302.Guttman, M. et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458, 223–227 (2009).

303.Zheng, G. X. Y., Do, B. T., Webster, D. E., Khavari, P. A. & Chang, H. Y. Dicer-microRNA-Myc circuit promotes transcription of hundreds of long noncoding RNAs. Nat Struct Mol Biol 21, 585–590 (2014).

304.Almada, A. E., Wu, X., Kriz, A. J., Burge, C. B. & Sharp, P. A. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. Nature 499, 360–363 (2013).

305.Austenaa, L. M. I. et al. Transcription of Mammalian cis-Regulatory Elements Is Restrained by Actively Enforced Early Termination. Molecular Cell 60, 460–474 (2015).

306.Reinhardt, J. A. et al. De Novo ORFs in Drosophila Are Important to Organismal Fitness and Evolved Rapidly from Previously Non-coding Sequences. PLOS Genetics 9, e1003860 (2013).

307.Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S. & Lander, E. S. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. Cell 154, 240–251 (2013).

308.Wang, K. C. & Chang, H. Y. Molecular mechanisms of long noncoding RNAs. Mol Cell 43, 904–914 (2011).

309.Guttman, M. et al. lincRNAs act in the c controlling pluripotency and differentiation. Nature 477, 295–300 (2011).

310. Ng, S.-Y., Johnson, R. & Stanton, L. W. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. EMBO J 31, 522–533 (2012).
311. Yang, Y. W. et al. Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. eLife 3, e02046 (2014).
312. Schmitz, K.-M., Mayer, C., Postepska, A. & Grummt, I. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev 24, 2264–2269 (2010).
313. Wang, P. et al. The STAT3-binding long noncoding RNA lnc-DC controls human dendritic cell differentiation. Science 344, 310–313 (2014).
314. Chu, H.-P. et al. TERRA RNA Antagonizes ATRX and Protects Telomeres. Cell 170, 86-101.e16 (2017).
315. Kino, T., Hurt, D. E., Ichijo, T., Nader, N. & Chrousos, G. P. Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Sci Signal 3, ra8 (2010).
316. Clemson, C. M. et al. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol Cell 33, 717–726 (2009).
317. Tripathi, V. et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell 39, 925–938 (2010).
318. Gonzalez, I. et al. A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature. Nat Struct Mol Biol 22, 370–376 (2015).
319. Wang, Y. et al. Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. Dev Cell 25, 69–80 (2013).
320. Lapidot, M. & Pilpel, Y. Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. EMBO Rep 7, 1216–1222 (2006).
321. Katayama, S. et al. Antisense transcription in the mammalian transcriptome. Science 309, 1564–1566 (2005).
322. Chen, J. et al. Over 20% of human transcripts might form sense-antisense pairs. Nucleic Acids Res 32, 4812–4820 (2004).
323. Prescott, E. M. & Proudfoot, N. J. Transcriptional collision between convergent genes in budding yeast. Proc Natl Acad Sci U S A 99, 8796–8801 (2002).
324. Yap, K. L. et al. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Mol Cell 38, 662–674 (2010).

325.Dinger, M. E. et al. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res 18, 1433–1445 (2008).

326.Tsai, M.-C. et al. Long noncoding RNA as modular scaffold of histone modification complexes. Science 329, 689–693 (2010).

327.Hastings, M. L., Milcarek, C., Martincic, K., Peterson, M. L. & Munroe, S. H. Expression of the thyroid hormone receptor gene, erbAalpha, in B lymphocytes: alternative mRNA processing is independent of differentiation but correlates with antisense RNA levels. Nucleic Acids Res 25, 4296–4300 (1997).

328.Carrieri, C. et al. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. Nature 491, 454–457 (2012).

329. Mlynarczyk, S. K. & Panning, B. X inactivation: Tsix and Xist as yin and yang. Current Biology 10, R899–R903 (2000).

330. Lin, N. et al. An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. Mol Cell 53, 1005–1019 (2014).

331.Jiapaer, Z. et al. LincU Preserves Naïve Pluripotency by Restricting ERK Activity in Embryonic Stem Cells. Stem Cell Reports 11, 395–409 (2018).

332.Bond, A. M. et al. Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. Nat Neurosci 12, 1020–1027 (2009).

333.Scarola, M. et al. Epigenetic silencing of Oct4 by a complex containing SUV39H1 and Oct4 pseudogene lncRNA. Nat Commun 6, 7631 (2015).

334.Koerner, M. V., Pauler, F. M., Huang, R. & Barlow, D. P. The function of non-coding RNAs in genomic imprinting. Development 136, 1771–1783 (2009).

335.Pandey, R. R. et al. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell 32, 232–246 (2008).

336.Sleutels, F., Zwart, R. & Barlow, D. P. The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature 415, 810–813 (2002).

337.Lee, J. T., Davidow, L. S. & Warshawsky, D. Tsix, a gene antisense to Xist at the X-inactivation centre. Nat Genet 21, 400–404 (1999).

338.Navarro, P. et al. Molecular coupling of Xist regulation and pluripotency. Science 321, 1693–1695 (2008).

339.Navarro, P. et al. Molecular coupling of Tsix regulation and pluripotency. Nature 468, 457–460 (2010).

340.Ohhata, T. et al. Dynamics of transcription-mediated conversion from euchromatin to facultative heterochromatin at the Xist promoter by Tsix. Cell Reports 34, 108912 (2021).

341. Heurtier, V. et al. The molecular logic of Nanog-induced self-renewal in mouse embryonic stem cells. Nat Commun 10, 1109 (2019).

342. Microfabricated Device for High-Resolution Imaging of Preimplantation Embryos | SpringerLink. https://link.springer.com/protocol/10.1007/978-1-0716-0958-3_2.

343. Wu, K. et al. SETDB1-Mediated Cell Fate Transition between 2C-Like and Pluripotent States. Cell Reports 30, 25-36.e6 (2020).

344. Fakhr, E., Zare, F. & Teimoori-Toolabi, L. Precise and efficient siRNA design: a key point in competent gene silencing. Cancer Gene Ther 23, 73–82 (2016).

345. Wang, L. et al. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res 41, e74 (2013).

346. Kang, Y.-J. et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic Acids Res 45, W12–W16 (2017).

347. Deng, Q., Ramskold, D., Reinius, B. & Sandberg, R. Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells. Science 343, 193–196 (2014).

348. Karlic, R. et al. Long non-coding RNA exchange during the oocyte-to-embryo transition in mice. DNA Res. 24, 129–141 (2017).

349. Karimi, M. M. et al. DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs. Cell Stem Cell 8, 676–687 (2011).

350. Montavon, T. et al. Complete loss of H3K9 methylation dissolves mouse heterochromatin organization. Nat Commun 12, 4359 (2021).

351. Miao, H. et al. A long noncoding RNA distributed in both nucleus and cytoplasm operates in the PYCARD-regulated apoptosis by coordinating the epigenetic and translational regulation. PLoS Genet 15, e1008144 (2019).

352. Jagus, R., Joshi, B., Miyamoto, S. & Beckler, G. S. In Vitro Translation. Current Protocols in Cell Biology 00, (1998).

353. Bouniol, C., Nguyen, E. & Debey, P. Endogenous Transcription Occurs at the 1-Cell Stage in the Mouse Embryo. Experimental Cell Research 218, 57–62 (1995).

354. Greenberg, M. V. C. et al. Transient transcription in the early embryo sets an epigenetic state that programs postnatal growth. Nat Genet 49, 110–118 (2017).

355. Probst, Aline. V. et al. A Strand-Specific Burst in Transcription of Pericentric Satellites Is Required for Chromocenter Formation and Early Mouse Development. Developmental Cell 19, 625–638 (2010).

356. Sankar, A. et al. KDM4A regulates the maternal-to-zygotic transition by protecting broad H3K4me3 domains from H3K9me3 invasion in oocytes. Nat Cell Biol 22, 380–388 (2020).
357. Iyyappan, R. et al. Oocyte specific lncRNA variant Rose influences oocyte and embryo development. Non-coding RNA Research 6, 107–113 (2021).