

THESIS / THÈSE

MASTER EN BIOCHIMIE ET BIOLOGIE MOLÉCULAIRE ET CELLULAIRE

Les mécanismes responsables de la déméthylation globale du génome des cellules souches embryonnaires femelles de souris

HUART, CAMILLE

Award date:
2017

Awarding institution:
Université de Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Faculté des Sciences

**LES MECANISMES RESPONSABLES DE LA DEMETHYLATION GLOBALE DU GENOME
DES CELLULES SOUCHES EMBRYONNAIRES FEMELLES DE SOURIS**

**Mémoire présenté pour l'obtention
du grade académique de master 120 en biochimie et biologie moléculaire et cellulaire**

Camille HUART

Janvier 2017

Université de Namur
FACULTE DES SCIENCES
Secrétariat du Département de Biologie
Rue de Bruxelles 61 - 5000 NAMUR
Téléphone: + 32(0)81.72.44.18 - Téléfax: + 32(0)81.72.44.20
E-mail: joelle.jonet@unamur.be - <http://www.unamur.be>

Les mécanismes responsables de la déméthylation globale du génome des cellules souches embryonnaires femelles de souris

HUART Camille

Résumé

Une étude a montré une diminution de la méthylation de l'ADN dans les cellules souches embryonnaires femelles de souris lorsque celles-ci sont comparées à leurs homologues mâles. L'objectif du présent mémoire est de trouver les mécanismes qui sont impliqués dans cette hypométhylation de l'ADN. Plusieurs expériences ont été mises en place. Premièrement, nous avons voulu vérifier si les lignées de cellules souches embryonnaires femelles et mâles de notre laboratoire (gC1 et E14TG2A) présentent un niveau de méthylation similaire à ce qui a été publié dans plusieurs études. Pour se faire, nous avons réalisé des études de pyroséquençage afin de comparer le niveau de méthylation de l'ADN entre les cellules mâles et femelles. Nous avons également réalisé des analyses d'expression de l'ARN de plusieurs gènes pouvant avoir une importance majeure sur le niveau de méthylation de l'ADN. Deuxièmement, nous avons essayé de générer une lignée cellulaire de cellules souches embryonnaires mâles avec un rapporteur de méthylation. Cette lignée cellulaire nous aurait permis de suivre la méthylation de l'ADN au cours du temps afin de réaliser un « screen » de l'ADN génomique de souris contenant des gènes impliqués dans le processus de déméthylation. Finalement, nous avons essayé de développer une stratégie afin de retirer l'un des centres de l'inactivation du chromosome X dans des cellules souches embryonnaires femelles, afin d'étudier si cette région pouvait être impliqué dans la déméthylation de l'ADN des cellules femelles.

Mémoire de master 120 en biochimie et biologie moléculaire et cellulaire

Janvier 2017

Promoteur: Professeur Olivier De Backer

Université de Namur
FACULTE DES SCIENCES
Secrétariat du Département de Biologie
Rue de Bruxelles 61 - 5000 NAMUR
Téléphone: + 32(0)81.72.44.18 - Téléfax: + 32(0)81.72.44.20
E-mail: joelle.jonet@unamur.be - <http://www.unamur.be>

Mechanisms responsible for global DNA demethylation in female mouse embryonic stem cells

HUART Camille

Summary

A study showed a decrease of DNA methylation in female mouse embryonic stem cells when being compared to their male counterparts. The objective of this present work was to identify the mechanisms responsible for this DNA hypomethylation. Several experiments have been designed. First, we wanted to characterize our female and male embryonic stem cell line (gC1 and E14TG2A) to see if these cell lines have similar levels of DNA methylation to what has been described in several publications. To do so, we carried out a pyrosequencing study to assess the levels of DNA methylation in male and female embryonic stem cells. We also developed an experiment to analyse the levels of expression of several genes which could be implicated in the levels of DNA methylation. Second, we tried to generate a male embryonic stem cell line with a reporter of endogenous methylation. This cell line would have allowed us to follow changes in DNA methylation over time, in order to « screen » a genomic library of mouse DNA containing genes which are potentially implicated in the process of DNA demethylation in female embryonic stem cells. Finally, we developed a strategy to delete one of the X inactivation center in female mouse embryonic stem cells, to observe if this region is implicated in the DNA demethylation of female cells.

Mémoire de master 120 en biochimie et biologie moléculaire et cellulaire

Janvier 2017

Promoteur: Professeur Olivier De Backer

Thank you to Professor Olivier De Backer for welcoming me in his laboratory, giving me numerous advices and suggestions during my master's thesis.

.....

Thank you to Professor Damien Hermand, Professor Jean-Yves Matroule, Mrs. Emeline Puissant and Mrs. Géraldine Sana for taking the time to read this document.

.....

Thank you to Olivier for sharing his knowledge with me, being so enthusiastic, patient and simply for being a great mentor.

.....

Thank you to Mrs. Elodie Falisse and Mrs. Enora Flamion for their precious help to design the pyrosequencing experiment.

.....

Thank you to Axelle, Cédric, Coraline, Dominique, Eve, Géraldine, Lucie, Manon, Marie, Olivier and Typhanie for helping me out and giving me such a supportive and friendly environment to work in.

.....

Thank you to my parents and my grandparents for encouraging me, helping me, and most importantly believing in me during my master's degree.

.....

Thank you to Alexis, Alice, Chloé, Louise, Olivia, Pauline and Sophie for being such great friends.

.....

Thank you to Mrs. Noël, Mrs. Bertoluzza and Mr. Christians, three of my high school teachers.

.....

Thank you to all the Professors at the University of Namur for their time, availability and giving me such a great learning experience during these five years.

.....

Thank you to sci-hub.cc for giving me access to all my favourite scientific articles at any time of the day.

TABLE OF CONTENTS

A. ABBREVIATIONS	1
B. INTRODUCTION	3
1. DNA METHYLATION	3
1.1. <i>The three families of DNA methyltransferases</i>	3
1.2. <i>Imprinting</i>	3
1.3. <i>Epigenetic reprogramming</i>	4
1.4. <i>Random X inactivation</i>	5
2. HISTONE MODIFICATIONS	6
3. DNA METHYLATION AND FEMALE MOUSE ES CELLS	6
4. GENES POTENTIALLY IMPLICATED IN DNA DEMETHYLATION OF FEMALE ES CELLS	9
5. MARKERS OF THE DNA DEMETHYLATION OF FEMALE ES CELLS	9
6. THE CRISPR-Cas9 SYSTEM	10
7. OBJECTIVES OF THIS PROJECT	12
C. RESULTS	13
1. CHARACTERISATION OF THE FEMALE gC1 AND MALE E14TG2A ES CELL LINES	13
1.1. <i>Assessing the presence of two X chromosomes in the female gC1 ES cell line</i>	13
1.2. <i>Methylation status of female and male ES cells</i>	14
1.3. <i>Expression analysis of several genes with a possible implication in the DNA demethylation of female mouse embryonic stem cells</i>	18
2. UNBIASED SCREEN FOR IDENTIFYING DNA DEMETHYLATION PROMOTING GENE(S)	18
3. DELETION OF THE X INACTIVATION CENTER IN XX ES CELLS	22
D. DISCUSSION AND PERSPECTIVES	24
E. MATERIALS AND METHODS	27
F. REFERENCES	36
G. SUPPLEMENTARY DATA	40

A. ABBREVIATIONS

5 mC	5 methylcytosine
5-azadC	5 aza-2'deoxyctidine
C	Cytosine
Cas9	CRISPR associated protein 9
Cas9n	CRISPR associated protein 9 nickase
CpG	Cytosine-phospho-guanine
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DMR	Differently Methylated Region
Dnmt	DNA methyltransferase
Dnmt3L	DNA methyltransferase 3-like
DSB	Double strand break
eGFP	Enhanced GFP
ELISA	Enzyme linked immunosorbent assay
ES cell	Embryonic stem cell
FBS	Fetal bovin serum
FISH	Fluorescent in situ hybridization
GSK3	Glycogen Synthase Kinase-3
H	Histidine or HpaII
HDR	Homology Directed Repair
HygroR	Hygromycin B resistance gene
IAP	Intracisternal A particle retrotransposon
ICM	Inner cell mass
ICR	Imprinting control regions
IRES	Internal ribosome entry site
LIF	Leukemia inhibitory factor
M	MspI
MEK	Mitogen-Activated Protein Kinase Kinase
mRNA	Messenger RNA
NeoR	Neomycin resistance gene

NHEJ	Non Homologous End Joining
PAM	Protospacer recognition motifs
PGC	Primordial germ cell
PuroR	Puromycin resistance gene
R	Arginine
sgRNA	Single guide RNA
SNP	Single nucleotide polymorphism
T	Thymine
U	Uracil
XCI	X chromosome inactivation
XIC	X inactivation center

B. INTRODUCTION

The genome corresponds to the complete set of genetic information found within a single cell of an organism. In the mouse, the genome is stored in the DNA. The epigenome refers to chemical modifications added to DNA and histones^{1,2}. These modifications can among other things, influence gene expression and aid in the maintenance of genomic stability^{1,3,4}.

1. DNA methylation

Methylation of the fifth carbon of cytosine (C) is the best characterized epigenetic modification. The mammalian genome codes for three families of DNA methyltransferases (*Dnmt1*, *Dnmt2*, *Dnmt3*) which establish and maintain DNA methylation^{2,5,6}. DNMT1 and DNMT3 possess two domains: one N-terminal and one C-terminal, respectively referred to as the “regulatory” and “catalytic” domain⁷. DNMT2 only possess the “catalytic” domain⁷.

In the mammalian genome, cytosine methylation mainly occurs at cytosines followed directly by a guanine which form what is commonly called a CpG dinucleotide - cytosine-phosphoguanine. CpGs are not evenly distributed across the genome^{8,9}. They tend to form “CpG islands”, which are short regions of 0.5 to 4 kb in length, they deviate from the average genomic pattern by being GC-rich and CpG-rich¹⁰. They are often located in promoter regions and play a role in gene regulation². Generally, “CpG islands” associated with a promoter region are devoid of DNA methylation^{2,11}. The switch from unmethylated state to methylated state is known to turn off transcription of the associated genes.

1.1. The three families of DNA methyltransferases

Dnmt1 is responsible for the maintenance of DNA methylation by reproducing the DNA methylation patterns present on the original strand onto the newly synthesized DNA strand. This process provides a way for DNA methylation patterns to be transmitted after DNA replication¹⁰.

Dnmt2 is a RNA methyltransferase, although its “catalytic” domain is similar to that of other Dnmt, it doesn't seem to be implicated in DNA methylation⁷.

Dnmt3a and *Dnmt3b* are generally considered as *de novo* DNA methyltransferases, even though recent studies have shown that they might be implicated in the maintenance of DNA methylation as well⁵. They are predominantly expressed during early developmental stages when the establishment of *de novo* DNA methylation patterns occurs. Recently, a new member of the Dnmt3 family has been discovered, namely DNA methyltransferase 3-like (*Dnmt3L*). This gene doesn't seem to display a methyltransferase activity, but appears to be essential to the establishment of methylation patterns⁷.

1.2. Imprinting

DNA methylation plays an important role in imprinting, an epigenetic phenomenon which restricts the expression of a gene to one of the two parental alleles¹². The number of imprinted genes identified so far in the mouse is around 150. Imprinted genes are often found in clusters,

though genes from one cluster are not necessarily expressed from the same parental allele. Each cluster has an imprinting control region (ICR) on which parental allele-specific DNA methylation occurs to establish imprinting¹³.

In this work, we tried to analyse the methylation profile of two imprinted genes: *H19* and *Igf2r*. The ICR which controls the expression of *H19* is normally methylated on the paternal allele and unmethylated on the maternal allele, resulting in the expression of the maternal allele. Conversely, the ICR of *Igf2r* is methylated on the maternal allele and unmethylated on the paternal allele, leading to transcription of the paternal allele. The profiles of DNA methylation of these genes was supposed to be analysed in both female (XX) and male (XY) embryonic stem cells.

1.3. Epigenetic reprogramming

In mammalian cells, DNA methylation patterns are seen as relatively stable. They have to be however, profoundly reprogrammed to restore developmental potency¹⁴. These epigenetic reprogramming events occur through passive and active DNA demethylation. Passive DNA demethylation leads to the decrease of DNA methylation levels due to a failure of DNMT1 to maintain methylation after DNA replication. While, active DNA demethylation actively transformed 5 methylcytosines (5 mC). The first global epigenetic reprogramming occurs after fertilization. The cells of the early embryo undergo rapid DNA demethylation, which reaches its lowest point around E3.5 in the mouse^{14,15}. A few genomic regions escape this reprogramming, such as intracisternal A particle retrotransposons (IAPs), centromeric repeats and imprinted genes^{14,16}. It has been demonstrated that imprinted genes are protected from DNA demethylation to preserve the parent-of-origin specific gene expression¹⁴. IAPs and centromeric repeats respectively remain methylated to prevent retrotransposition and genomic instability¹⁷. After this wave of demethylation, the pluripotent cells of the inner cell mass (ICM) of the blastocyst progressively re-methylate, and differentiate into specific somatic cell types¹⁴. The second epigenetic reprogramming occurs in the primordial germ cells (PGCs) in the developing embryo to reset imprinting¹⁸ (**Figure 1**).

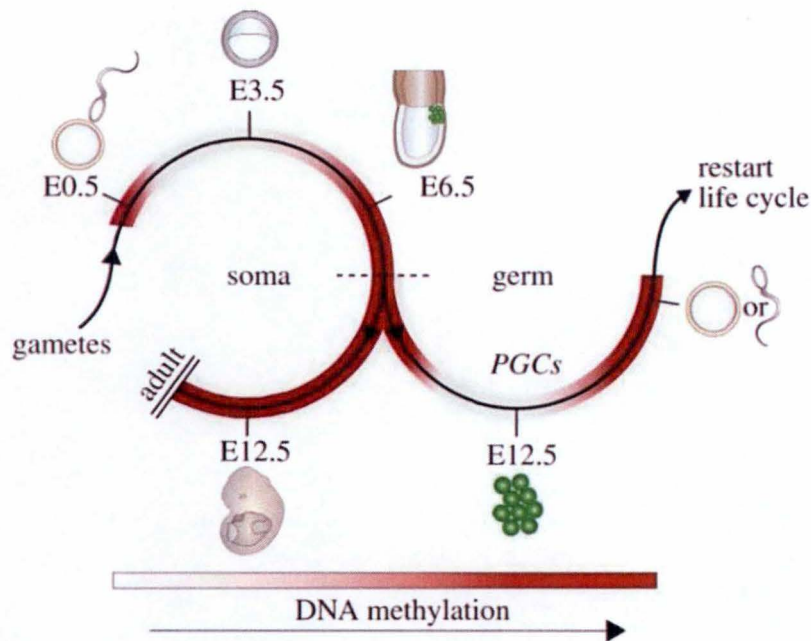


Figure 1 - Illustration of DNA methylation dynamics during the mammalian life cycle showing when and where epigenetic reprogramming takes place. After fertilization, both the paternal and maternal genome get demethylated. The cells of the ICM progressively re-methylate to reach a peak around E6.5. At E6.5, a fraction of cells differentiate into PGCs which get demethylated to start a new life cycle ¹⁹.

1.4. Random X inactivation

This epigenetic process consists in random silencing of one of the two X chromosomes in mammalian female cells. X chromosome inactivation ensures that expression of X-linked genes is balanced between XX and XY cells ²⁰. The paternal X chromosome is inactivated in extra-embryonic tissues while, in embryonic tissues, X inactivation can affect the paternal or maternal X chromosome ²¹. The X inactivation centre (XIC) located on the X chromosome, is necessary and sufficient for X inactivation to occur ²⁰⁻²². The presence of this region was revealed in mouse embryos and embryonic cells which carried translocated or truncated X chromosomes. Indeed, it was noticed that random X chromosome inactivation could only be triggered when two XICs were present ²³. *Xist*, a non-coding RNA which lies at the center of XIC, triggers XCI by coating the entire X chromosome, leading to chromosome-wide silencing ²¹. Transcription of *Xist* is negatively regulated by *Tsix*, an antisense RNA which is transcribed across the *Xist* locus ^{20,22}. Before XCI, *Tsix* is highly expressed from both X chromosomes, effectively repressing the transcription of *Xist*. Downregulation of *Tsix* on the future inactivated X chromosome allows initiation of XCI by enabling the transcription of *Xist*, and coating of one X chromosome. Once XCI is established, it is maintained through mitosis in all daughter cells ²⁴.

2. Histone modifications

Histones undergo epigenetic modifications such as methylation, acetylation and phosphorylation. Some of these modifications prevent the transcription of genomic repetitive sequences²⁵. A histone variant named H3.3, is often associated with heterochromatin which is a tightly packed form of DNA. The complex ATRX/DAXX is essential for the deposition of H3.3 at these heterochromatin regions. The enrichment in ATRX/DAXX/H3.3 also correlates with the presence of H3K9me3¹, a histone modification which is important for DNA methylation and genome stability²⁶. Indeed, knocking out of the histone methyltransferase, SETDB1 or of KAP-1 responsible for the recruitment of SETDB1, leads to a reduction of H3K9me3 at endogenous retroviruses in mouse ES cells. The reduction of H3K9me3 leads to the expression of several endogenous retroviruses classes, which increases genomic instability²⁷. In the genome of ES cells, regions of DNA capable of maintaining DNA methylation such as some transposons, heterochromatic repeats and differently methylated regions (DMR) of imprinted genes are associated with H3K9me3^{28,29,30}.

3. DNA methylation and female mouse ES cells

Mouse ES cells are derived from the ICM of blastocysts before random XCI occurs. It has been demonstrated that the presence of two active X chromosomes in female ES cells correlates with a reduced level of DNA methylation when compared to male ES cells³¹ (**Figure 2**). A global analysis demonstrated that 35% of the CpG sites are methylated in female ES cells, while 70% of these sites are methylated in male ES cells³¹.

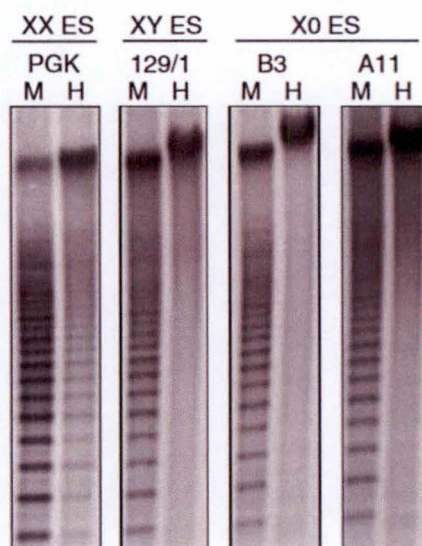


Figure 2 – Southern Blot analysis on XY ES cells (129/1) and XX ES cells (PGK) on major satellite sequences. DNA was digested with *MspI* (M) – a DNA methylation insensitive enzyme; and with *HpaII* (H) – a DNA methylation sensitive enzyme. *HpaII* digestion is increased in XX ES cells indicating that DNA of these cells is hypomethylated³¹.

¹ Methylation of lysine on H3

This loss of DNA methylation also affected regions which are generally considered to be resistant to demethylation, such as imprinted genes and IAPs^{15,28,31}. Moreover, their paper also reported difficulties to maintain female ES cells in culture due to an unstable karyotype: the cells tend to lose one of the two X chromosomes³¹. This instability has been proposed to be a consequence of DNA hypomethylation³¹. Since DNA methylation is restored in cells that lose one of their X chromosome (X0), the presence of two active X chromosomes must be responsible for the demethylation of XX ES cells³¹. DNMT3A, and to a lesser extent DNMT3B, are downregulated in XX ES cells when compared to XY ES cells, while DNMT1 expression is similar in both XX and XY ES cells^{31,32} (**Figure 3**). Ectopic expression of DNMT3A and DNMT3B in female ES cells restores to some extent the levels of DNA methylation indicating that these enzymes play a role in the decrease of DNA methylation observed in these cells³¹. It was later elucidated that the DNA hypomethylation observed in XX ES cells is in some part due to a decreased in transcript levels of *Dnmt3a* and *Dnmt3b*³².

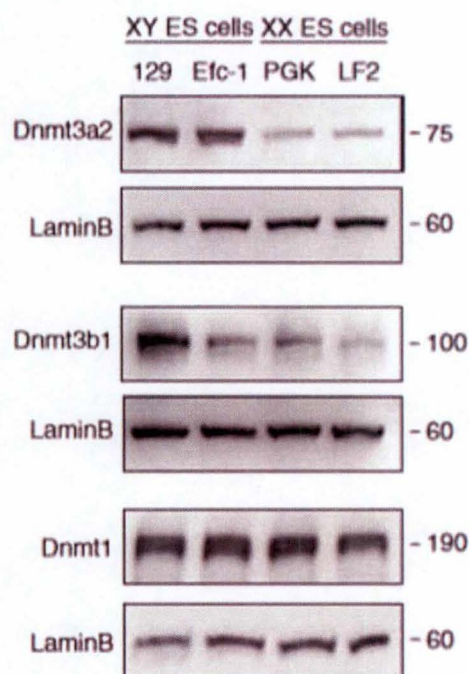


Figure 3 – Western Blot analysis of DNA methyltransferases on cell extract from XY ES cells (129 and Efc-1) and XX ES cells (PGK and LF2). LaminB was used as a loading control. Levels of DNMT3A2 (the major isoform of DNMT3A in ES cells) is lower in XX than in XY ES cell lines. The levels of DNMT3B1 (the major isoform of DNMT3B in ES cells) also seemed to be lower in XX cells, but it is subject to change depending on the different cell lines³¹.

The level to which female ES cells lose DNA methylation seems to be dependent on culture conditions. ES cells are commonly grown in a medium containing fetal bovine serum (FBS) and leukemia inhibitory factor (LIF), which maintains pluripotency through activation of the STAT3 signalling^{33,34}. In the pre-implanted embryo, ES cells are referred to as 'naive', and they become 'primed' for differentiation once the blastocyst is implanted. The cells which are cultured in medium containing FBS and LIF recapitulate this 'primed' state³⁵.

Recent studies use a serum-free medium, also named the 2i medium, which contains 2 inhibitors named PD0325901 and CHIR99021 which respectively target the mitogen-activated protein kinase (MEK) and the glycogen synthase kinase-3 (GSK3). This medium maintains the 'naïve' pluripotency state by inhibiting both the MAPK and GSK3 pathways and by activating the AKT pathway^{33,34} (**Figure 4**). Both XX and XY ES cells exhibit a decrease in global DNA methylation in the serum-free medium^{28,32,36}. After several passages in the 2i medium, the expression of *Dnmt3a* and *Dnmt3b* is reduced. This similarity between the effects of the presence of 2 active X chromosomes and the use of the 2i medium points to common mechanisms³². Actually, it has been shown that the presence of two active X chromosomes parallels the effect of the 2i medium, through the inhibition of the MAPK and GSK3 pathways and the stimulation of the AKT pathway. This might serve to slow down differentiation as long as one of the X chromosome has not been inactivated³².

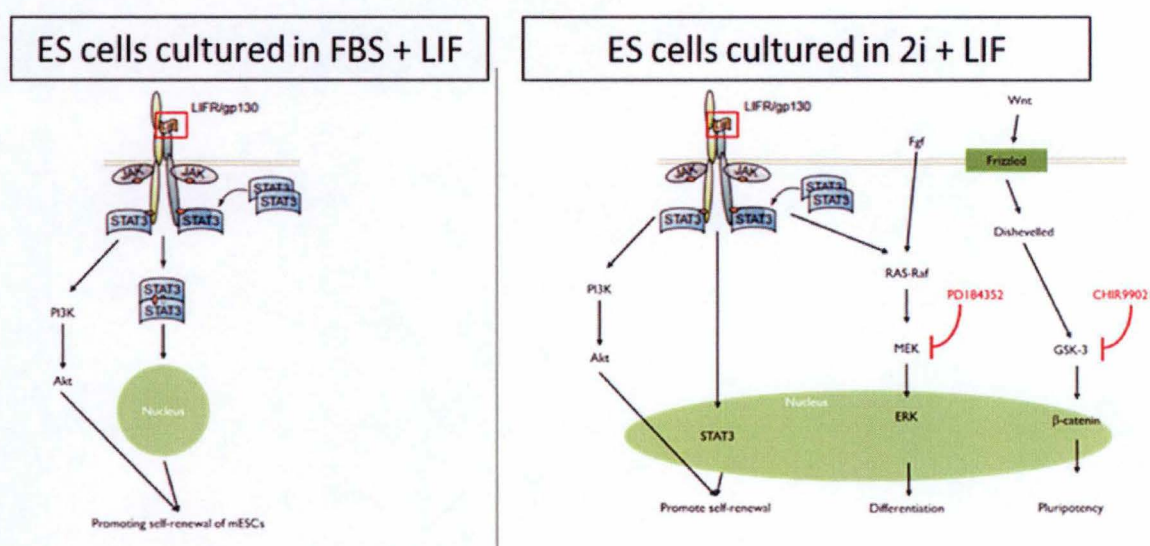


Figure 4 – Illustration of the different signalisation pathways involved in the repression of differentiation by the inhibitors used in ES cells culture. (A) Activation of the AKT pathway promotes self-renewal of ES cells. (B) Inhibition of MAPK and GSK3 pathways lead respectively to preventing the cells from differentiating while enhancing pluripotency. The AKT pathway is also activated in 2i medium.

However, the pathways implicated in the DNA hypomethylation of XX ES cells discovered by the usage of the 2i medium, can't explain all the differences between the DNA methylation status of XX and XY ES cells. First, the levels of DNA methylation is further reduced in female ES cells than in male ES cells when cultivated in the 2i medium²⁸. Second, only female ES cells seem to lose methylation on imprinted regions and IAP elements^{28,36}. Furthermore, male ES cells knockout for *Dnmt3a* and *Dnmt3b* exhibit higher DNA methylation levels than female ES cells, and these male cells conserve DNA methylation at imprinted regions³⁷. The cause of these differences between male and female ES cells remains to be identified.

4. Genes potentially implicated in DNA demethylation of female ES cells

In the present work, we developed a strategy to investigate the expression of several X-linked genes in XX and XY ES cells that could be involved in the loss of DNA methylation of female ES cells, based on their functions described in the literature.

Mecp2, a gene located on the X chromosome, is known to bind to methylated cytosines to mediate transcriptional repression of the associated genes^{38,39}. We hypothesised that *Mecp2* could be implicated in the demethylation process observed in XX cells even if it has never been identified as a regulator of the levels of DNA methylation.

The X-linked gene, *Dusp9* is an inhibitor of the MAPK pathway whose expression leads to a downregulation in the expression of *Dnmt3a* and *Dnmt3b*. In XX ES cells, knockdown and overexpression of *Dusp9* only had minor effects on DNA methylation³². However, contradictory results using XY ES cells have also been obtained where it was observed that overexpression of *Dusp9* lead to reduced DNA methylation⁴⁰.

Another X-linked gene which could potentially be implicated is *Eras*. It is an activator of the PI3K/AKT pathway³². As mentioned previously, AKT signalling controls the pluripotent state of the celles.

Atrx, located on the X chromosome, could be involved in the demethylation process through its aforementioned role in depositing H3.3 which allows the maintenance of H3K9me3 heterochromatin in the genome. The regions of the genome of ES cells which are capable of maintaining DNA methylation, are associated with H3K9me3. We hypothesised that the mechanism involved in the deposition of H3K9me3 could be impaired in XX ES cells.

Dicer-deficient cells display a global DNA demethylation similar to female ES cells^{41,42}. *Dicer*, a member of the RNase III family, is implicated in RNA interference by cleavage of double stranded RNA. The global DNA demethylation observed in Dicer-deficient cells was attributed to a downregulation of the miR-290 family^{41,42}. To our knowledge, analysis expression of *Dicer* hasn't been investigated in female ES cells. We would like to characterize its expression to investigate if *Dicer* could be implicated in the phenotype of female ES cells.

5. Markers of the DNA demethylation of female ES cells

In the present work, we also wanted to analyse the expression of several genes which could be use as "markers" of DNA demethylation in ES cells. These markers could help us validate our quantitative PCR analysis.

Prdm14, a known pluripotency marker, is upregulated in ES cells cultivated in serum free medium and is involved in the downregulation of *de novo* methyltransferases^{32,36}. It has already been shown that expression of *Prdm14* is influenced by the presence of two active X chromosomes³².

The expression of several *Magea* genes – we chose to focus on *Magea5* and *Magea8* – is upregulated in the demethylated XX ES cells³².

6. The CRISPR-Cas9 system

We used the CRISPR/Cas9 system in our project to perform genome editing to introduce a reporter of endogenous methylation in the genome of XY ES cells and to attempt to delete one of the two X inactivation centers in XX ES cells. In both cases, a DNA double-stranded break (DSB) has to be introduced in ES cells. We chose to use the CRISPR/Cas9 system which is based on RNA-guided nuclease which was originally discovered in bacteria and archaea where it provides adaptive immunity, since it offers a variety of advantages (**Figure 5**)^{43,44}.

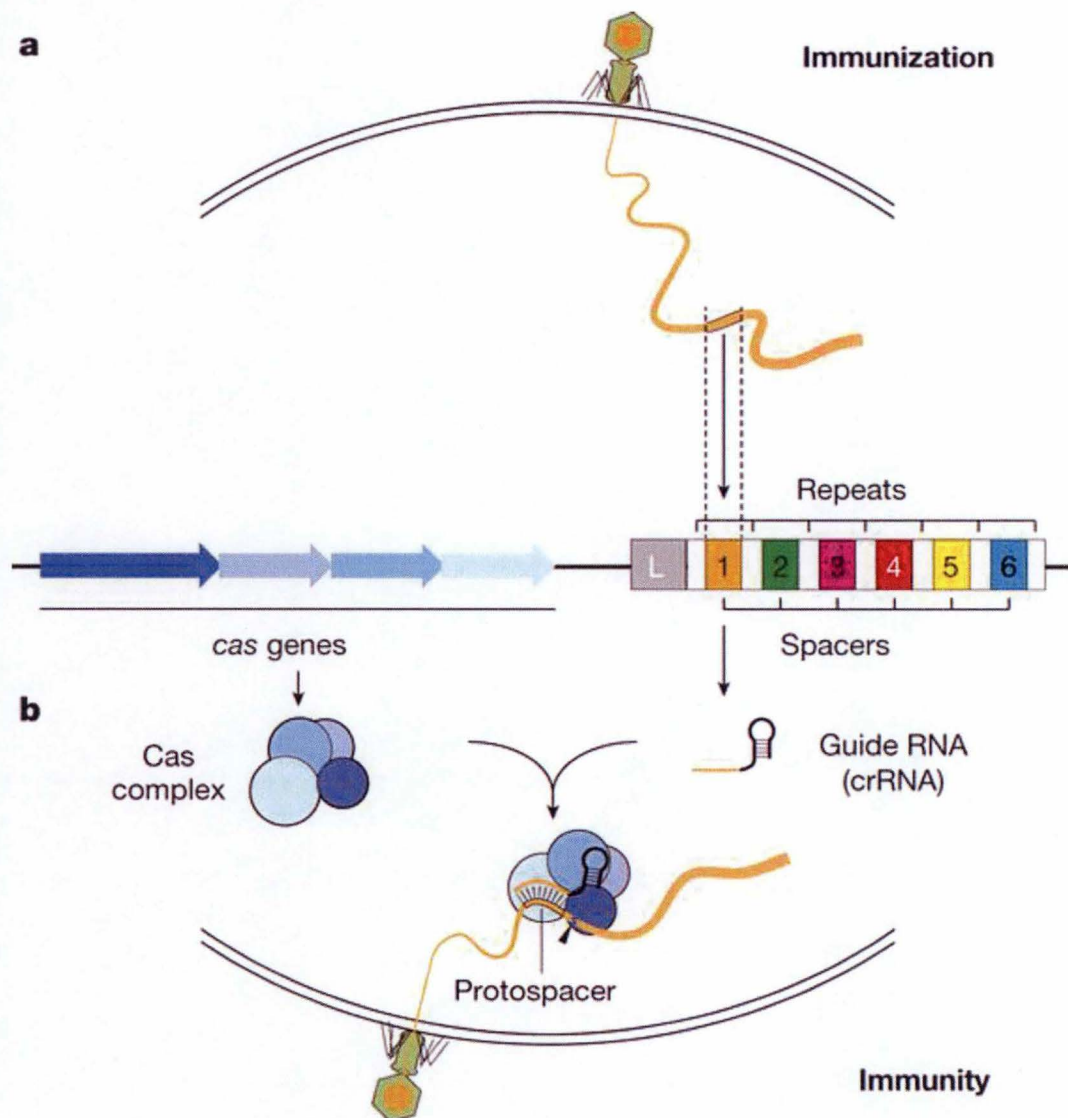


Figure 5 – The CRISPR/Cas9 immunity provides a defence mechanism against viruses and foreign DNA. (A) The CRISPR array is composed of repeated elements (represented by the white boxes) interspaced by the spacers (represented by the coloured boxes) which originates from virus or plasmids. The spacers are acquired during an infection stage, this phase is being referred to as ‘immunization’. (B) The CRISPR array is transcribed into the guide RNA (crRNA), whose function is guide the Cas9 endonuclease to protospacers which will result in the cleavage of the foreign DNA⁴⁵.

Three types of CRISPR-Cas systems have been discovered: Type I, Type II, and Type III ⁴³. The Type II CRISPR system composed of the *Streptococcus pyogenes* Cas9 nuclease and a RNA guide is being used as a genome engineering tool ⁴³. This nuclease, combined with the guide RNA, is able to generate targeted DSBs in the genome of mammalian cells as long as the DNA sequence which is targeted is immediately followed by a 5'-NGG protospacer adjacent motif (PAM). The DSBs generated by this system are preferentially repaired by the error-prone Non-Homologous End Joining (NHEJ), which often induces insertions or deletions, these DSBs can also be repaired by Homology Directed Repair (HDR) when a recombination template (which can be endogenous or introduced in the cells) is present (**Figure 6**).

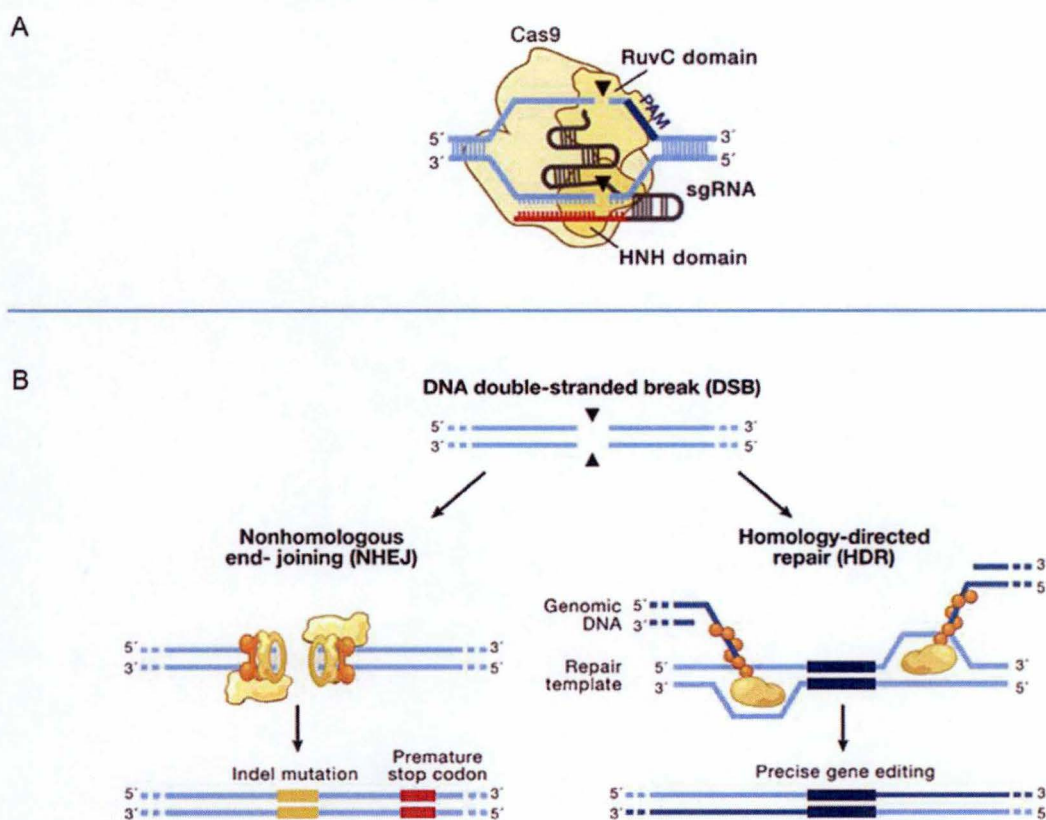


Figure 6 – The CRISPR/Cas9 system for genome editing. (A) Illustration of the Cas9 nuclease which localizes to a specific DNA sequence via the guide RNA represented in red, directly followed by a 5'-NGG PAM motif (in blue). (B) Illustration of the repair pathways: non-homologous end joining or homology-directed repair involved in repairing a double-strand breaks ⁴⁶.

A potential downside of the CRISPR/Cas9 system is its off-target activity. Although the target specificity is ensured by the guide RNA (which is 20 bp long), single mismatches between the guide and the targeted DNA are deleterious for cleavage efficiency when these mismatches are placed in the 8-14 bp region immediately upstream of the PAM, while they are tolerated elsewhere in the guide sequence ⁴⁷. A solution to reduce off-target activity is to use the SpCas9 nickase which generates a nick on the DNA strand of the targeted sequence. The SpCas9 nickase can be combined with a pair of guide RNAs to introduce simultaneous nicks. The cells

perceive double nicks as a DSB, which is repaired either by NHEJ or by HDR⁴⁸. Yet another strategy was developed to improve specificity, which is to use the so-called “enhanced specificity” nuclease. The latter binds more weakly to DNA due to the introduction of mutations in the SpCas9 domain to stabilize the non-targeted DNA strand. The “enhanced” specificity nuclease maintains robust on-target activity resulting in reduced off-target activity while on-target activity remains unchanged⁴⁹.

7. Objectives of this project

The final objective of our project was to shed light on the mechanisms by which the presence of two active X chromosomes influences DNA methylation in female ES cells.

Our first hypothesis is that an X-linked gene is responsible for this phenotype. To identify this gene, we wanted to screen a cosmid mouse genomic library by transfecting it into XY ES cells. We were hoping that some genes would promote DNA demethylation in the transfected cells. To trace DNA methylation in the transfected cells, we decided to insert a reporter of endogenous methylation developed by Stelzer et al., 2015. This construct had to be stably integrated near the 5' region of the DMR of *H19*.

Our second hypothesis is that DNA demethylation in XX ES cells could be explained by the presence of two active X inactivation centers. The female cells could detect their presence, and prepare for XCI by establishing a specific epigenetic environment, which would lead to DNA hypomethylation by among other things, downregulating the expression of *de novo* methyltransferases. To test this hypothesis, we wanted to delete the X inactivation centre on one X chromosome using the CRISPR/Cas9 system.

C. RESULTS

1. Characterisation of the female gC1 and male E14TG2A ES cell lines

To understand the causes of DNA hypomethylation in female ES cells, we decided to compare a female cell line (gC1) to a male cell line (E14TG2A) since the levels of DNA methylation can fluctuate between the different cell lines³¹. In our laboratory, we only had the male cell line, E14TG2A which had been derived from 129/OLA. The female gC1 ES cell line was derived from a 129Sv/M. m castaneus F1 female embryo, and was acquired from the laboratory of Professor P. Clerc⁵⁰. The first part of the project was to determine if the aforementioned published results could be reproduced on those 2 cell lines, to ensure that those cells could be used as a model to study the differences in DNA methylation levels between male and female ES cells.

1.1. Assessing the presence of two X chromosomes in the female gC1 ES cell line

Since the phenotype of female ES cells depends on the presence of two active X chromosomes, we wanted to ensure that the gC1 ES cell line conserves the two X chromosomes, especially because female ES cells are at risk of losing one X after prolonged cell culture³¹. We analysed the karyotype of these cells using Giemsa stain. In most cells analysed, 40 chromosomes were observed which is the expected number for murine cell (**Figure 7**). As the mouse chromosomes are difficult to distinguish due to limited differences in length, we also ensured that gC1 cells did not contain a Y chromosome. As expected, no PCR product could be obtained from our XX ES cells for *Zfy*, a sequence specific of the Y chromosome (**Figure 7**).

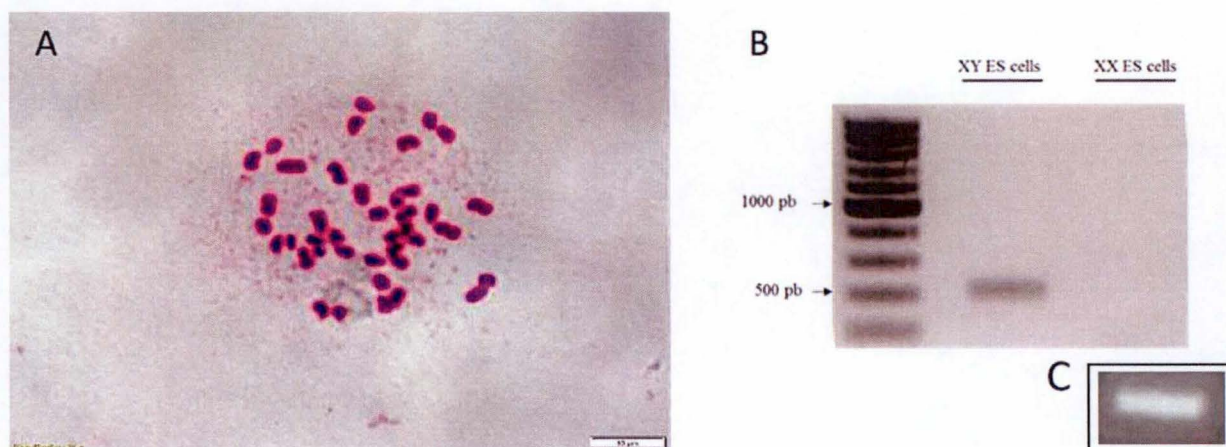


Figure 7 – (A) Karyotype of the gC1 ES cell line using Giemsa stain (Magnification of 60X). (B) Sexing PCR which detects the *Zfy* gene located on the Y chromosome. XY ES cells were used as a positive control. A band at the expected size can be observed when using genomic DNA from XY ES cells, while no DNA amplification is possible when using genomic DNA from XX ES cells. (C) Simultaneously, a PCR on *Igf2r* located on chromosome 17 using DNA from XX ES cells, was run to serve as a positive control.

1.2. Methylation status of female and male ES cells

As mentioned before, the genome of female ES cells is globally hypomethylated when compared to male ES cells. We assessed DNA methylation levels of gC1 and E14TG2A cells. To this purpose, we chose to determine the levels of DNA methylation of murine intracisternal A-particle (IAP) retrotransposons, and of two imprinted genes: *H19* and *Igf2r*. These sequences had to be analysed using pyrosequencing coupled to bisulfite treatment. This method permits to analyse DNA methylation of a 200-bp fragment. Genomic DNA is treated with bisulfite: the unmethylated cytosines are converted into uracils (U), while the methylated cytosines are protected by their methyl group. The treatment with bisulfite is followed by a PCR which converts the U into thymines (T). The pyrosequencing analysis uses a single-stranded PCR product which is biotinylated. This strand is isolated after denaturation, and is hybridized with a sequencing primer. The sequencing primer and the single-stranded PCR product are incubated with DNA polymerase, ATP sulfurylase, luciferase, apyrase, adenosine 5' phosphosulfate and luciferin. The DNA polymerase incorporates the nucleotides one by one to the sequencing primer, by following a sequence of injection which has to be defined prior to carrying out the experiment. The pyrosequencing analysis will take the form of a pyrogram, in which the methylated C (mC) and unmethylated C are respectively represented by C and T peaks. The C-peak-to-T-peak ratio is proportional to the methylation level at each CpG analysed⁵¹. The CpG positions are referred to as variable positions.

We started by assessing the level of DNA methylation of IAPs, which are sometimes used to estimate global DNA methylation. These endogenous retroviral sequences are part of a class of transposable elements which are resistant to DNA demethylation^{17,28}. The methylation state of IAPs was analysed by pyrosequencing in XX and XY ES cells by following the experimental design of Rowe et al., 2013^{52,53}.

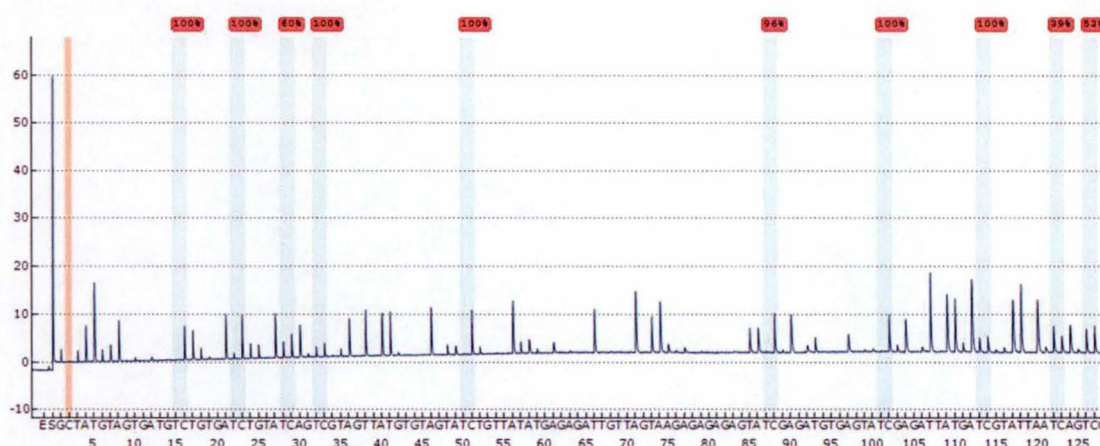


Figure 8 – Pyrogram obtained from XX ES cells cultivated in serum-free medium. Pyrogram includes 10 CpG sites. E represents the enzyme while S represents the substrate.

This pyrogram can't be analysed due to low peak height, baseline drift and important differences between the heights of peaks for almost all positions to what is expected. The software uses non-variable peaks (peaks that are not a part of CpG) as references, the software

can predict the height of the peaks since they are supposed to be identical between all the samples tested. In this case, variations are visible between what is predicted and what is observed. We hypothesised that the sequence being dispensed, doesn't correspond to the sequence of IAPs in our cell lines since we can't observe any signal for most of these non-variable peaks. Rowe's experiments were designed for sequences of IAPs from a BALB/c strain. To test this hypothesis, the PCR products prior to bisulfite treatment was sequenced by Sanger sequencing for gC1 and E14TG2A. We observed that these sequences were quite different from each other's (**Figure 9**).

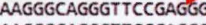


Figure 9 – Alignment of the different sequences obtained for IAPs. The star (*) shows when the six sequences are similar, whereas, a point (.) or a blank shows when they are different due to a SNP or polymerase errors. 5 sequences were amplified from XX ES cells and IAP-ref corresponds to the sequence used by Rowe et al., 2013.

Using the same method, we were supposed to assess the levels of DNA methylation at the differently methylated region (DMR) of *Igf2r* and *H19*. It is particularly important to verify the levels of DNA methylation at the DMR of *H19* for our project since we want to integrate our construct to trace DNA methylation in XY ES cells in this region. We respectively used the sequences determined by Zvetkova et al., 2005 for *H19* and by Sato et al., 2003 for *Igf2r* to design our experiment ^{31,54}. After having encountered problems in our analysis of DNA methylation of the IAPs, we sequenced the two regions before carrying out a pyrosequencing analysis to make sure there wasn't any polymorphism (**Figure 10. Figure 11**).

IGF2R-2-XX TCCGGCTTGGTTCGGAGCAATCCGGTTGTGCCGTGATCCTTGGTTGTGCTGAGTTGCGG
 IGF2R-5-XX TTCGGCTTGGTTCGGAGCAATCCGGTTGTGCCGTGATCCTTGGTTGTGCTGAGTTGCGG
 IGF2R-4-XX TTCGGCTTGGTTCGGAGCAATCCGGTTGTGCCGTGATCCTTGGTTGTGCTGAGTTGCGG
 IGF2R-1-XX TTCGGCTTGGTTCGGAGCAATCCGGTTGTGCCGTGATCCTTGGTTGTGCTGAGTTGCGG
 IGF2R-3-XX TTCGGCTTGGTTCGGAGCAATCCGGTTGTGCCGTGATCCTTGGTTGTGCTGAGTTGCGG
 IGF2R-ref-Sato ---ccgaggggttcggagcaattccggttggcgcgtgatccttggttgtgctgagttgccc
 IGF2R-XY -----TGTTCCGAGCAATCCGGTTGTGCCGTGATCCTTGGTTGTGCTGAGTTGCGG

 IGF2R-2-XX TGAGGGAAAGGGAAGGGAAGCTCAGAGGGTTCCGAGCTATCCTGAGGGTGCAGAGCTGC
 IGF2R-5-XX TGAGGGAAAGGGAAGGGAAGCTCAGAGGGTCCGAGCTATCCTGAGGGTGCAGAGCTGC
 IGF2R-4-XX TGAGGGAAAGGGAAGGGAAGCTCAGAGGGTCCGAGCTATCCTGAGGGTGCAGAGCTGC
 IGF2R-1-XX TGAGGGAAAGGGAAGGGAAGCTCAGAGGGTCCGAGCTATCCTGAGGGTGCAGAGCTGC
 IGF2R-3-XX TGAGGGAAAGGGAAGGGAAGCTCAGAGGGTCCGAGCTATCCTGAGGGTGCAGAGCTGC
 IGF2R-ref-Sato tgagggaagggaagggaagctcagagggttcgagagctatcctgagggtgcgaagctgc
 IGF2R-XY TGAGGGAAAGGGAAGGGAAGCTCAGAGGGTCCGAGCTATCCTGAGGGTGCAGAGCTGC

 Sequencing primer 
 IGF2R-2-XX ACAAGGGCAGGGTTCGAGGGTTCGCGGCACGTGCTGAGTCCCGGTGCCGCGCTGCCCA
 IGF2R-5-XX ACAAGGGCAGGGTTCGAGGGTTCGCGGCACGTGCTGAGCCCCGGTGCCGCGCTGCCCA
 IGF2R-4-XX ACAAGGGCAGGGTTCGAGGGTTCGCGGCACGTGCTGAGTCCC-GTGCCGCGCTGCCCA
 IGF2R-1-XX ACAAGGGCAGGGTTCGAGGGTTCGCGGCACGTGCTGAGCCCCGGTGCCGCGCTGCCCA
 IGF2R-3-XX ACAAGGGCAGGGTTCGAGGGTTCGCGGCACGTGCTGAGTCCCGGTGCCGCGCTGCCCA
 IGF2R-ref-Sato acaagggcaggggttcgaggggttcgcggcacgtgctgagctcccggtgccgcgctgccca
 IGF2R-XY ACAAGGGCAGGGTTCGAGGGTTCGCGGCACGTGCTGAGCCCCGGTGCCGCGCTGCCCA

 IGF2R-2-XX AAGGTTCCGAGGGTTTTAATGCGATTCCGGTTGGGCTGTGATTCTGGTTATGCCAAGTTG
 IGF2R-5-XX AAGGTTCCGAGGGTTTTAATGCGATTCCGGTTGGGCTGTGATTCTGGTTATGCCAAGTTG
 IGF2R-4-XX AAGGTTCCGAGGGTTTTAATGCGATTCCGGTTGGGCTGTGATTCTGGTTATGCCAAGTTG
 IGF2R-1-XX AAGGTTCCGAGGGTTTTAATGCGATTCCGGTTGGGCTGTGATTCTGGTTATGCCAAGTTG
 IGF2R-3-XX AAGGTTCCGAGGGTTTTAATGCGATTCCGGTTGGGCTGTGATTCTGGTTATGCCAAGTTG
 IGF2R-ref-Sato aaggttcgaggggttttaattgcaattccggttgggctgtgattctggttatgccaaagtgc
 IGF2R-XY AAGGTTCCGAGGGTTTTAATGCGATTCCGGTTGGGCTGTGATTCTGGTTATGCCAAGTTG

 IGF2R-2-XX CGCGAGGGGAGGGGAAGAGACAGCTCGAAGGATCCGAGGATCCTGAGGGTGCAGAACTG
 IGF2R-5-XX CGCGAGGGGAGGGGAAGAGACAGCTCGAAGGATCCGAGGATCCTGAGGGTGCAGAACTG
 IGF2R-4-XX CGCGAGGGGAGGGGAAGAGACAGCTCGAAGGATCCGAGGATCCTGAGGGTGCAGAACTG
 IGF2R-1-XX CGCGAGGGGAGGGGAAGAGACAGCTCGAAGGATCCGAGGATCCTGAGGGTGCAGAACTG
 IGF2R-3-XX CGCGAGGGGAGGGGAAGAGACAGCTCGAAGGATCCGAGGATCCTGAGGGTGCAGAACTG
 IGF2R-ref-Sato cgcgaggggaggggaagagacagctcggaggaattccgaggtatcctgaggggtgcagaaactg
 IGF2R-XY CGCGAGGGGAGGGGAAGAGACAGCTCGAAGGATCCGAGGATCCGAGGATCCTGAGGGTGCAGAACTG

 IGF2R-2-XX CACAAGGGGAGGATTCGAAGGGTTCTGTGATCAGGGCCAAACGCTCAAAAGTGCCATGTTA
 IGF2R-5-XX CACAAGGGGAGGATTCGAAGGGTTCTGTGATCAGGGCCAAACGCTCAAAAGTGCCATGTTA
 IGF2R-4-XX CACAAGGGGAGGATTCGAAGGGTTCTGTGATCAGGGCCAAACGCTCAAAAGTGCCATGTTA
 IGF2R-1-XX CACAAGGGGAGGATTCGAAGGGTTCTGTGATCAGGGCCAAACGCTCAAAAGTGCCATGTTA
 IGF2R-3-XX CACAAGGGGAGGATTCGAAGGGTTCTGTGATCAGGGCCAAACGCTCAAAAGTGCCATGTTA
 IGF2R-ref-Sato cacaaggggagggattcgaaggggttctgtgatcagggccaaacgctcaaaagtgcattgtta
 IGF2R-XY CACAAGGGGAGGATTCgaagttcka-----cccc

Figure 10 – Alignment of the different sequences obtained for IGF2R. The star (*) shows when the seven sequences are similar whereas, the green triangle represents a SNP. 5 sequences were generated from DNA of XX ES cells, one sequence was generated from DNA of XY ES cells. Igf2r-ref-Sato corresponds to the sequence used by Sato et al., 2003.

	Sequencing primer 
H19-ref-Zvetkova	acatgctacattcacacgagcatccaggaggcataagaattctgcaaggagaccatgcc-
H19-XY	-----gagctmt-tCMCGAGCATCCAGGAGGCATAAGAATTCTGCAAGGAGACCATGCC
H19-XX	-----CTMTTCACGAGCATCCAGGAGGCATAAGAATTCTGCAAGGAGACCATGCC
	* . * * *
H19-ref-Zvetkova	tattcttggacgtctgctggaatcagttgtggggtttatacgcgggagttgtggcccggt
H19-XY	TATTCTTGACGTCTGCTG-AATCAGTTGTGGGTTTATACGCGGGAGTTGCCGCGTGGT
H19-XX	TATTCTTGACGTCTGCTG-AATCAGTTGTGGGTTTATACGCGGGAGTTGCCGCGTGGC
	***** ** *
H19-ref-Zvetkova	ggcagcaaaatcgattgcgccaaacctaagagccccccacccctggatttgaattca
H19-XY	GGCAGCAAAATCGATTGCGCCAAACCTAAAGAGCCCCCACCCTGGTATTGGAATTCA
H19-XX	GGCAGCAAAATCGATTGCGCCAAACCTAAAGAGCCCCCACCCTGGTATTGGAATTCA

H19-ref-Zvetkova	caaatggcaatgctgtgggtcacccaagttcagtacctcaggggggtcacaaatgccact
H19-XY	CAAATGGCAATGCTGTGGGTCAACCAAGTTCAGTACCTCAGGGGGGTACAAATGCCACT
H19-XX	CAAATGGCAATGCTGTGGGTCAACCAAGTTCAGTACCTCAGG-GGGTCACAAATGCCACT

H19-ref-Zvetkova	aggggggcaggacacatgcattttctaggctgggtacctcgtggactcggactcccaaatc
H19-XY	AGGGGGGCAGGACACATGCATTTTCTAGGCTGGTACCTCGTGGACTCGGACTCCCAAATC
H19-XX	AGGGGGGCAGGACACATGCATTTTCTAGGCTGGTACCTCGTGGACTCGGACTCCCAAATC

H19-ref-Zvetkova	aacaaggtcggcttactctctgcaaagaatcctttgtgtgtaaagaccagggttgccgc
H19-XY	AACAAGGTCGGCTTACTCTCTGCAAAGAATCCTTTGTGTGTAAGACCAGGGTTGC-CGC
H19-XX	AACAAGATCGGCTTACTCTCTGCAAAGAATCCTTTGTGTGTAAGACCAGGGTTGC-CGC
	***** *
H19-ref-Zvetkova	acggcggcagtggaagtctcgtacatcgcagtcctaaaacggat-tgcaactga---ttga
H19-XY	ACGGCGGCAGTGAGGCCTCGTACAK-----
H19-XX	ACGGCGGCAGTGAAGTCTCGTACAAAGCCGAATTCAGCACACTGGCGGCCGTTACTAGT
	***** *

Figure 11 – Alignment of the different sequences obtained for H19. The star (*) shows when the sequences are similar whereas, a point (.) or a blank shows when they are different due to a SNP or polymerase errors. 1 sequence was generated from DNA of XX ES cells, 1 sequence was generated from DNA of XY ES cells. H19-ref-Zvetkova corresponds to the sequence used by Zvetkova et al., 2005.

We observed two single nucleotide polymorphisms (SNP) between the seven sequences of *Igf2r*. However, for *H19*, we can't always tell if the differences between the sequences analysed are the consequence of a SNP or an error related to the use of low fidelity polymerase (DreamTaq) since we only determined two sequences (one from XX ES cells and one from XY ES cells). Moreover, we are missing one allele for *H19* since the XX ES cells were derived from a hybrid mouse. Due to a lack of time, we couldn't carried out the pyrosequencing analysis for *Igf2r* nor *H19*, but we would need to order new sequencing primers which would be located between SNPs, to eliminate problems related to the sequence of injection.

1.3. Expression analysis of several genes with a possible implication in the DNA demethylation of female mouse embryonic stem cells

We designed a quantitative PCR experiment to confirm that when our cell lines were cultured into serum-free medium, the cells were in a 'naïve' pluripotent state. To do so, we compared the expression of *Nanog* and *Oct4* between cells cultivated in serum to cells cultivated in serum-free medium. These genes are known as two pluripotency markers which are upregulated in serum-free medium^{28,32}. Moreover, we wanted to assess the expression of *Dnmt1*, *Dnmt3a* and *Dnmt3b* to check if the expression of *Dnmt1* was comparable between gC1 and E14TG2A cell line and if there was a downregulation of *de novo* DNA methyltransferases in the gC1 cell line. We designed cDNA-specific PCR primers for all these genes and tested them on cDNA from female ES cells cultivated either in serum or in serum-free medium. We tested the efficiency of each pair of primers. Since each PCR cycle is supposed to double the DNA quantity present in the reaction, there should be a linear relationship between Ct values and starting cDNA amounts when expressed in binary logarithm, with a value around 1 for the R coefficient. When it wasn't the case, the primers were re-designed.

The results obtained don't follow what is reported in the aforementioned literature. We observed that the expression of *Nanog* and *Oct4* was upregulated in XY and XX ES cells cultured in serum medium (**Supplementary Figure 1**, **Supplementary Figure 2**). *Dnmt1* is upregulated in XX ES cells cultured in serum medium (**Supplementary Figure 3**). The highest expression for *Dnmt3a* was observed in XY and XX ES cells cultured in serum medium (**Supplementary Figure 4**). Expression of *Dnmt3b* appears to be more important in XX ES cells cultured in serum medium (**Supplementary Figure 5**). These results can be easily explained by the important variations of Ct for the endogenous control (*GAPDH*) between the different samples tested and by the variations of Ct between the duplicates for gene tested. At the time of writing this document, we are still trying to optimize our experimental design for RT-qPCR.

2. Unbiased screen for identifying DNA demethylation promoting gene(s)

To identify the putative X-linked gene(s) implicated in the demethylation process occurring in XX ES cells, we wanted to perform an unbiased screen using a mouse genomic cosmid library. With this strategy, we were hoping to identify genes capable to induce demethylation when transfected into XY ES cells. In order to monitor DNA methylation levels of the transfected cells, we used a reporter of DNA methylation. The reporter of endogenous methylation was derived from a construct made by Stelzer et al, 2015. This construct contains the SNRPN promoter which is sensitive to the state of DNA methylation of adjacent sequences. In other words, adjacent sequences influence the methylation status of the SNRPN promoter. This latter controls the expression of an enhanced GFP (eGFP) and the hygromycin B selection marker (HygroR). If the sequence adjacent to the SNRPN promoter is methylated, eGFP and HygroR won't be expressed and if the adjacent sequence is unmethylated, eGFP and HygroR should be expressed. This construct will be inserted by using the CRISPR/Cas9 near the DMR of *H19* on

the paternal allele which is methylated in XY ES cells, but is expected to be demethylated in XX ES cells (**Figure 12**).

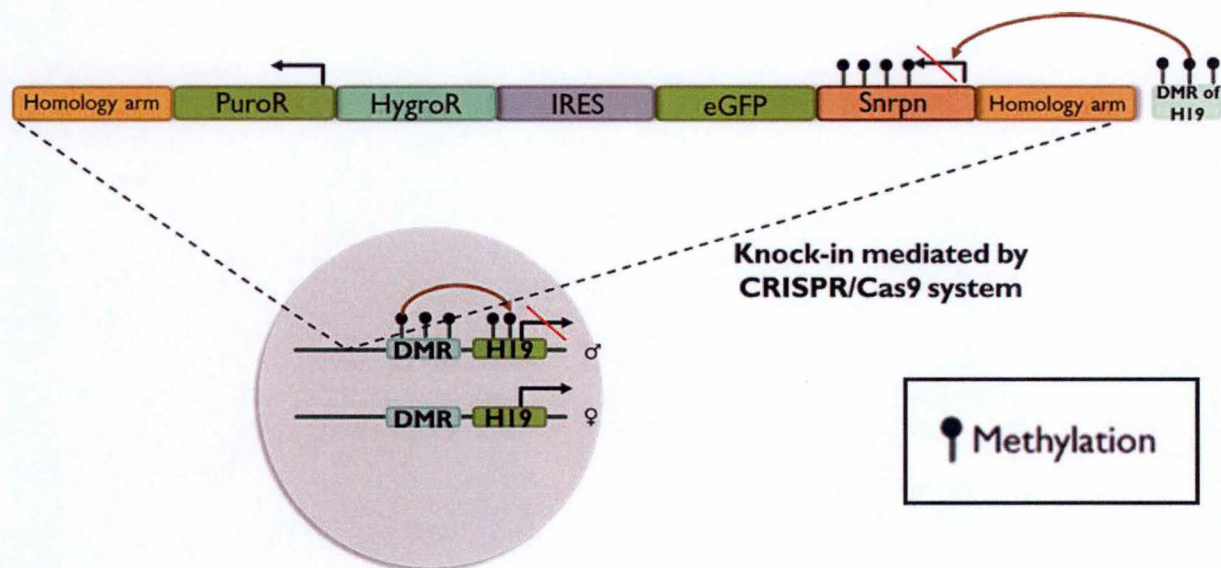


Figure 12 – Schematic representation of the CRISPR/Cas9 strategy used to stably integrate the reporter of endogenous methylation.

The reporter of endogenous methylation also named, targeting vector was constructed in three steps. We inserted an IRES-Hygro cassette downstream from the eGFP sequence. The IRES (internal ribosome entry site) sequence allows the initiation of translation independently of the cap-structure of RNA messengers. By adding this cassette to the construct, whenever the SNRPN promoter gets demethylated, eGFP and HygroR will be expressed. The XY ES cells which get demethylated after the introduction of cosmid will be followed by eGFP expression. The XY ES cells will also be selected by their resistance to hygromycin B. The IRES-Hygro cassette was obtained through PCR amplification using primers flanked with restriction sites for MluI, to allow its insertion downstream of eGFP. The homology arms used by Stelzer et al., 2015 to target *GAPDH* were replaced to target the DMR of *H19*. Restriction sites for SbfI and MfeI were added on the PCR primers used to amplify the 5' homology arm (5' HA), and the PCR primers used to amplify the 3' homology arm (3' HA) contain restrictions for AscI and FseI, to allow its insertion. Both homology arms were amplified from genomic mouse DNA (**Figure 13**).

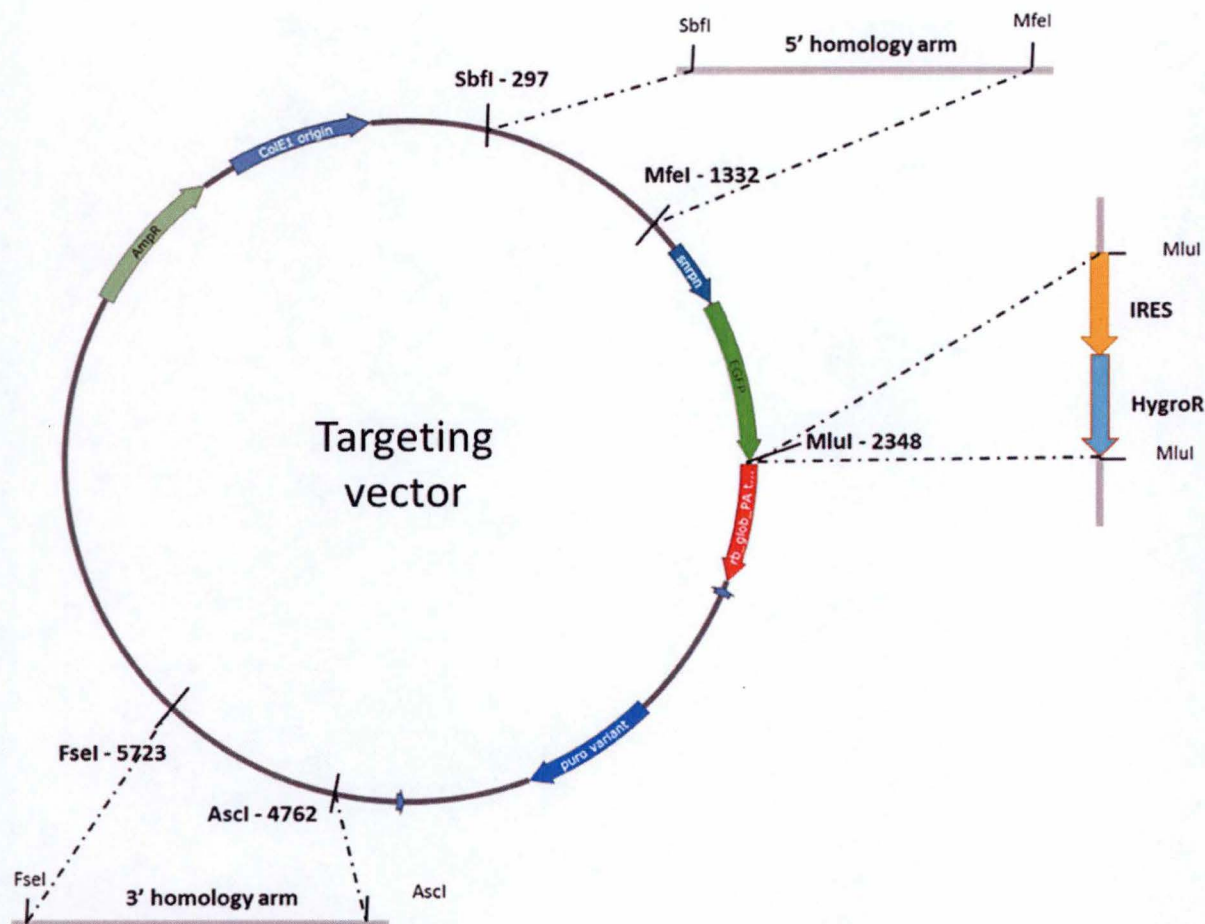


Figure 13 - Schematic representation of the targeting vector (reporter of endogenous methylation) for CRISPR/Cas9-mediated HDR.

To introduce breaks in the DMR of H19, we chose to use a modified version of the Cas9 named the D10A mutant nickase. As mentioned above, this version of the Cas9 is more precise since it generates single strand breaks, commonly called nicks. Therefore, 2 single guide RNAs (sgRNA) are needed to generate a doublestranded break (DSB). These guides were designed using the online “CRISPR Design Tool” (<http://crispr.mit.edu/>, Zhang Lab), and cloned in pX335 which expresses the D10A mutant nickase. We hope that the DSB generated will be repaired through HDR with the targeting vector. Both homology arms used are about 1 kb long, which is sufficient for CRISPR-mediated HDR⁴³. To increase the efficiency of HDR, the sequences of both homology arms were chosen to be close to the cutting sites of Cas9⁴³. It was demonstrated that the efficiency of HDR significantly drops if homology arms are far away from the DSB site.

XY ES cells have been transfected with the Cas9 expressing plasmids and the targeting vector. The cells were kept under puromycin selection for a week to obtain cells with a stably integrated reporter. No XY ES cell was able to resist to the puromycin selection. We therefore verified the sequence of the puromycin resistance gene (PuroR) by Sanger sequencing from the GAPDH-CpG-TV, the original plasmid from Stelzer et al., 2015 (the plasmid was received from Addgene). We found that there was a single base modification in the PuroR sequence.

This mutation leads to change in amino acid from an arginine (R) to a histidine (H), which could account for the absence of resistance (**Figure 14**).

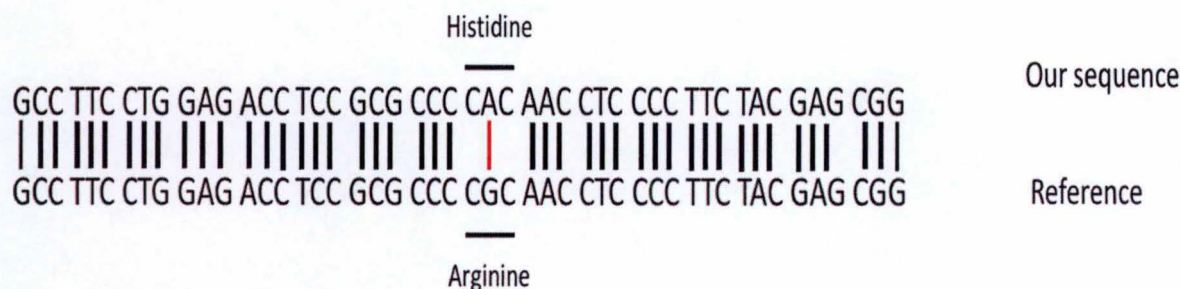


Figure 14 – The PuroR for the GAPDH-CpG-TV contains a single base modification. The single base mutation leads to a change in amino acid which doesn't confer the puromycin resistance in XY ES cells.

We decided to repair this mutation by targeted mutagenesis using the “triple-PCR method”. In brief, 2 fragments are amplified which share an overlapping sequence. This overlap allows those fragments to anneal to each other in the third PCR, thereby generating a single fragment. The point mutation is corrected thanks to a mismatching base inserted in both primers (**Figure 15**).

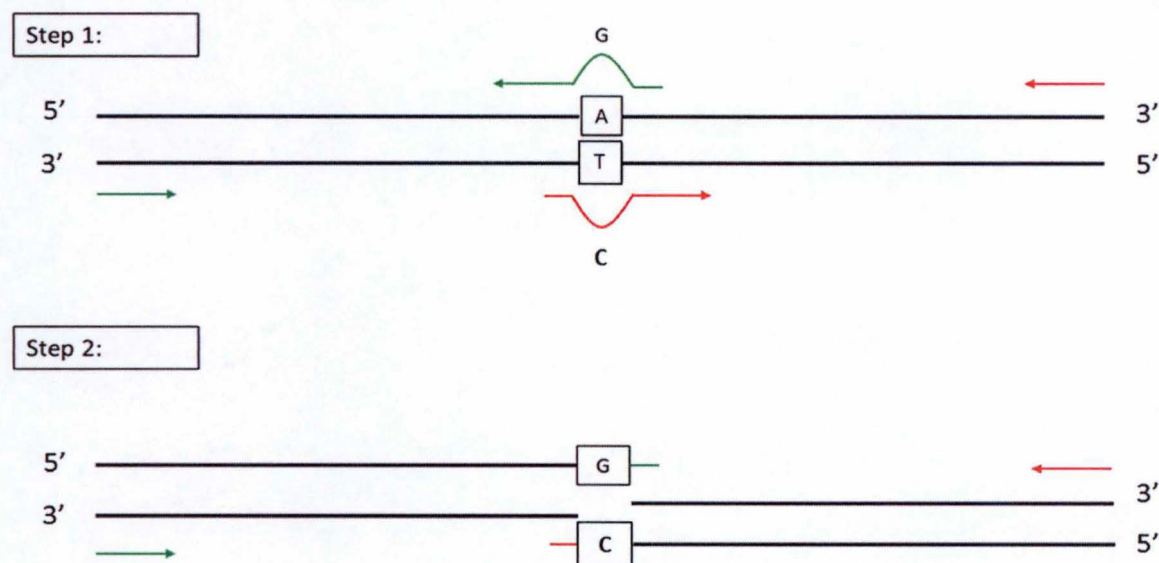


Figure 15 - The principle of the triple-PCR used to repair the mutated repair template.

We optimized the PCR conditions to amplify one of the two fragment with Q5 Polymerase to increase fidelity. At the time of writing this document, we are still trying to amplify the second fragment to start the second step of the triple PCR to synthesize a single fragment permitting us to repair the PuroR.

3. Deletion of the X inactivation center in XX ES cells

As aforementioned, female ES cells contain two active X chromosomes, we wanted to test our hypothesis that the X inactivation center could be implicated in the hypomethylation of XX ES cells. We attempted to delete one of the XIC using the CRISPR/Cas9 technology. The largest deletion reported in mouse ES cells using the CRISPR/Cas9 system corresponds to an entire gene of 65 kb⁵⁵. This is much smaller than the XIC, which is approximatively 10 to 20 Mb long.

The “enhanced specificity” Cas9 and 2 sgRNAs (instead of one sgRNA) on each side of the XIC were used, thereby potentially generating 4 DSBs. To increase the chances of HDR, we designed a repair template which contains a neomycin selection marker (NeoR) flanked by two homology arms (named 5' HA and 3' HA). The selection marker permits to select cells in which deletion of one XIC occurred. This latter is flanked by LoxP sites, meaning that it can be removed after deletion of one XIC. We also placed a diphtheria toxin A (DTA) gene in the pBlueScript SK- (serving as the targeting vector) on the 3' side of the 3' HA. This gene is an inhibitor of protein synthesis, it provides negative selection for the cells where random insertion of the repair template occurred.

Originally, the template had to be constructed using the In-Fusion HD Cloning technology in a “one-step” reaction. In brief, this system works by fusing homology sequences of 15-bp between DNA fragments. 15 bp were respectively added on each extremities of 5' HA and 3' HA, these 15 bp were designed to either be homologous to the pBlueScript SK- or to the NeoR gene (Figure 16).

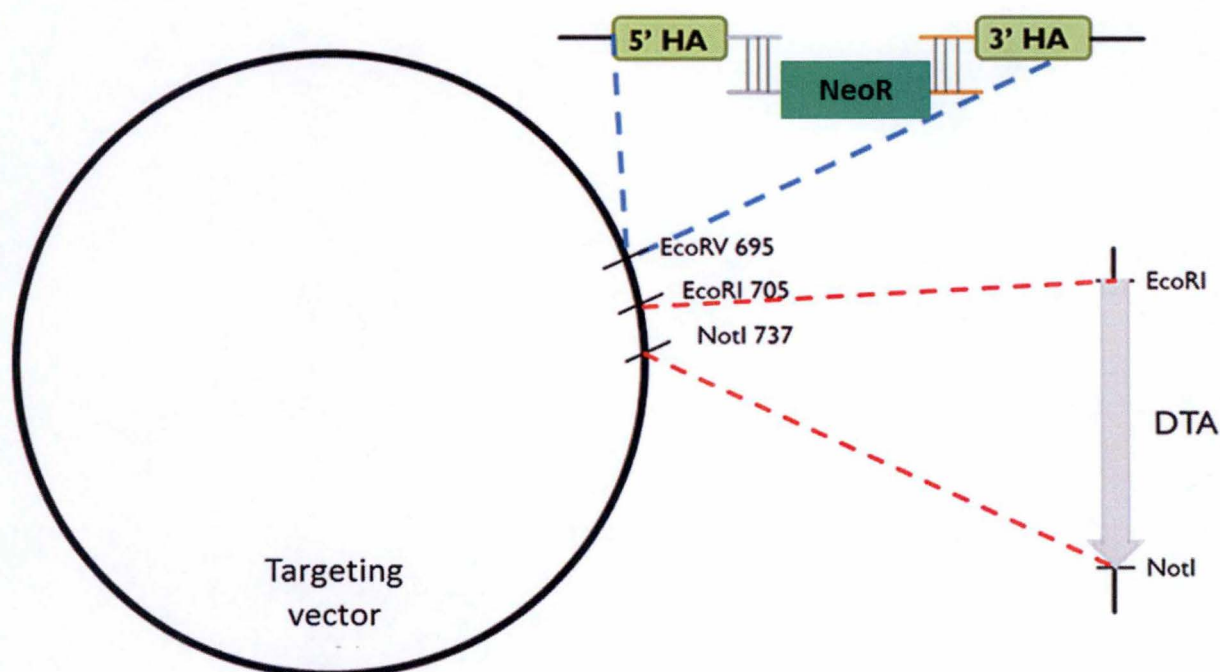


Figure 16 – Strategy I to construct a repair template for CRISPR/Cas9-mediated HDR to delete the X inactivation center using the In-Fusion HD Cloning method.

The “one-step” reaction never worked out in our hands. Indeed, the “one-step” reaction was electroporated into electro-competent bacteria, which were plated onto agar plates. Several colonies were screened with the primers: M13F and M13R, but we couldn’t amplify a fragment of the correct size. We hypothesized that the colonies obtained, were ampicillin resistant due to the presence of a non-restricted pBlueScript SK-.

The strategy was adapted to integrate the three fragments in the pBlueScript SK- separately with the In-Fusion HD Cloning technology. We wanted to clone the 5’ HA by using an EcoRV restriction site. The NeoR cassette was to be inserted in 3’ of the 5’ HA after AfeI restriction. The 3’ HA was to be added in 3’ of the NeoR, after a second AfeI restriction. PCR primers were modified accordingly to bear the appropriate 15 bp of homology sequences and AfeI restriction sites were added. The DTA gene was inserted in the receiving plasmid similarly to what was described above (**Figure 17**). Once again, this strategy with the use of the In-Fusion HD Cloning system didn’t work in our hands. We tried to electroporate the In-Fusion which was supposed to ligate the targeting vector and the 5’ HA, but never obtained any colonies.

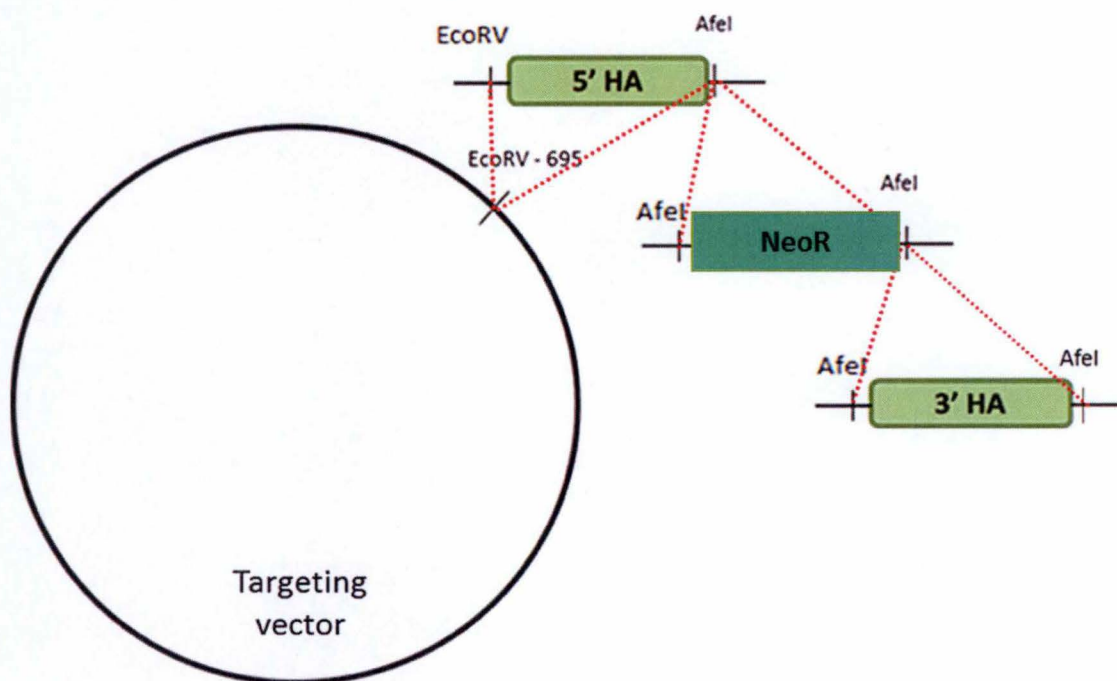


Figure 17 - Strategy II to construct a repair template for CRISPR/Cas9-mediated HDR to delete the X inactivation center using the In-Fusion HD Cloning method.

Since we couldn’t obtain the template vector in time, we attempt to delete one XIC using only the 4 sgRNAs. XX ES cells have been transfected with the 4 CRISPR plasmids and will be shortly plated for screening. To identify the cells with the desired deletion, we designed 2 pairs of primers to perform a nested PCR. These cells will then, be divided into pools in order to extract genomic DNA.

D. DISCUSSION AND PERSPECTIVES

We have started to set up tools which will help us study the mechanisms leading to loss of DNA methylation in XX ES cells.

We demonstrated that the gC1 cell line obtained from another laboratory, possesses 40 chromosomes, and is not contaminated with male cells. This, however, doesn't guarantee the presence of the two active X chromosomes. It would be interesting to ensure that they are still present and active by using a FISH analysis (fluorescent in situ hybridization). DNA FISH is often used in cytogenetics to detect deletions, duplications and chromosomes. RNA FISH is used to assess gene expression on single cell basis. FISH is based on the use of probes which are complementary to a sequence of interest and are bound to a fluorescent dye. To distinguish between the active and inactive X chromosome, RNA FISH which detects *Xist* RNA is often used^{56,57}. As aforementioned, in undifferentiated XX ES cells, both X chromosomes are active and *Xist* is expressed at low levels from both X chromosomes. *Xist* RNA is visible as two "dots" in each cell. Whereas, *Xist* is highly expressed when random X inactivation occurs, and it progressively coats the X chromosome undergoing inactivation. *Xist* RNA forms a cloud around the future inactive X chromosome⁵⁸. The experiment can be improved by using DNA-RNA FISH to detect both X chromosomes and *Xist* RNA⁵⁸. Indeed, female ES cells tend to lose one of their X chromosome, this method allows for analysis of cells with 40 chromosomes only.

In order to further characterize the ES cell lines, we wanted to assess the levels of DNA methylation at IAPs, *H19* and *Igf2r*, through pyrosequencing. Unfortunately, our experiment failed within the time which had been allocated, due to technical reasons. Indeed, we wanted to analyse IAPs as markers of global DNA methylation, but the diversity of sequences amplified by PCR after the bisulfite treatment, renders impossible any pyrosequencing analysis. Indeed, the pyrosequencing technique does not tolerate divergence from the sequence of reference. It can be explained by the presence of 1000 copies of IAPs in the mouse genome. Several classes of IAPs have been identified, they differ to deletions and insertions of various lengths⁵⁹⁻⁶². There are other ways to assess global DNA methylation such as commercial DNA methylation quantification kits. These kits are all based on enzyme-linked immunosorbent assay (ELISA) meaning that, first genomic DNA is immobilized on wells with high DNA affinity. Then, the primary antibody directed against 5 mC is added, followed by a labelled secondary antibody. The final step consists in adding the enzyme's substrate to determine the amount of methylated DNA⁶³. Another method to quantify global DNA methylation is by Southern blot analysis with methyl-sensitive restriction enzymes. Using this method, Zvetkova et al., 2005 assess DNA methylation at mouse major satellite repeats and at repetitive C-type retroviral sequences. They digested genomic DNA of XX and XY ES cells either with *MspI* (M), a methylation-insensitive enzyme or with *HpaII* (H), a methylation-sensitive enzyme. Both restriction enzymes recognize the following sequence: CCGG. However, *HpaII*, but not *MspI*, restriction activity is inhibited by methylation of cytosine meaning that C^{5m}CGG can't be cut by *HpaII*. In light of the problems encountered with IAPs, we decided to sequence *H19* and *Igf2r*. Several SNPs were discovered, which means that we should design the sequencing primers consequently, to avoid those SNPs. Other pyrosequencing experiments should be

attempted to determine if both of these imprinted are hypomethylated in the gC1 cell and if the E14TG2A cell line is suitable for the introduction of the reporter of endogenous methylation.

Surprisingly, our qPCR analysis revealed that *Oct4* and *Nanog* were upregulated both in XY and XX ES cells when placed in serum medium. These results were unexpected since these pluripotent genes are supposed to be upregulated in serum-free medium^{28,32,36}. We wonder whether the expression levels inferred from these data are accurate. At the moment, we can't conclude that the gC1 cell line follows what has been published. We think that RNA quality can be improved, and it might account for the variations in the Ct values. To standardize our extraction process, we are going to use the QIAcube. If we managed to obtain conclusive results, we could validate our cellular model. We could also characterize the expression of several candidates with a possible implication in the control of DNA methylation. If the expression of the candidates is different between female and male ES cells, we could overexpress several of these candidates in male or female ES cells to observe the possible impact on DNA methylation. We could also develop strategies using the CRISPR/Cas9 to knock out the expression of several of these candidates in female or male ES cells.

Based on the work of Stelzer et al., 2015, we have started to develop a strategy using the CRISPR/Cas9 to insert a reporter of endogenous methylation in the 5' region of the *H19* imprinted gene. This reporter contains a missense mutation in the puromycin resistance gene. In our XY ES cell line, this gene doesn't confer the puromycin resistance to the transfected cells. Once, the mutation will have been repaired, the cells will be transfected. These cells will be selected using puromycin during an entire week to allow for stable integration. The cells will have to be checked for corrected insertion, in this case, the reporter has to integrate on the paternal allele of *H19*. A simple way to check is by analysing eGFP expression. If the reporter integrated in the maternal allele, the ES cells will express eGFP (since the ICR is unmethylated). Whereas if the reporter integrated in the paternal allele, the transcription of eGFP is repressed. After obtaining a clone with the correctly integrated construction, we will have to test whether eGFP is transcribed once the DMR of the paternal allele of *H19* is demethylated. To do so, the cells could be treated with a demethylating drug such as 5-aza-2'deoxyctidine (5-azadC) to demethylate the DMR of *H19* which should lead to the transcription of eGFP. If we managed to obtain a clone, we could move on to the transfection of the cells with the mouse genomic library.

A cosmid genomic DNA library is composed of DNA fragments representing the entire genome of an organism, in this case, the mouse. The genome of a mouse is composed of $2,5 \times 10^9$ bp, whereas, the size of a cosmid is approximatively of 30×10^3 bp. In order to represent the entire genome of a mouse, the library has to approximately contain 83 333 cosmids. In general, several cosmids are incorporated into a cell. To determine the number of cosmids which can be transfected at once: we should estimate the number of cells which have been transfected and the number of cosmids which have entered the cells. These experiments will allow us to estimate the number of cosmids which can be screened with a single transfection.

We designed a strategy to test if the presence of two X inactivation centers has an impact on the levels of DNA methylation of female ES cells. We couldn't assemble the targeting vector within the time which had been granted. In order to delete one X inactivation center, we

transfected 4 sgRNAs to guide the “enhanced specificity” Cas9. The XX ES cells were easily transfected, at the time of writing this document, XX ES cells are in culture to be transfected. Cells will be diluted to search for deletion of the XIC on one X chromosome by nested PCR. If we managed to obtain cells with a deletion of one XIC, we will have generated the longest deletion ever reported through CRISPR/Cas9. These cells should offer interesting perspectives for studying random X inactivation, as well as further developing the CRISPR/Cas9 technology. We could use these cells to test the implication of XIC is implicated in the hypomethylation of XX ES cells, using karyotyping, pyrosequencing and RT-qPCR as done for characterisation of the gC1 ES cells.

Some genomic regions of ES cells are associated with H3K9me3, as mentioned previously, they tend to resist better to DNA demethylation. We hypothesised that the deposition mechanism for histone modification might be impaired in XX ES cells leading to a reduced levels of H3K9me3. This hypothesis could first be verified by doing a Western Blot on total cell lysate from XX and XY ES cells to compare the level of H3K9me3.

In conclusion, much remains to be investigated in regards to the mechanisms controlling the DNA demethylation observed in mouse XX ES cells.

E. MATERIALS AND METHODS

1. Cell culture

Female gC1 ES cells were derived from a 129Sv/M. m castaneus F1 female embryo at 2,5 day post-coitum, these cells were a gift from Professor P. Clerc of the Institut Pasteur, Paris (P. Clerc et al., 2014). We also used male ESTG2A ES cells. Both cell line were grown on 0.1% gelatin-coated flasks containing either serum or serum-free (2i) medium.

The serum medium contained Glasgow's MEM (Gibco, 11710035) supplemented with 10% FBS (Millipore, ES-009-B), 1 mM MEM non-essential amino acids (Gibco, 11140-050), 1 mM sodium pyruvate (Lonza, BE13-115E), 0.1 mM β -mercaptoethanol (Gibco, 31350010) and 1 mM Penicillin/Streptomycin (Lonza, DE17-602E). Before use, 50 μ l of LIF 10^6 units/ml was added (ESGRO, ESG1106) per 50 mL of serum medium. The serum-free medium was composed of a 1:1 mixture of DMEM/F12 (Gibco, 11320-033) and Neurobasal medium (Gibco, 21103-049) supplemented with 1% of Glutamax 100X (Gibco, 35050-061), 0.5% of N2-supplement 100X (Gibco, 17502-048), 1% of B27-supplement 50X (Gibco, 17504-044), 0.5 mM of β -mercaptoethanol (Gibco, 31350010), 1 mM sodium pyruvate (Lonza, BE13-115E), 1 mM Penicillin/Streptomycin (Lonza, DE17-602E), 1 mM MEM non-essential amino acids (Gibco, 11140-050), 50 μ l of LIF 10^6 units/ml (ESGRO, ESG1106), 3 μ M of Gsk3 inhibitor CT-99021 (Sigma, SML 1046) and 1 μ M of MEK inhibitor PD0325901 (Sigma P20162).

When necessary, cells were frozen in FBS + 10% DMSO (Sigma, D2650).

2. Transfections

Transfections of the ES cells were carried out using FuGENE® HD Transfection Reagent (Promega, E2311) and Lipofectamine® 2000 (ThermoFisher, 11668019) according to the manufacturer's instructions.

One day prior transfection, the cells were counted and plated at a density of 1×10^4 cells per cm^2 on 0.1% gelatin-coated flask to reach optimum density for transfection. When transfection was carried out with FuGENE HD, 3,3 μ g of plasmid DNA was mixed in Opti-MEM® Reduced Serum Medium (Gibco) in a volume of 158 μ l for a single well of 6-well plate. When transfection was carried with Lipofetamine 2000, 0,8 μ g of plasmid DNA was mixed in Opti-MEM® Reduced Serum Medium (Gibco) in a volume of 50 μ l for a single well of 6-well plate. When several plasmids had to be transfected, an equimolar ratio of DNA was calculated. Selection using appropriate antibiotic (1 μ g/ml of puromycin or 100 μ g/ml hygromycin B) started 24 hours post-transfection.

3. Karyotyping of ES cells

0.4×10^6 XX ES cells were incubated for 1h30 in 0.05 μ g/ml of colcemid KaryoMAX (Gibco, 15210-040). Cells were then collected and centrifuged at 2000 rpm. To breakdown the plasma membrane, 0.075M of KCl (Merck, 104936) was added. The samples were fixed in 3:1 methanol:acetic acid for 30 minutes. Then, ES cells were centrifuged for 5 minutes at 2000 rpm. The fixation process was repeated three times. After the last centrifugation, 500 μ l of 3:1

methanol:acetic acid was added to the samples. The microscope slides were cleaned with 3:1 methanol:acetic acid, and left on ice before spreading the suspensions on the slides using a Pasteur pipette. The slides were left at room temperature until they were completely dried. Chromosomes were stained with Giemsa's azuo eosin methylene blue solution (Merck, 1092042500) for 15 minutes before washing with deionized water.

4. RNA extraction and RT-qPCR

Total RNA from cultured ES cells was isolated using the TRIzol® Reagent (Amresco, N580), according to the manufacturer's instructions. After RNA extraction, 5 µg of total RNA for each condition was reverse transcribed with GoScript™ Reverse Transcription System (Promega, A5000), according to the manufacturer's instructions with random RT primers.

The primers used for the qPCR were designed with Primer BLAST and are described in Table 1. Linear relationship between Ct values and cDNA concentration expressed in log₂ were checked for all primer pairs. All real-time PCR reactions were performed in duplicates with SYBR Green (Thermo Scientific, K0222) on a StepOnePlus™ Real-Time PCR System (Applied Biosystems, 4376600) using a standard run.

5. DNA extraction

Cells were lysed overnight at 55°C in the following solution: Tris-HCl pH = 8.0 100 mM (Sigma, T6066), EDTA 5 mM (Roth, X986.2), SDS 0,2% (Sigma-Aldrich, L5750), NaCl 200 mM (VWR Chemicals, 27810.295) and 100 µg/mL of proteinase K was added prior usage of lysis solution. After lysis, DNA is precipitated with EtOH, and resuspended in water.

6. Vector constructions

Generalities

PCR products were amplified using High-Fidelity Q5 Polymerase (NEB, M0491S) using primers described in Table 1. The products were purified using the Wizard® SV Gel and PCR Clean-Up (Promega, A9281). When appropriate, the receiving plasmid was dephosphorylated using Shrimp Alkaline Phosphatase (NEB, M0371S). After restriction, vectors and inserts were purified as done classically, using phenol-chloroform and precipitated with ethanol. Ligation were carried out with T4 ligase (NEB, M02025) following a ratio of 3:1 of insert to vector, and left overnight at 18°C. After each ligation step, the constructs were electroporated into DH5α bacteria. Those were spread out on agar plates and selected with proper antibiotic. The day after electroporation, colonies were PCR screened using DreamTaq Green PCR Master Mix (ThermoFisher, K1081). Plasmid DNA was extracted using GeneJET Plasmid Miniprep Kit (ThermoFisher, K0502) for small culture volumes or NucleoBond PC (Macherey-Nagel) for larger volumes. Each fragment added in the constructs was verified by Sanger sequencing (Genewiz®)

Construction of the reporter of endogenous methylation

The plasmid GAPDH-CPG-TV was a gift from Professor Rudolf Jaenisch (Addgene plasmid #70148) ⁶⁴. The IRES-Hygromycin fragment was PCR-amplified using plasmid pIRES1hyg. The product was inserted into a pCR2.1-TOPO-TA cloning vector, and then, transferred

downstream of the eGFP gene in GAPDH-CPG-TV after MluI restriction (Promega, R0198S). Both 1 kb 5' homology arm and 1 kb 3' homology arm were PCR-amplified. Fragments were ligated into a pCR2.1-TOPO-TA cloning vector. They were then transferred in the repoter plasmid after restriction with SbfI (NEB, R0642S) and MfeI (NEB, R0589S) for 5' homology arm, and restriction with AscI (NEB, R0558S) and FseI for 3' homology arm (NEB, R0588S). Both homology arms were cloned into GAPDH-CPG-TV after its digestion with SbfI, MfeI, AscI and FseI to allow insertion.

Construction of the repair template for XIC deletion

The PGKdtabpA plasmid was a gift from Shahragim Tajbakhsh from the Pasteur Institute, Paris. The DTA gene was transferred from PGKdtabpA to pBlueScript SK- following NotI (NEB, 3189) and EcoRI (NEB, R3101) restriction. The pBlueScript SK- was digested with EcoRV (NEB, R3195) to allow insertion of the kanamycin/neomycin gene flanked by homology arms. The kanamycin/neomycin selection marker was PCR-amplified from the plasmid pL1L2GT-IRES. Homology arms were PCR-amplified from mouse genomic DNA. Fragments were purified before being ligated using In-Fusion HD Cloning Kit (Clontech, 011614).

Following what is described above, the DTA gene was transferred into the pBlueScript SK- with the same restriction enzymes. The pBlueScript SK- was digested with EcoRV to allow the insertion of the first homology arm (named 5' HA). The 5' HA was PCR-amplified from genomic DNA which had been extracted from XX ES cells. This construct was then, digested with AfeI (NEB, R0652S) to allow the insertion of the kanamycin/neomycin selection marker which was PCR-amplified from the plasmid pL1L2GT-IRES. This construct was once again digested with AfeI to allow the insertion of the second homology arm (named 3' HA). The 3' HA was PCR-amplified as described for 5' HA.

7. DNA Gel Electrophoresis

DNA fragments were separated by electrophoresis on agar gels (1% or 2%) which was stained with ethidium bromide. The agar gels were prepared by dissolving agarose in TAE 1X buffer. If needed, DNA was mixed with 6X Orange DNA loading dye before the migration. In order to estimate fragments length, GeneRuler 100 bp or 1 kb were used.

8. RNA Gel Electrophoresis

To prepare the gel, 1 g of agarose was dissolved in 72 ml of DEPC-treated water. After cooling down, 10 ml of MOPS 10X running buffer and 18 ml of 37% formaldehyde was added. 3 µg of RNA was mixed with 10 µl of Loading Dye and 10 µl of a mix composed of 20 µl of MOPS 10X, 36 µl of formaldehyde and 100 µl of formamide. The samples were then, denatured at 65°C during 15 minutes. The samples were loaded onto the gel, and MOPS 1X was used as the running buffer.

9. CRISPR/Cas9 mediated genome engineering

To create two DSBs in the imprinted gene, H19, the pX335-U6-Chimeric_BB-CBh-hSpCas9n (D10A) was acquired from Addgene (plasmid reference #42335). As for the deletion of XIC,

the eSpCas9 (1.1) was also acquired from Addgene (plasmid reference #71814). To assemble the CRISPR/Cas9 vectors, the protocol of Ran et al., 2013, was followed. Before being cloned into the pX335 or eSpCas9 (1.1), the sgRNAs were annealed with their complementary sequence to form a duplex. The sgRNAs used for our project, are described in Table 1. The sgRNAs were designed through an online CRISPR/Cas9 Design Tool (<http://crispr.mit.edu/>), they were designed on a 20 bp basis. To insert the oligonucleotides, both plasmids were restricted with BbsI. Additional bases were added to the oligonucleotides on 5' to create overhangs compatible with the ends generated by restriction with BbsI, it corresponds to the following pattern:

5'- CACC-(G)NNNNNNNNNNNNNNNNNNNN
(C)NNNNNNNNNNNNNNNNNNNN-AAAC -5'

10. Analysis of DNA demethylation using pyrosequencing coupled to bisulfite treatment

The protocol of Silvestre et al., 2015, was followed to design our pyrosequencing experiment. The primers used for pyrosequencing are described in Table 1. 400 ng of DNA for each conditions was bisulfite converted using EZ DNA Methylation-Gold™ Kit (Zymo research, D5006) according to the manufacturer's instructions. After bisulfite treatment, amplification of the sequences of interest was done with PyroMark PCR Kit (Qiagen, 978703). PCR products were verified on 1% agarose gels for each experiment before immobilization on 24-well plates using a Vacuum Prep Workstation. Analysis were done with the PyroMark Q24, and results were analysed using PyroMark Q24 Advanced.

Before running the analysis, the 'expected' sequence has to be given to the PyroMark Q24, these sequences are given in Table 2.

Usage	Name of the primers	Forward (5' -> 3') <i>Complementary sequences are in bold</i>	Reverse (5' -> 3') <i>Complementary sequences are in bold</i>
Expression analysis via qPCR			
	Beta-actin	GGCTGTATTCCCCTCCATCG	CCAGTTGGTAACAATGCCATGT
	GAPDH	CGTGCCGCCTGGAGAA	GATGCCTGCTTCACCACCTT
	ATRX	ACGATCCAGCACTATCTGCG	ACTTGTTTCCACTCATGGGCT
	Dicer	GTGCTGCAGTAAGCTGTGCTA	TTATGGTCCAGAGCTGCTTCC
	Dnmt1	AGATCCCAGAGCAATCGAGG	CAAGTCTTTGAGCCTGCCATC
	Dnmt3a	GAGCCGCCTGAAGCCC	CCTGTTCCCTCTCCTTCCTTTTCG
	Dnmt3b	TGCCAGACCTTGGAACCTC	ATTGTTTCCTGAAAGAAGGCC
	Mage A5	GAAAGGAGTTCGCCTTGCC	TCCTCCTCGGATGTCTCCA
	Mage A8	TGCCATTTGAGGTCCCTGGT	TTGGATGTTATGGGAGTCAGCC
	MeCP 2	TGGTAAAACCCGTCCGAAA	TTCTCCCTGAGCCCTAACA
	Oct4	GGAGCACGAGTGGAAGCAAC	GCTTTCATGTCCTGGGACTCCTC
	Nanog	TCTGCTACTGAGATGCTCTGCAC	GGTAGAAGAATCAGGGCTGCCT
	Prdm14	TGAACGCTGTGGAAAGGTGT	TCCCACAGGTTGAACACAGG
Vectors construction			
Amplification of the IRES-Hyg ^R cassette	IRES-Hyg ^R	ACGACGACGCGTTTCGCCCCCTC TCCCTCCC	ACGACGACGCGTTTCCTTTGCCCT CGGACGAG
First strategy - Amplification of the 5' homology arm for insertion of the reporter of endogenous methylation	5' Homology arm (HA) H19	ATTCCTGCAGGCATCAGTGCCC ATGTCCTGTT	ATTCAATTGTGGACAGTGAGCAG GAAAGC
First strategy - Amplification of the 3' Homology arm H19 for insertion of the reporter of endogenous methylation	3' HA H19	ATTGGCGCGCCTGGACAGTGAG CAGGAAAGC	ATTGGCCGGCCCCATCAGTGCCCAT GTCCTGTT
Triple PCR to repair the puromycin resistance (1)	Fragment 1	5' ccgcgccccgcaacctccccctctacgag	5' gaatgagctggcccttaatttg

Triple PCR to repair the puromycin resistance (2)	Fragment 2	5' gggtaggggagggcgcttt	5' ggggaggttcgggggcgagggtctcca
First strategy - Amplification of the 5' HA to delete the Xic	5' HA XIC	ATCGATAAGCTTGATAAGAACC AAGCGATTCTGCC	TATCCGGATCCAGATTCTAGCAG TTACCTCAAATAACCT
First strategy - Amplification of the 3' HA to delete the XIC	3' HA XIC	ATATCTAGACCCAGCGTACTCA AATGGCATCATGGGC	CTGCAGGAATTCGATACACCTTTC AGTTCCTCATGTTC
First strategy - Amplification of Kan ^R /Neo ^R to delete XIC	Kan ^R /Neo ^R	TCTGGATCCGGAATAACTTCG	GCTGGGTCTAGATATCTCGAC
Second strategy - Amplification of 5' HA to delete XIC	5' HA XIC	5' ATCGATAAGCTTGATAAGAACC AAGCGATTCTGCC	5' CTGCAGGAATTCGATAGCGCTTTC TAGCAGTTACCTCAAATAACCT
Second strategy - Amplification of Kan ^R /Neomycin ^R to delete XIC	Kan ^R /Neo ^R	5' TAACTGCTAGAAAGCTCTGGATC CGGAATAACTTCG	5' CAGGAATTCGATAGCGCTGCTGGG TCTAGATATCTCGAC
Second strategy - Amplification of 3' HA to delete XIC	3' HA	5' CAGGAATTCGATAGCGGGTCTCA TAGACAGACTCTCCA	5' TCTAGACCCAGCAGCacaccttcagttccca ttgttc
Nested-PCR screen for the deletion of the XIC in XX ES cells (1)	Nested PCR 1	5' GTATACCAGGCTGGCCTTAA	5' CCCAAGCTGTTGGTAAATGCC
Nested-PCR screen for the deletion of the XIC in XX ES cells (2)	Nested PCR 2	5' AGGTTTCGTTGGAACAACAGT	5' TCTGTGGGGTGAGTTTCAGG

Second strategy to delete one of the XIC in XX ES cells – 3' HA			
Guide of the nickase to H19 imprinted gene	sgRNA1 – H19 – UP	CACCGCATTGAATGGACAGTGA GC	
Guide of the nickase to H19 imprinted gene	Complementary sequence to sgRNA1 H19	AAACGCTCACTGTCCATTCAAT GC	
Guide of the nickase to H19 imprinted gene	sgRNA2 – H19 – UP	CACCCAAAAGTGCTGTGACTAT ACAGG	
Guide of the nickase to H19 imprinted gene	Complementary sequence to sgRNA2 H19	AAACCCCTGTATAGTCACAGCAC TTTTG	
Guide of eSpCas9	sgRNA named 3' HA (1) – UP	CACCGGTGTATGGCGCATTCCC TC	
Guide of eSpCas9	Complementary sequence to sgRNA 3' HA (1)	AAACGAGGGAATGCGCCATAC ACC	
Guide of eSpCas9	sgRNA named 3' HA (2) – UP	CACCGAATTCGAAAAGGGCGGT TA	
Guide of eSpCas9	Complementary sequence to sgRNA 3' HA (2)	AAACTAACCGCCCTTTTCGAAT TC	
Guide of eSpCas9	sgRNA named 5' HA (1) – UP	CACCGTGCTTTTGAATTGCAGT AAG	
Guide of eSpCas9	Complementary sequence to sgRNA 5' HA (1)	AAACCTTACTGCAATTCAAAAAG CAC	
Guide of eSpCas9	sgRNA named 5' HA (2) – UP	CACCGTGCACTGGCAGAATC GCT	
Guide of eSpCas9	Complementary sequence to sgRNA 5' HA (2)	AAACAGCGATTCTGCCAGTGTG CAC	
Verification of insertion of sgRNAs into pX335 or eSpCas9 plasmid	pX335-Rev		TACCGTAAGTTATGTAACGGG
Bisulfite treatment coupled to pyrosequencing to analyse DNA methylation of ES cells			
Amplification of bisulfite converted DNA to analyse DNA	H19	TAGGAGGTATAAGAATTTTGTAAGGAGAT	ATTCCAATACCAAAAATAAAAAAACTCTT

methylation at H19 imprinted gene			
Amplification of bisulfite converted DNA to analyse DNA methylation at Igf2r imprinted gene	Igf2r	GGGAAAGGGAAGGGAAAGTT	TCCTCCCCTTATACAATTTACACC
Amplification of bisulfite converted DNA to analyse DNA methylation at Igf2r imprinted gene	5' Igf2r	AGGGGATTAGGAGGGTTTAG	TCCTCCCCTTATACAATTTACACC
Amplification of bisulfite converted DNA to analyse DNA methylation at IAPs LTR	IAPs	GGTTTTGGAATGAGGGATTTT	CTCTACTCCATATACTCTACCTTC
Sexing PCR			
Amplification Zfy (region specific to the Y- chromosome) to observe eventual contamination of XX ES cells	Zfy	GTAGGAAGAATCTTTCTCATGCT G	TTTTTGAGTGCTGATGGGTGACGG

Table 1 – Sequence of the primers

Pyrosequencing analysis to assess global DNA methylation using IAPs in XX and XY ES cells

Y corresponds to C or T

Sequence analysed for mouse IAPs LTR before bisulfite treatment
CACTGCAGTTGAGCCGGCTACCCGACGTCGGTTAGTCCCTCACTGTGCAGGAT CCGCTTTATATTGAGAGGATTTTTTCCCTGCCCCAAAGCAAAAGAGAGAGAGA ACGAAGAATGTGAGAACGAGGACTTCTACATTCGTTATTTCAAACGGCGTCT T
Sequence analysed for mouse IAPs LTR after bisulfite treatment
TATTGTAGTTGAGTYGGTTATTYGAYGTYGGTTAGTTTTTTATTGTGTAGGATT YGTTTTATATTGAGAGGATTTTTTTTTTTGTTTTAAAGTAAAAGAGAGAGAGAAY GAAGAATGTGAGAAYGAGGATTTTATATTYGTTATTTTAAAYGGYGT
Injection sequence for PyroMark Q24
GCTATGTAGTGATGTCTGTGATCTGTATCAGTCGTAGTTATGTGTAGTATCTGT TATATGAGAGATTGTTAGTAAGAGAGAGAGATATCGAGATGTGAGTATCGAGAT TATGATCGTATTAATCAGTCG

Table 2 - The reference sequence which was used to carry out our pyrosequencing experiment.

F. REFERENCES

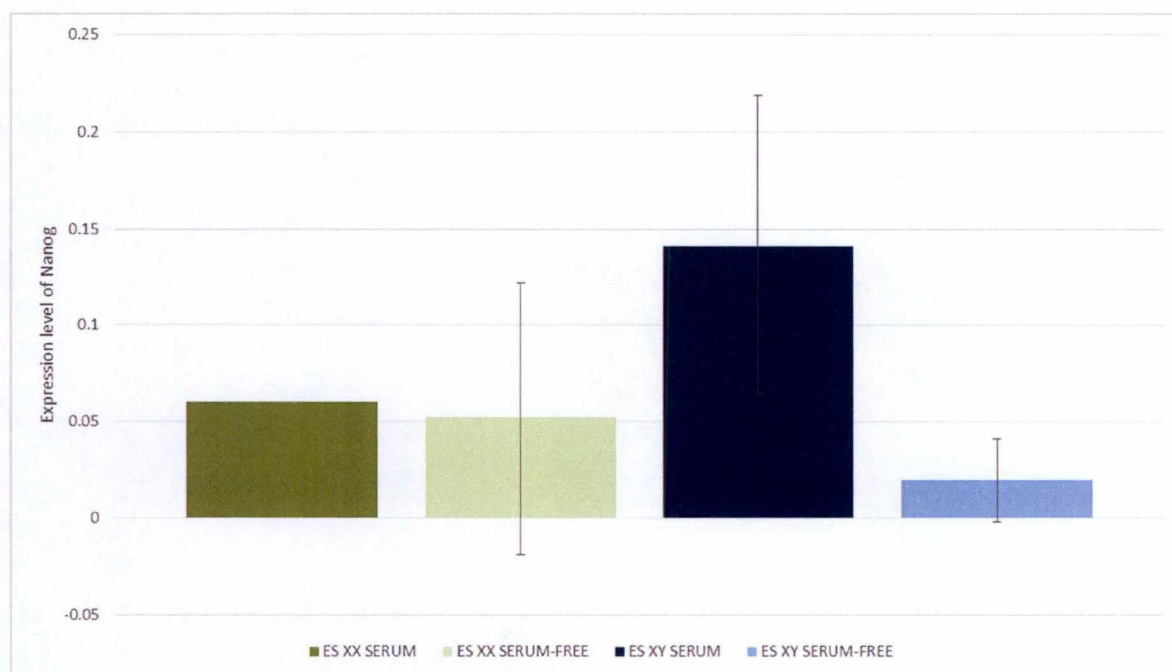
1. Holliday, R. Epigenetics: A Historical Overview. *Epigenetics* **2294**, 76–80 (2006).
2. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression : how the genome integrates intrinsic and environmental signals. *Nat. Genet.* **33**, 245–254 (2003).
3. Felsenfeld, G. A Brief History of Epigenetics. *Cold Spring Harb. Lab. Press* (2014). doi:10.1101/cshperspect.a018200 Cite this article as Cold Spring Harb Perspect Biol 2014;6:a018200 1
4. Putiri, E. L. & Robertson, K. D. Epigenetic mechanisms and genome stability. *Clin. Epigenetics* 299–314 (2011). doi:10.1007/s13148-010-0017-z
5. Denis, H., Ndlovu, M. N. & Fuks, F. Regulation of mammalian DNA methyltransferases: a route to new mechanisms. *Nat. Publ. Gr.* **12**, 647–656 (2011).
6. Robertson, K. D. DNA methylation, methyltransferases, and cancer. *Oncogene* (2001).
7. Siedlecki, P. & Zielenkiewicz, P. Mammalian DNA methyltransferases. *Acta Biochim. Pol.* **53**, 245–256 (2006).
8. Latham, T., Gilbert, N. & Ramsahoye, B. DNA methylation in mouse embryonic stem cells and development. *Cell Tissue Res.* 31–55 (2008). doi:10.1007/s00441-007-0537-9
9. Saitou, M., Kagiwada, S. & Kurimoto, K. Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. *Development* **0000**, (2012).
10. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* 6–21 (2002). doi:10.1101/gad.947102.6
11. Vinson, C. & Chatterjee, R. CG methylation. *Epigenomics* **4**, 655–663 (2013).
12. Barlow, D. P. & Bartolomei, M. S. Genomic Imprinting in Mammals. *Cold Spring Harb. Perspect. Biol.* **31**, 493–525 (2014).
13. Bartolomei, M. S. Genomic imprinting : recognition and marking of imprinted loci. *Curr. Opin. Genet. Dev.* **22**, 72–78 (2013).
14. Li, E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat. Publ. Gr.* **3**, 662–673 (2002).
15. Grabole, N. *et al.* DNA methylation dynamics during the mammalian life cycle. *Philos. Trans. R. Soc. B Biol. Sci.* **14**, (2013).
16. Green, K., Dean, W., Reik, W. & Morgan, H. D. Epigenetic reprogramming in mammals. *Hum. Mol. Genet.* **14**, 47–58 (2005).
17. 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).
18. Bagci, H. & Fisher, A. G. DNA Demethylation in Pluripotency and Reprogramming : The Role of Tet Proteins and Cell Division. *Stem Cell* **13**, 265–269 (2013).
19. Hackett, J. A. & Surani, M. A. DNA methylation dynamics during the mammalian life cycle. *Philos. Trans. R. Soc. B Biol. Sci.* 1–8 (2012).
20. Lee, J. T. & Bartolomei, M. S. Review X-Inactivation , Imprinting , and Long Noncoding RNAs in Health and Disease. *Cell* **152**, 1308–1323 (2013).

21. Augui, S., Nora, E. P. & Heard, E. Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat. Publ. Gr.* **12**, 429–442 (2011).
22. Galupa, R. & Heard, E. ScienceDirect X-chromosome inactivation : new insights into cis and trans regulation. *Curr. Opin. Genet. Dev.* **31**, 57–66 (2015).
23. Sohaila, R. Non-random X-chromosome inactivation in mouse X-autosome translocation embryos — location of the inactivation centre. *J. Embryol. Exp. Morphol.* **22**, 1–22 (1983).
24. Plath, K., Mlynarczyk-evans, S., Nusinow, D. A. & Panning, B. Xist RNA and the mechanism of X chromosome inactivation. *Annu. Rev. Genet.* **36**, 233–278 (2002).
25. Portela, A. & Esteller, M. review Epigenetic modifications and human disease. *Nat. Publ. Gr.* **28**, 1057–1068 (2010).
26. Voon, H. P. J. & Wong, L. H. New players in heterochromatin silencing : histone variant H3 . 3 and the ATRX / DAXX chaperone. *Nucleic Acids Res.* **44**, 1496–1501 (2016).
27. Rebollo, R. *et al.* Retrotransposon-Induced Heterochromatin Spreading in the Mouse Revealed by Insertional Polymorphisms. **7**, (2011).
28. Habibi, E. *et al.* Short Article Whole-Genome Bisulfite Sequencing of Two Distinct Interconvertible DNA Methylomes of Mouse Embryonic Stem Cells. *Cell Stem Cell* 360–369 (2013). doi:10.1016/j.stem.2013.06.002
29. Pe, R. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *Elife* 1–30 (2016). doi:10.7554/eLife.11418
30. Voon, H. P. J., Hughes, J. R., Higgs, D. R. & Gibbons, R. J. ATRX Plays a Key Role in Maintaining Silencing at Interstitial Heterochromatic Loci and Imprinted Article ATRX Plays a Key Role in Maintaining Silencing at Interstitial Heterochromatic Loci and Imprinted Genes. *Cell* **11**, 405–418 (2015).
31. Zvetkova, I. *et al.* Global hypomethylation of the genome in XX embryonic stem cells. *Nat. Genet.* **37**, 1274–1279 (2005).
32. 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).
33. Welham, M. J., Kingham, E., Sanchez-ripoll, Y., Kumpfmüller, B. & Storm, M. Controlling embryonic stem cell proliferation and pluripotency : the role of PI3K- and GSK-3- dependent signalling. *Biochem. Soc. Trans.* **39**, 674–678 (2011).
34. Burdon, T., Smith, A., Savatier, P. & Smith, A. Signalling , cell cycle and pluripotency in embryonic stem cells. *Trends Cell Biol.* **12**, 432–438 (2002).
35. Tosolini, M. & Jouneau, A. Acquiring Ground State Pluripotency : Switching Mouse Embryonic Stem Cells from Serum / LIF Medium to 2i / LIF Medium. *Methods Mol. Biol.* (2015). doi:10.1007/7651
36. Hackett, J. A. *et al.* Synergistic mechanisms of DNA demethylation during transition to ground-state pluripotency. *Stem Cell Reports* **1**, 518–531 (2013).
37. 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).

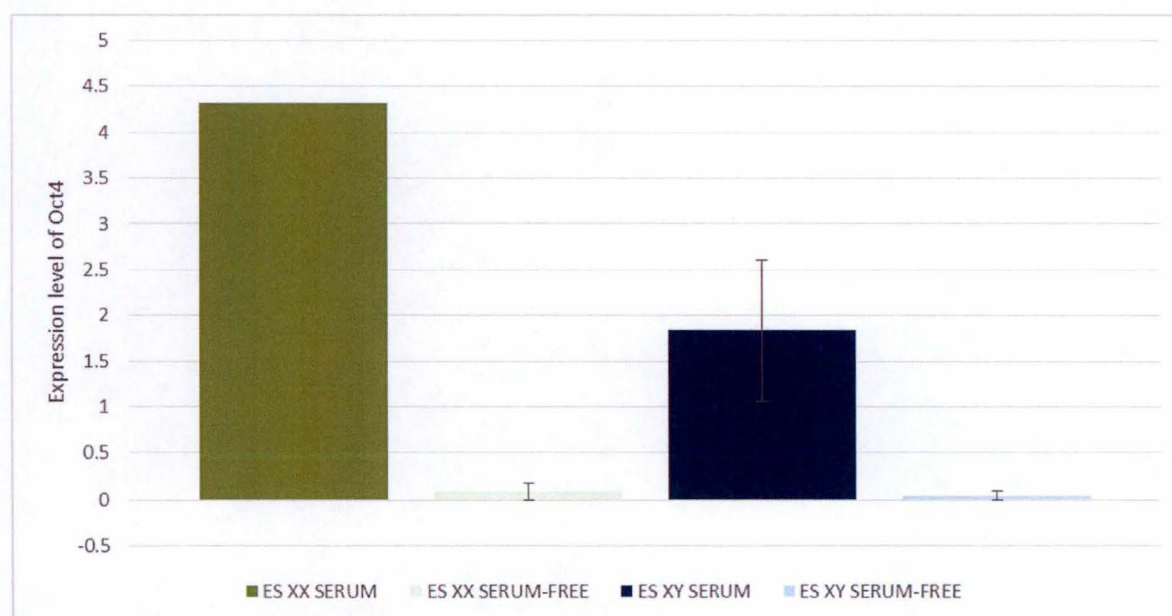
38. Zimmermann, C. A., Hoffmann, A., Raabe, F. & Spengler, D. Role of Mecp2 in Experience-Dependent Epigenetic Programming. *Genes (Basel)*. **6**, 60–86 (2015).
39. Adkins, N. L. & Georgel, P. T. MeCP2 : structure and function. *Biochem. Cell Biol.* **11**, 1–11 (2011).
40. Webster, J. O. Genome-Scale DNA Methylation analysis in distinct pluripotent stem cells. (2014).
41. Nesterova, T. B. *et al.* Dicer regulates Xist promoter methylation in ES cells indirectly through transcriptional control of Dnmt3a. *Epigenetics Chromatin* **21**, 1–21 (2008).
42. Sinkkonen, L. *et al.* MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* **15**, 259–267 (2008).
43. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
44. Barrangou, R. & Marraffini, L. A. CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Mol. Cell* **54**, 234–244 (2014).
45. Marraffini, L. A. CRISPR-Cas immunity in prokaryotes. *Nature* **526**, 55–61 (2015).
46. Hsu, P. D., Lander, E. S. & Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **157**, 1262–1278 (2014).
47. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 1–8 (2013). doi:10.1038/nbt.2647
48. Ran, F. A. *et al.* Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificit. *Cell* **154**, 1380–1389 (2014).
49. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science (80-.)*. **351**, 84–88 (2016).
50. Maenner, S. & Chu, E. T. Live Cell Imaging of the Nascent Inactive X Chromosome during the Early Differentiation Process of Naive ES Cells towards Epiblast Stem Cells. *PLoS One* 1–20 (2014). doi:10.1371/journal.pone.0116109
51. Gut, I. G. DNA methylation analysis by pyrosequencing. *Nat. Protoc.* **2**, 2265–2275 (2007).
52. Rowe, H. M. *et al.* TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. 452–461 (2013). doi:10.1101/gr.147678.112.In
53. Rowe, H. M., Friedli, M., Offner, S., Verp, S. & Mesnard, D. De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs / KAP1 and ESET. *Development* **140**, 519–29 (2013).
54. Sato, S., Yoshimizu, T., Sato, E. & Matsui, Y. Erasure of Methylation Imprinting of Igf2r During Mouse Primordial Germ-Cell Development. **50**, 41–50 (2003).
55. Zhang, L., Jia, R., Palange, N. J. & Satheka, A. C. Large Genomic Fragment Deletions and Insertions in Mouse Using CRISPR / Cas9. *PLoS One* **10**, 1–14 (2015).
56. Chaumeil, J., Augui, S., Chow, J. C. & Heard, E. in *The Nucleus: Volume 1: Nuclei and Subnuclear Components* **1**, 297–308 (2008).
57. Chaumeil, J. & Heard, E. X-Chromosome Inactivation in Mouse Embryonic Stem

- Cells : Analysis of Histone Modifications and ... *Methods Enzymol.* (2004). doi:10.1016/S0076-6879(03)76027-3
58. Barakat, T. S. & Gribnau, J. Combined DNA-RNA Fluorescent In situ Hybridization (FISH) to Study X Chromosome Inactivation in Differentiated Female Mouse Embryonic Stem Cells. *J. Vis. Exp.* 1–6 (2014). doi:10.3791/51628
59. Shen-ong, G. L. C. & Cole, M. D. Differing Populations of Intracisternal A-Particle Genes in Myeloma Tumors and Mouse Subspecies. *J. Virol.* **42**, 411–421 (1982).
60. Dupressoir, A. & Heidmann, T. Expression of intracisternal A-particle retrotransposons in primary tumors of oncogene-expressing transgenic mice. *Oncogene* **14**, 2951–2958 (1997).
61. Dupressoir, A. & Heidmann, T. Germ Line-Specific Expression of Intracisternal A-Particle Retrotransposons in Transgenic Mice. *Mol. Cell. Biol.* **16**, 4495–4503 (1996).
62. Lueders, K. K. & Mietz, J. A. Nucleic Acids Research Structural analysis of type H variants within the mouse intracisternal A-particle sequence family Nucleic Acids Research. *Nucleic Acids Res.* **14**, 1495–1510 (1986).
63. Kurdyukov, S. & Bullock, M. DNA Methylation Analysis : Choosing the Right Method. *MDPI Biol.* 1–21 (2016). doi:10.3390/biology5010003
64. Stelzer, Y. *et al.* Tracing Dynamic Changes of DNA Methylation at Resource Tracing Dynamic Changes of DNA Methylation at Single-Cell Resolution. *Cell* **163**, 218–229 (2015).

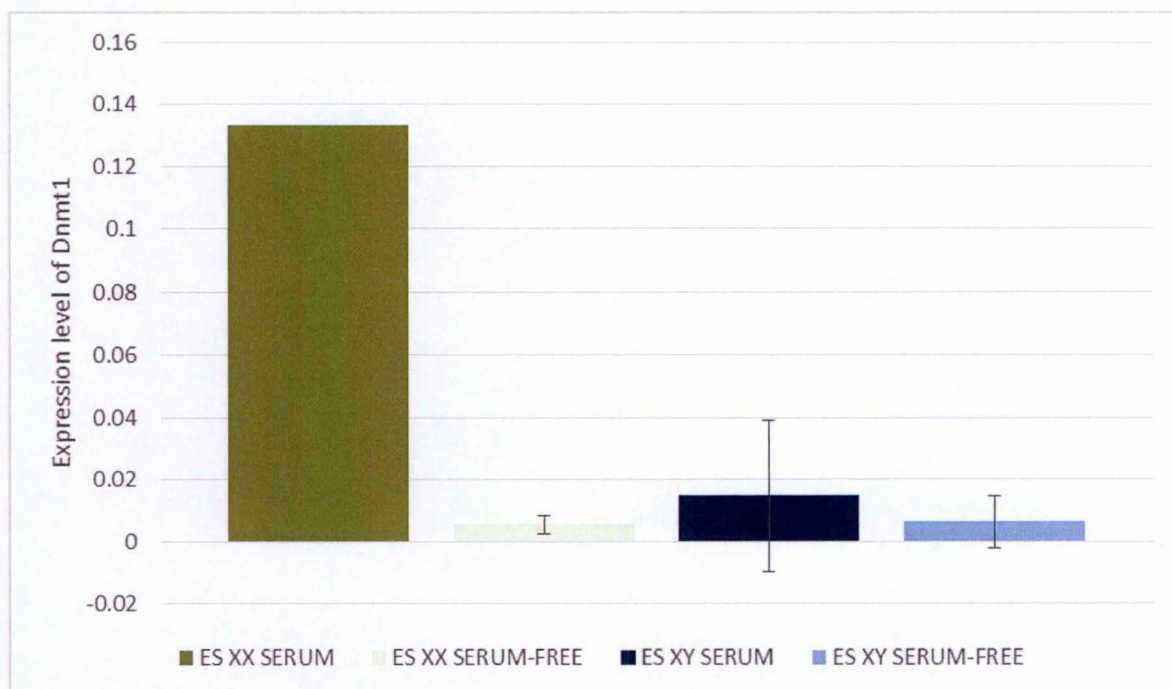
G. SUPPLEMENTARY DATA



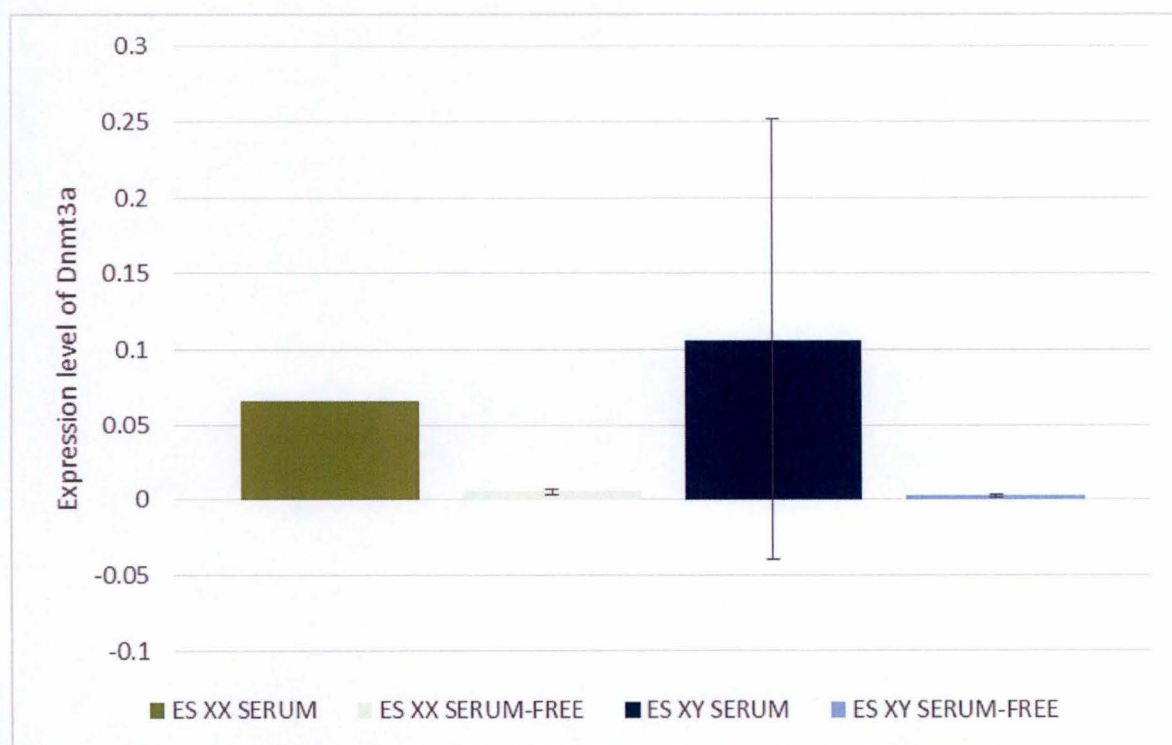
Supplementary Figure 1 - Expression of Nanog in XY and XX ES cells. RT-qPCR on messenger RNA, normalized by GAPDH.



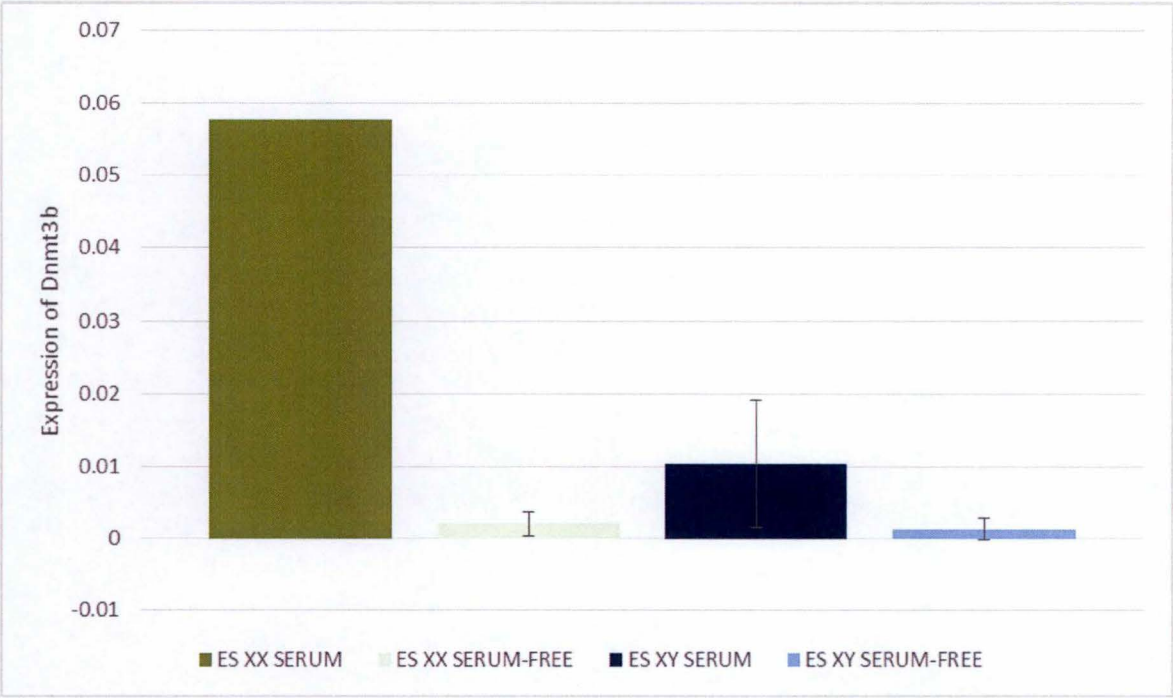
Supplementary Figure 2 - Expression of Oct4 in XY and XX ES cells. RT-qPCR on messenger RNA, normalized by GAPDH.



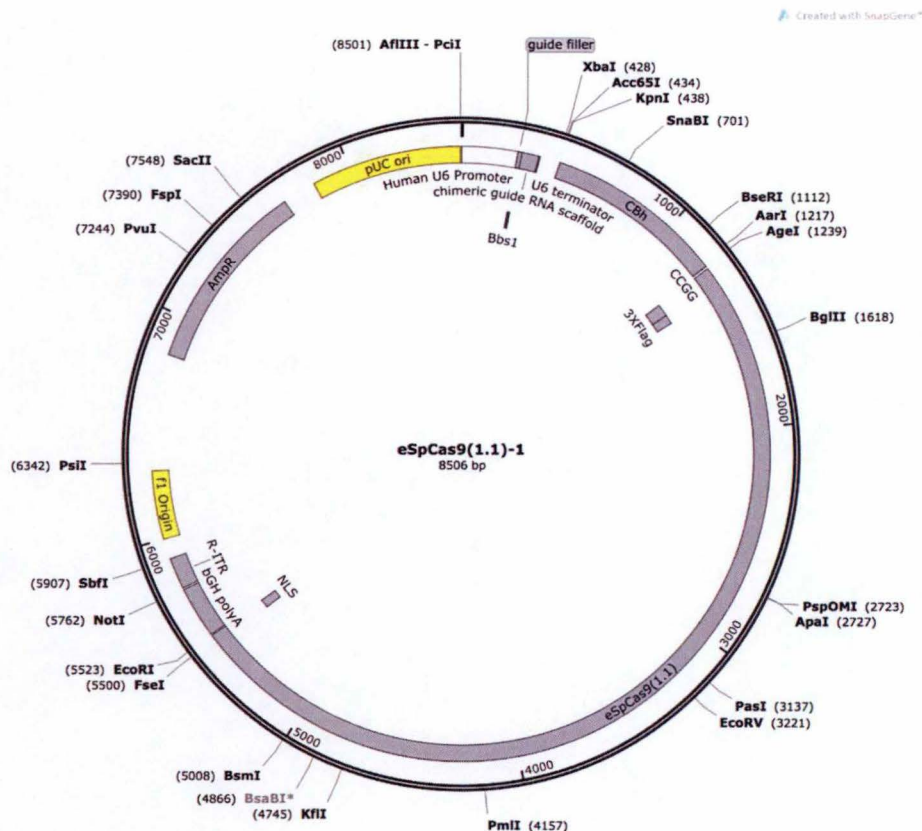
Supplementary Figure 3 - Expression of Dnmt1 in XY and XX ES cells. RT-qPCR on messenger RNA, normalized by GAPDH.



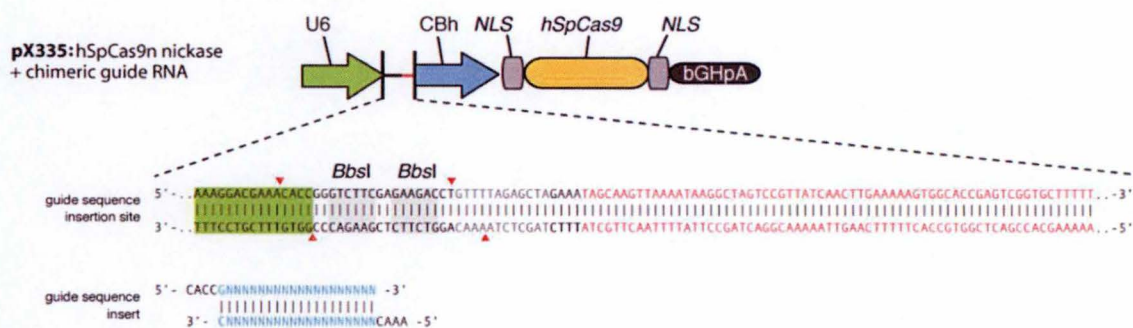
Supplementary Figure 4 - Expression of Dnmt3a in XY and XX ES cells. RT-qPCR on messenger RNA, normalized by GAPDH.



Supplementary Figure 5 - Expression of Dnmt3b in XY and XX ES cells. RT-qPCR on messenger RNA, normalized by GAPDH.



Supplementary Figure 6 - Graphical map of the eSpCas9 used to delete the X inactivation center in XX ES cells (Addgene.com).



Supplementary Figure 7 - Graphical map of the pX335 used to insert the reporter of endogenous methylation in XY ES cells (Addgene.com)