

**THÈSE** présentée par :

**Ambre BENDER**

soutenue le : 23 Novembre 2017

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Aspects moléculaires et cellulaires Biologie

**Méthylation de l'ADN & Identité Cellulaire :**  
**Fonctions de la méthylation de l'ADN dans les**  
**lignages gamétiques et hématopoïétiques chez la**  
**souris.**

**THÈSE dirigée par :**

**Dr WEBER Michaël**

Directeur de Recherche, Laboratoire Biotechnologie et  
Signalisation Cellulaire, UMR7242

**RAPPORTEURS :**

**Dr BOURC'HIS Deborah**

Directrice de Recherche, Institut Curie, U934,  
UMR3215

**Dr ARNAUD Philippe**

Directeur de Recherche, Institut de Génétique  
Reproduction et Développement, UMR6286

---

**AUTRES MEMBRES DU JURY :**

**Dr TAVIAN Manuella**

Chargée de Recherche, Etablissement Français du  
Sang, U949

# Sommaire

---

<b>REMERCIEMENTS</b>	<b>5</b>
<b>ABREVIATIONS</b>	<b>9</b>
<b>INTRODUCTION</b>	<b>11</b>
<b>1. LA METHYLATION DE L'ADN</b>	<b>15</b>
a. Un contexte phylogénique	16
b. Un contexte nucléotidique	18
c. Un contexte génomique	20
<b>2. LES ACTEURS DE LA METHYLATION DE L'ADN</b>	<b>23</b>
a. DNMT1	23
b. UHRF1	26
c. DNMT2	29
d. DNMT3	30
a) DNMT3A et DNMT3B	30
b) DNMT3L	32
c) DNMT3c	33
<b>3. LA DEMETHYLATION DES CYTOSINES</b>	<b>35</b>
a. La déméthylation passive	35
b. La déméthylation active	35
<b>4. TECHNIQUES D'ETUDES DE LA METHYLATION DES CYTOSINES</b>	<b>38</b>
a. Les enzymes de restriction	40
b. Conversion au bisulfite de Sodium	40
c. Les méthodes d'analyses à l'échelle génomique	42
<b>5. ROLE DE LA METHYLATION DE L'ADN AU COURS DU DEVELOPPEMENT</b>	<b>45</b>
a. Régulation de la pluripotence	47
b. Les régions soumises à empreintes parentales (ICRs)	51
c. L'inactivation du chromosome X	54
d. L'inactivation des éléments transposables	55
e. Régulation génique	57
<b>6. LES PATHOLOGIES ASSOCIEES A LA DEREGLATION DE LA METHYLATION DE L'ADN</b>	<b>59</b>
a. Les cancers	59
b. Les pathologies associées aux mutations des gènes TETs	61
c. Les pathologies associées aux mutations des gènes DNMTs	63
d. Les autres pathologies associées à des défauts de méthylation de l'ADN	64
e. Les pathologies associées à des défauts de méthylation des ICRs	65
<b>MATERIELS ET METHODES</b>	<b>66</b>

<b>I.</b>	<b>LA METHYLATION DE L'ADN AU COURS DU DEVELOPPEMENT EMBRYONNAIRE DES CELLULES PRIMORDIALES GERMINALES (PGCs)</b>	<b>76</b>
A.	Introduction	76
	1. Le développement des PGCs	76
	2. Reprogrammation épigénétique des PGCs : que sait-on ?	78
	3. Objectifs du projet	81
B.	Résultats	83
	1. Isolation des PGCs	83
	2. Dynamique du méthylome au cours du développement des PGCs	83
	a. Cinétique globale .....	83
	b. La cinétique de déméthylation des PGCs suggère une déméthylation passive	85
	c. Certaines régions se déméthylent plus tardivement que le génome .....	87
	d. La phase de reméthylation .....	89
	e. Les PGCs mâles accumulent de la méthylation non-CG .....	91
	3. Analyse des régions résistantes à la vague de déméthylation	93
	a. Quelles régions résistent à la vague de déméthylation des PGCs ? .....	93
	b. Cinétique des éléments transposables au cours du développement des PGCs.	95
	4. DNMT1 : un acteur dans le maintien de la méthylation au cours du développement des PGCs	99
	a. Génération de souris avec une inactivation conditionnelle de Dnmt1 dans les PGCs .....	99
	b. Dnmt1 est nécessaire pour le maintien de la méthylation des transposons dans les PGCs.....	101
	5. UHRF2 : un nouvel acteur de la méthylation de l'ADN dans les PGCs	104
	a. Stratégie d'obtention de PGCs Uhrf2 <sup>-/-</sup> .....	106
	b. Uhrf2 est nécessaire pour la méthylation des transposons dans les PGCs	106
	c. Uhrf2 n'est pas impliqué dans la méthylation de l'ADN dans les cellules somatiques.....	108
	d. L'absence d'Uhrf2 n'induit pas la réexpression des transposons dans les PGCs	110
	e. Analyse préliminaire du phénotype reproductif des souris Uhrf2 <sup>-/-</sup> .....	110

**DISCUSSION ET PERSPECTIVES : PGCS**

**RESULTATS**

<b>II.</b>	<b>LA METHYLATION DE L'ADN AU COURS DU DEVELOPPEMENT EMBRYONNAIRE MURIN DES CELLULES SOUCHES HEMATOPOÏÏTIQUES (HSCs)</b>	<b>121</b>
------------	--	------------

A.	Introduction	121
	1. Le développement des HSCs	121
	2. HSCs et Méthylation de l'ADN, que sait-on ?	124
	3. Objectifs du projet	126
B.	Résultats	128
	1. Isolation des HSCs	128
	2. Dynamique du méthylome des HSCs au cours du développement	128
	a. Cinétique globale .....	128
	b. La phase précoce d'acquisition de la signature hématopoïétique .....	132
	c. La phase tardive d'acquisition de la signature hématopoïétique .....	134
	3. Dnmt3a et Dnmt3b sont acteurs dans la mise en place de la signature de méthylation des HSCs	136
	a. Dnmt3b est l'acteur principal de la méthylation précoce des HSCs .....	136
	b. Génération de modèles de souris avec une inactivation conditionnelle des gènes Dnmt3a/b dans les HSCs au cours du développement .....	138
	c. Dnmt3a et Dnmt3b participent à la mise en place des profils de méthylation des HSCs de la moelle osseuse .....	142
	d. L'inactivation conditionnelle de Dnmt3a dans les HSCs semble influencer la survie des animaux .....	143
<b><u>DISCUSSION ET PERSPECTIVES : HSCS</u></b>		<b>144</b>
<b><u>BIBLIOGRAPHIE</u></b>		<b>149</b>
<b><u>ANNEXE 1 : LIGNEES MURINES UTILISEES</u></b>		<b>183</b>
<b><u>ANNEXE 2 : LISTE DES ANTICORPS UTILISES EN CYTOMETRIE DE FLUX</u></b>		<b>186</b>
<b><u>ANNEXE 3 : PRIMERS DE QPCR ET DE PCR BISULFITE POUR COBRA</u></b>		<b>187</b>
<b><u>ANNEXE 4 : PRIMERS DE GENOTYPAGE</u></b>		<b>188</b>
<b><u>ANNEXE 5 : TABLEAU DES STATISTIQUES DE RRBS DU PROJET PGCS</u></b>		<b>189</b>
<b><u>ANNEXE 5 : TABLEAU DES STATISTIQUES DE RRBS DU PROJET HSCS</u></b>		<b>190</b>
<b><u>ANNEXE 6 : PROTOCOLES RRBS</u></b>		<b>191</b>
<b><u>REVUE: DNA METHYLATION: AN IDENTITY CARD FOR BRAIN CELLS</u></b>		<b>191</b>



# Remerciements

---

Je tiens tout d'abord à remercier les membres du jury, Déborah BOURC'HIS, Manuella TAVIAN et Philippe ARNAUD qui ont accepté de juger mon travail de thèse mais aussi de décaler la soutenance.

Ensuite, je tiens à dire un grand merci à mon directeur de thèse, Michaël WEBER alias « le chef », sans qui cette aventure n'aurait pas été possible et qui aurait pu s'achever. Il m'a permis d'apprendre énormément de choses et de m'épanouir. Outre le fait que je pense qu'il s'agit d'une personne ayant un excellent niveau scientifique, c'est aussi une personne au grand cœur. Lors des moments les plus sombres qui se sont déroulés au cours de ma thèse, il a su trouver les mots pour me redonner le sourire et me remotiver. Il ne m'a jamais abandonné et, bien au contraire, il m'a soutenu au-delà de mes attentes. Ainsi, Michaël, je te remercie du plus profond de mon cœur pour tout ce que tu m'as apporté.

Lorsque je suis arrivée dans le laboratoire de Michaël, l'équipe venait d'être fraîchement montée et nous n'étions que 4. Parmi les premiers arrivants, j'ai rencontré notre célèbre Sylvain. Un post-doc atypique issu d'un croisement entre un bio-informaticien et un scientifique. C'est lui qui m'a introduit dans le domaine des PGCs et la bio-informatique. Nous avons vécu beaucoup de choses ensemble et, malgré son petit caractère de cochon, il m'a appris beaucoup tant au niveau de la science qu'au niveau de la vie. Je remercie ce petit filou qui me manque beaucoup depuis son départ.

Puis, toujours dans le thème de profil atypique, j'ai fait la connaissance d'Elouan. Lui aussi un sacré croisement ! Un quart scientifique, un quart motard, un quart militaire et un quart « prude ». C'était notre « TopGun » à nous sur tous les points. J'ai eu le bonheur de pouvoir partager mon bureau avec lui ainsi que des parties de basket improvisées dans notre bureau avec le reste de l'équipe. Merci beaucoup mon p'tit Elouan pour ces années passées à tes côtés, ta patience et ta bienveillance.

Viens ensuite un homme tout aussi particulier, Ghislain. Ce troubadour, muni à l'époque d'un piercing, est devenu une des personnes qui m'est la plus chère à ce jour. Lorsque je l'ai rencontré, tout comme les autres hommes de l'équipe, il me semblait très calme, voir trop calme. Mais c'était sans compter sur Gigi et sa folie. J'ai passé les meilleures années de ma thèse à ses côtés. Nous avons expérimenté des heures et des heures ensemble, se soutenant mutuellement dans nos thèses, avec toujours un très bon fond musical pour rythmer tout cela, même lorsque

monsieur chantait à tue-tête prince Ali ^^. Mon cher Gigi, je ne trouve pas de mots pour t'exprimer ma reconnaissance pour tout ce que tu as fait pour moi et pour te remercier de tous ces merveilleux moments que nous avons partagés. Tu es une personne en or et un ami hors-pair, et pour cela, je t'aime très fort. Et en plus de tout cela, tu m'as fait connaître ton âme sœur, un sacré bout de femme : LA ROSACE !!!!!

Ha, cette petite Rose-Marie, si mimi mais si féroce. C'est elle qui m'a introduit dans le monde de l'hématopoïèse et qui m'a appris la dissection de moelle osseuse de souris au mortier avec son air sadique^^. C'est aussi elle (et mon cher et tendre) que Ghislain et moi attendions des heures et des heures après avoir fini au labo. Au cours des dernières années, j'ai appris à la connaître et tout comme Ghislain, elle fait partie des gens qui comptent le plus pour moi aujourd'hui. Pendant mes soucis de santé, elle a été l'un de mes piliers et je pense qu'elle le restera. Elle n'a pas hésité à venir travailler auprès de moi après sa thèse pour me servir de bras droit quand le mien ne fonctionnait plus. Sa bonne humeur, son humour, bref, qui elle est, a été et est toujours un bol de bonheur pour moi. Et, là encore, je n'ai pas de mots pour te dire à quel point je te remercie pour tout ce que nous avons partagé ensemble, mis à part que tout comme ton homme, je t'aime très fort.

Puis le laboratoire s'est agrandi et a accueilli pleins de belles personnes que je suis heureuse de connaître : Andrea, Judith, Michaël (alias minimike), Thomas, Sirine, Hala, Manon, Jitesh, Patrick et Anaïs.

La première c'est l'ESPAGNOL. Merci beaucoup ma p'tite Andrea d'avoir choisi la paillasse à côté de moi et d'avoir ce merveilleux accent qui ensoleille nos journées. Merci pour ta bonne humeur et les éclats de rire que nous avons partagé, notamment à Heidelberg ^^ (je me passe de commentaires, car toi-même tu sais de quoi je parle). La prochaine à soutenir c'est toi !!!! Alors bonne chance et sache que je serai toujours là pour toi si tu as besoin comme tu l'as été pour moi.

Je tiens aussi à remercier Judith qui m'a permis d'évoluer personnellement. Nos dernières conversations m'ont beaucoup appris sur la vie. Ne baisse jamais les bras et reste motivée car à cœur vaillant rien d'impossible.

Mon cher Mini-mike... Je me rappellerai toujours de la première fois où tu es venu au laboratoire...et comment tu as su maîtriser la bête, alias Bourrino. Merci pour ton aide en bio-informatique, ta disponibilité, ta patience, et surtout pour ta bonne humeur. Merci aussi de ne pas m'avoir dénoncé pour harcèlement lors de la rédaction de ma thèse.

Merci aussi à la p'tite Tomate !!!!! Toujours de bonne humeur, vaillant et avec un sacré humour noir. Ne change pas !!! Et garde le cap, car toi aussi tu as une thèse à rédiger^^

Puis, je tiens à dire un grand merci à Sirine, ma stagiaire. J'ai vraiment été heureuse d'avoir pu être ton encadrante malgré mes soucis de santé qui ne m'ont pas permis d'en faire autant pour toi comme je l'aurais voulu. Tu es une très belle personne qui est très doué en science. Avec mes absences répétées, tu as su te gérer et gérer nos projets d'une main de maître. Amuse-toi bien en thèse !!!!

Et enfin, sont arrivés les deniers Hala, Manon, Jitesh, Patrick et Anaïs.

Je tiens tout d'abord à remercier Hala, une post-doc libanaise passionnée. Merci pour tous tes bons conseils et l'harmonie que tu sèmes au laboratoire. Ta passion pour la science est contaminante et ta bonne humeur au travail est un vrai plaisir. A chaque fois que j'ai eu besoin de quoique ce soit, tu étais toujours disponible pour moi, notamment pour ma thèse. Merci pour tout mais aussi pour ces belles conversations que nous avons eu concernant la science et la vie.

Puis, il y a eu la bretonne qui a rejoint le groupe. Une autre passionnée de science mais aussi de crêpes et de beurre salé. Merci beaucoup pour avoir repris en charge l'animalerie et d'avoir pris soin de nos petites bêtes. Merci aussi de t'être occupé de Sirine lorsque je n'étais pas là. Et un autre merci pour le projet UHRF2 que tu as repris !!!!! Tu vas cartonner pour ta thèse !!!

Et, il y a eu le post-doc népalais, Jitesh. Thank you a lot Mister gloves!!!!!! Your consummation in gloves is proportional to your good mood and your kindness. And thank you too, for taking care of me when I was in my wheelchair. I hope to you all the best, and I'm very happy to come back to the lab in order to more know you. I forgot ... thank you for disabling me even more than I was already with a Pasteur pipet^^

Ensuite, il y a eu Patrick, alias Patoche et Anaïs. Je n'ai pas pu passer beaucoup de temps à vos côtés mais pour le peu que j'ai vu de vous, sachez que je trouve que vous êtes de très belles personnes. Merci Patrick, pour le temps que tu m'as accordé à discuter de la vie et à partager ton opinion. Merci Anaïs pour l'aide que tu m'as fourni pour ma thèse.

Vous tous, merci. J'ai hâte de revenir au labo et d'en apprendre plus sur vous car vous m'avez l'air de personnes exceptionnelles et je suis heureuse d'avoir pu croiser votre route.

Merci aussi à tous les autres membres de l'IREBS que j'ai pu côtoyer. Avec un remerciement tout particulier à Marc (mon chauffeur particulier), Anne et Aurélia (mes voisines de bureau, qui savent vous faire rire et qui sont les personnes les plus polies que je n'ai jamais rencontré). Un grand merci aussi à Sylvie et Claire qui ont été là lorsque ce n'était pas facile et qui ont toujours su m'aiguiller. Merci aussi à Sophie, Stéphanie et Colette pour leur aide précieuse.

Je tiens aussi à remercier Bruno, Françoise et Valérie, qui m'ont aidé tout au long de ma thèse avec mes questions de sciences. Merci aussi à Jean-Luc, le grand chef, qui m'a aidé dans mon devenir professionnelle et qui a su trouver les mots à chacune de mes questions. Je tiens aussi à remercier Maria-Elena et Michel, mes deux précédents mentors qui m'ont poussé à faire une thèse et qui ne m'ont pas laissé tomber lorsque j'avais besoin d'eux au cours des dernières années.

Je tiens aussi à remercier très particulièrement Claudine et Murielle qui m'ont aidé et accompagnée tout au long de mes tris cellulaires, sans compter leur bonne humeur toujours au poste et nos fameux cafés.

Enfin, je tiens à remercier ma douce moitié qui a vécu ces dernières années auprès de moi et sans qui l'aventure aurait pu se terminer. A chaque fois, il a su trouver les mots et me donner la force d'avancer quel que soit la situation, quelques soient les problèmes m'affectant. Merci pour ta patience (et je sais qu'il t'en a fallu beaucoup), ta gentillesse et ton amour inébranlable. Je t'aime.

Et si j'ai oublié quelqu'un je m'en excuse, mais je sais que je le ou la remercie aussi. Au cours des dernières années l'ensemble des personnes que j'ai rencontré m'ont permis de grandement évoluer. Que mes relations avec ces personnes furent bonnes ou moins bonnes, elles m'ont permis de grandir philosophiquement parlant et de ce fait, je ne peux que les remercier.

Ambre BENDER

## Abréviations

---

4mC :	N4-méthylcytosine
5mC :	5-méthylcytosine
5hmC :	5-hydroxy-méthylcytosine
5fC :	5-formylcytosine
5caC :	5-carboxylcytosine
6mA :	N6-méthyladénosine
ADNg :	ADN génomique
AdoMet :	SAM, S-adenosyl-L-méthionine
AGM :	Aorta-gonad-mesonephros
BER :	Base Excision Repair
COBRA :	Combined Bisulfite Restriction Analysis
PGC :	Primordial Germ Cells (Cellules germinales primordiales)
CpG :	Dinucléotide CG (C-phosphate-G)
CGI :	CpG island (îlot CpG)
CLP :	Common lymphoid progenitor
CMP :	Common myeloid progenitor
DMR :	Differentially methylated region (Région de méthylation différentielle)
DNMT :	ADN méthyltransférase
ECREM :	Epigenetic Code Replication Machinery
EMP :	Erythroid-macrophage progenitors
FPKM :	Fragments Per Kilobase of transcript per Million mapped reads
GMP :	Granulocyte-macrophage progenitors
HDAC :	Histone Désacétylase
HSC :	Cellules souches hématopoïétiques (Hematopoietic Stem Cells)
IAP :	Intracisternal A-particle

ICF : Immunodeficiency centromeric region instability and facial anomalies  
ICR : Imprinting Control Region  
LSK : Lin<sup>-</sup>Sca1<sup>+</sup>cKit<sup>+</sup>.  
LT-HSC : Long term hematopoietic stem cell  
MBD-CAP : Methyl Binding Domain Capture  
MBP : Methyl Binding Proteins  
MDS : Myelodysplastic syndrome  
MedIP : Methylated DNA immunoprecipitation  
MEP : Megakaryocyte-erythroid progenitor  
MPN : Myeloproliferative syndrome  
MPP : Multipotent progenitor  
MS-PCR : Methylation-specific PCR  
RRBS : Reduced Representation Bisulfite Sequencing  
SAM : S-adenosyl-L-methionine (AdoMet)  
SRA : Set and Ring Associated domain  
ST-HSC : Short term hematopoietic stem cell  
TDG : Thymine DNA Glycosylase  
TEs : Eléments Transposables (Transposable Elements)  
TET : ten-eleven translocation  
TTD : Tandem Tudor Domaine  
TSS : Transcription Start Site  
WGBS : Whole Genome Bisulfite Sequencing

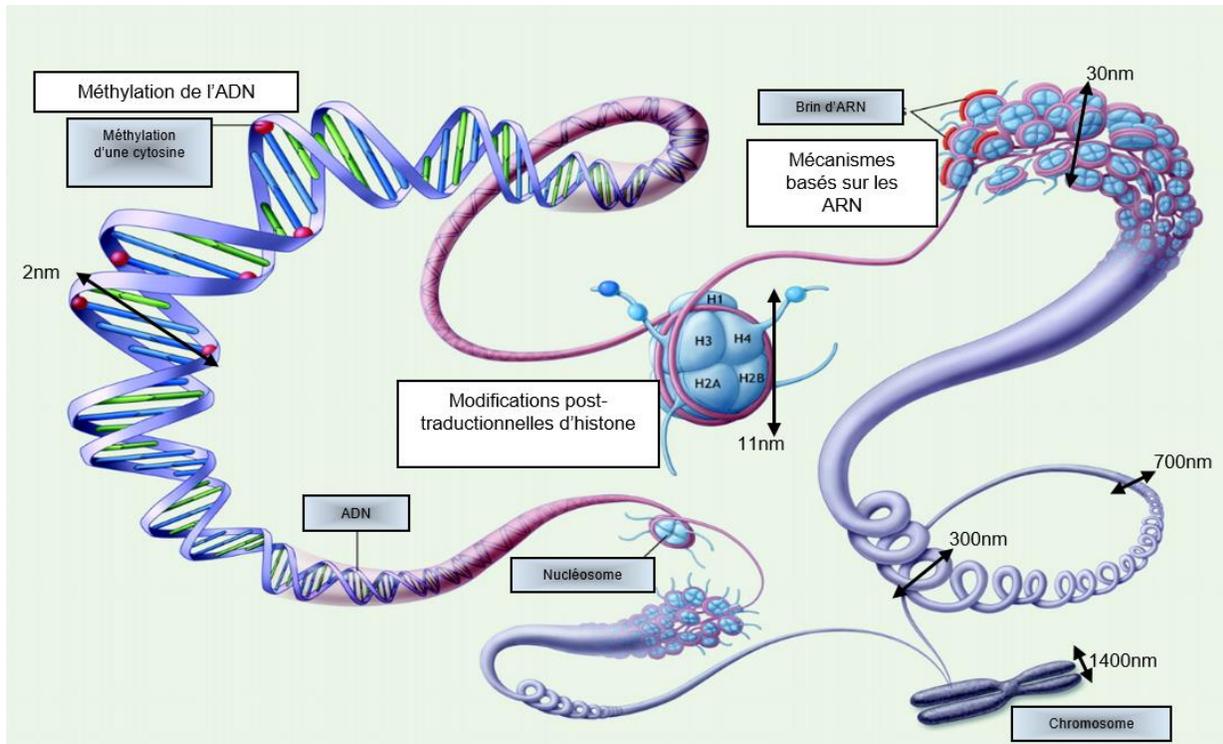
## Introduction

---

Toutes les cellules d'un organisme possèdent le même code génétique et le même nombre de gènes. Pourtant, lors de leur différenciation, les cellules vont acquérir des profils d'expression des gènes et des fonctionnalités caractéristiques qui leur sont propres et nécessaires. Nous savons que l'expression des gènes est dépendante de l'information génétique encodée dans la séquence de la molécule d'ADN (acide désoxyribonucléique), qui est constituée de nucléotides, eux-mêmes formés d'une base nucléique (adénine (A)–cytosines (C)–guanine (G)– thymine (T)), lié à un ose (ici, le désoxyribose) et à un groupe phosphate. Cependant, il existe une régulation supérieure, plus fine et dynamique de l'expression génique : les modifications épigénétiques. Ainsi, l'existence d'épigénomes distincts explique pourquoi le même génotype peut générer différents phénotypes.

Mais qu'est-ce que l'épigénétique ?

Etymologiquement, le mot épigénétique provient du grec ancien ἐπί, épí, « au-dessus de », et de génétique. On peut faire remonter l'origine du concept à *Aristote* (384-322 av. J.-C.). Il croyait en une épigénèse et nommait ainsi le développement d'un œuf informe de façon graduelle aboutissant à un organisme aux tissus différenciés. Néanmoins, c'est à *Conrad Waddington* que l'on attribue l'invention du terme « épigénétique », en 1942, pour nommer la branche de la biologie étudiant les implications entre les systèmes *gènes + environnement* et leurs *produits*, donnant naissance au phénotype d'un individu. Ainsi, le domaine de l'épigénétique vient combler des lacunes du modèle génétique postulant une équivalence unique entre phénotype et génotype qui ne pouvait expliquer tous les phénomènes liés à la différenciation cellulaire. C'est au XXI<sup>ème</sup> siècle qu'est apparu la définition la plus courante de l'épigénétique, c'est-à-dire, l'étude de l'ensemble des modifications de l'expression génique transmissibles d'une génération à l'autre, sans altération des séquences nucléotidiques, et ayant un caractère réversible (*Berger et al., 2009*). Personnellement, j'apprécie la définition imagée que *Thomas Jenuwein* (Institut Max Planck d'immunobiologie et épigénétique – Fribourg – Allemagne,) a donnée lors d'une interview. Il compare la génétique et l'épigénétique au même titre que la différence qu'il existe entre l'écriture d'un livre et sa lecture. Une fois que le livre est écrit, le texte (les gènes ou l'information stockée sous forme d'ADN) seront les mêmes dans tous les exemplaires distribués au public. Cependant, chaque lecteur aura une interprétation légèrement différente de l'histoire, qui suscitera en lui des émotions et des projections personnelles au fil des chapitres. D'une manière très comparable, l'épigénétique permettrait plusieurs lectures d'une matrice fixe (le livre ou le code génétique),



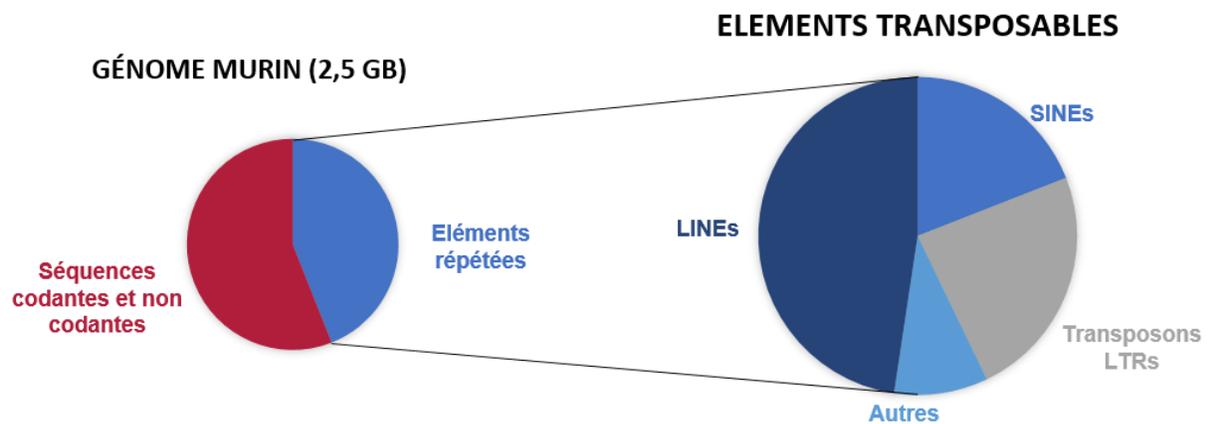
**Figure 1 : Condensation et modifications épigénétique de la chromatine**

L'ADN chromosomique (1400nm) est compacté dans le noyau à l'aide des histones. L'ADN, qui est une molécule chargée négativement, adhère très fortement aux histones qui elles sont chargées positivement, et forment ainsi les nucléosomes (11nm). Chaque nucléosome est composé de dimères d'histone H3, H4, H2A et H2B, le tout maintenu par l'histone H1. Les histones peuvent subir des modifications post-traductionnelles qui permettent de compacter et de décompacter la chromatine. Ensemble, les nucléosomes forment la fibre chromatinienne (30nm). D'autres modifications chimiques peuvent intervenir afin de réguler l'expression des gènes, il s'agit de la méthylation de l'ADN, ici représentée par les boules rouges. (Figure adaptée de Matouk et al. 2008)

donnant lieu à diverses interprétations, selon les conditions dans lesquelles on interroge cette matrice.

La taille de l'ADN contenu dans une seule cellule humaine représente environ deux mètres de long et est contenu dans un noyau qui ne fait que quelques micromètres. L'épigénétique agit en premier lieu dans l'organisation même de notre ADN. En effet, la molécule d'ADN est enroulée dans une structure nucléoprotéique appelée chromatine. Elle s'enroule à l'échelle de 146 paires de bases autour d'octamères d'histones, formant ainsi le nucléosome. Les histones créant le cœur du nucléosome sont les dimères d'histones H3, H4, H2A et H2B. L'histone H1, dites histone de liaison, se lie à l'ADN entre les nucléosomes permettant une structure complexe composée d'histones et d'ADN appelée la fibre chromatinienne (Luger et al., 1997) (**Figure 1**). Il existe deux types de chromatine, l'euchromatine et l'hétérochromatine, qui correspondent à des états plus ou moins condensés de la chromatine qui influencent la transcription des gènes. L'euchromatine, qui présente un faible taux de compaction, est considérée comme un état permissif pour la transcription au contraire de l'hétérochromatine qui présente un niveau plus élevé de compaction. L'hétérochromatine peut elle-même être subdivisée en deux états : l'hétérochromatine constitutive et l'hétérochromatine facultative. L'hétérochromatine constitutive est la forme la plus compacte, elle contient très peu de gènes et est composée de nombreux éléments répétés qui doivent rester réprimés. L'hétérochromatine facultative, en revanche, contient de nombreux gènes et peut inactiver ou activer certains de ces gènes en fonction du contexte environnemental de la cellule. Elle contient ainsi des gènes du cycle cellulaire, des gènes métaboliques, ou encore des gènes soumis à l'empreinte parentale. Durant l'embryogénèse, l'hétérochromatine facultative se met en place de manière tissu-spécifique pour réprimer certains gènes et favoriser l'expression d'autres gènes nécessaires au développement des différents types cellulaires. La mise en place ou le retrait de l'hétérochromatine facultative implique de nombreuses modifications épigénétiques (Woodcock and Ghosh, 2010). Enfin, les fibres de chromatine peuvent se compacter en donnant lieu au niveau ultime de compaction que tout le monde connaît : le chromosome métaphasique (1 400nm) (**Figure 1**).

Si l'on regarde plus en détails la séquence des génomes des organismes Eucaryotes supérieurs, on se rend compte que seule une infime partie est exprimée, et ce en fonction du stade développemental et du type cellulaire. En se référant à la dernière mise à jour d'*Ensembl* (Décembre 2016), nous savons par exemple que le génome de la souris comporte environ  $3.4 \times 10^9$  nucléotides et compte environ 30 000 gènes codant. Le reste du génome regroupe les introns situés dans les gènes, les régions intergéniques, les pseudogènes, et les régions répétées



**Figure 2 : Composition du génome murin**

Le génome de la souris est composé de 56% de séquences uniques comprenant introns, exons... et d'environ 44% de séquences répétées qui doivent être stablement réprimés afin d'assurer la stabilité du génome. Parmi ces séquences répétées, il y a 20% de LINEs, 8% de SINEs, 10% de transposons LTRs (ERVs), et 4% de séquences répétées autres (satellites, répétitions simples...).

(**Figure 2**). Dans les régions répétées, on retrouve les répétitions en tandem (satellites et microsatellites) et les éléments transposables (TEs) qui représentent au moins 44% du génome et qui sont subdivisés en deux classes (*Wicker et al., 2007*). La classe I se compose des transposons à LTRs (Long Terminal Repeat), tels que les ERVs (Endogenous Retroviruses) et les IAPs (Intracisternal A-particle), des LINEs (Long Interspersed Elements) et les SINEs (Short Interspersed Elements). La classe II comprend les transposons ADN qui, contrairement à la classe I, ne nécessitent pas d'intermédiaire ARN pour se dupliquer dans le génome. Grâce aux diverses modifications épigénétiques, ces régions sont stablement réprimées pour protéger l'intégrité du génome et celle de l'organisme. Les marques épigénétiques et leur dynamisme permettent également de réguler l'expression spatio-temporelle de certains gènes pour qu'ils soient actifs dans un type cellulaire et à un moment donné au cours du développement embryonnaire, alors que ces mêmes gènes doivent être réprimés dans d'autres types cellulaires ou à l'état adulte. Ces modifications épigénétiques comprennent les modifications post-traductionnelles d'histones (ex : l'acétylation, la méthylation, la sumoylation...), les modifications chimiques de l'ADN (ex : la méthylation de l'ADN), le remplacement d'histones par des variant d'histones (ex : H1.1, H3.1, H2AX...), le remodelage de la chromatine par des complexes ATP-dépendant, ou l'utilisation de petits ou long ARN non codants capables de dégrader les transcrits ou de réguler l'expression des gènes (**Figure 1**).

Au cours de ma thèse, je me suis plus particulièrement intéressée au rôle de la méthylation de l'ADN dans la régulation épigénétique des génomes eucaryotes, et plus précisément lors de la mise en place de deux lignages cellulaires chez la souris : les cellules germinales et les cellules hématopoïétiques. La méthylation de l'ADN a été découverte il y a plusieurs décennies et constitue aujourd'hui un paradigme de régulation épigénétique chez les eucaryotes. Dans cette introduction, je présenterai la distribution la méthylation de l'ADN et les facteurs moléculaires qui modulent ses profils. Brièvement, je discuterai aussi les mécanismes de déméthylation de l'ADN. Enfin, je vous exposerai les différentes techniques que nous avons à notre disposition afin d'étudier la méthylation de l'ADN pour finalement aborder son rôle dans le développement embryonnaire et la progression de certaines pathologies

## 1. La méthylation de l'ADN

On attribue la découverte de la méthylation de l'ADN, c'est-à-dire l'existence d'un groupement méthyle (-CH<sub>3</sub>) directement sur une base d'ADN, à *Johnson et Coghill*, en 1925, qui ont mis en évidence la présence d'une cytosine modifiée dans l'acide tuberculinique. Par la suite, dans les années 1950, *Rollin Hotchkiss* met en évidence l'existence naturelle de cytosines

méthylées chez les eucaryotes. Et ce n'est qu'à la fin du vingtième siècle que la méthylation de l'ADN va être considérée comme étant impliquée dans la régulation génique et la différenciation cellulaire de par sa fonction répressive (Compere and Palmiter, 1981; Holliday and Pugh, 1975).

La méthylation de l'ADN peut avoir lieu soit sur les cytosines, soit sur les adénines. La plus connue se nomme la 5-méthylcytosine (5mC) et consiste en l'ajout d'un groupement  $-CH_3$  sur le cinquième carbone de la cytosine. On l'appelle communément « méthylation de l'ADN » sans aucune précision car c'est l'une des modifications les plus étudiées de par son abondance au niveau du génome des mammifères et des plantes (Iyer et al., 2011). Cette dernière est associée à une répression à long terme de la transcription et se propage à travers la mitose, suggérant une potentielle « mémoire » de la 5mC (Holliday and Pugh, 1975; Riggs, 1975). Cependant il existe également une méthylation de l'ADN au niveau du quatrième azote de la cytosine, la N4-méthylcytosine (4mC), et au niveau du sixième azote de l'adénine, la N6-méthyladénine (6mA). La 4mC a été identifiée principalement chez les archées et les bactéries thermophiles dans les années 1980 (Janulaitis et al., 1983), où elle est impliquée dans la stabilité de la molécule d'ADN. La 6mA a pendant longtemps été classée comme une modification de l'ADN spécifique des procaryotes (Vanyushin et al., 1968). Bien qu'elle ne soit pas nécessaire pour la viabilité des procaryotes, elle joue un rôle crucial dans la régulation de la réplication et la transcription de l'ADN ainsi que dans le système de réparation et de défense cellulaire. Nous savons aujourd'hui que cette modification n'est pas exclusive aux procaryotes et qu'elle est aussi présente chez des eucaryotes tels que *Drosophila Melanogaster* et *Caenorhabditis Elegans* (Greer et al., 2015; Zhang et al., 2015) ou encore chez *Xenopus Laevis* (Kozioł et al., 2016; Wu et al., 2016). Sa présence a également été suggérée à faible niveau dans les cellules embryonnaires de souris (Wu et al., 2016), cependant cette découverte est soumise à débat (Schiffers et al., 2017). A la différence de la 5mC qui est associé à un état répressif, la 6mA est plutôt associé à l'activation génique.

Durant mon doctorat, je me suis plus particulièrement intéressée à la 5mC que je nommerai par la suite tout simplement méthylation de l'ADN. Ainsi, je vais vous décrire son contexte phylogénique, nucléotidique et sa distribution génomique.

### **a. Un contexte phylogénique**

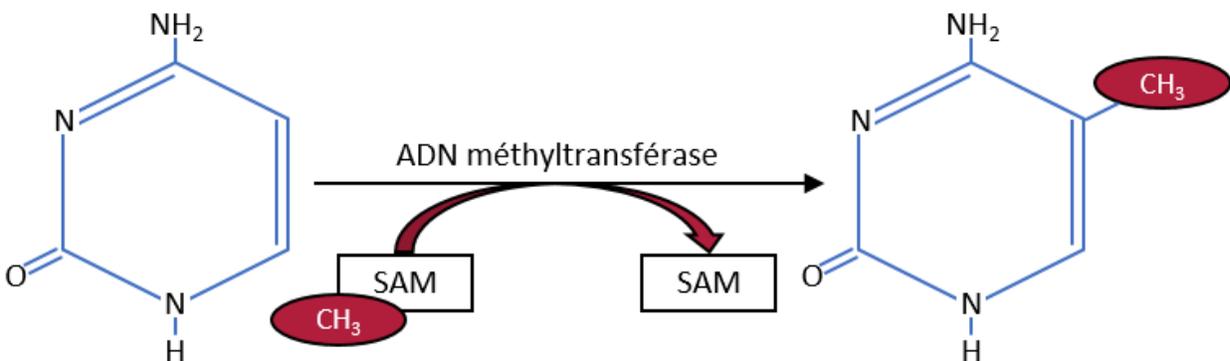
La 5mC est observée chez la plupart des vertébrés, invertébrés, plantes et champignons. Néanmoins, certains organismes ne présentent pas cette modification dans leur génome tel que *Drosophila Melanogaster* et *Caenorhabditis Elegans*, pour ne donner que deux exemples. En

Espèces	Taille du génome	Répartition génomique 5mC	Contexte 5mC	Séquences cibles	6mA
Plantes					
<i>Arabidopsis Thaliana</i>	125Mb	Mosaïque	CG, CHG, CHH	-Corps de gènes -Séquences répétées	?
<i>Oryza sativa</i>	400Mb				?
Champignons					
<i>Neurospora crassa</i>	47Mb	Mosaïque	CNN	-Séquences répétées	?
<i>Coprinopsis Cinerea</i>	37,5Mb		CG, CNN		?
Invertébrés					
<i>Caenorhabditis elegans</i>	97Mb	X	X	X	✓
<i>Drosophila melanogaster</i>	175Mb	X	X	X	✓
<i>Apis mellifera</i>	1,5Gb	Mosaïque	CG	-Corps de gènes	?
Vertébrés					
<i>Mus musculus</i>	2,4Gb	Globale	CG, et parfois non-CG	-Corps de gènes -Séquences répétées -Régions intergéniques	?
<i>Homo sapiens</i>	3,2Gb				?

**Figure 3 : Répartition de la méthylation de l'ADN dans divers génomes**

La répartition de la 5mC est mosaïque chez les plantes, les champignons et les invertébrés (pour ceux qui en possèdent). En revanche, la 5mC est distribuée de manière globale dans tout le génome chez les vertébrés. Certains invertébrés ne possèdent pas de 5mC mais des niveaux détectables de 6mA.

H = A, C ou T et N = A, C, G ou T.



**Figure 4 : Réaction de méthylation d'une cytosine**

Une DNMT (DNA Méthyltransférase) transfère un groupement méthyl ( $\text{CH}_3$ ), venant d'un cofacteur appelé SAM (S-Adénosyl-L-méthionine ou encore AdoMet), sur le carbone 5 de la cytosine. La nouvelle base est alors nommée 5mC (5-méthyl-cytosine).

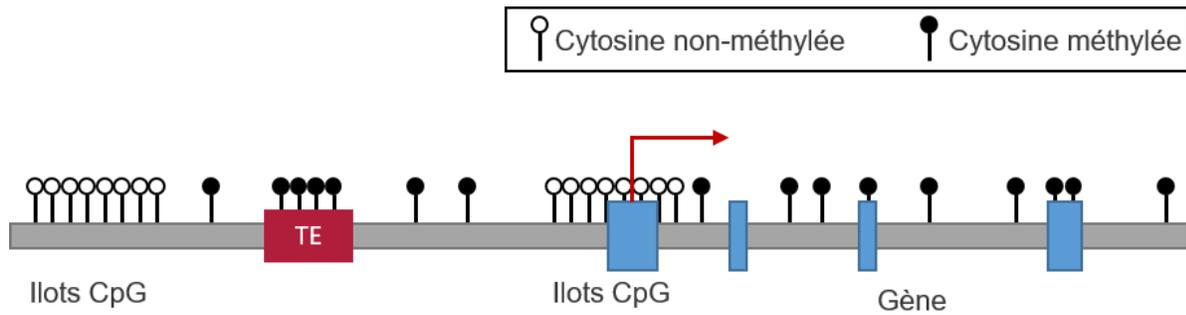
effet, ces derniers présentent une méthylation au niveau des adénines mais sont totalement dépourvus de 5mC (**Figure 3**). Quand on y regarde de plus près, on s'aperçoit que les profils de méthylation des cytosines varient considérablement entre les différents génomes en fonction de la cible de la méthylation, du contexte nucléotidique et de la répartition génomique (**Feng et al., 2010; Zemach et al., 2010**).

L'histoire évolutive de la méthylation de l'ADN semble donc très complexe et les raisons de son apparition restent encore à élucider. Néanmoins, une des hypothèses serait qu'elle soit apparue en même temps que l'augmentation relative de la taille des génomes, comme un nouveau mécanisme de surveillance de l'ADN non codant (**Figure 3**).

### ***b. Un contexte nucléotidique***

Comme je l'ai déjà dit plus haut, la méthylation des cytosines de l'ADN consiste en l'ajout d'un groupement méthyle au niveau du cinquième carbone des cytosines. Cette modification chimique est induite par les enzymes de la famille des ADN méthyltransférases dont on parlera dans le chapitre suivant. Ces enzymes ont la capacité de transférer un groupement méthyle à l'aide du cofacteur SAM (S-adenosyl-L-méthionine ou encore appelé AdoMet) (**Figure 4**).

Chez les vertébrés, la méthylation des cytosines a lieu presque exclusivement dans un contexte de dinucléotides CpG (Cytosine-phosphate-Guanine) (**Figure 3**) (**Feng et al., 2010; Lister et al., 2008, 2009**). Au contraire chez les plantes et les champignons, elle a lieu dans un contexte de dinucléotides CpG ou de trinuécléotides CHG ou CHH (La lettre H représentant un A, C ou T) [méthylation non-CpG]. Cependant, bien que la méthylation chez les mammifères soit principalement de type CpG, il existe de la méthylation non-CpG. De récentes études démontrent que la méthylation non-CpG est présente à des niveaux élevés dans des cellules ES (embryonic stem cells) (**Feng et al., 2010; Lister et al., 2008, 2009**), des neurones et des cellules gliales (**Varley et al., 2013; Xie et al., 2012b; Lister et al., 2013**) ainsi qu'au niveau des ovocytes (**Tomizawa et al., 2011**), principalement au niveau de séquences CpA. Les neurones et les ovocytes accumulent la méthylation non-CpG au cours de leur développement (**Guo et al., 2014b; Tomizawa et al., 2011; Xie et al., 2012b**). Concernant les autres cellules somatiques, la méthylation non-CpG est quasiment absente et ne représentent que 0,02% de l'ensemble des cytosines méthylées. Contrairement aux sites CpG, les sites CpA ne sont pas palindromiques et donc la méthylation CpA ne peut pas se copier au cours des divisions cellulaires. C'est pourquoi, malgré le fait que la fonction et les mécanismes impliqués dans ce type de méthylation ne soient pas encore élucidés, plusieurs chercheurs pensent que cette méthylation non-CpG est une résultante de l'hyperactivité non-spécifique des enzymes de méthylation autour des sites CpG qui devient visible justement dans des cellules qui ne se divisent pas comme les neurones ou les



**Figure 5 : Répartition génomique de la méthylation des cytosines chez les mammifères**

La figure représente schématiquement le profil de méthylation typique du génome chez les mammifères. La plupart du génome est pauvre en CpG et hyper méthylé. A l'inverse, les régions denses en CpG appelées « ilots CpG » sont généralement non-méthylés. TE = Eléments transposables.

ovocytes. D'autres ont cependant observé que la méthylation non-CpG est corrélée à l'expression spécifique de certains gènes et tissus, ce qui suggère un rôle régulateur (Barrès et al., 2009; Lister et al., 2011; Ma et al., 2014). Enfin, la méthylation non-CpG est aussi fréquemment observée dans de nombreuses pathologies (Kinoshita et al., 2000; Nielsen et al., 2001; Woodcock et al., 1999; Zhang et al., 1998) mais l'impact de la méthylation de ces dinucléotides est encore mal connu.

### **c. Un contexte génomique**

A l'inverse du génome de la plupart des plantes, invertébrés et champignons qui montrent un profil de méthylation plutôt « mosaïque » où seuls des éléments génomiques spécifiques sont ciblés, la méthylation de l'ADN couvre l'ensemble du génome des vertébrés (Suzuki and Bird, 2008; Zemach et al., 2010). Dans les cellules somatiques de ces derniers, environ 80 % des CpG sont méthylés, ce qui fait que la méthylation apparaît comme l'état par défaut des CpGs (Bird, 2002; Lister et al., 2009; Stadler et al., 2011).

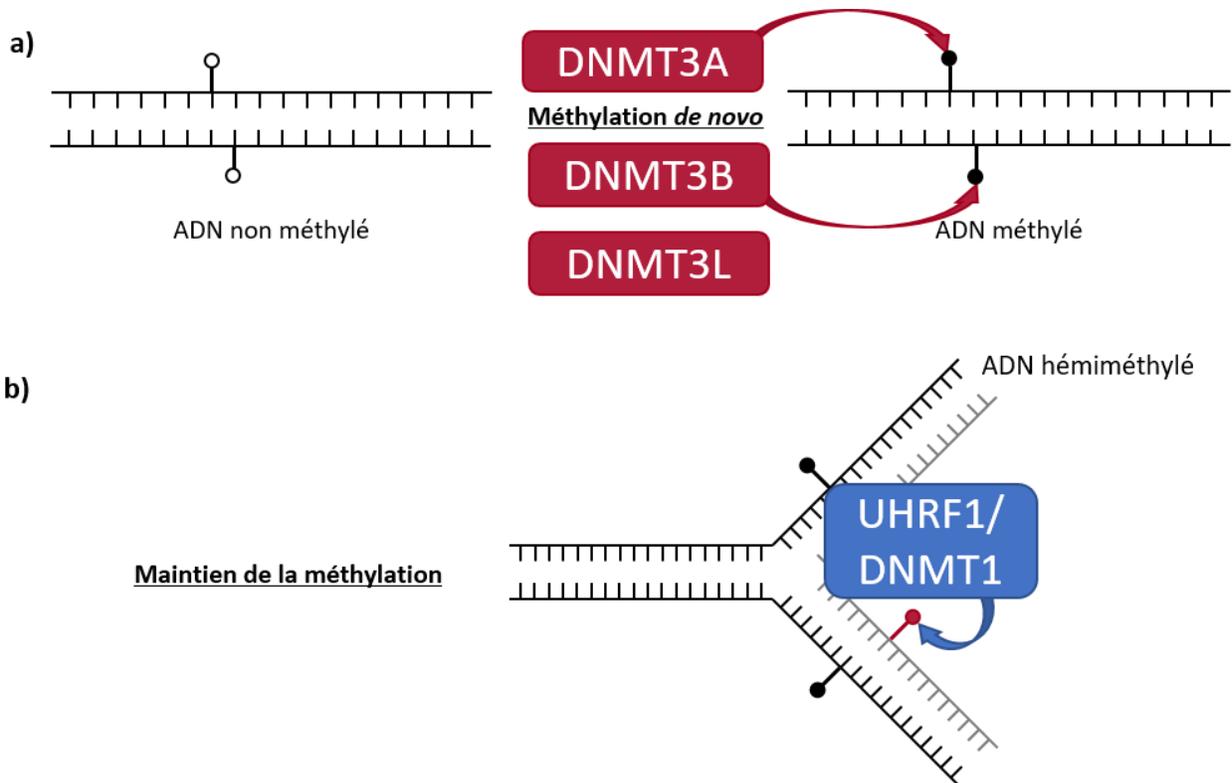
Néanmoins, il existe des exceptions à cet état de méthylation par défaut dans les génomes vertébrés avec l'existence de séquences résistantes à la méthylation de l'ADN appelées « îlots CpG » ou encore CGI (CpG island) (Bird et al., 1985; Cooper et al., 1983). Bien qu'il existe plusieurs définitions des îlots CpG dans la littérature (Gardiner-Garden and Frommer, 1987), de manière générale ceux-ci se définissent en 3 points : 1) il s'agit de séquences de plus de 200pb, 2) le nombre de guanines et de cytosines contenus dans ces séquences est supérieur à 50 %, 3) le ratio de CpG observées sur le nombre de CpG attendus est supérieur à 0,6 (Figure 5). Si on ne tient pas compte des séquences répétées qui sont fréquemment riches en CpG, il y a environ 25 000 îlots CpG dans le génome humain et 75 % font moins de 850pb de long (Lander et al., 2001). Ces séquences sont considérées comme des unités majeures de la régulation génique puisqu'environ 50 % des îlots CpG sont localisés dans les promoteurs de gènes et 25 % au niveau des corps de gènes. Réciproquement, environ 60 à 70 % des gènes possèdent un îlot CpG dans leur région promotrice (Illingworth et al., 2010; Saxonov et al., 2006) et ces îlots possèdent des caractéristiques précises qui ont été conservé au fil de l'évolution (Antequera, 2003; Long et al., 2013a).

La majorité des îlots CpG sont constitutivement non méthylés dans tous les types cellulaires chez l'homme et la souris. C'est probablement le cas également chez tous les vertébrés même si cela reste à démontrer formellement par des expériences de cartographie exhaustive de la méthylation de l'ADN dans plusieurs tissus chez d'autres espèces. Les CGIs sont enrichies au niveau de l'euchromatine et associée à des marques permissives d'histone (comme par exemple H3K4me3). Les expériences menées par Long et al., 2016, au cours desquelles ils ont

inséré un chromosome humain au sein d'un génome murin, révèlent que les promoteurs associés aux CGIs sont réfractaires à la 5mC indépendamment de l'espèce de l'hôte. Ceci démontre que la séquence d'ADN joue un rôle central, et ce au travers de l'évolution, dans le maintien d'un stade hypométhylé. Plusieurs études ont démontré que certains agencements et motifs nucléotidiques fonctionneraient comme des signaux moléculaires pour certaines protéines, au niveau de régions spécifiques, afin d'inhiber la méthylation de l'ADN (Krebs et al., 2014; Lienert et al., 2011; Mendenhall et al., 2010; Thomson et al., 2010; Wachter et al., 2014). Cependant les mécanismes moléculaires qui connectent la séquence des CGI avec la protection contre la méthylation des cytosines sont encore très mal connus. Une possibilité est que la liaison de facteurs de transcription tels que SP1 ou VEZF1 à des séquences spécifiques dans les CGIs protégeraient l'ADN de la 5mC (Brandeis et al., 1994; Dickson et al., 2010; Krebs et al., 2014; Lienert et al., 2011; Long et al., 2013b; Macleod et al., 1994; Voo et al., 2000). Il a également été proposé que la famille des protéines à domaine CXXC, qui reconnaît spécifiquement l'ADN contenant des CpG non méthylés, soit impliquée dans la protection des CGI contre la méthylation. Parmi les membres de cette famille, une étude récente a démontré que l'inactivation de FBXL10/KDM2B induit une hyperméthylation d'un grand nombre d'îlots CpG dans les cellules ES de souris (Boulard et al., 2015). Enfin, il est possible que la séquence d'ADN influence des variants d'histones ainsi que des marques d'histone spécifiques, tels que la présence de H2AZ ou H3K4me3, qui joueraient aussi un rôle dans le maintien de cette hypométhylation.

Dans les cellules somatiques, seuls 2 à 4% des îlots CpG sont méthylés (Auclair and Weber, 2012; Meissner et al., 2008) et la majorité d'entre eux est localisée dans les régions inter- et intra- géniques (**Figure 5**). Le nombre d'îlots méthylés diminue jusqu'à 1 % dans les gamètes mâles (Smallwood et al., 2011).

Alors que la séquence génomique influence la présence de méthylation de l'ADN, inversement la méthylation de l'ADN peut également influencer la séquence du génome. En effet, la présence de cette modification chimique augmente la fréquence des mutations spontanées dans l'ADN. Cela s'explique par l'existence de réactions de déamination spontanée des cytosines. La déamination d'une cytosine non méthylée se traduit par la création d'un uracile qui va facilement être reconnu et enlevé par les mécanismes de réparation de l'ADN impliquant le système BER (Base Excision Repair). Par contre, la déamination de cytosines méthylées induit une conversion en thymine, ce qui est plus difficile à reconnaître et à réparer pour les systèmes de réparation de l'ADN. A l'échelle des temps évolutifs, cela a conduit à une conversion progressive des cytosines en thymines dans le génome, ayant pour conséquence une déplétion globale du nombre de site CpG dans les génomes des mammifères (Lander et al., 2001; Swartz



**Figure 6 : Mise en place et maintien de la méthylation par les DNMT**

**a)** Les enzymes DNMT3A et DNMT3B catalysent la réaction de méthylation *de novo*. Dans certains types cellulaires, DNMT3L forme un complexe avec DNMT3A/B et stimule leur activité catalytique. **b)** Lors de la réplication, la méthylation de l'ADN est maintenue à l'aide de l'enzyme DNMT1. UHRF1 reconnaît l'ADN hémiméthylé et favorise le recrutement de DNMT1 au niveau du site de réplication pour méthyler le brin néosynthétisé (en gris).

et al., 1962). Il existe cependant une exception à cette perte globale de CpGs dans les génomes des mammifères avec la présence des îlots CpG qui, puisqu'ils sont résistants à la méthylation de l'ADN, n'ont pas subi cette perte évolutive des CpGs (Bird, 1986; Bird et al., 1985). Ce lien entre méthylation de l'ADN et perte évolutive des CpG a pu être démontré notamment en comparant les profils de méthylation de l'ADN et l'histoire évolutive du génome humain en comparaison avec celui des grands singes (Cohen et al., 2011; Weber et al., 2007).

## 2. Les acteurs de la méthylation de l'ADN

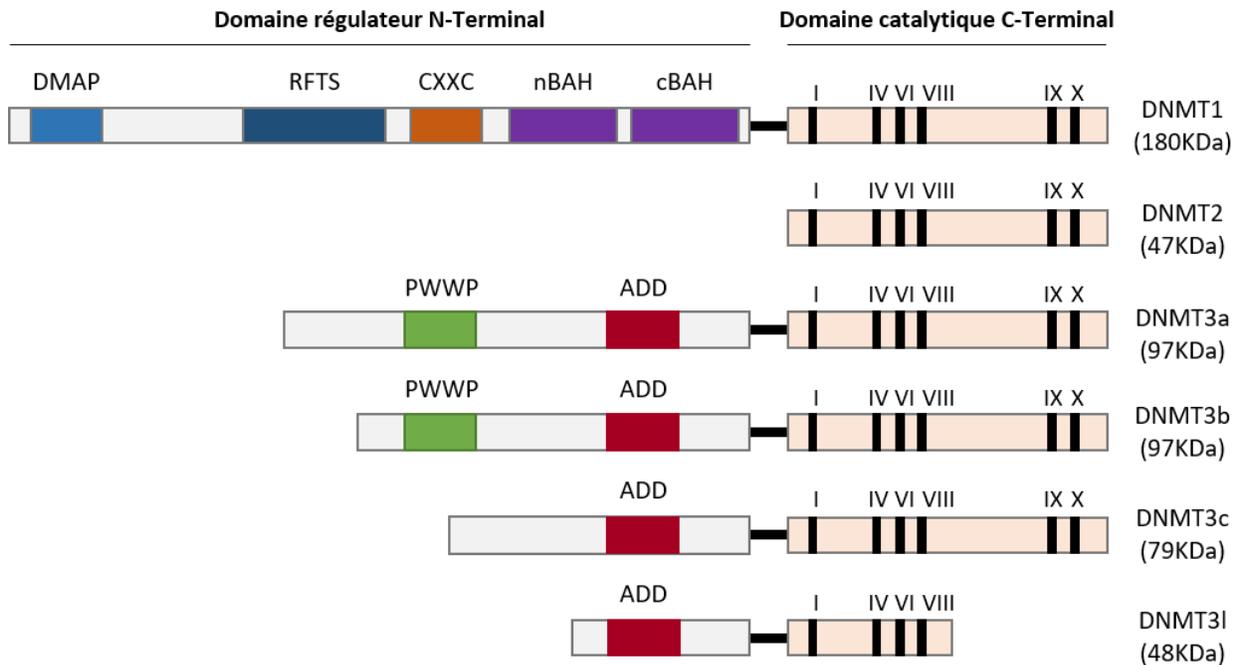
Les ADN méthyltransférases (DNMT) sont les enzymes qui catalysent le transfert d'un groupe méthyle (CH<sub>3</sub>) à partir du donneur SAM sur le carbone 5 de la cytosine, créant la 5-méthylcytosine. Selon les enzymes, cette réaction de méthylation de la cytosine peut participer à un processus de méthylation *de novo* de l'ADN ou simplement à un processus de maintien de la méthylation de l'ADN au cours de la réplication (Figure 6).

Chez les eucaryotes, on dénombre selon les espèces jusqu'à 6 enzymes au total dans la famille des DNMT : *Dnmt1*, *Dnmt2*, *Dnmt3a*, *Dnmt3b*, *Dnmt3c*, *Dnmt3L*. Les DNMT eucaryotes sont de larges protéines partageant une grande homologie avec leurs cousines procaryotes, et semblent utiliser le même mécanisme catalytique afin de déposer de groupement méthyl sur les cytosines. A ce propos, une grande partie de nos connaissances à propos de cette réaction enzymatique vient d'études menées sur des DNMT bactériennes (Cheng and Roberts, 2001).

Le domaine catalytique (C-terminal) de ces enzymes est très conservé entre elles (les motifs les plus conservés sont les motifs I, IV, VI, IX et X), alors qu'il existe beaucoup plus de variabilité au niveau du domaine régulateur N-terminal de ces enzymes (Figure 7). Dans le domaine catalytique, le motif IV correspond à l'activité catalytique de l'enzyme et la poche pouvant contenir le cofacteur SAM est formée par les motifs I et X principalement. Le plus petit domaine compris entre les motifs VIII et IX est responsable de la reconnaissance à l'ADN. La grande variabilité qui existe entre les différentes DNMT au niveau du site N-terminal semble expliquer les différentes propriétés biochimiques et les différentes fonctions biologiques de ces dernières. Ainsi, basée sur leur similarité fonctionnelle et structurale, les DNMT mammifères sont regroupées en trois classes : *Dnmt1*, *Dnmt2* et *Dnmt3*.

### a. DNMT1

Le premier gène eucaryotique d'ADN méthyltransférase, *Dnmt1*, a été cloné à partir de cellules murines en 1988 par *T. Bestor* (Bestor et al., 1988). C'est la plus grande des DNMT avec un poids moléculaire de 180 KDa (Figure 7). Le locus *Dnmt1* possède plusieurs sites de démarrage de la transcription (Mertineit et al., 1998; Rouleau et al., 1992). En effet, il existe un



### Figure 7 : Architecture des ADN méthyltransférases

Les enzymes DNMTs disposent d'un domaine catalytique en C-terminal. Elles possèdent également différents domaines régulateurs en N-terminal qui leur permettent d'interagir avec l'ADN ou des protéines. DNMT3L possède un domaine catalytique tronqué et n'a donc pas d'activité catalytique pour méthyler l'ADN. DNMT1 a de nombreux domaines régulateurs : un domaine DMAP (DNA methyltransferase associated protein) et un domaine RFTS (replication foci targeting sequence). Grâce à ces domaines, DNMT1 peut être recrutée à la fourche de réplication de l'ADN. Elle possède un domaine CXXC, riche en cystéines permettant la liaison de DNMT1 à l'ADN hémiméthylés, et un domaine nBAH et un domaine cBAH (Bromo-adjacent Homology). DNMT3A et DNMT3B contiennent un domaine PWWP contenant un motif riche en prolines et tryptophanes, leur donnant la possibilité d'interagir avec certaines histones modifiées. Enfin, les protéines DNMT3 ont un domaine ADD (ATRX-DNMT3-DNMT3L) riche en cystéines et de type doigt de Zinc leur permettant d'interagir aux queues d'histones H3 non méthylées sur la lysine 4.

promoteur spécifique des cellules somatiques qui s'active peu de temps après l'implantation de l'embryon, et qui exprime le transcrit à partir d'un exon dit spécifique des cellules somatiques : l'exon 1s. C'est ce long transcrit de 1620aa que l'on appelle communément *Dnmt1* (Tucker et al., 1996; Yoder et al., 1996). *Dnmt1* est exprimé constitutivement dans les cellules prolifératives et de façon ubiquitaire dans les tissus somatiques au cours du développement chez la souris. Il existe un autre promoteur en amont qui contrôle l'expression d'un transcrit à partir d'un exon spécifique des ovocytes : l'exon 1o. Ce promoteur alternatif code pour une protéine tronquée, nommée *Dnmt1o*, où il manque 118aa en N-terminal par rapport à la forme somatique (Mertineit et al., 1998). Ces deux protéines, DNMT1 et DNMT1o, possède le même domaine catalytique mais diffèrent donc au niveau de la région régulatrice N-terminale.

Les propriétés enzymatiques de DNMT1 ont été étudiées par des approches à la fois *in vivo* et *in vitro*. Bien que DNMT1 puisse méthyler à la fois les CpG non méthylés et hémiméthylés *in vitro*, son activité sur des sites hémiméthylés est de 5 à 50 fois plus élevée que celle sur des cytosines non méthylées (Hitt et al., 1988; Pradhan et al., 1999; Ruchirawat et al., 1987; Yoder et al., 1997a). Par ailleurs, DNMT1 est une protéine nucléaire dont la localisation change de manière dynamique en fonction du cycle cellulaire. Elle se retrouve dans le nucléoplasme pendant les phases G1 et G2, et elle est associée à la fourche de réplication lors de la phase S (Leonhardt et al., 1992). Cette association avec les sites de réplication et son activité préférentielle sur les CpG hémiméthylés a suggéré un rôle dans le maintien de la méthylation de l'ADN aux cours de la division cellulaire, en déposant un groupement méthyl sur le brin D'ADN néosynthétisé. Il a également été montré que DNMT1 interagit directement avec PCNA, une composant central de la machinerie de méthylation, cependant cette interaction n'est pas strictement requise pour l'activité de DNMT1 (Schermelleh et al., 2007).

En accord avec un rôle global de maintien de la méthylation du génome, l'inactivation de *Dnmt1* dans les cellules ES et chez la souris conduit à une déméthylation globale du génome, mais cette délétion n'a que peu d'effets sur la méthylation *de novo* mesurée par exemple sur de l'ADN rétroviral nouvellement intégré (Lei et al., 1996; Li et al., 1992). Chez la souris, le mutant *Dnmt1* présente une létalité embryonnaire à E9.5 avec une perte de méthylation globale sur de nombreux éléments génomiques. En particulier la perte de la forme somatique de *Dnmt1* induit une perte du maintien de la méthylation et une expression anormale des gènes soumis à empreintes (Hirasawa et al., 2008; Li et al., 1993). L'inactivation de la forme ovocytaire *Dnmt1o* conduit également à une déméthylation et perte d'expression allélique des gènes soumis à empreinte, ce qui suggère que cette enzyme pourrait jouer un rôle de maintenance dans l'embryon

préimplantatoire (Howell et al., 2001). De plus, l'inactivation de *Dnmt1* conduit à une activation ectopique du chromosome X (Bhatnagar et al., 2014; McGraw et al., 2013; Sado et al., 2000) et la réactivation de transposons comme les IAPs notamment (Gaudet et al., 2004; Jackson-Grusby et al., 2001a; Maksakova et al., 2008; Walsh et al., 1998).

La surexpression de DNMT1 en l'absence des ADN méthyltransférases *de novo* DNMT3A et DNMT3B (dont on parlera dans un autre chapitre) ne permet pas de restaurer la méthylation *de novo* dans les cellules ES murines (Chen et al., 2003). Néanmoins, il faut noter que d'autres études suggèrent un rôle coopératif entre ces trois enzymes (Fatemi et al., 2002; Kim et al., 2002) et, malgré une préférence pour les sites hémiméthylés, plusieurs études suggèrent que DNMT1 pourrait jouer un rôle dans la méthylation *de novo*, que ce soit à partir d'expériences menées *in vitro* (Arand et al., 2012; Biniszkiwicz et al., 2002; Fatemi et al., 2002; Lorincz et al., 2002; Takagi et al., 1995; Vertino et al., 1996; Vilkaitis et al., 2005) ou *in vivo* (Jeltsch and Jurkowska, 2014; Jones and Liang, 2009). L'inactivation de *Dnmt1* dans les ovocytes, lorsque ceux-ci acquièrent la méthylation du génome mais ne se divisent plus, induit une baisse légère de la méthylation du génome, ce qui suggère à nouveau une contribution possible de DNMT1 à la méthylation *de novo* (Shirane et al., 2013). Tout ceci amène une complexité supplémentaire concernant la véritable fonction de DNMT1.

## b. UHRF1

De façon générale, la communauté scientifique s'accorde à dire que DNMT1 est l'enzyme majeure de la maintenance de la méthylation des CpG au cours de la réplication chez les mammifères. Néanmoins, seule, elle ne semble pas capable de déposer le groupement méthyle au niveau du bras d'ADN néosynthétisé car elle a besoin d'un partenaire, UHRF1, dont nous allons parler dans ce chapitre.

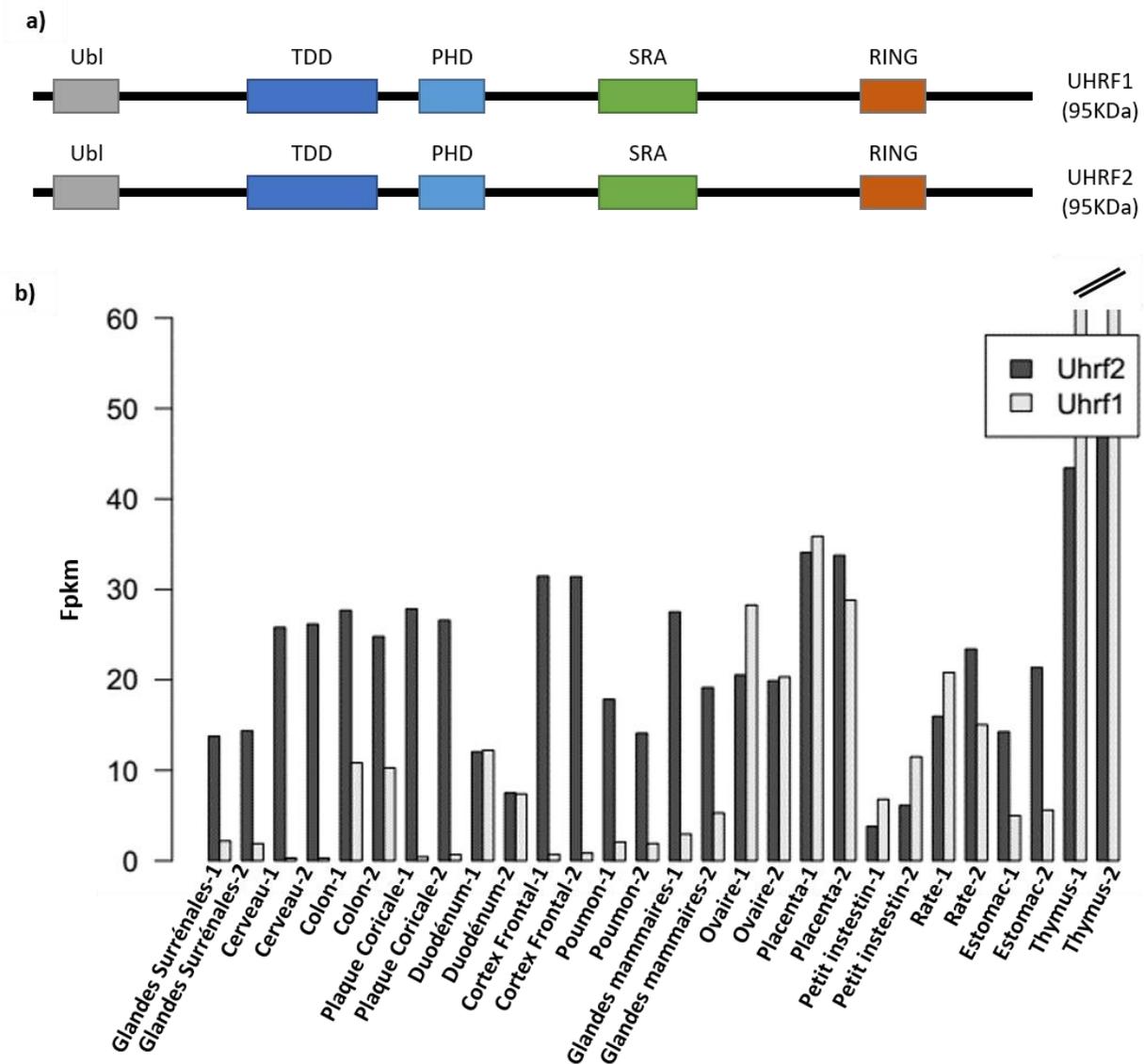
Le gène *Uhrf1* a été découvert chez la souris à la fin du XX<sup>ème</sup> siècle et chez l'humain dans les années 2000 (Fujimori et al., 1998; Hopfner et al., 2000). Ce gène possède d'autres noms, tels que *Np95* et *ICBP90*. A cette époque, UHRF1 est alors décrit comme un facteur de régulation de l'expression de la topoisomérase II $\alpha$  et est également lié au cycle cellulaire. Ce n'est qu'en 2007 qu'UHRF1 est reconnu comme une protéine partenaire de DNMT1 et impliquée dans la régulation du maintien de la méthylation de l'ADN (Bostick et al., 2007).

*Uhrf1* code pour une protéine nucléaire de 95 kDa qui possède un domaine UBL (ubiquitin-like), deux sites de localisation nucléaire, deux domaines à doigt de zinc (PHD et RING finger), deux domaines TTD (Tandem Tudor Domain), ainsi qu'un domaine SRA (Set and Ring Associated Domain) (**Figure 8**) (Fujimori et al., 1998; Hopfner et al., 2000).

Hormis un rôle actif d'*Uhrf1* dans la régulation de la topoisomérase II $\alpha$  (Trotzler et al., 2004) ainsi que celle du cycle cellulaire qui le désigne comme un gène suppresseur de tumeur (Jeanblanc et al., 2005; Mousli et al., 2003), de nombreuses études ont par la suite révélé l'importance de la protéine UHRF1 dans la propagation du code épigénétique. Effectivement, cette protéine permettrait la formation d'un gros complexe macromoléculaire appelé ECREM (Epigenetic Code Replication Machinery), responsable de la transmission du code épigénétique au cours du cycle cellulaire (Bronner et al., 2007; Kim et al., 2009; Sharif et al., 2007). En effet, UHRF1 est capable de reconnaître les dinucléotides CpG hémiméthylés grâce à son domaine SRA (Bostick et al., 2007; Greiner et al., 2015; Qian et al., 2008; Rottach et al., 2010; Sharif et al., 2007). UHRF1 peut aussi se lier à DNMT1 et colocalise avec DNMT1 au niveau des sites de réplication de l'ADN (Achour et al., 2008; Berkyurek et al., 2014; Bostick et al., 2007; Felle et al., 2011; Sharif et al., 2007). Il en a été déduit un modèle selon lequel UHRF1 favoriserait le recrutement de DNMT1 au niveau de l'ADN répliqué pour copier la méthylation de l'ADN au cours de la division cellulaire (Bostick et al., 2007; Sharif et al., 2007) (**Figure 6b**). Par ailleurs, il a également été montré que UHRF1 ubiquityle l'histone H3 grâce à l'activité E3 ligase de son domaine RING (Citterio et al., 2004), ce qui constituerait une plateforme de recrutement de DNMT1 via un nouveau motif de liaison à l'ubiquitine dans la protéine DNMT1 (Nishiyama et al., 2013; Qin et al., 2015). L'activité E3 ubiquitine ligase de UHRF1 pourrait également réguler la stabilité de la protéine DNMT1 (Felle et al., 2011; Qin et al., 2011). Enfin, d'autres études suggèrent que l'interaction avec UHRF1 stimule également l'activité enzymatique de DNMT1 par un mécanisme allostérique (Bashtrykov et al., 2014).

Outre DNMT1, UHRF1 interagit également avec HDAC1 (Unoki et al., 2004) et G9A (Kim et al., 2009), une histone méthyltransférase qui catalyse la méthylation de H3K9. De plus, UHRF1 a la capacité de lire le code histone avec une préférence pour les marques H3R2me0, H3K4me0, H3K4me3 et H3K9me2/3 via ses domaines PHD-TTD (Hashimoto et al., 2010; Lallous et al., 2011; Rajakumara et al., 2011; Rothbart et al., 2012; Xie et al., 2012a). Quant au domaine UBL qui possède une homologie de 35% avec l'ubiquitine, sa fonction n'est pas encore claire.

Enfin, les embryons présentant une délétion du gène *Uhrf1* possèdent un phénotype très similaire à celui de *Dnmt1* avec une létalité embryonnaire couplée à une hypométhylation globale du génome (Sharif et al., 2007), ce qui soutient le rôle d'UHRF1 dans le maintien des profils de méthylation de l'ADN. De plus, une récente étude ayant réalisé une délétion conditionnelle d'*Uhrf1* dans des cellules souches neuronales montre que cela induit une réexpression des éléments transposables, normalement stablement réprimés et par les marques d'histone H3K9me3 et par la méthylation de l'ADN (Ramesh et al., 2016).



**Figure 8 : Comparaison de UHRF1 et UHRF2**

**a) Structure des protéines UHRF1 et UHRF2.** Le domaine SRA (SET and Ring finger-Associated) permet la reconnaissance de l'ADN hémi-méthylé. Le domaine TDD (Tudor Tandem Domain) et le domaine PHD (Plant Homeo Domain) permettent la reconnaissance de la marque d'histone H3K9me3, caractéristique de l'hétérochromatine. Le domaine RING (Really Interesting New Gene) possède une activité E3 ligase qui permet d'ubiquitinyler les histones sur la lysine 23 de l'histone H3 (H3K23). Le domaine UBL (Ubiquitin-Like) possède une grande homologie avec l'ubiquitine mais sa fonction est encore mal connue. **b) Expression de *Uhrf1* et *Uhrf2* dans les tissus de souris.** Les valeurs sont données en FPKM (fragment of transcript per million mapped reads) et représentent le niveau d'expression de *Uhrf1* et *Uhrf2* mesurée par RNA-seq en duplicatas dans les organes de souris (données recueillies par le laboratoire du Dr. Gingeras).

Finalement, UHRF1 a un paralogue: UHRF2 (NIRF) qui possède les mêmes domaines ainsi qu'une homologie de 64% (**Figure 8-a**) (Bronner et al., 2007; Mori et al., 2012). Peu de choses sont connues sur la fonction des protéines UHRF2. Par contre, nous savons que le gène *Uhrf1* est principalement exprimé dans des cellules souches en cours de prolifération, alors que *Uhrf2* est plus exprimé dans les cellules différenciées (Pichler et al., 2011; Sen et al., 2010) (**Figure 8-b**). Cette différence d'expression suggère une fonction différente d'UHRF2 au niveau de la méthylation de l'ADN. En effet, cette dernière n'est pas capable de compenser la perte de méthylation dans des cellules présentant une délétion d'*Uhrf1*. De plus, de récentes études ayant réalisé une délétion du gène *Uhrf2* chez la souris suggèrent que *Uhrf2* a un rôle dans le fonctionnement des cellules neuronales mais peu d'influence sur les profils de méthylation global du génome (Chen et al., 2017a; Liu et al., 2017). Cependant aucune cartographie quantitative du méthylome n'a encore été réalisée dans les souris *Uhrf2*<sup>-/-</sup> et il reste encore beaucoup de chose à étudier concernant les fonctions possibles d'*Uhrf2* dans la régulation de la méthylation de l'ADN.

### c. DNMT2

Le gène *Dnmt2* a été identifié par plusieurs groupes indépendamment en 1998 lors de recherche d'EST sur des bases de données (Okano et al., 1998a; Van den Wyngaert et al., 1998; Yoder and Bestor, 1998). Il code pour une protéine de 391aa chez l'humain et de 415aa chez la souris. C'est la DNMT la plus conservée dans la famille des gènes *Dnmt* puisque des homologues ont été identifiés dans un grand nombre d'espèces eucaryotes. Structurellement parlant, elle est aussi plus proche de ses cousines procaryotes que les autres DNMT puisqu'il lui manque le domaine de régulation N-terminal (**Figure 7**). Malgré une conservation des X motifs et sa capacité à lier l'ADN (Dong et al., 2001), DNMT2 ne présente aucune activité catalytique de méthylation de l'ADN *in vitro* et ainsi, ne semble pas avoir de fonction au niveau de la méthylation de l'ADN (Okano et al., 1998a). Néanmoins, la délétion du résidu Ser permet de créer une protéine active qui méthyle les cytosines dans un contexte CCWGG (W = A ou T) (Pinarbasi et al., 1996).

De plus, la délétion du gène *Dnmt2* dans des cellules ES murines n'a aucun effet sur les profils de méthylation ou sur l'habilité à faire de la méthylation *de novo* sur de l'ADN rétroviral (Okano et al., 1998a). Enfin, *Dnmt2* n'est pas essentiel au développement des mammifères puisque les souris homozygotes pour une délétion de *Dnmt2* sont viables et fertiles avec aucun gros défaut observable (Okano et al., 1998a).

Plus récemment, il a été démontré que même si cette enzyme ne semble pas avoir de fonction au niveau de la méthylation de l'ADN, elle est importante pour la méthylation des cytosines 38 au niveau de l'anticodon des ARNtAsp et a depuis été renommée *Trdmt1* (tRNA aspartic acid (D) méthyltransférase 1) (Ehrenhofer-Murray, 2017; Goll et al., 2006; Müller et al.,

2015). Cette méthylation aurait pour fonction de protéger les ARNt des clivages par des ribonucléases (Schaefer et al., 2010), et ainsi réguler de manière plus fine la synthèse protéique (Tuorto et al., 2012, 2015).

#### d. DNMT3

##### a) DNMT3A et DNMT3B

Elles ont été découvertes peu de temps après la découverte de *Dnmt1*. En effet, il y a eu un long débat afin de savoir comment se mettait en place la méthylation de l'ADN et comment cette dernière était maintenue au cours des générations. C'est l'expérience avec la délétion de *Dnmt1* dans des cellules ES murines qui a permis de répondre à cette question. En effet, malgré une hypométhylation globale du génome dû à la perte de DNMT1, ces cellules présentaient tout de même un niveau résiduel de méthylation. De plus, ces cellules avaient toujours la capacité de méthyler *de novo* puisque l'ajout d'ADN proviral ectopique induisait systématiquement la méthylation de ce dernier malgré l'absence de DNMT1, ce qui suggérait l'existence d'autres enzymes responsables de la méthylation *de novo* dans les génomes mammifères (Lei et al., 1996).

C'est ainsi qu'ont été découvertes, en même temps chez l'humain et la souris, les enzymes DNMT3A et DNMT3B en 1998 (Okano et al., 1998b). Elles sont considérées comme des ADN méthyltransférases dites « *de novo* », c'est-à-dire qu'elles peuvent déposer un groupement méthyle sur une molécule d'ADN préalablement non méthylée (Figure 6). Ceci est soutenu par le fait qu'elles sont très fortement exprimées dans les cellules ES murines, dans les étapes précoces du développement de l'embryon ainsi que lors du développement des cellules souches germinales au moment de la mise en place des profils de méthylation du génome. Par contre, leur expression est réduite dans les tissus somatiques différenciés après la naissance (Okano et al., 1998b). Une délétion totale de ces deux gènes chez la souris conduit à une létalité embryonnaire. C'est le cas aussi pour la délétion seule de *Dnmt3b*, alors que la délétion seule de *Dnmt3a* induit une létalité post-naissance à l'âge de 3-4 semaines (Okano et al., 1999). De plus, les protéines recombinantes DNMT3A et DNMT3B sont capables de méthyler les cytosines au niveau des dinucléotides CpG sur différents substrats d'ADN natif ou synthétique *in vitro* et elles ne présentent aucune préférence pour l'ADN hémiméthylé (Aoki et al., 2001; Okano et al., 1998b). L'expression ectopique de ces deux gènes au niveau de cellules mammifères ou au niveau de drosophile augmente le niveau de méthylation de l'ADN (Chen et al., 2003; Hsieh, 1999; Lyko et al., 1999), tandis que la délétion d'un de ces deux gènes ou mêmes des deux bloque la méthylation *de novo* des cellules ES, des embryons précoces, ainsi que celle des gènes soumis à empreinte au niveau des cellules germinales (Kaneda et al., 2004; Okano et al., 1999). En

résumé, ces données démontrent que ces enzymes jouent des rôles fondamentaux lors de la mise en place des profils de méthylation au cours du développement, ce dont on parlera au chapitre 5.

Elles sont fortement exprimées durant l'embryogénèse. Lors des stades précoces du développement embryonnaire, c'est DNMT3B qui est préférentiellement exprimée, et elle est la principale responsable de l'acquisition de la méthylation de l'ADN au cours de l'implantation du blastocyte (Auclair et al., 2014; Borgel et al., 2010). DNMT3A est quant à elle plutôt exprimée dans les stades embryonnaires tardifs et dans les cellules différenciées (Watanabe et al., 2002). Le gène *Dnmt3a* code deux isoformes : DNMT3A et DNMT3A2 qui font respectivement 908 et 689 aa chez la souris. Ces deux isoformes possèdent le même domaine catalytique mais leur différence réside dans leur domaine N-terminal. Cette différence semble influencer leur localisation cellulaire, en effet DNMT3A se retrouve localisé au niveau de l'hétérochromatine, alors que DNMT3A2 est diffus dans le nucléoplasme, excluant les nucléoles et l'hétérochromatine (Bachman et al., 2001; Chen et al., 2002). Le gène *Dnmt3b* code pour cinq isoformes chez l'Homme (DNMT3B1-DNMT3B5) dont la plus longue comporte 853 aa. Contrairement aux isoformes de DNMT3A, seules les formes 1 et 2 des DNMT3B montrent une activité catalytique, les autres présentant une perte partielle des domaines IX et X de leur partie C-terminale (Aoki et al., 2001; Chen et al., 2003; Okano et al., 1998b). Concernant ces isoformes inactifs, une hypothèse serait qu'elles jouent un rôle de dominant négatif pour la régulation de l'activité de DNMT3B, notamment en modifiant la capacité des variants actifs de DNMT3B à se lier à l'ADN, à s'homo-dimériser et à interagir avec DNMT3L (Saito et al., 2002; Van Emburgh and Robertson, 2011). Les isoformes DNMT3A2 et DNMT3B1 s'expriment fortement durant l'embryogénèse puis subissent une répression jusqu'à devenir indétectable dans les tissus somatiques alors que les autres isoformes montrent une expression modérée, tissu-spécifique, à l'exception de DNMT3A qui reste faiblement détectable de manière ubiquitaire (Chen et al., 2002).

La partie N-Terminales de DNMT3A et DNMT3B est plus courte que celle de DNMT1 mais les domaines qui les composent sont similaires, malgré des loci bien distincts. DNMT3A pèse 100 KDa et DNMT3B, 130 KDa. Chacune d'elle possède un domaine PWWP, un domaine ADD (ATRX-DNMT3-DNMT3L) et un domaine catalytique en C-terminal (Figure 7). Le domaine ADD est une région riche en cystéine qui permet l'interaction entre protéines et plus précisément, ce domaine permet l'interaction des DNMT3 avec les queues des histones H3K4 non méthylées (Guo et al., 2015b; Ooi et al., 2007; Otani et al., 2009; Zhang et al., 2010). Cette interaction stimule l'activité de DNMT3A car il stabilise la conformation active de l'enzyme. De plus, ce domaine permet aussi l'interaction avec de nombreuses protéines appartenant à la

machinerie épigénétiques telles que SUV39H1 (Fuks et al., 2003), SETDB1 (Li et al., 2006), ou encore EZH2 (Viré et al., 2006). Le domaine PWWP qui est riche en motifs prolines et tryptophanes, reconnaît quant à lui spécifiquement la marque d'histone H3K36me3 et permettrait de cibler l'activité enzymatique de DNMT3A et DNMT3B au niveau des corps de gènes transcriptionnellement actifs (Baubec et al., 2015; Dhayalan et al., 2010). Par ailleurs, cette reconnaissance serait également essentielle aux enzymes DNMT3 pour leur ciblage de la 5mC au niveau des zones péri-centromériques de la chromatine (Chen et al., 2004; Ge et al., 2004). Ainsi, les domaines ADD et PWWP des protéines DNMT3 assurent une interaction forte avec la chromatine et une régulation de l'activité des enzymes par les marques d'histone H3K4 et H3K36me3. Néanmoins, l'activité enzymatique de ces deux protéines est dépendant d'un co-facteur nommé DNMT3L.

#### b) DNMT3L

DNMT3L (DNMT3-like) a été isolé en 2000 à partir de l'analyse de bases de données de séquences du génome humain (Aapola et al., 2000). Son homologue murine a été identifié par la suite en 2001 (Aapola et al., 2001; Hata et al., 2002). Cette protéine de 421 aa chez la souris possède une séquence homologue à celle de DNMT3A et de DNMT3B, mais sans un domaine catalytique complet ni de domaine PWWP (Figure 7). Ainsi, elle n'a pas la capacité de méthyliser l'ADN (Hata et al., 2002). Bien qu'elle n'ait pas de domaine méthyltransférase fonctionnel, son absence conduit à une hypométhylation des ICRs et de plusieurs éléments répétés, couplés à une hypométhylation globale du génome, dans les cellules germinales (Bourc'his and Bestor, 2004; Hata et al., 2002; Kato et al., 2007; Shirane et al., 2013; Smallwood et al., 2011; Zamudio et al., 2015). En conséquence, l'inactivation de *Dnmt3L* conduit à des catastrophes méiotiques et une stérilité complète chez les mâles (Bourc'his et al., 2001). Chez les femelles, l'absence de DNMT3L n'affecte pas le développement des ovocytes mais les descendants des femelles *Dnmt3L*<sup>-/-</sup> ne sont pas viables en raison notamment de l'absence de mise en place des empreintes de méthylation maternelles (Bourc'his and Bestor, 2004; Hata et al., 2002). Néanmoins, la délétion de *Dnmt3L* chez la souris n'a aucune incidence sur le développement embryonnaire et la viabilité, ce qui suggère que son action est limitée aux cellules germinales.

DNMT3L est abondante dans les embryons précoces ainsi que dans les cellules germinales (Bourc'his and Bestor, 2004; Guenatri et al., 2013). Elle co-localise avec les protéines DNMT3A et DNMT3B dans les cellules mammifères (Hata et al., 2002) mais pas avec DNMT1. DNMT3L est connue pour interagir avec les domaines catalytiques de DNMT3A et DNMT3B afin de stimuler leur activité enzymatique (Chedin et al., 2002; Chen et al., 2005; Jia et al., 2007). Enfin, il a été démontré que DNMT3L interagit aussi avec la protéine HDAC1 afin de renforcer la

répression de la transcription, ce qui pourrait également contribuer à l'établissement des empreintes génomiques de méthylation par DNMT3L dans les gamètes (Aapola et al., 2002). Ceci éclaire le fait que la méthylation de l'ADN n'est probablement pas l'unique mécanisme de répression qui régule le phénomène des empreintes génomiques parentales, mais qu'elle est vraisemblablement coordonnée avec des modifications des histones.

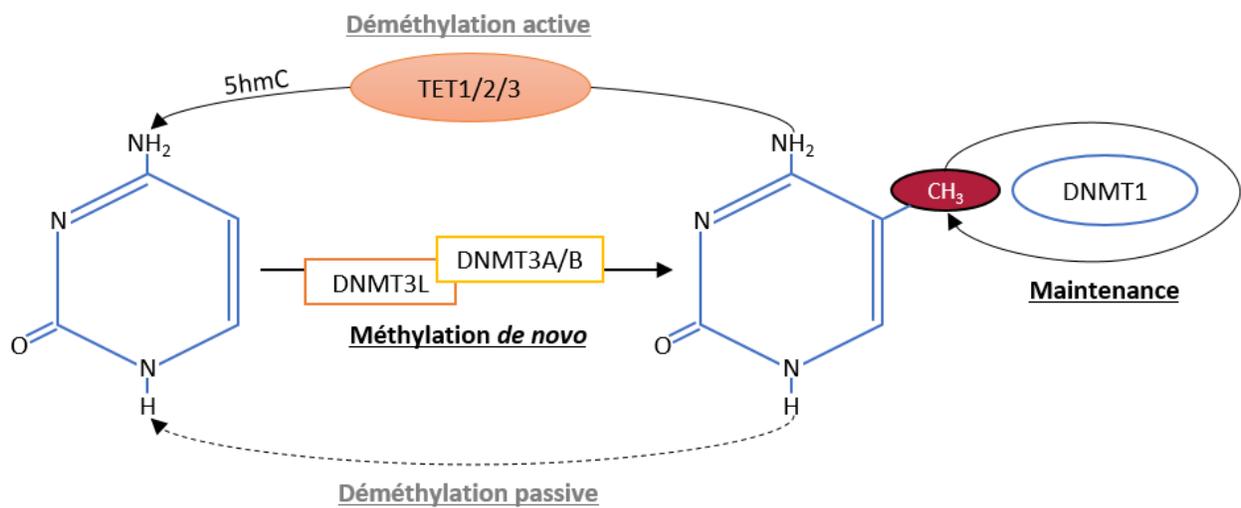
### c) DNMT3c

Ce gène a été découvert très récemment par Barau et al., 2016 et correspond au pseudogène *Gm14490*. Localisé 9Kb en amont du gène *Dnmt3b*, il fut tout d'abord considéré comme un pseudo gène étant donné que sa transcription et qu'un possible ORF n'avaient jusqu'alors pas été décrits. *Gm14490* est spécifiquement exprimé dans les cellules germinales mâles exclusivement. De plus, il existe deux isoformes. La petite est spécifique des testicules post-natals tandis que la longue isoforme est exprimée dans les cellules germinales embryonnaire. Le pic d'expression à E16,5 de la longue isoforme coïncide avec la mise en place de la méthylation *de novo* au niveau des gamètes mâles (Molaro et al., 2014).

L'inactivation de *Dnmt3c* chez la souris induit un phénotype semblable à celui de l'inactivation de *Dnmt3L* (Jain et al., 2017; Zamudio et al., 2015). Les animaux *Dnmt3c*<sup>-/-</sup> présentent un développement somatique normal. Néanmoins, les souris mâles ont une taille de testicules réduite due à une azoospermie liée à un arrêt de la spermatogénèse en méiose I au stade pachytène. De plus, tout comme pour les souris *Dnmt3L*<sup>-/-</sup>, une surexpression des transposons est observée dans les testicules, plus spécifiquement les transposons LINE et les « jeunes rétrotransposons » (Barau et al., 2016). On parle de « jeunes rétrotransposons » pour désigner les transposons apparus « récemment » dans le génome au cours de l'évolution.

En regardant de plus près l'organisation des domaines de DNMT3c, on remarque que cette dernière est très similaire à la famille des DNMT3s (Figure 7). En effet, DNMT3c possède 70% d'identité avec DNMT3B et, tout comme DNMT3L, elle ne possède pas de domaine PWWP ce qui la rend incapable de reconnaître les marques d'histone H3K36me3. Sa similitude avec DNMT3B explique que cette nouvelle enzyme soit capable de méthyler l'ADN. Dans leur travaux (Barau et al., 2016), ils ont notamment démontré qu'elle méthyle spécifiquement les promoteurs des « jeunes rétrotransposons » dans la lignée gamétique mâle, probablement en remplacement de DNMT3B puisqu'il a été montré que l'inactivation de DNMT3B n'affecte pas le développement et la méthylation des cellules germinales (Kaneda et al., 2004).

Beaucoup de choses restent à être élucidées à propos de cette nouvelle protéine, mais il semblerait qu'elle affecte spécifiquement les rétrotransposons actifs tout comme le fait le système de méthylation dépendant des piRNA lors de la spermatogénèse foetal



### **Figure 9 : Acteurs principaux de la méthylation de l'ADN**

La méthylation *de novo* est dirigée par DNMT3A/B et leur cofacteur DNMT3L. DNMT1 permet le maintien de la méthylation au niveau des fourches de réplication lors des divisions cellulaires. Lorsque l'activité de DNMT1 au niveau des fourches de réplication est inhibée, la 5mC n'est plus maintenue au cours des divisions cellulaires ce qui conduit à une dilution progressive de la méthylation. C'est ce qu'on appelle la déméthylation passive. La déméthylation active met en jeu des voies enzymatiques, notamment au travers des enzymes TET1/2/3.

(Manakov et al., 2015; Molaro et al., 2014). Cela suggère que DNMT3C pourrait être l'enzyme recrutée par la voie des piRNAs au niveau des transposons. Enfin, bien qu'il semble que DNMT3c est indispensable pour la répression stable des éléments répétés chez les gamètes murines, cette enzyme n'est pas présente chez tous les mammifères comme par exemple chez l'Homme et le lapin (Barau et al., 2016).

### 3. La déméthylation des cytosines

Comme noté dans l'introduction, l'une des caractéristiques des marques épigénétiques est qu'elles sont réversibles. Ainsi, alors que les DNMT méthylent l'ADN, il existe des mécanismes permettant de déméthylater l'ADN. Dans la littérature ces mécanismes se classent généralement en deux catégories : la déméthylation dite "passive" et la déméthylation dite "active". Cette capacité que possède les organismes à méthyler et déméthylater l'ADN est indispensable à la régulation dynamique de la méthylation de l'ADN dans les différents lignages cellulaires, et donc de ce fait au développement embryonnaire et à la survie de l'organisme. Chez la souris par exemple, nous savons qu'il existe deux phases de reprogrammation de la méthylation de l'ADN : une programmation au niveau du blastocyste et une reprogrammation lors de la spécification des cellules germinales primordiales. Je parlerai plus en détails de ces deux événements plus tard. Voyons d'abord les mécanismes de déméthylation de l'ADN.

#### a. La déméthylation passive

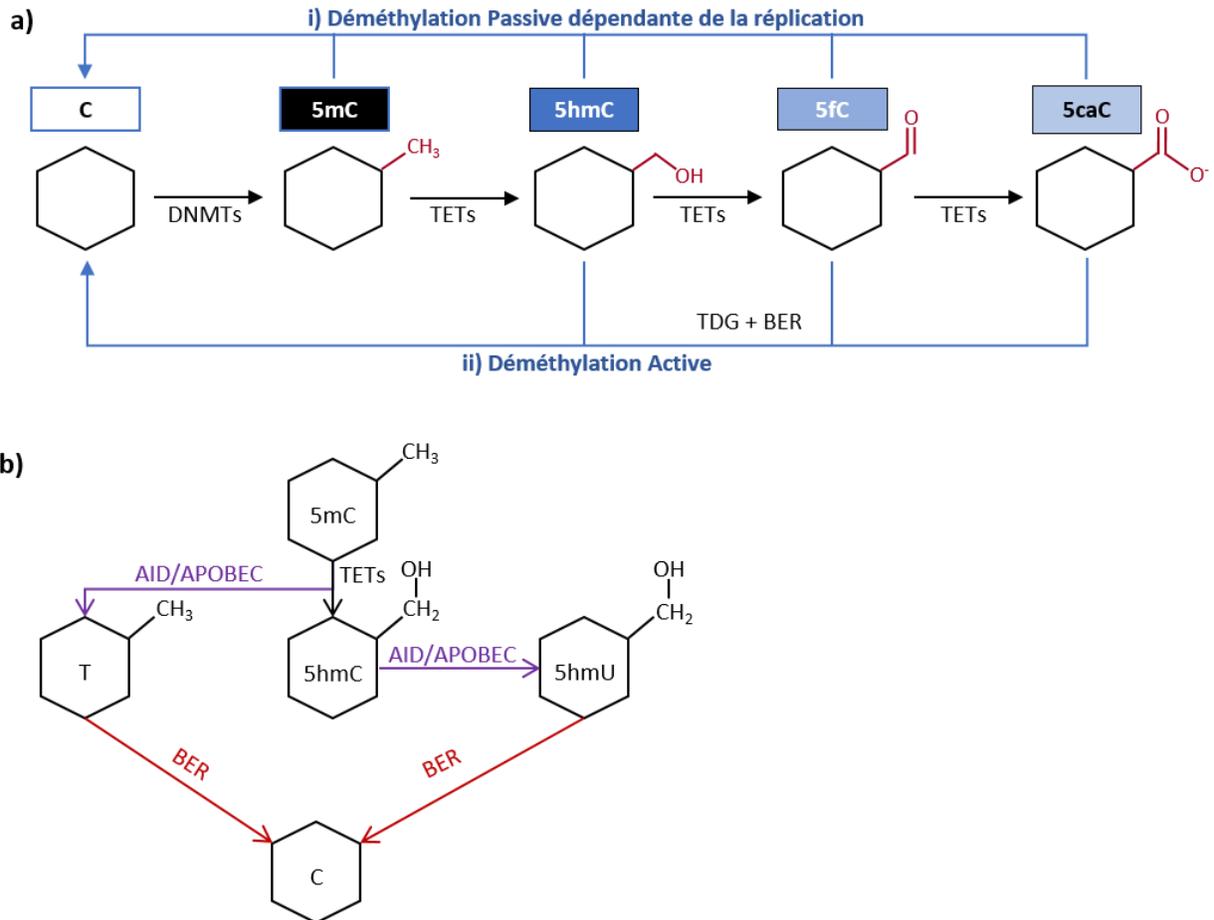
La déméthylation passive consiste en l'exclusion de DNMT1 au niveau de la fourche de réplication (**Figure 9**). Ceci a pour conséquence que DNMT1 ne peut plus assurer la copie des profils de méthylation de l'ADN au cours des divisions cellulaires, induisant une dilution progressive de la 5mC. Ce processus semble impliqué lors de la vague de déméthylation des cellules primordiales germinales dont nous parlerons plus tard ainsi que dans l'embryon, juste après la fécondation.

#### b. La déméthylation active

La déméthylation active consiste en la mise œuvre de réactions enzymatiques de suppression du groupement méthyl au niveau des cytosines, indépendamment de la réplication de l'ADN au cours de la division cellulaire.

- **Les enzymes TET**

En 2009, deux études ont confirmé l'existence d'une nouvelle modification des cytosines dans le génome mammifère : l'hydroxyméthyl-cytosine (5hmC), une forme oxydée de la 5mC



**Figure 10 : Les voies enzymatiques de déméthylation de l'ADN**

**a)** Les enzymes TET1/2/3 convertissent la 5mC en 5hmC, en 5fC puis en 5caC, par oxydation successive. Ces dérivés oxydés ne sont pas reconnus par DNMT1 et peuvent donc être dilués lors de la réplication passivement (i). Alternativement les formes oxydées peuvent être excisées par une glycosylase (TDG) et remplacées par une C lors du processus de réparation par le système BER. **b)** Les bases 5mC et 5hmC peuvent être déaminées par les enzymes de la famille des déaminases AID/APOBEC, formant ainsi les bases T ou 5hmU qui pourraient être excisées par des glycosylase et réparées par le système de réparation BER.

(Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). La conversion d'une cytosine méthylée en cytosine hydroxyméthylée se fait par le biais d'enzymes appartenant à la famille des TET (ten-eleven translocation) : TET1, TET2 et TET3. Ces enzymes, qui sont des  $\alpha$ -kétoglutarate( $\alpha$ KG) et Fe(II)-dépendant dioxygénases, ont la capacité d'oxyder les cytosines méthylées et de les convertir en cytosines hydroxyméthylées (Tahiliani et al., 2009). Les protéines TET possèdent un domaine CXXC qui favorise la liaison aux CpG non méthylés. En accord avec le fait que le domaine CXXC préfère les CpG non méthylés, des études de cartographie par ChIP-seq ont montré que les TET se fixent plus préférentiellement au niveau des îlots CpG (Williams et al., 2011; Wu et al., 2011). TET1 et TET3 possèdent des domaines CXXC en N-terminal, alors que TET2 interagit avec une protéine possédant un domaine CXXC, la protéine IDAX. Néanmoins, la capacité des TET à être recrutés dans le génome au travers de leurs domaines CXXC et leur contribution à la déméthylation de l'ADN reste encore sujet de débats et de découvertes.

Les trois membres de la famille TET chez les mammifères sont impliqués dans de nombreux processus biologiques tels que la régulation des gènes au cours développement embryonnaire ou encore lors de cancers (Ito et al., 2010; Rasmussen and Helin, 2016). Chacune de ces enzymes possède un profil d'expression qui lui est propre et dépendant d'un type cellulaire. Par exemple, chez la souris, TET1 et TET2 sont exprimés lors du développement précoce au niveau de la masse cellulaire interne et pendant l'implantation, alors que TET3 est exprimé dans les ovocytes et dans le zygote jusqu'au stade 2 cellules. Il est d'ailleurs suggéré que TET3 contribue à la déméthylation des deux génomes parentaux au cours du développement précoce de l'embryon (Shen et al., 2014). Toujours chez les souris, la délétion totale de *Tet1* induit une diminution de la taille des souris (Dawlaty et al., 2013) et la perte totale de *Tet2* conduit à l'apparition de leucémies à 20 semaines (Moran-Crusio et al., 2011). Enfin, la délétion totale de *Tet3* chez la souris induit une mort autour de la naissance pour des raisons qui ne sont pas encore élucidées (Gu et al., 2011b).

L'hydroxyméthylation des cytosines catalysée par les enzymes TET peut initier une réaction de déméthylation de l'ADN selon plusieurs mécanismes passifs ou actifs (Figure 9-10). Pour la voie passive, des expériences *in vitro* ont démontré qu'UHRF1 peut reconnaître la 5hmC avec autant d'affinité que la 5mC. En revanche ceci n'est pas le cas de DNMT1 qui présente une faible affinité de liaison à la 5hmC (Frauer et al., 2011; Hashimoto et al., 2012). Ainsi, non reconnue par DNMT1, la 5hmC ne peut plus servir de substrat pour copier la méthylation au cours de la réplication, ce qui conduit à une perte de la méthylation au fil des divisions cellulaires (Figure 10-a). Les voies actives de déméthylation par les enzymes TET impliquent une série de réactions chimiques en cascade. Les enzymes TET peuvent transformer la 5hmC par oxydation successive

en 5-formylcytosine (5fC) et en 5-carboxylcytosine (5caC). Ces dérivés oxydatifs non reconnus par DNMT1 peuvent eux aussi disparaître passivement. Mais ils peuvent aussi être spécifiquement reconnus et excisés par l'enzyme TDG (Thymine DNA Glycosylase), puis réparés par un C à l'aide du système de réparation de l'ADN BER (Base-Excision Repair) (He et al., 2011) (**Figure 10-a**). Il est aujourd'hui reconnu que TDG possède une forte activité glycosylase sur des substrats contenant de la 5fC et 5caC. En accord avec ce modèle, l'inactivation de TDG induit une accumulation de 5fC et 5caC à un grand nombre de sites génomiques dans les cellules embryonnaires de souris (Raiber et al., 2012; Shen et al., 2013). De plus les embryons de souris mutants *Tdg*<sup>-/-</sup> présentent une hyperméthylation de certains promoteurs de gènes (Cortázar et al., 2011; Cortellino et al., 2011). Par ailleurs d'autres glycosylases que TDG, comme par exemple UNG2 (Xue et al., 2016) ou NEIL1/2 (Schomacher et al., 2016), pourraient également participer à la déméthylation en excisant de l'ADN les produit d'oxydation des enzymes TET.

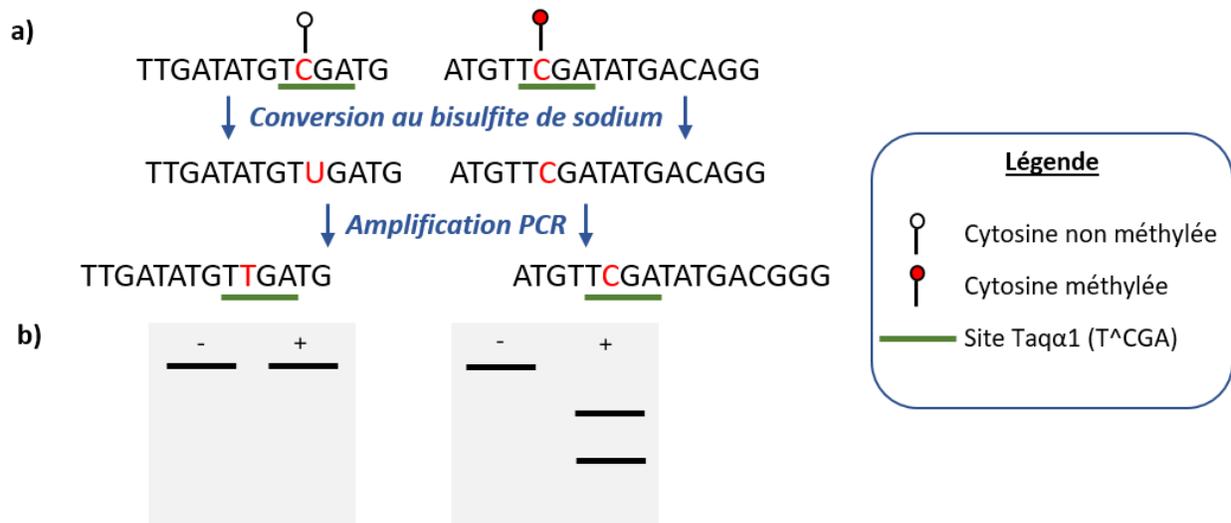
En résumé, il existe plusieurs voies de déméthylation de l'ADN et il est fort probable que toutes ces voies passives ou actives soient complémentaires pour réguler les profils de méthylation de l'ADN au cours du développement embryonnaire et dans les tissus adultes (Kohli and Zhang, 2013).

- **Les enzymes AID/APOBEC**

Une autre façon de déméthylar la 5mC implique des déaminases de la famille AID/APOBEC. Comme leur nom le suggère, ces enzymes peuvent réaliser une réaction de déamination des 5mC, remplaçant la Cytosine par une Thymine (**Figure 10-b**). Ceci a pour conséquence un mésappariement G:T qui pourra être réparé en G:C par le système de réparation BER. Ces enzymes sont aussi capables de déaminer les 5hmC, en créant cette fois-ci une Uracile oxydée qui pourra à son tour être excisée par le TDG ou par l'Uracil-DNA glycosylase 1 (SMUG1) puis être finalement repérée par le système BER (Guo et al., 2011) (**Figure 10-b**). Il a été proposé que AID participe à la vague de déméthylation globale dans les cellules germinales chez la souris (Popp et al., 2010), cependant cette découverte est controversée et le rôle des déaminases dans le déméthylation de l'ADN en conditions physiologiques au cours du développement reste encore à démontrer avec plus de précision.

#### **4. Techniques d'études de la méthylation des cytosines**

De nos jours, il existe de nombreuses techniques permettant d'étudier la méthylation des cytosines. Chacune de ces méthodes possèdent des avantages et des inconvénients. Ainsi, le plus important est de choisir la technique qui permettra de répondre à une question biologique



**Figure 11 : La conversion au bisulfite de sodium et l'approche COBRA**

**a)** La conversion au bisulfite de sodium convertit uniquement les cytosines non méthylées en uracile. Après PCR, les cytosines non méthylées (devenues des uraciles) sont remplacées par des Thymines.

**b)** Principe de la méthode COBRA (Combinated Bisulfite Restriction Anaslysis): après conversion au bisulfite de sodium, le locus d'intérêt est amplifié par PCR et le produit de PCR est digéré avec une enzyme de restriction qui contient un CG dans son site de restriction (par exemple Taq $\alpha$ 1) puis déposé sur gel d'agarose pour observer les profils de digestion. Si l'enzyme coupe cela signifie que le CpG présent dans le site de restriction était méthylé.

donnée mais aussi en accord avec d'autres facteurs tels que (i) la quantité de l'ADN à étudier, (ii) la sensibilité et la spécificité de la méthode, (iii) les compétences bio-informatiques à disposition pour l'analyse des données, (iv) le coût.,

Les méthodes de détection de la méthylation de l'ADN se regroupent en trois catégories. La méthylation de l'ADN peut être étudiée directement parce qu'elle crée ou fait disparaître un site de restriction enzymatique. L'ADN méthylé peut aussi être sélectionné/enrichi par immunoprécipitation avec des anticorps anti 5mC (ou anti 5hmC) ou des protéines reconnaissant spécifiquement l'ADN méthylé. Enfin, la conversion au bisulfite de sodium de l'ADN permet de « fixer » l'état de méthylation de l'ADN en convertissant les cytosines non méthylées en uraciles, et le rend donc analysable par des techniques bas débit comme la PCR ou la digestion enzymatique, mais aussi par des technologies haut débit comme le séquençage massif ou les puces dédiées. Dans cette introduction, je ne parlerai que des principales méthodes utilisées mais il en existe énormément comme vous pourrez le voir dans les revues ci-après (Kurdyukov and Bullock, 2016; Olkhov-Mitsel and Bapat, 2012).

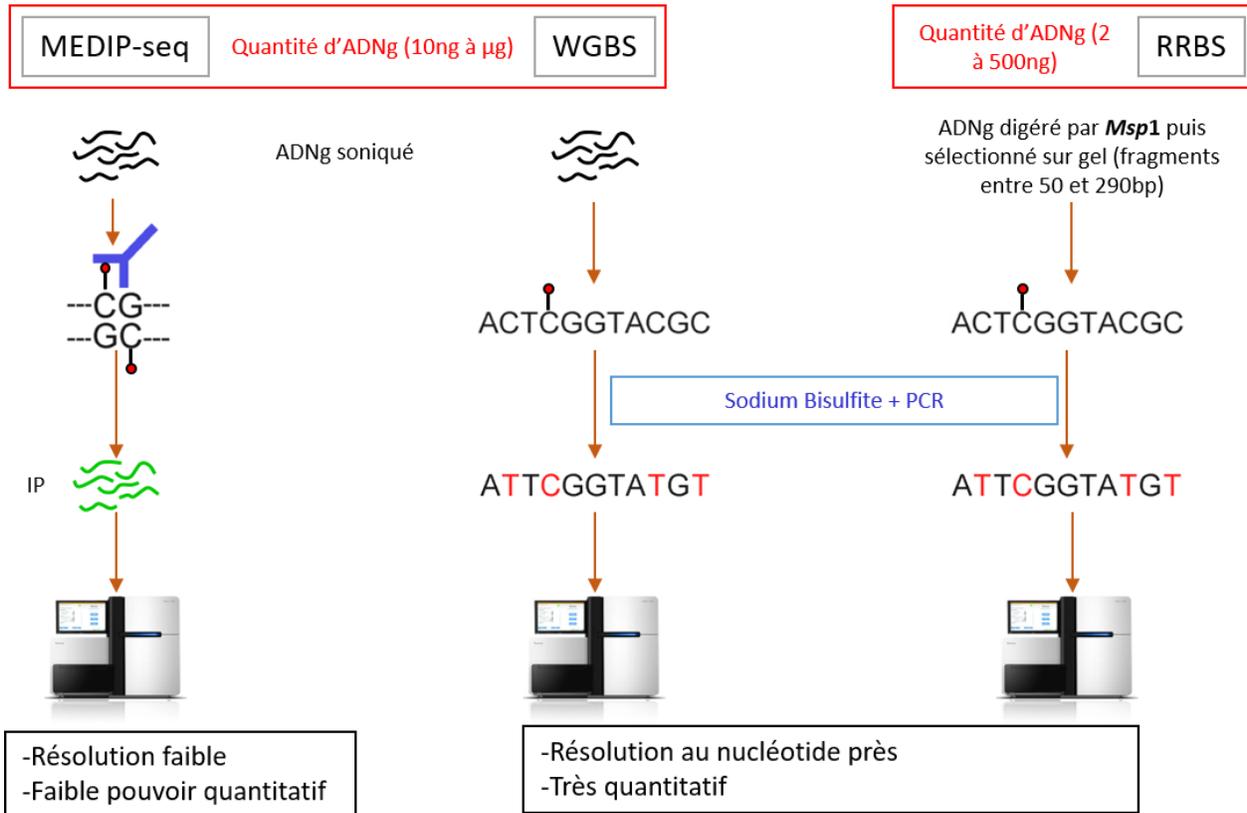
#### **a. Les enzymes de restriction**

Cette méthode consiste en l'utilisation d'enzymes de restriction sensibles à la méthylation de l'ADN (par exemple *HhaI*, *HpaII*, *NotI*, *McrBC* etc...). Pour révéler le statut de méthylation d'un locus spécifique, les digestions peuvent être couplées à plusieurs méthodes : la PCR, le Southern blot, ou des gels 2 dimensions comme dans le cas de la technique du RLGS (Restriction Landmark Genome Scanning) qui est l'une des premières approches développées pour cartographier la méthylation de l'ADN à grande échelle.

Les désavantages majeurs de ces approches sont qu'elles nécessitent généralement des quantités importantes d'ADN, qu'elles offrent une couverture non complète et biaisée du génome du fait de la distribution non homogène des sites de restriction et qu'il est nécessaire que le site de restriction soit présent dans la séquence ADN à étudier. De plus, ces techniques sont peu quantitatives et n'offre qu'une information qualitative sur l'état de méthylation des séquences étudiées.

#### **b. Conversion au bisulfite de Sodium**

La bisulfitation de l'ADN permet de transformer, par désamination hydrolytique, une cytosine en uracile si cette dernière n'est pas méthylée. Au contraire, les cytosines méthylées sont résistantes à la conversion au bisulfite de sodium, ce qui permet de « fixer » l'état de méthylation dans la séquence de l'ADN (**Figure 11a**). La conversion au bisulfite de sodium peut



**Figure 12 : Technique d'étude de la méthylation de l'ADN à haut débit**

**MeDIP-seq:** Methylated DNA Immunoprecipitation. Après fragmentation de l'ADNg par sonication, les fragments d'ADN contenant la 5mC, sont immunoprécipités avec un anticorps anti-5mC, puis séquencés à haut débit.

**WGBS:** Whole Genome Bisulfite Sequencing. L'ADNg est fragmenté et converti au bisulfite de sodium, puis séquencé à haut débit pour cartographier les bases C qui ont résisté à la conversion.

**RRBS:** Reduced Representation Bisulfite Sequencing. La technique de lecture de la méthylation est identique à celle du WGBS, néanmoins l'ADNg est préalablement digéré avec l'enzyme *MspI* (C<sup>^</sup>CGG), ce qui permet d'enrichir en fragments d'ADN contenus dans les îlots CpG et de réduire les coûts de séquençage.

être couplée à de nombreuses approches expérimentales pour quantifier la méthylation à l'échelle du nucléotide.

Une de ces approches consiste à réaliser une étape d'amplification par PCR de la région d'intérêt après conversion bisulfite, et de mesurer l'état de méthylation de chaque cytosine par clonage et séquençage du produit de PCR suivi d'un alignement avec la séquence de référence non convertie. Outre sa résolution et son aspect quantitatif à la résolution du nucléotide, l'intérêt de cette approche est qu'elle permet également d'accéder à l'état de méthylation de molécules individuelles (Frommer et al., 1992).

Il est aussi possible de digérer le produit de PCR par une enzyme de restriction contenant un CpG dans son site de restriction telle que l'enzyme *TaqI*α (T<sup>α</sup>CGA). Cette technique s'appelle le COBRA (Combined Bisulfite Restriction Analysis) (Xiong and Laird, 1997) (Figure 11b). L'enzyme ne coupera que si le CpG initialement présent est méthylé, c'est à dire résistant à la conversion en UpG lors du traitement au bisulfite de sodium. Après migration sur un gel d'agarose, on peut visualiser si le ou les sites de restriction présents sur la région amplifiée étaient méthylés ou non. L'intérêt de cette méthode est qu'elle permet un criblage rapide et à faible coût de la méthylation d'une région d'intérêt, néanmoins elle est limitée par la nécessité que le site de restriction soit présent et bien placé dans la région d'intérêt à étudier.

### ***c. Les méthodes d'analyses à l'échelle génomique***

Au fil des années, de nouvelles technologies permettant de cartographier la méthylation de l'ADN à l'échelle génomique ont été développées. Ces cartographies permettent d'accéder à une information beaucoup plus complexe, de découvrir de nouvelles régions différenciellement méthylées, ainsi que de suivre la reprogrammation de la méthylation sur le génome en fonction des stades de développement ou des changements environnementaux. Encore une fois, je ne vous présenterai pas toutes les techniques existantes décrites dans plusieurs revues récentes (Bock, 2012; Kurdyukov and Bullock, 2016), mais je me limiterai à celles qui sont le plus utilisées : le MeDIP (Methylated DNA immunoprecipitation), le WGBS (Whole Genome Bisulfite Sequencing) ainsi que celle que j'ai mise en place dans le laboratoire et adaptée à la vue des faibles quantités d'ADN disponible pour mes projets : le RRBS (Reduced Representation Bisulfite Sequencing).

- **MeDIP (Methylated DNA immunoprecipitation)**

Cette technique consiste à utiliser un anticorps dirigé contre la 5mC (Weber et al., 2005), après sonication de l'ADN génomique (Figure 12). Le principe repose sur le fait que l'ADN immunoprécipité (IP) et l'ADN non traité (INPUT) sont couplés à des fluorochromes et co-hybridés

sur des puces à ADN (Weber et al., 2005), ou bien séquencés par séquençage à haut-débit (Down et al., 2008). Cette méthode permet de mesurer la méthylation de l'ADN avec une résolution d'environ 250 paires de bases. Ses inconvénients majeurs sont qu'elle nécessite beaucoup d'ADN et qu'elle est peu quantitative ainsi que fortement biaisée par la densité en CpG. C'est pourquoi les technologies combinant le traitement au bisulfite de sodium et le séquençage haut-débit sont désormais privilégiées (voir ci-dessous).

Une approche similaire est le **MBD-CAP** (Methyl Binding Domain Capture). Cette technique utilise des protéines capables de se lier à l'ADN méthylé telles que MeCP2 pour immunoprécipiter l'ADN méthylé. De la même manière que le MeDIP, cette méthode permet d'obtenir des régions fortement méthylées mais n'apporte pas de résolution à la résolution du nucléotide (Rauch et al., 2008).

- **RRBS (Reduced Representation Bisulfite Sequencing)**

C'est la technique que j'ai le plus utilisée tout au long de ma thèse et que j'ai mise au point dès mon arrivée au laboratoire à partir du protocole de Gu (Gu et al., 2011a). J'ai adapté le protocole afin qu'il soit compatible avec mes projets pour lesquels les quantités d'ADN génomique disponibles sont très faibles. Ainsi, j'ai réussi à l'adapter jusqu'à une quantité d'ADNg de départ de 2ng alors que Gu l'avait optimisé pour une quantité d'ADNg de départ entre 10 et 300ng. De plus, j'ai aussi optimisé la sélection de taille des fragments (jusqu'à 300bp au lieu de 220bp) pour améliorer la couverture génomique par rapport au protocole d'origine. J'ai rédigé une version détaillée du protocole pour le site Epigenesys (voir **Annexe**) et le protocole est également détaillé dans le chapitre 'matériels et méthodes'.

Cette technique consiste à réaliser une conversion au bisulfite de sodium suivi d'un séquençage haut débit sur une fraction du génome sélectionnée par digestion avec l'enzyme de restriction *MspI* (C<sup>A</sup>CGG). Après digestion par l'enzyme *MspI* et ligation d'adaptateurs méthylés, une sélection de taille est effectuée par migration sur un gel d'agarose pour isoler les fragments d'ADN compris entre 50 et 300bp (**Figure 12**). La sélection de fragments de petite taille générés par l'enzyme *MspI* permet d'enrichir en fragments génomiques riches en CpG, c'est-à-dire les îlots CpG au niveau des promoteurs mais aussi les séquences répétées. Ces fragments sont ensuite ligués à des adaptateurs Illumina et convertis en bibliothèques de séquençage haut-débit pour un séquenceur Illumina HiSeq. Les fragments séquencés sont alignés sur le génome de référence et la méthylation est estimée pour chaque cytosine en rapportant le nombre de fragments séquencés contenant un C (c'est-à-dire un C initialement méthylé qui a résisté à la conversion) à cette position par rapport au nombre total de fragments séquencés à cette position. Cette

technologie permet de mesurer la méthylation d'environ 1,5 millions de CpG dans le génome de souris avec une résolution au nucléotide près. Ainsi, le RRBS couvre la quasi-totalité des îlots CpG du génome ainsi que de manière plus sporadique les corps de gènes, les éléments transposables et les régions intergéniques. Elle constitue une excellente méthode pour estimer de manière quantitative la méthylation globale du génome et des éléments de régulation à un coût bien plus faible que le WGBS (environ 300€). La technique RRBS présente également l'avantage d'être robuste et reproductible sur de faibles quantités de cellules. D'ailleurs une équipe a amélioré cette technique pour cartographier l'ADN méthylé à partir de cellules uniques (Guo et al., 2013).

- **WGBS (Whole Genome Bisulfite Sequencing)**

Cette technique permet de générer des méthylomes d'un génome entier à la résolution du nucléotide (Figure 12). Cette technique peut couvrir 95% du génome alors que le RRBS n'en couvre qu'environ 4%. Par rapport au RRBS, les deux principales différences sont dans la technique de fragmentation de l'ADN et le séquençage. Dans le cadre du WGBS, l'ADN est simplement soniqué et le séquençage s'effectue sur le génome entier, ce qui nécessite une profondeur de séquençage bien plus importante avec plus de 300 millions de lectures par échantillon contre 20 millions pour le RRBS. Ainsi, elle permet la mesure de la méthylation au niveau des régions intra et intergéniques, sites inaccessibles par RRBS. De par sa couverture complète du génome, le WGBS constitue la méthode de référence pour l'analyse du méthylome car elle apporte bien plus d'informations que le RRBS, cependant son coût prohibitif (environ 2 000€) en limite encore l'utilisation.

Tout comme le RRBS, cette technologie a été améliorée afin de diminuer la dégradation d'ADN du au traitement au bisulfite (Miura et al., 2012), permettant l'utilisation de quantités beaucoup plus réduites d'ADN de départ. Depuis, ce protocole a même été adapté pour pouvoir analyser la méthylation de l'ADN sur cellule unique (Smallwood et al., 2014).

- **Technologies dérivées du séquençage bisulfite**

Le désavantage majeur des technologies utilisant la conversion au bisulfite de sodium est qu'elles ne permettent pas de distinguer les différentes formes oxydées de la 5mC, en particulier la méthylation (5mC) de l'hydroxyméthylation (5hmC) de l'ADN. En effet, les cytosines hydroxyméthylées sont résistantes à la conversion au bisulfite de sodium au même titre que les cytosines méthylées, alors que les bases 5fC et 5caC sont efficacement déaminées par le bisulfite de sodium et séquençées comme un T au même titre que la base C. Ceci a conduit la communauté à réfléchir au développement de nouvelles technologies permettant de quantifier

spécifiquement la présence de plusieurs formes modifiées de la cytosine. Ces dernières années, de nombreuses technologies ont été développées pour cartographier la présence de 5mC, 5hmC, 5caC et 5fC dans le génome. En voici quelques exemples :

- **L'oxBS-seq** (Oxidative Bisulfite Sequencing) permet d'analyser la présence des cytosines hydroxylées (5hmC). Les 5hmC sont spécifiquement oxydées par traitement chimique puis converties au bisulfite de sodium. Ce traitement ne modifie donc pas les 5mC. Le séquençage en parallèle d'échantillons traités en BS et oxBS permet de déterminer par soustraction la localisation des modifications 5hmC et 5mC à la résolution du nucléotide (Booth et al., 2012).

- **Le TAB-seq** (TET-assisted Bisulfite Sequencing) consiste à modifier chimiquement les 5hmC par glycosylation pour les protéger d'une oxidation in vitro par les protéines TET. Les 5mC et cytosines sont ensuite converties au bisulfite de sodium en thymines, permettant d'identifier spécifiquement où sont localisées les 5hmC (Yu et al., 2012).

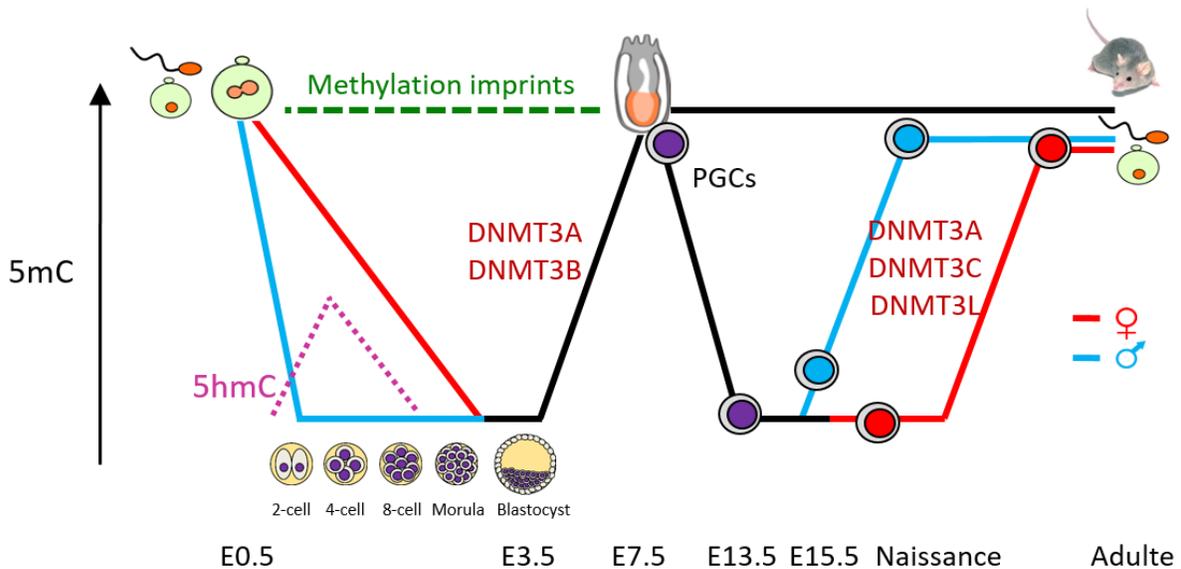
- **Le MAB-seq** (M.SssI methylase-assisted bisulfite sequencing) qui permet directement de cartographier la présence de 5fC/5caC dans le génome (Wu et al., 2014). Cette technologie consiste à méthyler artificiellement les C des CpG avec l'enzyme M.SssI puis à convertir l'ADN avec le bisulfite de sodium qui ne va donc déaminer que les bases 5fC et 5caC qui ont été protégées de l'action de M.SssI. Le séquençage haut débit permettra donc de révéler directement la présence des bases 5fC et 5caC en T dans le génome, sans pouvoir cependant les distinguer entre elles.

- **Le CLEVER-Seq** (chemical-labeling-enabled C-to-T conversion sequencing) qui permet de cartographier la 5fC à la résolution du nucléotide par séquençage haut débit sans conversion au bisulfite (Zhu et al., 2017).

## 5. Rôle de la méthylation de l'ADN au cours du développement

Au début de cette introduction, j'ai décrit la méthylation de l'ADN, les acteurs moléculaires qui la contrôlent ainsi que les techniques permettant d'étudier sa distribution génomique. Cette méthylation n'est pas présente chez tous les organismes. Alors pourquoi ce mécanisme est-il préservé chez certains eucaryotes ? Quels événements développementaux ou cellulaires nécessitent une régulation par la méthylation de l'ADN ?

L'intérêt des chercheurs pour la 5mC a permis de montrer que sa fonction est très dépendante des organismes d'études. Par exemple, chez les champignons, la 5mC a été décrite comme étant impliqué dans la répression des éléments répétés (Selker, 2002). Chez les plantes, elle est nécessaire à la répression stable des gènes répétés et des rétrotransposons (Martienssen and Colot, 2001). Enfin, chez les mammifères, elle est aussi nécessaire à la répression stable des



**Figure 13 : Reprogrammation de la méthylation de l'ADN au cours du développement murin**  
 Après la fécondation, les génomes parentaux se démethylent rapidement jusqu'à atteindre une hypométhylation globale dans le blastocyste. La déméthylation paternelle est associée à une augmentation transitoire du taux de 5-hydroxyméthylcytosine qui disparaît à nouveau dans le blastocyste. Durant cette vague de déméthylation, certaines régions répétées ainsi que les régions soumises à l'empreinte génomique résistent à la vague de déméthylation. Après l'implantation de l'embryon (E3,5), la méthylation de l'ADN reprend place sur tout le génome grâce aux enzymes DNMT3A et DNMT3B. Le génome restera ensuite méthylé dans tous les tissus adultes. Seules les PGCs (Primordial Germ Cells) contenues dans l'épiblaste subissent une nouvelle vague de déméthylation entre E8.5 et E13.5. Ensuite les cellules germinales réacquièrent la méthylation dans l'embryon pour les gamètes mâles (en bleu) et après la naissance pour les gamètes femelles (en rouge). L'acquisition de la méthylation dans les gamètes dépend surtout de DNMT3A, DNMT3C et du cofacteur DNMT3L.

rétrotransposons mais elles possèdent d'autres fonctions indispensables à la survie de l'individu, telles que la régulation de l'expression des gènes au cours du développement ou ceux soumis à l'empreinte parentale, et la répression d'un des deux chromosomes X chez la femelle (Walsh et al., 1998). Enfin, chez le Xénope par exemple, l'inactivation du gène *Dnmt1* réactive l'expression des gènes zygotiques conduisant à une létalité précoce au cours du développement (Stancheva and Meehan, 2000). En somme, il semble que la fonction ancestrale de la méthylation de l'ADN soit de protéger l'organisme contre l'activité des transposons et ainsi d'assurer la stabilité génomique, mais que la méthylation de l'ADN a été détournée au cours de l'évolution pour réaliser d'autres fonctions de régulation chez certains organismes. Mais il reste encore beaucoup à découvrir quant aux rôles exacts de la méthylation de l'ADN, les séquences génomiques qu'elle cible ainsi à la mise en place de cette dernière.

#### **a. Régulation de la pluripotence**

La méthylation de l'ADN est relativement stable dans les tissus somatiques (Hon et al., 2013), cependant les profils de méthylation du génome sont considérablement reprogrammés au cours du développement en relation avec la perte et l'acquisition de la pluripotence (Figure 13). L'une des deux vagues de reprogrammation épigénétique a lieu lors du développement embryonnaire précoce, juste après la fécondation et jusqu'au stade blastocyste (E3,5 chez la souris) (Dean et al., 2001). Une seconde vague de reprogrammation a lieu dans les cellules germinales primordiales (PGCs : Primordial Germ Cells) entre E7,5 et 17,5 (Hemberger et al., 2009; Sasaki and Matsui, 2008; Seisenberger et al., 2013). Ces deux phénomènes permettent un remodelage de l'état épigénétique de l'ADN et de la chromatine, et de ce fait, la répression ou l'expression de certains gènes impliqué dans la pluripotence et le développement embryonnaire. Dans ce paragraphe, je discute les vagues de reprogrammation épigénétiques qui ont lieu dans le développement préimplantatoire et les cellules germinales.

#### **La vague de reprogrammation dans l'embryogenèse précoce**

Tandis que les spermatozoïdes possèdent un génome fortement méthylés, entre 80 et 90% (Kobayashi et al., 2012; Oswald et al., 2000; Smith et al., 2012), le génome de l'ovocyte est quant à lui méthylé à environ 45% (Howlett and Reik, 1991; Smallwood et al., 2011). Quelques heures après la fécondation, les deux génomes des pronoyaux mâle et femelle subissent une déméthylation globale (Santos et al., 2002). Le génome paternel est rapidement déméthylé dans le zygote précoce (Wang et al., 2014b). L'enzyme TET3 est grandement enrichie dans le pronoyau paternel et convertit un bon nombre des bases 5mC en 5hmC, ce qui favoriserait la déméthylation globale du génome (Gu et al., 2011b). Nous ne savons pas encore exactement comment cette

déméthylation a lieu, mais les études suggèrent la présence à la fois des mécanismes actifs et passifs de la déméthylation, avec notamment l'implication du système BER (Hajkova et al., 2010; Santos et al., 2013). L'absence de TET3 n'induit qu'une faible augmentation de la méthylation, de l'ordre de 10%, suggérant que d'autres mécanismes comme une dilution passive de la 5mC pendant la réplication participent à la déméthylation du génome paternel (Guo et al., 2014a; Peat et al., 2014).

Le génome maternel semble résistant à la déméthylation rapide dans le zygote et subit une déméthylation progressive (Guo et al., 2014a; Wang et al., 2014b). Il a été proposé que cela s'explique en partie par l'action de la protéine STELLA qui protège le génome maternel de l'hydroxyméthylation (Nakamura et al., 2007, 2012). La déméthylation progressive pourrait être causée par l'absence de maintien de la méthylation pendant les premiers cycles de divisions cellulaires, probablement dû à l'exclusion de DNMT1 du noyau (Hirasawa et al., 2008; Howell et al., 2001). A partir du stade deux cellules jusqu'au stade blastocyste, la 5mC est diluée passivement au cours des divisions successives et atteint un niveau global de méthylation d'environ 20% dans le blastocyste (**Figure 13**). A cette étape, l'ensemble du génome est hypométhylé.

Les profils de méthylation spécifiques des cellules différenciées sont ensuite progressivement établis par les enzymes DNMT3 après l'implantation de l'embryon, et finalement maintenus lors des divisions cellulaires de l'embryon par l'enzyme DNMT1. En effet, le travail de thèse de l'ancien doctorant de mon laboratoire, Ghislain Auclair (Auclair et al., 2014), a démontré par RRBS que le blastocyste se reméthylye globalement à hauteur de 50% le premier jour après l'implantation entre E4,5 et E5,5. Par la suite, le niveau de méthylation continue à augmenter progressivement jusqu'à E7,5 (**Figure 13**). Pendant cette période la grande majorité des 25,000 îlots CpG du génome de souris sont protégés de la méthylation de l'ADN. Les ~700 îlots CpG qui se retrouvent reméthylysés sont associés aux gènes développementaux ainsi qu'aux gènes gamétiques (Auclair et al., 2014; Borgel et al., 2010). Il a aussi mis en exergue que cette reméthylation requière préférentiellement l'enzyme DNMT3B, puisque l'absence de DNMT3A en comparaison avec celle de DNMT3B n'a que des conséquences limitées sur l'établissement de la 5mC. Il est probable que la protéine DNMT3L n'a pas de rôle important dans l'établissement de la méthylation du génome dans l'embryon, cependant certaines études n'écartent pas son implication (Arand et al., 2012; Guenatri et al., 2013; Ooi et al., 2010).

Cette déméthylation globale du génome après la fécondation jouerait un rôle dans la ré-acquisition d'un état épigénétique compatible avec la pluripotence dans l'embryon précoce. En effet, certains gènes clefs de la pluripotence tels que *Oct4* et *Nanog* sont méthylés dans les gamètes matures. Par conséquent, ils ont besoin d'être épigénétiquement remaniés afin de

permettre leur réactivation transcriptionnelle dans les embryons préimplantatoires (Farthing et al., 2008). Cependant les cellules ES hyperméthylées restent pluripotentes, tout comme l'embryon du poisson zèbre qui reste hyperméthylé tout au long de son développement. En effet chez le poisson zèbre le méthylome du génome paternel reste inchangé, démontrant que l'hypométhylation n'est pas obligatoirement requise pour acquérir un stade dit pluripotent en fonction des espèces (Jiang et al., 2013; Potok et al., 2013). Une autre hypothèse serait que cette déméthylation sert de garde-fou dans l'éventualité de défauts de méthylation hérités des gamètes (Heard and Martienssen, 2014; Seisenberger et al., 2013).

Néanmoins, certaines régions résistent toute même à cette vague de déméthylation telle que les régions de contrôle de l'empreinte (ICRs, Imprinted Controlled Regions) dont je parlerai dans le prochain sous chapitre, et des éléments transposables tels que les LINES, LTR et IAP (Lane et al., 2003; Smith et al., 2012) (que je discuterai dans le sous chapitre I\_5\_d). Les mécanismes permettant de maintenir ces régions méthylées sont encore mal connus. Toutefois, il semblerait que la protéine Stella (PGC7) joue un rôle dans cette protection. Cette protéine est capable d'interagir avec la modification d'histone H3K9me2 et protégerait l'ADN méthylé contre l'oxydation induite par les enzymes TET (Nakamura et al., 2012). Une autre protéine, ZFP57, qui reconnaît l'ADN par un doigt de zinc et possède un domaine de répression KRAB, est également requise pour maintenir la méthylation des ICR durant ces stades préimplantatoires (Li et al., 2008; Quenneville et al., 2011).

### **La phase de reprogrammation germinales**

Dans ce paragraphe, je n'aborderai que les grandes lignes étant donné que l'un de mes projets de thèse consiste en l'étude de cette reprogrammation. Ainsi, j'expliquerai de manière plus approfondie ce sujet dans la phase d'introduction de ma première partie des résultats.

Chez la souris, les PGC proviennent de cellules précurseurs de l'épiblaste à E7,25 (Ginsburg et al., 1990). Tout d'abord localisés au niveau de l'allantoïde chez l'embryon à E8, les PGC migrent jusqu'à rejoindre les gonades en développement à E11,5 (Anderson et al., 2000). A partir de ce stade, les cellules vont nicher dans les gonades et suivre leur programme développemental. Les cellules de l'épiblaste ayant déjà acquis leur programme développemental somatique, elles présentent de ce fait une méthylation globale du génome y compris sur les promoteurs à îlots CpG des gènes gamétiques. Ainsi, il est nécessaire de reprogrammer les cellules germinales afin qu'elles réacquièrent leur caractère pluripotent et puissent activer le programme gamétique.

La déméthylation des PGC se déroule graduellement entre E9.5 et E12.5. Il a été démontré que la maintenance de la méthylation de l'ADN n'est plus garantie dû à une exclusion de UHRF1 du noyau et l'absence de recrutement de DNMT1 aux sites de réplication de l'ADN (Kagiwada et al., 2013; Seisenberger et al., 2012). Ainsi, le niveau global de méthylation du génome passerait de 70 % dans l'épiblaste à E6,5 à 20–30 % dans les PGC à 10,5 (Kobayashi et al., 2013; Seisenberger et al., 2012) (**Figure 13**). Le niveau global de méthylation du génome des PGC atteint un minimum de 5 à 7 % à E13,5, ou leur génome est globalement déméthylé. Parallèlement à ce qui se passe dans le blastocyste, certaines séquences, en particulier des éléments transposables, résistent partiellement à cette vague de déméthylation globale dans les PGCs (Guibert et al., 2012). Il a été démontré que cette déméthylation globale au niveau du génome des cellules germinales est aussi conservée chez l'humain (Eguizabal et al., 2016; Guo et al., 2015a; Tang et al., 2015). Ceci indique que la reprogrammation des cellules germinales est une des caractéristiques générales du développement des mammifères.

Dans les cellules germinales mâles, cette hypométhylation globale est un état très succinct au cours du développement murin. La méthylation *de novo* du génome prend place immédiatement à E14,5 dans les gamètes mâles (**Figure 13**). Les PGC mâles se retrouvent en arrêt mitotique peu de temps après E13,5, et ne se divisent plus jusqu'à ce qu'elles deviennent des spermatogonies, c'est-à-dire quelques jours après la naissance entre 3 et 4 jours. L'inactivation conditionnelle de *Dnmt3b* dans les cellules gamétiques mâles ne semble pas induire de défaut développemental, au contraire de l'absence de DNMT3A ou DNMT3c qui induisent une baisse de méthylation dans les spermatogonies, une réactivation des éléments transposables, un arrêt méiotique et l'apoptose des cellules germinales (Barau et al., 2016; Bourc'his and Bestor, 2004; Kaneda et al., 2004). DNMT3L est également absolument nécessaire à la reméthylation des PGCs puisque l'inactivation de *Dnmt3L* empêche l'acquisition de la méthylation et le développement normal des spermatozoïdes (Bourc'his and Bestor, 2004). Ces données indiquent donc que les trois protéines DNMT3A, DNMT3C et DNMT3L coopèrent pour établir le méthylome des gamètes mâles.

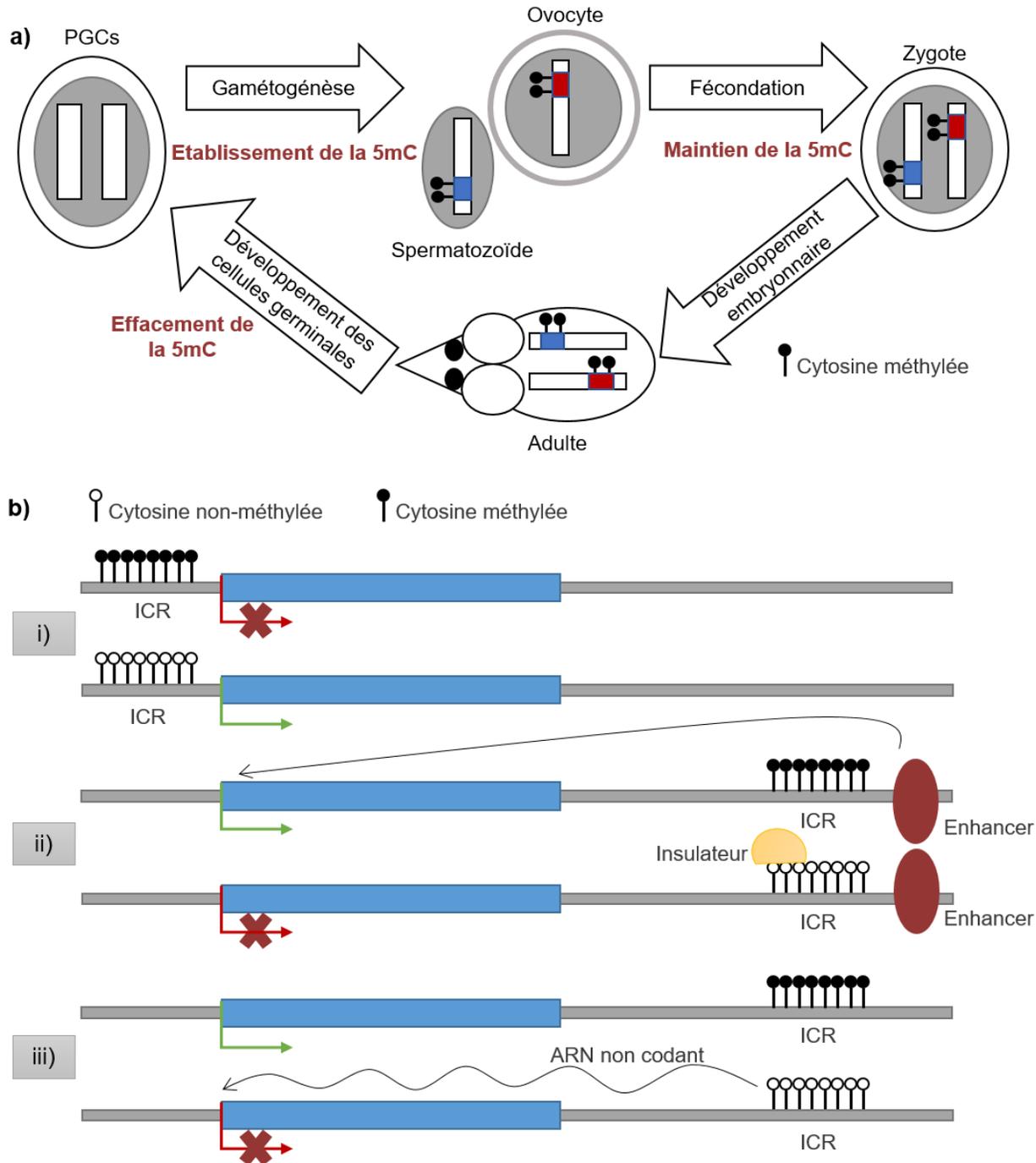
A E12,5, les PGCs femelles prennent une trajectoire développemental différente des gamètes mâles. Elles rentrent en prophase de méiose I à E13,5 et s'arrêtent à la fin de la métaphase I, peu de temps avant la naissance. Jusqu'à la puberté, les ovocytes resteront bloqués à ce stade de méiose (Ewen and Koopman, 2010) et resteront déméthylés pendant toute cette période (**Figure 13**). La reméthylation des ovocytes intervient seulement lors de la maturation de ces derniers et atteint un niveau de méthylation d'environ 40 % (Kobayashi et al., 2012; Smallwood et al., 2011). Le niveau de méthylation du génome des ovocytes corrèle avec la

transcription, et la plupart des 5mC sont localisées au niveau des corps de gènes. À l'inverse des spermatozoïdes où la majorité du génome est méthylé à l'exception des îlots CpG, ce sont les régions intergéniques qui sont globalement hypométhylées chez les ovocytes. D'autre part, les îlots CpG localisés dans les unités de transcription sont fréquemment méthylés dans les ovocytes, et c'est d'ailleurs ce mécanisme de méthylation des îlots CpG dépendant de la transcription qui est responsable de la mise en place des empreintes maternelles de méthylation au niveau des ovocytes (Chotalia et al., 2009; Veselovska et al., 2015). Cette méthylation est totalement dépendante de DNMT3A et DNMT3L (Shirane et al., 2013; Smallwood et al., 2011). En absence de DNMT3A ou DNMT3L, le développement à terme des ovocytes n'est pas affecté mais les embryons issus de ces ovocytes ne peuvent pas supporter un développement embryonnaire complet à cause des défauts de méthylation au niveau des régions de contrôle de l'empreinte parentale (Bourc'his et al., 2001; Kaneda et al., 2004). À l'inverse, la délétion des gènes *Dnmt3b* ou *Dnmt1* dans les ovocytes n'a que peu d'effet sur le méthylome des ovocytes (Shirane et al., 2013).

#### ***b. Les régions soumises à empreintes parentales (ICRs)***

Au début des années 1980, des expériences sur la transplantation nucléaire chez les souris à partir de pronoyaux a mis en évidence que le développement d'embryons de souris ne peut se faire à partir uniquement d'un génome diploïde paternel ou maternel (Barton et al., 1984; McGrath and Solter, 1984; Surani et al., 1984). Ces expériences ont abouti à des défauts développementaux des embryons, et ont permis de souligner l'importance de chacun des deux génomes, paternel et maternel, de façon non équivalente, pour le développement embryonnaire. Ces découvertes ont conduit à la découverte de l'existence de gènes soumis à empreinte génomique parentale. Ainsi, on dit d'un gène qu'il est soumis à empreinte génomique parentale lorsque, chez les organismes diploïdes, la copie héritée de la mère et la copie héritée du père ne sont pas exprimées de la même manière. En règle générale, l'une des deux copies du gène est totalement éteinte alors que l'autre est active.

C'est au début des années 90 que les premiers gènes soumis à empreinte parentale ont été découverts : *Igf2r*, *Igf2* et *H19* (Barlow et al., 1991; Bartolomei et al., 1991; DeChiara et al., 1991). On compte actuellement chez la souris environ 150 gènes soumis à empreinte présentant une expression mono-allélique strict en fonction de l'origine parentale, et la majorité d'entre eux sont également soumis à empreinte chez l'Homme. Environ 80 % des gènes soumis à une empreinte parentale sont regroupés en cluster de plusieurs centaines de kilobases, suggérant une régulation commune de l'ensemble de ces gènes. Chaque groupe est contrôlé par une séquence essentielle régulatrice appelée ICR (imprinting control region), contenant au moins une



**Figure 14: Régulation épigénétique de l’empreinte génomique parentale**

**a) Cycle de vie des régions de contrôle de l’empreinte parentale (ICRs).** Les ICR paternelles (bleu) et maternelles (rouge) acquièrent une méthylation différentielle pendant la gamétogénèse qui se transmet aux cellules de l’embryon. Cette méthylation est maintenue et contrôlée pendant l’embryogénèse et dans les tissus somatiques. Cependant, la méthylation des ICRs est effacée dans les PGCs juste avant la spécification sexuelle puis remis en place plus tard au cours du développement des cellules germinales. **b) Modèle de régulation par les ICRs.** i) Le promoteur du gène est réprimé par la 5mC. ii) La 5mC régule l’activité d’un insulateur bloquant l’accès à un enhancer. Iii) La 5mC régule la transcription d’un ARN non codant de grande taille qui induit une répression en *cis* des gènes du locus.

région différentiellement méthylée (DMR : differentially methylated region) entre les deux allèles. Cette méthylation différentielle est établie lors de la gamétogenèse dans les lignées germinales des parents, puis maintenue pendant la fusion des gamètes et durant le développement embryonnaire où elle va contrôler l'expression monoallélique des gènes (Arnaud, 2010) (Figure 14a). Il est intéressant de noter que selon les loci, les DMRs peuvent être méthylées soit sur l'allèle exprimé, soit sur l'allèle non exprimé, ce qui suggère une grande complexité des mécanismes de régulation.

La régulation des gènes soumis à empreinte par les ICRs peut se faire de différentes manières. Ainsi, si l'ICR se trouve directement au niveau du promoteur du gène, l'allèle méthylé est alors l'allèle réprimé. Dans le cas où l'ICR régule l'expression de plusieurs gènes au sein d'un grand cluster, il peut jouer la fonction d'insulateur ou de promoteur d'un ARN non codant qui réprime l'expression de plusieurs gènes éloignés en *cis* (Arnaud, 2010) (Figure 14b).

Bien que l'ensemble des mécanismes par lesquels les empreintes gamétiques sont établies reste encore peu compris, les études réalisées sur *Dnmt3L* prouvent l'importance cruciale de la méthylation de l'ADN dans la mise en place de l'empreinte génomique parentale (Bourc'his et al., 2001; Hata et al., 2002; Vlachogiannis et al., 2015). Comme je l'ai expliqué dans l'introduction, la protéine DNMT3L partage une forte homologie avec DNMT3A et DNMT3B, mais elle ne possède pas d'activité méthyltransférase. L'inactivation du gène *Dnmt3L* entraîne un « effet maternel » sur le développement embryonnaire. En effet, les embryons issus d'ovocytes *Dnmt3L*<sup>-/-</sup> présentent des défauts au niveau du tube neural et du placenta, conduisant à une mort embryonnaire à E10,5. L'analyse des profils de méthylation des femelles *Dnmt3L*<sup>-/-</sup> montrent qu'elles n'ont pas la capacité d'établir les empreintes de méthylation d'origine maternelle au niveau de leur ovocytes, ce qui induit la perte de l'expression mono-allélique des gènes soumis à empreinte maternelle dans la descendance. La même chose est observée avec des femelles *Dnmt3a*<sup>-/-</sup> (Kaneda et al., 2004). Ces études suggèrent que la famille des enzymes DNMT3 est cruciale pour la mise en place de l'empreinte génomique parentale dans les gamètes. D'autres études démontrent aussi l'implication de *Dnmt1* pour le maintien de l'empreinte parentale dans l'embryon. Chez les embryons déficients pour *Dnmt1*, les gènes *Igf2* et *Igf2r*, qui sont normalement exprimés à partir de l'allèle paternel et maternel respectivement, sont réprimés sur les deux allèles, alors que le gène *H19*, qui est normalement exprimé à partir de l'allèle maternel, se trouve transcrit de façon bi-allélique (Li et al., 1993). En résumé cela démontre que la méthylation de l'ADN est nécessaire pour initier l'empreinte parentale dans les gamètes et la maintenir au cours du développement embryonnaire.

De façon un peu plus globale, d'autres facteurs épigénétiques coopèrent avec la méthylation de l'ADN pour réguler l'empreinte parentale. Par exemple, l'un des prérequis à la mise en place de la 5mC est l'élimination des marques d'histone H3K4me2 par les enzymes KDM1A et KDM1B dans les ovocytes (Stewart et al., 2015). L'acquisition de la méthylation d'ADN sur les îlots CpG qui constituent les ICR dépend aussi de la transcription qui traverse les îlots CpG et de la marque H3K36me3 qui lui est associée. Si on prend comme exemple le locus *Gnas*, la mise en place de la méthylation au niveau de l'ICR *Gnas*-XL nécessite la transcription à partir du promoteur *Nesp* qui se localise en amont du locus *Gnas*. (Chotalia et al., 2009; Fröhlich et al., 2010; Williamson et al., 2011).

Les gènes soumis à empreinte sont impliqués dans divers processus biologiques tels que le développement embryonnaire, la formation du placenta, la croissance fœtale et post-natale ainsi qu'au niveau des comportements adultes. Chez l'humain, l'altération de l'expression des gènes soumis à empreinte est liée à plusieurs syndromes métaboliques tels que le syndrome de Prader-Willi ou le syndrome d'Angelman par exemple (Butler, 2009). La perte de l'empreinte, c'est-à-dire l'expression bi-allélique ou la répression de un ou plusieurs gènes soumis à empreinte, est aussi fréquemment observée dans les cancers (Jelinic and Shaw, 2007).

### ***c. L'inactivation du chromosome X***

Chez les femelles mammifères, les cellules possèdent deux chromosome X tandis que les mâles n'en possèdent qu'un seul exemplaire. Afin de normaliser l'expression du chromosome X entre les individus mâles et femelles, une des copies du chromosome X est donc inactivée chez les cellules femelles. Cette compensation de dosage chromosomique est réalisée par l'inactivation transcriptionnelle et permanente de l'un des deux chromosomes X chez la femelle (Lyon, 1961). Ceci est vitale pour la survie de l'organisme (Takagi and Abe, 1990).

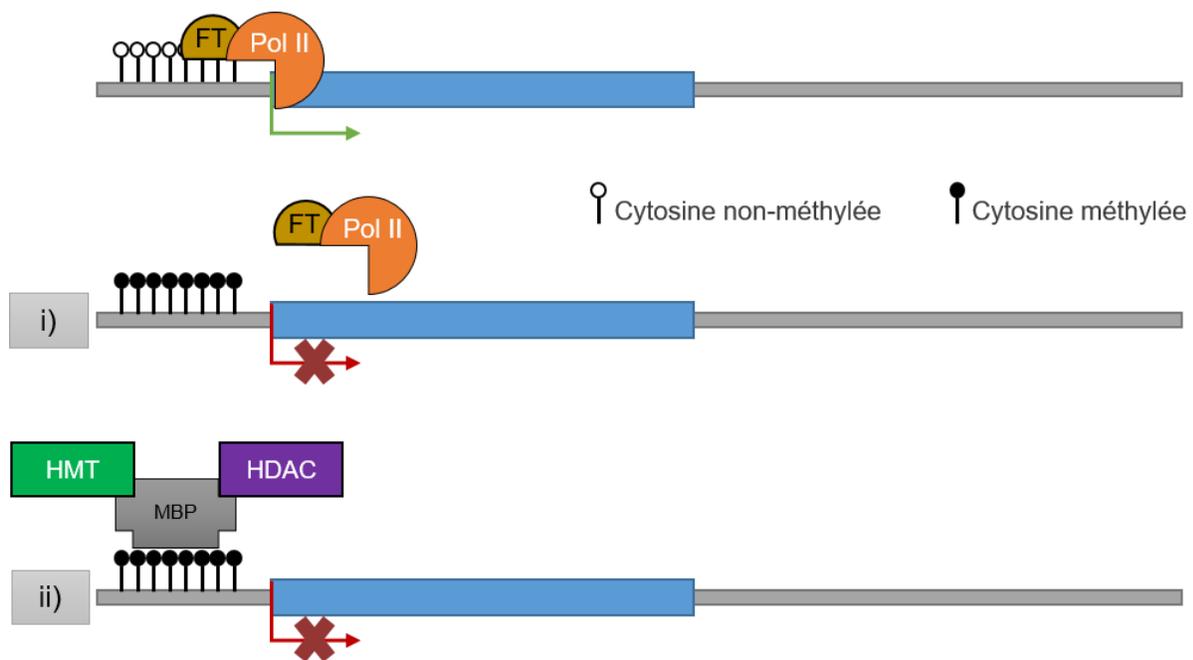
L'inactivation du chromosome X s'établit très tôt durant l'embryogénèse entre E3,5 et E8,5 et dépend de l'expression d'un ARN régulateur non codant appelé Xist (Augui et al., 2011). L'expression de Xist au niveau de l'un des deux chromosomes X induit une cascade d'événements épigénétiques en *cis* incluant le recrutement de protéines de la famille Polycomb, de modifications d'histones, du variant d'histones macro H2A et de la méthylation de l'ADN sur les îlots CpGs. Une fois inactivé, le chromosome X reste inactif dans toutes les cellules somatiques. Le rôle des DNMT sur la mise en place et/ou le maintien de la méthylation de l'ADN du chromosome X a été étudié. Ces études démontrent que DNMT3B est la seule enzyme responsable de la mise en place de la méthylation des îlots CPG sur le chromosome X inactif (Auclair et al., 2014; Gendrel et al., 2012). D'autre part, l'acquisition de cette méthylation au niveau des îlots CpGs est un événement tardif qui arrive après l'inactivation des gènes (Prissette et al.,

2001). De plus, l'initiation de l'inactivation du chromosome X a lieu normalement dans des embryons déficients pour les *de novo* méthyltransférases DNMT3A et DNMT3B (Sado et al., 2004). En revanche l'inhibition de la méthylation de l'ADN peut affecter le maintien de l'inactivation du chromosome X dans les cellules somatiques (Csankovszki et al., 2001; Sado et al., 2000). Collectivement ces données suggèrent que la méthylation de l'ADN participerait à la stabilité à long terme de la répression plutôt qu'à l'initiation du processus d'inactivation du chromosome X.

#### **d. L'inactivation des éléments transposables**

Une des fonctions majeures de la méthylation de l'ADN chez les eucaryotes est de contrôler l'activité des éléments mobiles et la stabilité génomique. Les génomes des eucaryotes supérieurs sont constitués d'une majorité de séquences d'ADN non codantes qui composent les introns, les séquences intergéniques et les séquences répétées (Figure 2). Certaines de ces séquences répétées, telles que les gènes des ARN ribosomiques ou les séquences répétées associées au centromère, jouent un rôle important dans l'organisme. Cependant, les autres séquences répétées telles que les éléments transposables, les séquences virales et les éléments parasites de l'ADN peuvent avoir des effets délétères sur la stabilité et l'intégrité du génome. C'est pourquoi la plupart des eucaryotes ont développé des mécanismes épigénétiques pour contrôler et limiter l'activité des transposons. Parmi ces mécanismes, la méthylation de l'ADN a acquis au cours de l'évolution un rôle majeur de défense contre ces éléments répétés mobiles.

L'idée que la méthylation de l'ADN soit le principal mécanisme de défense de l'intégrité du génome nous vient majoritairement des analyses issues des génomes des plantes et des champignons, où les éléments mobiles sont les cibles privilégiées de la méthylation de l'ADN. En effet, chez les plantes par exemple, la méthylation des éléments transposables limite l'expression de ces éléments mobiles et empêche leur propagation au sein du génome (Yoder et al., 1997b). Chez les mammifères, plusieurs études prouvent que l'absence de la méthylation de l'ADN induit une réactivation transcriptionnelle des éléments mobiles, même s'il reste à démontrer leur mobilité. Ainsi, les embryons déficients pour *Dnmt1* et *Uhrf1* montrent une élévation importante de l'expression d'éléments mobiles tels que les IAPs (Sharif et al., 2007), tout comme les gamètes déficients pour la protéine DNMT3L (Bourc'his and Bestor, 2004). Dans le génome humain, les SINEs et les LINEs sont réactivés lorsque la méthylation de l'ADN est réduite (Liu et al., 1994; Woodcock et al., 1997). Par ailleurs, il a aussi été proposé que la méthylation des séquences répétées limite également leur potentiel de recombinaison par homologie de séquences, favorisant ainsi la stabilité du génome (Maloisel and Rossignol, 1998). En effet, le traitement des cellules mammifères avec un inhibiteur de méthyltransférases, la 5-AZA-2'-deoxycytidine (azaCdR), provoque souvent des translocations chromosomiques.



**Figure 15 : La méthylation de l'ADN, un répresseur de l'expression génique**

L'îlot CpG non méthylé est accessible aux facteurs de transcription (FT) et à la polymérase (Pol II). Lorsque l'îlot CpG est méthylé, ce dernier perd sa capacité à lier les facteurs de transcription (i). Il peut aussi devenir la cible des methyl-binding proteins (MBP) qui lient l'ADN méthylé et recrutent à leur tour les histones déacétylases (HDAC) et les histones méthyltransférases (HMT) qui stimulent la formation d'un état chromatinien inactif (ii).

Bien que les transposons soient réprimés dans les tissus différenciés, ces derniers se retrouvent exprimés de manière très contrôlée dans l'embryon précoce et contribuent à la régulation du génome (Fadloun et al., 2013). Les TEs qui contribuent le plus au développement embryonnaire sont les ERVs de Classe III (MaIR et MERVL) et représentent les plus anciens ERVs du génome mammifère. Par exemple, l'expression de MaIR est particulièrement élevée dans l'ovocyte non fécondée et représentent 13% du transcriptome (Peaston et al., 2004). Quant aux MERVLs, ce sont les premiers à être transcrit lors de l'activation du génome zygotique (Kigami et al., 2003; Macfarlan et al., 2012). De plus, leur expression influence celle d'environ 300 gènes de par la formation de transcrits chimériques entre les exons et MERVL LTR. La plupart de ces gènes sont spécifiques du stade embryonnaire 2 cellules, suggérant que leur activation est importante pour réguler le développement embryonnaire. Ainsi, l'expression des MERVL est considérée comme une marque de totipotence. A l'inverse, les IAPs possèdent un pic d'expression au niveau du stade blastocyste et sont rapidement réprimés après ce stade (Peaston et al., 2004), mais peu de choses sont connus quant à leur implication dans la régulation du transcriptome.

Enfin, tant chez l'humain que chez la souris, environ 5 à 20% des sites d'accrochages des protéines NANOG et OCT4 sont localisés au niveau des éléments transposables (Kunarso et al., 2010). Durant les phases de reprogrammation dont nous avons discuté plus haut, ces sites servent probablement d'unité de régulation afin de définir au mieux l'expression du programme de pluripotence. Par exemple, il a été démontré chez l'humain que les rétrovirus endogènes primates aident par leur transcrits à mettre en place des cellules naïves proches des cellules ES (Wang et al., 2014a).

### ***e. Régulation génique***

Plusieurs études, dont la plupart chez les mammifères, ont mis en évidence le potentiel répressif de la méthylation de l'ADN sur l'expression des gènes. Il est aujourd'hui admis que la méthylation d'un promoteur à îlots CpG peut inhiber l'expression des gènes associés par plusieurs mécanismes. Elle peut directement réprimer la transcription en empêchant l'accès à l'ADN aux facteurs de transcription (Figure 15). L'ADN méthylé peut également induire le recrutement des methyl-binding proteins (MBPs), qui à leur tour vont recruter des complexes répresseur de modifications d'histones et de remodelage de la chromatine (Figure 15). Ce rôle répressif de la méthylation des cytosines a été démontré notamment dans le cas de l'inactivation du chromosome X et de la régulation des gènes soumis à empreintes.

Les avancées récentes de cartographie à grande échelle ont permis de mieux comprendre le rôle de la méthylation dans la régulation génique à l'échelle du génome. En effet, plusieurs

équipes ont cartographié la 5mC autour des sites d'initiation de la transcription (TSS), révélant ainsi que la plupart des promoteurs riches en CpG sont constitutivement hypométhylés, et ce, de manière indépendante à l'état d'activité des gènes qui lui sont associés. Cependant, 3-5 % des promoteurs riches en CpG sont méthylés dans les cellules somatiques (Illingworth et al., 2008; Weber et al., 2007). Une autre conclusion de ces travaux est que les promoteurs se comportent différemment face à la méthylation de l'ADN en fonction de la richesse en CpG. En effet, au niveau des promoteurs riches en CpG, le taux de méthylation corrèle très bien avec le taux de répression, tandis que pour les promoteurs pauvres en CpG, la méthylation peut aussi bien être associée à des gènes réprimés qu'exprimés (Weber et al., 2007).

Afin de comprendre la fonction de cette méthylation des promoteurs, la nature des gènes cible est très informative. Les cartographies du méthylome réalisées par mon équipe chez la souris montre que la méthylation des promoteurs à CGI cible préférentiellement les gènes gamétiques au cours de l'embryogenèse (Auclair et al., 2014). De plus les gènes gamétiques sont préférentiellement sur-exprimés dans les embryons de souris *Dnmt3b*<sup>-/-</sup> (Auclair et al., 2014) ou les cellules ES triplement invalidées pour les trois gènes *Dnmt1*, *Dnmt3a* et *Dnmt3b* (Karimi et al., 2011). Cela démontre que la méthylation de l'ADN est nécessaire pour maintenir une répression stable du programme gamétique dans les cellules de l'embryon.

En outre, plusieurs équipes ont étudié la méthylation des promoteurs au cours de la différenciation des cellules souches embryonnaires (ES) (Farthing et al., 2008; Meissner et al., 2008; Mohn et al., 2008). Ces études révèlent que la méthylation est rare dans les cellules ES et augmente pendant la différenciation sur un petit nombre de promoteurs, en particulier des gènes de régulation de la pluripotence, suggérant un rôle de la méthylation de l'ADN comme un moyen de répression de la pluripotence dans les cellules différenciées.

Ces dernières années des cartographies exhaustives de méthylation de l'ADN ont été réalisées dans de nombreux tissus chez l'homme et la souris (Bogdanović et al., 2016; Hon et al., 2013; Ziller et al., 2013). Ces données montrent que les profils de méthylation sont très variables entre types cellulaires, cependant les variations sont rares dans les promoteurs de gènes mais plus fréquentes dans les "enhancers" et les corps de gènes. Cela suggère un rôle important de la méthylation de l'ADN dans le maintien dans une destinée cellulaire précise. Ceci a été suggéré dans diverses cellules somatiques tel que, par exemple, les neurones et les cellules gliales, ou encore les cellules musculaires (Bogdanović and Lister, 2017; Laker and Ryall, 2016; Lister et al., 2013). Cependant il reste à démontrer si ces changements de méthylation sont une cause ou plutôt dans certains cas une conséquence des changements d'expression de gènes. Ainsi beaucoup de choses restent encore à étudier pour comprendre le rôle de la méthylation de l'ADN

dans l'établissement ou le maintien des programmes d'expression de gènes au cours de l'organogenèse.

## **6. Les pathologies associées à la dérégulation de la méthylation de l'ADN**

Comme je l'ai expliqué dans les chapitres précédents, la méthylation de l'ADN joue un rôle crucial dans la stabilité du génome et la régulation du transcriptome (Smith and Meissner, 2013). En conséquence, un défaut ou une altération dans la mise en place des profils de la 5mC peut induire des défauts de régulation et l'apparition de pathologies dues à une dérégulation de l'expression de certains gènes. Les cartographies des profils de méthylation de l'ADN dans les cellules cancéreuses ou provenant d'individus atteints de certaines pathologies montrent clairement que ceux-ci sont fortement remaniés en condition pathologique par rapport à des cellules saines. Ces dernières années, un grand nombre de chercheurs a mis en évidence un lien entre les pathologies (y compris les cancers) et la dérégulation de la mise en place de la méthylation de l'ADN. Dans ce chapitre, je vais discuter du lien entre la méthylation de l'ADN et les pathologies humaines.

### **a. Les cancers**

Des changements drastiques au niveau de l'épigénome accompagnent l'initiation ainsi que la progression des cancers. Il est connu depuis longtemps que les cellules cancéreuses présentent une hypométhylation globale, tant au niveau des régions ayant une faible densité en CpG qu'au niveau des éléments répétés et des rétrotransposons. L'hypométhylation des CpG est l'une des premières modifications découverte dans de nombreux cancers dans les années 80 (Feinberg and Vogelstein, 1983; Gama-Sosa et al., 1983). Cette hypométhylation globale peut être détectée précocement lors de la tumorigenèse et dans les hyperplasies bénignes mais elle a principalement lieu durant la progression tumorale et dans les lésions métastatiques plutôt que dans les tumeurs primaires (Li et al., 2014).

Cette déméthylation touche principalement les régions inter- et intragéniques qui contiennent des séquences répétées ainsi que des éléments transposables. La perte de la 5mC au niveau de ces régions peut favoriser l'expression ectopique de gènes ainsi que la transposition de ces éléments. Par exemple, l'hypométhylation au niveau du promoteur d'un transposons LINE-1 stimule un remaniement chromatinien du promoteur alternatif de *Met*, induisant l'activation de cet oncogène dans les cancers de la vessie (Wolff et al., 2010). Un autre exemple est celui des SINEs, une autre classe d'éléments transposables, qui ont été observés comme hypométhylés dans des leucémies de type AML (Acute Myeloid leukemia) (Saied et al., 2012). Les cartographies plus précises de méthylation par WGBS dans les cellules cancéreuses ont confirmé l'existence

de large domaines d'hypométhylation, pouvant faire plus de 100 kb, dans les cellules de cancer du sein ou du colon par rapport aux cellules normales (Berman et al., 2011; Hon et al., 2012).

Les études à échelle génomique ont démontré que les corps de gènes sont largement méthylés dans les cellules normales de mammifères. Alors que la 5mC inhibe l'initiation de la transcription, des études ont suggéré un rôle possible de la méthylation de l'ADN dans l'élongation de la transcription, la régulation de l'épissage ou la régulation de l'utilisation de promoteurs alternatifs (Kulis et al., 2013; Lou et al., 2014). De ce fait, la perte de la 5mC au niveau de ces régions pourrait favoriser le démasquage d'éléments régulateurs du génome tel que des enhancers, des exons alternatifs ou des promoteurs alternatifs. Une étude portant sur l'analyse à grande échelle du méthylome des cellules B d'individus sains ou atteints d'une leucémie chronique myéloïde a révélé une forte hypométhylation au niveau des corps de gènes, favorisant l'accessibilité d'enhancers (Kulis et al., 2012). L'hypométhylation d'enhancers intergénomiques et intragénomiques peut libérer des motifs spécifiques à des facteurs de transcriptions, et par conséquent induire des changements drastiques du transcriptome qui favorise le développement cancéreux (Aran and Hellman, 2013; Kulis et al., 2013).

En parallèle d'une hypométhylation à grande échelle, les cellules cancéreuses sont caractérisées par une hyperméthylation au niveau de nombreux promoteurs riches en CpG, certains contrôlant l'expression de gènes suppresseurs de tumeurs (Jones and Laird, 1999; You and Jones, 2012) et induisant de ce fait la répression de l'expression de ces gènes. La répression de gènes régulateurs du cycle cellulaire et de la réparation de l'ADN par une hyperméthylation est observée dans de nombreux types de cancer (Alvarez-Nuñez et al., 2006; Chiang et al., 2006; Costello et al., 1996; Sakai et al., 1991). Aujourd'hui il est unanimement considéré que l'inactivation épigénétique de gènes suppresseurs de tumeurs dans les cancers et au moins aussi fréquente que l'inactivation par mutations génétiques. La répression des gènes de réparation de l'ADN contribue hautement à l'instabilité génétique et à l'apparition de mutations génétiques. Par exemple, le gène codant l'enzyme O6-Méthylguanine-DNA méthyltransférase (MGMT), une enzyme de réparation de l'ADN responsable du nettoyage des groupements acétyles fixés à l'ADN, est fréquemment hyperméthylé dans de nombreux types de cancers. La suppression de cette protéine, dû à l'hyperméthylation de son promoteurs, induit une augmentation non négligeable de risque de mutations génétiques au niveau de gènes essentiels tels que *p53* et *Kras* (Silber et al., 2012). Par conséquent, *Mgmt* est devenu l'un des premiers biomarqueurs épigénétiques dans les cas de cancer.

L'hyperméthylation des îlots CpG peut aussi contribuer à la perte de régulation de l'expression des gènes soumis à empreintes. Lorsque le locus *IGF2/H19* possède une méthylation

aberrante, l'expression du facteur de croissance IGF2 se retrouve augmentée (Kaneda and Feinberg, 2005; Ravenel et al., 2001), ce qui peut contribuer au développement et à la progression des cancers, tels que les cancers colorectaux et gastriques.

En plus de sa fonction au niveau des promoteurs de gènes, la méthylation joue aussi un rôle très important dans la régulation des ARN non codants. Ces éléments sont des régulateurs cruciaux des processus cellulaires que ce soit au niveau de la prolifération, de la différenciation, et du développement (Esteller, 2011). Ainsi, une hyperméthylation aberrante peut résulter en une dérégulation des micros ARN et contribuer au développement cancéreux (Formosa et al., 2013; Lujambio et al., 2007).

En conclusion, l'épigénome et les processus cancéreux sont étroitement liés, ce qui conduit beaucoup de chercheurs à étudier l'utilisation possible d'inhibiteurs de méthylation en thérapie du cancer et à identifier de possibles biomarqueurs épigénétiques afin de mieux diagnostiquer et prévoir l'évolution tumorale et la réponse aux traitements. Pour renforcer davantage le rôle de la méthylation de l'ADN dans les cancers, les gènes *DNMT* ou *TET* sont fréquemment mutés dans certaines formes de cancers. Par exemple, le gène *Dnmt3a* est connu pour être muté et avoir un rôle important dans le processus de cancérisation lié à certaines formes de leucémies (Brunetti et al., 2017). Ces pathologies causées par des mutations dans les gènes *DNMT* ou *TET* sont décrites dans les paragraphes suivant.

### ***b. Les pathologies associées aux mutations des gènes TETs***

- ***TET1***

Le gène *TET1* a été le premier gène identifié comme étant fusionné au gène *MLL* (Mixed lineage leukemia) dans les cas de leucémies liées à la translocation des protéines MLL (Lorsbach et al., 2003; Ono et al., 2002). Il s'agit de la partie C-terminale du gène *TET1* qui fusionne avec la partie N-terminale du gène *MLL*. Le rôle de *TET1* dans les processus tumoraux reste encore très ambigu. Plusieurs études démontrent que le niveau de 5hmC diminue drastiquement et que la protéine TET1 est fréquemment sous-exprimée ou encore exclue du noyau dans les tumeurs solides (Jin et al., 2011; Müller et al., 2012). Par contraste, l'étude de Huang et al. a démontré que *TET1* est la cible directe du gène *MLL* et que son expression est significativement surrégulée dans les cas de leucémies liées aux réarrangements des protéines de MLL, conduisant à une augmentation de niveau de 5hmC (Huang et al., 2013). La même étude a démontré que TET1 joue un rôle essentiel dans le développement des leucémies liées au réarrangement de *MLL*, en coopérant avec les protéines MLL dans la régulation des gènes critiques du leucémogénèse tel que *HOXA9*, *MEIS1* et *PBX3* (Huang et al., 2013). Cette étude révèle ainsi la complexité des

événements épigénétiques de conversion de la 5mC en 5hmC et son impact sur le développement tumoral.

- TET2

Le gène *TET2* est fréquemment muté dans les tumeurs hématologiques malignes. Des mutations somatiques de ce gène sont observés à hauteur de 7 à 23% des cas d'AML (Acute Myeloid Leukemia) et dans 10 à 20% des cas de MPN (myeloproliferative neoplasms) et MDS (myelodysplastic syndrome) (Shih et al., 2012). Des mutations de *TET2* apparaissent aussi dans les lymphomes, avec une très forte fréquence (12-19%) dans les lymphomes spécifiques des cellules T (Couronne et al., 2012; Quivoron et al., 2011). Les mutations du gène *TET2* associées aux tumeurs myéloïdes occasionnent un dysfonctionnement de la protéine, et de ce fait, une diminution du niveau de 5hmC (Ko et al., 2010). Ces observations suggèrent que l'haplo insuffisance liée à *TET2* est le mécanisme majeur par lequel la mutation de *TET2* contribue à la leucémogénèse. En corrélation avec cette notion, les délétions hétérozygotes du gène *Tet2* chez la souris sont suffisantes pour induire un phénotype hématologique chez les animaux adultes (Zhao and Chen, 2013). Les mutations de *TET2* au niveau de son site catalytique semblent corrélés avec une hyperméthylation des CpG dans les pathologies myéloïdes. De nombreuses études ont démontré une hyperméthylation de promoteurs conduisant à une répression de l'expression de gènes régulant la maturation des cellules hématopoïétiques et leur différenciation. Des recherches sur la déficience du gène *TET2* menées chez les souris démontrent une expansion des HSC et une myéloprolifération, qui peuvent éventuellement conduire à des tumeurs myéloïdes. L'acquisition de mutation au niveau de *TET2* est considérée comme un événement précoce de la leucémogénèse (Quivoron et al., 2011).

- TET3

A ce jour, aucune mutation somatique de *TET3* a été enregistré dans la base de données COSMIC (Catalogue of Somatic Mutations in Cancer). Néanmoins, des études portant sur le cancer du côlon ont identifiées plusieurs mutations potentielles non-sens du gène *TET3*. Cependant, la probabilité de mutation de *TET3* dans les cancers du côlon est très basse. L'expression de *TET3* apparaît comme étant plus restreint au niveau des tissus adultes que pour les protéines *TET1* et *TET2* (Lorsbach et al., 2003). Des recherches menées chez la souris démontrent que la délétion hétérozygote de *TET3* ne conduit pas à des phénotypes visibles, alors que la délétion homozygote induit une létalité néonatale (Zhao and Chen, 2013).

### ***c. Les pathologies associées aux mutations des gènes DNMTs***

- *DNMT1*

Klein *et al.* (Klein *et al.*, 2011) ont identifié pour la première fois une mutation hétérozygote du gène *DNMT1* au niveau de l'exon 20 chez des patients atteints de la neuropathie HSAN de type IE (Hereditary sensory and autonomic neuropathy). Cette maladie autosomique dominante se traduit par une perte de la sensibilité à la douleur et aux températures dans les membres inférieurs, une perte de l'audition, des mutilations ulcéreuses ainsi qu'à une démence. Par la suite, Winkelmann *et al.* (Winkelmann *et al.*, 2012) ont découvert que des mutations de *DNMT1* étaient aussi présentes au niveau de l'exon 21 dans la maladie autosomique dominante de l'ataxie cérébrale. L'exon 20 et 21 de *DNMT1* code pour le domaine RFTS (replication foci-targeting sequence) qui régule l'accrochage de DNMT1 à la chromatine pendant la phase S et qui est responsable du maintien de cette association pendant la phase G2 et M. La mutation au niveau de ce domaine produit une protéine instable dont l'activité méthyltransférase se retrouve réduite et qui n'est plus capable de rester liée à la chromatine après la phase S, ce qui a pour conséquence une hypométhylation globale. *DNMT1* a aussi été impliquée dans des pathologies neurodégénératives telles que la maladie d'Alzheimer (Di Francesco *et al.*, 2015; Mastroeni *et al.*, 2013).

- *DNMT3A*

La leucémie aigüe myéloïde (AML : acute myeloid leukemia) est une pathologie génétique très hétérogène qui se caractérise par une prolifération anormale des cellules myéloïdes immatures. Il y a quelques années, différents groupes de chercheurs ont identifiés de manière indépendante que le gène *DNMT3A* est fréquemment muté dans cette pathologie à hauteur de 12 à 35% (Ley *et al.*, 2010; Yamashita *et al.*, 2010; Yan *et al.*, 2011).

Des mutations au niveau de *DNMT3A* sont aussi retrouvées, à plus faible fréquence, dans d'autres maladies hématologiques telles que le syndrome myélodysplasique (MDS), les tumeurs myéloprolifératives (MPN), les leucémies lymphoblastiques liées aux cellules T (T-ALL : T-Cell acute lymphoblastic leukemia) et les lymphomes angioimmunoblastiques liées elles aussi aux cellules T (AITL : angioimmunoblastic T-cell lymphoma). De plus, il a été observé que des mutations de *DNMT3A* sont présentes très fréquemment au niveau des cellules souches hématopoïétiques pré-leucémiques de patients atteints d'AML, suggérant un rôle précoce et prédominant de *DNMT3A* dans le développement de la maladie (Shlush *et al.*, 2014).

La majorité des mutations sur *DNMT3A* sont des mutations non-sens affectant un seul amino-acide en C-ter qui altère l'Arginine 882 en histidine la plupart du temps. Cette mutation ne

touche souvent qu'un seul des deux allèles et n'empêche pas l'expression du gène. Des études menées chez la souris ont montré que cette mutation n'exclue pas la mise en place des complexes protéiques avec DNMT3A ou DNMT3B, mais elle va, de par son interaction, inhiber les fonctions des protéines associées (Kim et al., 2013; Russler-Germain et al., 2014). En outre, il semblerait que cette mutation induit une prolifération aberrante des cellules souches et progénitrices hématopoïétiques chez les souris (Xu et al., 2014).

Si on regarde les profils de méthylation des cellules leucémiques, ces dernières sont évidemment aberrantes, bien que les cellules d'AML possédant la mutation de *DNMT3A* ne présentent pas un changement drastique au niveau de leur épigénome. Cependant, on observe une hypométhylation à des endroits ciblés tels que les promoteurs des gènes HOX. Bien que les changements d'expression liés à la perte de la 5mC contribue à l'AML, elle n'est pas suffisante à elle seule au processus de leucémie.

- *DNMT3B*

Le syndrome de l'ICF (Immunodeficiency, centromeric region instability and facial anomalies) est une maladie rare autosomale récessive. Les patients présentent de bas niveau d'immunoglobulines dans le sérum malgré la présence de cellules B et T, et meurent très jeunes, souvent à cause d'infections bactériennes ou virales (Ehrlich, 2003). Environ 60% des cas d'ICF sont associés à des mutations non-sens, homozygotes ou hétérozygotes, au niveau du site catalytique du gène *Dnmt3b* (Hansen et al., 1999; Okano et al., 1999; Xu et al., 1999). Ces mutations induisent une diminution de l'activité enzymatique de DNMT3B puisqu'elles affectent son association avec SAM, les groupements méthyles, l'ADN et les autres DNMTs (Moarefi and Chédin, 2011).

#### ***d. Les autres pathologies associées à des défauts de méthylation de l'ADN***

La mutation du gène *ATRX* induit un changement au niveau des profils de méthylation de l'ADN ribosomiques, des éléments répétés télomériques et spécifiques du chromosome Y (Gibbons and Higgs, 2000). La pathologie qui en découle se nomme la thalassémie alpha liée à l'X. Elle se traduit par à un visage caractéristique, des anomalies génitales, un retard de développement important s'associant avec un retard mental et des signes hématologiques de thalassémie alpha.

Le syndrome du X Fragile (Oostra and Willemsen, 2002) est quant à lui la conséquence de l'expansion et de la méthylation de la répétition du tri-nucléotides CGG au niveau de la région 5'UTR (Untranslated region) du gène *FRM*. Ceci réprime l'expression de ce gène et crée un site dit « Fragile » au niveau du chromosome X qui peut être visible en cartographie. Cette aberration

peut conduire à des caractéristiques physiques comme un visage allongé, des oreilles larges et décollées, de gros testicules (macro-orchidie). Cela entraîne aussi des difficultés intellectuelles très variables avec un retard cognitif moyen à sévère. On retrouve des caractéristiques comportementales proches de l'autisme avec des mouvements stéréotypés et une anxiété sociale. Les filles atteintes par cette maladie sont diversement affectées, avec généralement un retard cognitif moyen.

Une autre pathologie connue pour posséder un défaut de méthylation est le syndrome de Rett (Amir et al., 1999; Fasolino and Zhou, 2017). C'est une maladie génétique rare se développant chez le très jeune enfant, principalement la fille, et provoquant un handicap mental et des atteintes motrices sévères. Ce dernier est lié à une mutation du gène *MeCP2*, une protéine de la famille des MBPs (Methyl Binding Proteins) qui se lie aux cytosines méthylées et joue un rôle dans le contrôle de l'activité des gènes neuronaux.

#### ***e. Les pathologies associées à des défauts de méthylation des ICRs***

De nombreux syndromes héréditaires sont causés par une dérégulation épigénétique au niveau des ICRs, tel que le syndrome d'Angelman's, le syndrome de Prader-Willi ou encore le syndrome de Beckwith-Wiedemann (Brzezinski et al., 2017; Goldstone, 2004; Le Fevre et al., 2017; Maher and Reik, 2000; Manzardo and Butler, 2016; Nicholls et al., 1998). Si on prend l'exemple du dernier syndrome cité, ce dernier est dû à une déméthylation d'un groupe de gènes soumis à empreinte au niveau du chromosome 11p15.5 tel que *IGF2*, qui, à ce moment-là, se retrouve exprimé de façon biallélique, ou encore *CDKN1C* dont l'expression va se retrouver réprimé. Ceci a pour conséquence une hypertrophie des organes ainsi qu'un développement précoce de certaines tumeurs.

On observe aussi des défauts de méthylation des ICRs dans certaines cas de diabète tels que le TNDM (transient neonata diabetes mellitus) (Arthur et al., 1997; Babenko et al., 2006; Shield et al., 1997; Temple and Shield, 2002). La revue d'Elhamamsy explique plus en détails l'ensemble des pathologies associées au défaut de méthylation au niveau des ICRs (Elhamamsy, 2017)

# Matériels et méthodes

---

## Liste des lignées murines utilisées

Rosa26cre/GFP (Birling et al., 2012) (provenant de l'ICS, France)  
Oct4-GFP (Yoshimizu et al., 1999) (généreusement fournie par Robert Feil)  
Tnap-Cre (Lomeli et al., 2000) (provenant du laboratoire Jackson)  
Dnmt1-2lox (Jackson-Grusby et al., 2001a) (généreusement fournie par Franck Rosenbauer)  
Uhrf2 tm1b (Bradley et al., 2012) (provenant du CMMR, Canadian mouse mutant repository)  
Vav-Cre (Georgiades et al., 2002) (généreusement fournie par Olivier Bernard)  
Dnmt3a-2lox (Kaneda et al., 2004) (provenant du RIKEN BioResource Center, Japan)  
Dnmt3b-2lox (Dodge et al., 2005) (provenant du RIKEN BioResource Center, Japan)

Toutes les lignées de souris utilisées dans cette Thèse ont été maintenues par croisement avec des souris C57BL/6J provenant de Charles Rivers. Toutes les lignées ont été élevées à l'animalerie de l'unité BSC, en accord avec la réglementation en vigueur. Le détail des lignées est fourni dans l'**Annexe 1**.

## Préparation des échantillons biologiques

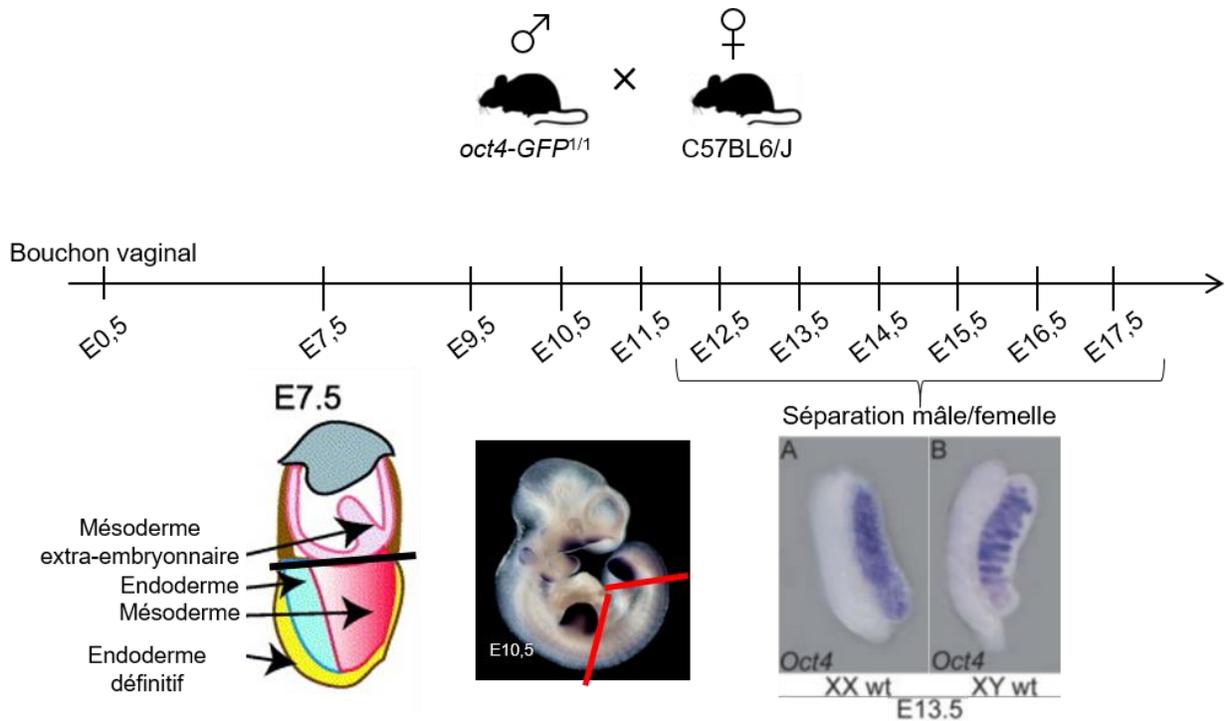
Pour toutes les dissections, le matin du bouchon vaginal correspond à E0.5 et toutes les dissections ont été réalisées à la même heure (9h).

### Isolation d'épiblastes à E7.5 et d'embryons à E8.5

L'ensemble des embryons sauvages a été obtenu grâce à des croisements de souris C57BL/6J provenant de Charles Rivers. Les embryons ont été disséqués manuellement dans du milieu de culture M2 sous un microscope binoculaire. À E7.5, l'épiblaste est séparé des tissus extra-embryonnaires par dissection manuelle tandis qu'à E8.5 l'embryon en entier est utilisé.

### Isolation des cellules germinales primordiales (PGCs) par le marqueur Oct4-GFP

L'étude de la méthylation des PGCs au cours du développement embryonnaire a été réalisée en isolant les cellules par tri cellulaire grâce à lignée de souris Oct4-GFP qui exprime le marqueur



**Figure A : Isolation des PGCs**

L'étude de la méthylation de l'ADN au niveau des cellules primordiales germinales a été réalisé entre E9.5 et E17.5. A E7.5, l'épiblaste a été disséqué en entier (au-dessus du trait noir) qui a servi de référence pour la méthylation des PGCs. De E9.5 à E11.5, seule la partie inférieure de l'embryon a été récupérée (entre les deux lignes rouge). A partir de E12.5, j'ai disséqué les gonades qui peuvent être sexées en fonction de leur anatomie : les gonades femelles sont longues et duveteuses tandis que les gonades mâles sont plus rondes et striées.

Les photos des gonades mâles et femelles à E13.5 ont été adapté de Pitetti JL, et al., 2013. La coloration bleue correspond au marqueur des cellules germinales primordiales Oct4.

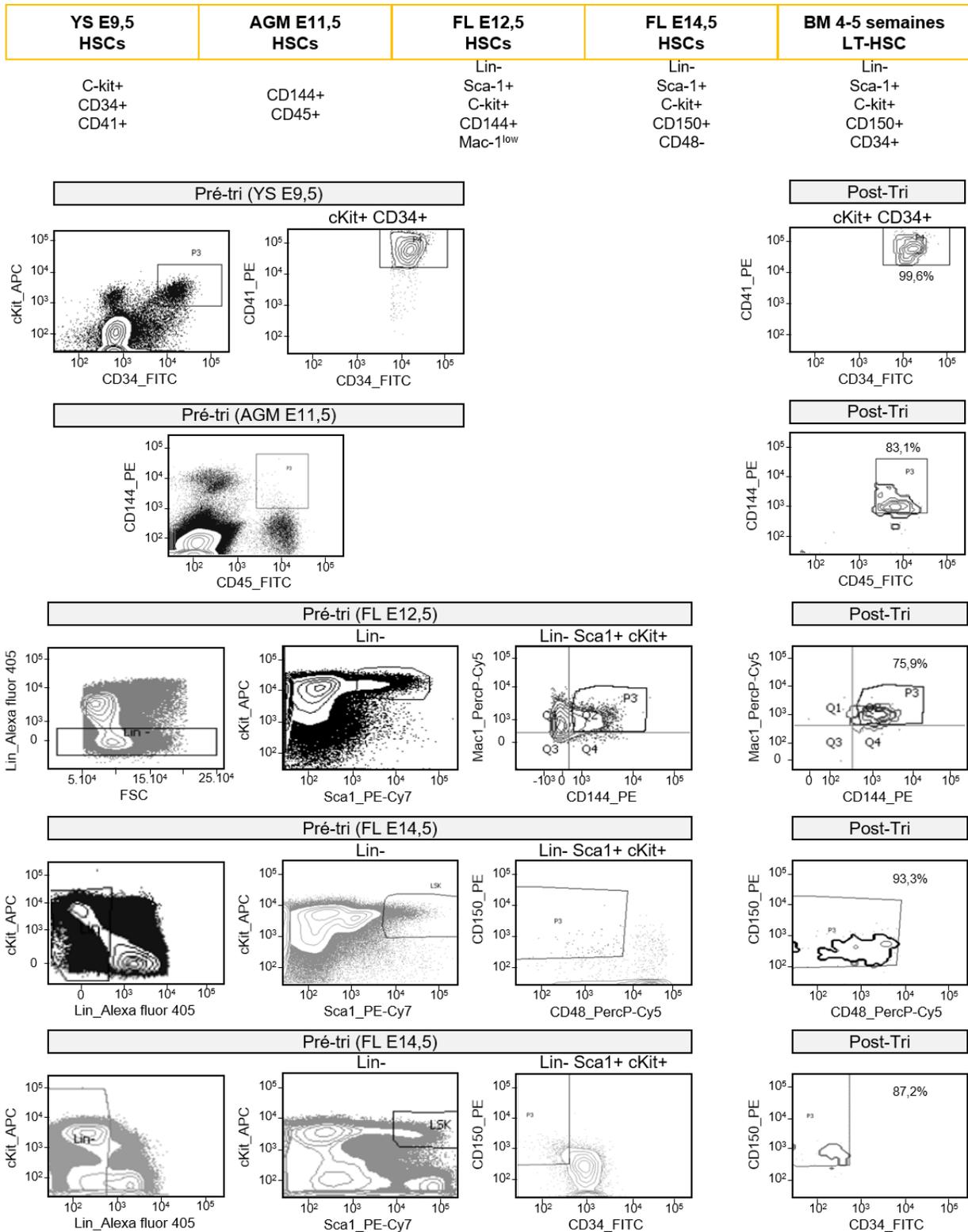
GFP spécifiquement dans les PGCs. De E9.5 à E11.5, seule la partie inférieure de l'embryon est récupérée lors de la dissection. À partir de E12. 5 jusqu'à E17.5, les gonades sont disséquées et les animaux sont sexés en fonction de l'anatomie de ces dernières (**Figure A**). L'ensemble des dissections a été réalisé sous microscope et dans du PBS 1X. Les cellules des gonades ont été dissociées par une incubation de 5 minutes à 37°C dans une solution contenant de la Trypsine 0.25% EDTA, 0.16mg/mL de DNase diluée dans du milieu M2 (Sigma-Aldrich) et 1/12 Vo d'Accumax (Millipore). Les cellules sont ensuite filtrées à 70 microns (MACS SmartStrainer, Miltenyi Biotec) et triées par cytométrie en Flux (FACSVantage, BD Biosciences) grâce à l'expression de GFP.

### Isolation des cellules germinales primordiales (PGCs) par le marqueur SSEA-1

Dans le but de créer le mutant conditionnel du gène *Dnmt1* dans les PGCs, j'ai croisé des souris *Dnmt1-2lox* avec la lignée murine TNAP-Cre (**Annexe 1**). Dans ces lignées j'ai constaté que les PGCs ne pouvaient pas être triées par le marqueur Oct4-GFP en raison d'une expression ectopique de GFP dans les cellules somatiques des gonades. En raison de ces soucis techniques, les PGCs ont été triées à partir d'un marqueur de surface (SSEA-1) naturellement exprimé par les PGCs, en réalisant un tri cellulaire avec un anticorps anti-SSEA-1 couplé au fluorochrome PE (Miltenyi Biotec, France). Après, obtention de la suspension cellulaire, les PGCs sont incubées 10 minutes à 4°C avec les anticorps SSEA-1-PE (**Annexe 2**) dilués dans du milieu M2 (Sigma, France), contenant 2µg/mL de DNase I (Thermo Fisher, France). Les cellules sont ensuite filtrées à 70 microns (MACS SmartStrainer, Miltenyi Biotec) et triées par cytométrie en Flux (FACSVantage, BD Biosciences).

### Validation de la pureté des PGCs

L'ADN génomique a été extrait par digestion à la protéinase K, suivie d'une extraction au phénol-chloroforme-isoamyle et d'une précipitation à l'éthanol. La pureté des PGCs a été vérifiée par COBRA (Xiong and Laird, 1997) à partir de l'ADN génomique provenant du stade E13. 5, sur les gènes *Ckt2* et *Dppa3*. Ces gènes sont connus pour être méthylés dans les cellules somatiques mais pas dans les PGCs. Les primers utilisés pour la réalisation du COBRA sont disponibles dans l'**Annexe 3**. En parallèle du COBRA, j'ai vérifié la pureté des PGCs triées par le test de marquage de la phosphatase alcaline. Pour ce test, les cellules GFP<sup>+</sup>, GFP<sup>-</sup>, CD15-PE<sup>+</sup> et CD15-PE<sup>-</sup> issues des tris cellulaires ont été centrifugées sur des lames de verres, puis le test de phosphatase



**Figure B : Protocoles de tris cellulaires utilisés pour isoler les cellules souches hématopoïétiques (HSCs)**

Les anticorps utilisés pour trier les HSCs ont été choisis selon la publication de McKinney-Freeman *et al.* 2012. Un pré-tri est d'abord réalisé suivi d'un second tri avec les cellules récoltées (post-tri) afin de vérifier l'efficacité de tri. Le post-tri n'a été réalisé que lors de la phase de mise au point des expériences en raison du faible nombre de cellules. YS=Yolk Sack (Sac Vitellin); AGM=Aorta-Gonad-Mesonephros; FL=Fetal Liver (Foie fœtal); BM=Bone Marrow (Moelle Osseuse).

alcaline a été effectué avec un kit de marquage commercial (Millipore, SCR004) en suivant le protocole fourni dans le kit.

### Isolation des cellules souches hématopoïétiques (HSC)

Les cellules souches hématopoïétiques (HSC) ont été triées par cytométrie en flux (FACS Arian SORP, BD Bioscience) en utilisant les mêmes marqueurs membranaires que ceux utilisés dans une publication du laboratoire de George Daley (McKinney-Freeman et al., 2012). Ainsi, nous avons récoltés des HSC embryonnaires à partir du sac vitellin (SV) à E9.5, de l'axe gonadomésonephrique (AGM) à E10.5, et du foie fœtal (FF) à E12.5 et E14.5. Enfin, nous avons récupérés les LT-HSCS (*Long term*) et les ST-HSCS (*Short Term*) à partir de la moëlle osseuse (MO) d'individus âgés de 4 semaines (**Figure B**). Les suspensions cellulaires provenant de la MO ont été réalisées à partir des fémurs, du plexus, des hanches et de la colonne vertébrale des souris qui ont été broyées à l'aide d'un mortier dans du PBS 1X, 0.5% BSA, 2mM EDTA. Les suspensions cellulaires provenant du foie fœtal ont été obtenues par passage successif du foie dans une seringue de 30G dans du PBS 1X, 0.5% BSA, 2mM EDTA, 20mM de DNase I (Sigma). Ensuite, les cellules ont subi une érythrolyse pendant 7 minutes à RT avec 0.83% de NH<sub>4</sub>Cl, puis elles sont filtrées à 70 microns et marquées avec les différents cocktails d'anticorps 15 minutes à 4°C. Pour les cellules provenant du sac vitellin et de l'AGM, les dissections ont été réalisées dans du PBS 1X/10% FBS et la suspension cellulaire obtenue par une incubation de 35 minutes à 37°C dans du PBS 1X /20% FBS/0.25% Collagenase (*Clostridium histolyticum*, Sigma). Après rinçage, les cellules ont été incubées 15 minutes à 4°C avec les anticorps correspondants. Les anticorps utilisés en cytométrie de flux proviennent tous d'eBiosciences et tous les clones ainsi que les marqueurs de fluorescence utilisés sont listés dans l'**Annexe 2**. Pour le tri des HSCs provenant de la MO, toutes les cellules ont été premièrement marquées avec un cocktail de marqueurs biotinylés (anti-Ter119, anti-CD3a, anti-CD4, anti-CD8a, anti-CD19, anti-IgM, anti-Gr1, CD11b) afin de dépléter nos échantillons en cellules matures. Les cellules du foie fœtal subissent la même déplétion sauf pour le marqueur CD11b. Des billes magnétiques couplées à de la streptavidine (MACS Miltenyi Biotec) sont ensuite incubées 15 minutes à 4°C avec les échantillons puis l'ensemble est passé à travers des colonnes magnétiques (colonne LS, MACS Miltenyi Biotec, France) selon le protocole de MACS Miltenyi Biotec. L'efficacité du tri est vérifiée grâce à un post-tri (**Figure B**) lors de la mise au point des différents tris. Par la suite, les post-tris ne sont plus réalisés en raison de la faible quantité de cellules triées. Pour les stades les plus critiques, c'est-

à-dire les HSCS provenant du SV et de l'AGM, un COBRA est réalisé sur les gènes *Gata2* et *Runx1* pour vérifier la pureté du tri.

### **Préparation d'échantillons ADN et ARN**

L'ADN génomique des PGCs a été isolé par une extraction classique au phenol:chloroforme:alcool isoamylique 25:24:1, suivi d'une précipitation à l'éthanol. L'ARN des PGCs a été extrait avec le kit « RNeasy Mini kit » (Qiagen). Pour les HSCs, j'ai extrait simultanément l'ADN et l'ARN des mêmes échantillons avec le kit « AllPrep DNA/RNA Mini Kit » (Qiagen).

### **Génotypage**

Les génotypages ont été réalisés à l'aide du kit « REDextract » de Sigma selon le programme de PCR suivant : 2 minutes 95°C, 30 cycles (30s 95°C, 30s TM, 1min 72°C), puis 2 minutes à 72°C. La liste des oligonucléotides est répertoriée dans l'**Annexe 4**.

### **Analyses Sanguines**

Le sang a été obtenu à partir de prélèvement retro-orbitaux et collecté dans des tubes enduit d'EDTA (Microvette, Startedt). La quantification de globules blancs, rouges, des plaquettes et de l'hémoglobine ont été réalisés par l'Institut Clinique de la Souris, Illkirch.

### **Préparation de bibliothèques RRBS**

#### Protocole de préparation des bibliothèques RRBS

Les bibliothèques RRBS ont été réalisées selon un protocole que j'ai optimisé à partir de celui publié par le laboratoire de Alexandre Meissner (Gu et al., 2011a). En résumé, nous avons digéré entre 5 et 100 ng d'ADN génomique pendant cinq heures avec l'enzyme *MspI* (Thermo Scientific, Waltham, MA, USA), réparé et ajouté une adénosine au niveau des sites de restriction avec l'enzyme Klenow Exo5' (Thermo Scientific), et enfin ajouté les adaptateurs méthylés à l'aide de la T4 ADN ligase (Thermo Scientific) dans du tampon Tango 1X. Les fragments compris entre 150 et 400 pb (l'insert auquel on a ajouté la taille des adaptateurs) ont été purifiés par électrophorèse sur un gel d'agarose 3 %, TBE 0.5X avec le kit d'extraction de gel MinElute (Qiagen). Par la suite, nous avons réalisé deux tours de conversion bisulfite avec le kit de EpiTect (Qiagen) selon le protocole fourni avec le kit. Les bibliothèques RRBS finales ont été amplifiées par PCR avec l'ADN polymérase PfUTurbo Cx hotstart (Agilent, Santa Clara, CA, USA) et des amorces PE Illumina indexées selon les conditions de PCR suivante : 95° C pendant 2 minutes, 14 à 18 cycles (95°C pendant 30 secondes, 65° C pendant 30 secondes, 72°

C pendant 45 secondes), 72°C pendant 7 minutes. Les librairies sont purifiées avec des billes magnétiques AMPure (Beckman Coulter, Brea, CA, USA), quantifiées avec un fluoromètre Qubit (Life Technologies, Carlsbad, CA, USA) et vérifiées en chargeant 10 ng de librairie sur un gel pré-coulé Criterion, 4–20 % (Biorad, Hercules, CA, USA) qui est révélé par la suite avec du SYBR Green. Une version complète et détaillée du protocole de RRBS est fournie dans l'**Annexe 6**.

#### Préparation des librairies RRBS pour le projet PGCs

Les librairies RRBS des PGCs WT au cours de la cinétique de reprogrammation ont été réalisées à partir de mélanges de cellules provenant de plusieurs embryons. Pour le stade E9,5, deux duplicatas ont été réalisés à partir de 125 embryons. Pour le stade E10,5, deux duplicatas ont été réalisés à partir de 97 embryons. Pour le stade E11,5, deux duplicatas ont été réalisés à partir de 31 embryons. Pour le stade E12,5, 15 paires de gonades femelles et 20 paires de gonades mâles ont été utilisées. Pour le stade E13,5, 19 paires de gonades femelles et 27 paires de gonades mâles ont été utilisées. Pour le stade E14,5, 4 paires de gonades femelles et 13 paires de gonades mâles ont été utilisées. Pour le stade E15,5, 6 paires de gonades femelles et 8 paires de gonades mâles ont été utilisées. Pour le stade E16,5, 8 paires de gonades femelles et 9 paires de gonades mâles ont été utilisées. Pour le stade E17,5, 8 paires de gonades femelles et 7 paires de gonades mâles ont été utilisées. Les librairies RRBS des PGCs délétées pour *Dnmt1* ou *Uhrf2*, ainsi que leurs contrôles provenant des mêmes portées, ont été réalisées à partir d'embryons uniques.

#### Préparation des librairies RRBS pour le projet HSCs

Les librairies RRBS des HSCs WT au cours du développement ont été réalisées à partir de mélanges de cellules de plusieurs embryons. Pour le stade E9,5, deux duplicatas ont été réalisés à partir de 40 et 18 sacs vitellins. Pour le stade E11,5, deux duplicatas ont été réalisés à partir de 12 et 30 AGM. Pour le stade E12,5, deux duplicatas ont été réalisés à partir de 29 et 20 foies foetaux. Pour le stade E14,5, quatre duplicatas ont été réalisés à partir de 20, 40, 28 et 24 foies foetaux. Pour l'étude de la moelle osseuse à l'âge adulte, trois duplicatas ont été réalisés à partir de 8, 4 et 11 souris femelles. Les librairies RRBS sur les cellules de l'AGM invalidées pour *Dnmt3a* ou *Dnmt3b*, ainsi que leurs contrôles provenant des mêmes portées, ont été réalisées à partir de mélanges de cellules provenant de plusieurs embryons (*AGM Dnmt3a+/+* en deux duplicatas sur 3 et 2 embryons ; *AGM Dnmt3a-/+* en deux duplicatas sur 9 et 1 embryons ; *AGM Dnmt3a-/-* en duplicatas sur 3 et 3 embryons ; *AGM Dnmt3b-/+* en deux duplicatas sur 6 et 11 embryons ; *AGM*

*Dnmt3b*<sup>-/-</sup> en duplicatas sur 4 et 3 embryons). Les bibliothèques RRBS de HSC de la moelle osseuse présentant une délétion totale ou conditionnelle de *Dnmt3a*, *Dnmt3b* ou *Dnmt3a/Dnmt3b*, ainsi que leurs contrôles WT ou hétérozygotes, ont été réalisées en plusieurs répliques à partir de cellules provenant d'animaux uniques. En raison du très faible nombre de LT-HSC que l'on peut isoler par animal, nous avons généré les bibliothèques RRBS à partir des cellules ST-HSC.

### **Séquençage et traitement des bibliothèques RRBS**

Les bibliothèques ont été séquençées de façon paired-end (2 x 75 pb) en multiplex dans un séquenceur Illumina HiSeq2500 chez Integragen SA (Evry, France) afin de générer en moyenne 30 millions de paires de reads par échantillon (**Annexe 5**). Le contrôle qualité a été fait avec FastQC (v0.11.12). Afin de retirer la séquence des adaptateurs ainsi que les séquences de mauvaise qualité, nous avons utilisé le logiciel Trim Galore (v0.3.7) et enlevé les bases en fin de séquences avec un Phred score en dessous de 20. Les séquences sont par la suite alignées sur le génome de la souris (version mm10) grâce au logiciel BSMAP (v2.74) avec les paramètres suivants : mode RRBS, un maximum de deux mésappariements de bases et une taille d'insertion entre 30 et 400 pb. Nous avons extrait le pourcentage de méthylation en comparant le nombre de Cs par rapport au nombre total de Cs et de Ts pour chaque position en utilisant methratio.py. Le ratio de méthylation est combiné pour les deux brins et filtré en fonction d'une profondeur minimale de séquençage de 8X. L'efficacité de conversion bisulfite est estimée par le calcul du nombre de C convertis en T au niveau des sites non-CpGs, qui est de plus de 99.5% en moyenne (**Annexe 5**). Les niveaux de méthylation sont observés avec le logiciel IGV.

### **Analyse bioinformatique des données RRBS**

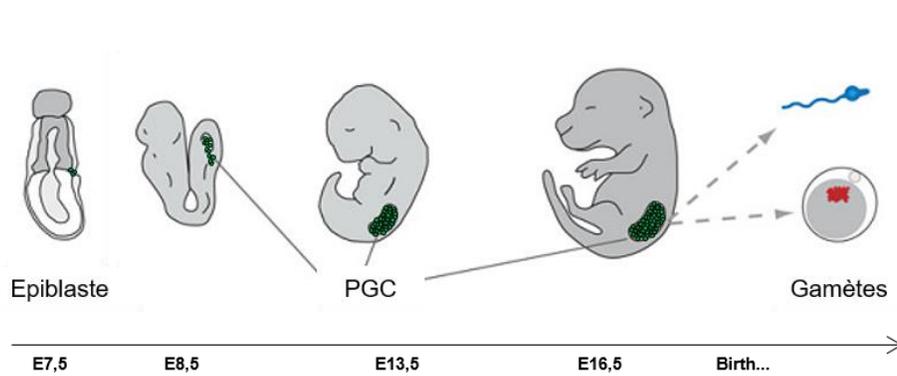
La séquence référence des gènes, des transposons et l'annotations des CGI a été obtenue à partir de l'annotation du génome mm10 sur le site UCSC. Les transposons sont filtrés de manière à avoir une taille minimum de 400pb. Les régions de contrôle de l'empreinte parentale (ICR) ont été récupérées à partir de la base de données « Wamidex imprinting database ». Pour analyser les différences de profils de méthylation entre échantillons, nous avons identifié des régions de méthylation différentielle (DMRs) avec l'algorithme eDMR (<http://code.google.com/p/edmr/>) en utilisant les critères suivants : un seuil minimum de différence de méthylation de 20%, au moins 3 CpGs dans les DMRs et une p-value inférieure à 0.01. Enfin, la recherche d'ontologie de gènes a été réalisée avec l'algorithme DAVID disponible sur internet (<https://david-d.ncifcrf.gov>).

## RT-qPCR

L'ARN a été rétro-transcrit avec le kit « Maxima first strand cDNA synthesis kit » (Thermo Scientific) en utilisant une combinaison d'oligo(dT)s et d'hexamères d'oligonucléotides aléatoires. Les réactions de RT-qPCR ont été réalisées avec le « KAPA SYBR FAST qPCR Kit » (Cliniscience) sur la machine « StepOnePlus realtime PCR system » (Life Technologies). Nous avons utilisé des cycles rapides de PCR: 95°C pendant 20 s, 40 cycles (95°C pendant 20 s, 60 ou 64°C pendant 30 s) suivi par une courbe de dissociation. Les mesures ont été réalisées en triplicata et l'expression des gènes a été normalisée par rapport à deux gènes de ménage (*Rpl13a* et *Actb*). En parallèle, nous amplifions systématiquement les contrôles RT- pour être certains de n'avoir aucune contamination d'ADN génomique. Les séquences des oligonucléotides utilisés sont répertoriées dans l'**Annexe 2**.

## RNA-Seq

Nous avons préparé les ARN avec le kit « AllPrep DNA/RNA Mini Kit » (Qiagen). Pour les différents projets de RNA-seq, les ARN sont préparés à partir d'animaux provenant de la même fratrie. L'intégrité des ARN a été vérifiée par Bioanalyzer (Agilent Technologie) et la préparation des librairies ARN ainsi que le séquençage ont été réalisés à l'IGBMC selon le protocole Ovation (pour les petites quantités de cellules) ou le protocole Ribo-Zero Gold (pour les quantités supérieures à 100 ng d'ARN). Le contrôle qualité a été fait avec FastQC et l'alignement sur le génome de la souris en mm10 avec le logiciel TopHat2 (V2.0.12). Dans l'optique de visualiser nos données, des fichiers BigWig ont été générés où le nombre de séquences est normalisé selon des fenêtres de taille variable avec bam2wig.py dans le complément de logiciel RSeqQC (v2.4) en utilisant les reads alignés de manière unique. Le nombre brut de séquences par gène a été calculé à partir des fichiers BAM dans le logiciel HTseqcount de la suite HTSeq (v0.6.0). Ces nombres sont ensuite normalisés et utilisés afin d'identifier les différences d'expression des gènes avec DESeq2 (v1.4.5). Les gènes sont désignés comme différentiellement exprimés lorsque la p-value entre les échantillons est inférieure à 0.05 et que le « fold change » est supérieur à 2.



**Figure 1 : Développement des cellules primordiales germinales (PGCs)**

Les PGCs, représentées ici en vert, forment tout d'abord un groupe de 40 cellules dans l'épiblaste (E7,5), à la base de l'allantoïde. Elles prolifèrent et migrent pendant l'embryogénèse jusqu'à atteindre les gonades à E11,5 où a lieu la détermination sexuelle. Ainsi, à partir de E12,5, on peut facilement différencier les gonades femelles des gonades mâles. A partir de E13,5, les PGCs arrêtent de proliférer pour entrer dans le programme méiotique.

# Résultats

---

## I. La méthylation de l'ADN au cours du développement embryonnaire des cellules primordiales germinales (PGCs)

### A. Introduction

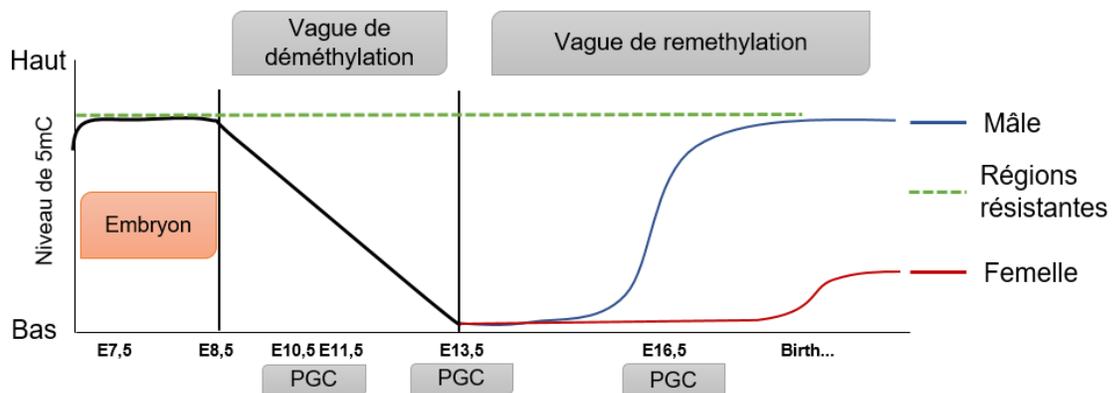
#### 1. Le développement des PGCs

Les cellules germinales primordiales (PGCs) sont les précurseurs embryonnaires des gamètes. Chez la souris, elles proviennent de l'épiblaste et forment initialement un amas de 40 cellules au niveau du mésoderme extra-embryonnaire à E7.25 (**Figure 1**) (Chiquoine, 1954; Ginsburg et al., 1990). Ces cellules sont originaires de six cellules de la partie la plus proximale de l'épiblaste (Lawson and Hage, 1994). Par la suite, les PGCs prolifèrent et migrent en direction des ébauches des futurs organes génitaux à partir de E11,5, c'est-à-dire au tout début de la détermination sexuelle (**Figure 1**) (Molyneaux et al., 2001; Richardson and Lehmann, 2010; Seki et al., 2007; Tam and Snow, 1981).

Leur migration vers les gonades est dirigée par l'environnement somatique (Wylie, 1999) qui est composé à la fois de signaux chimiotactiques, provenant des gonades en développement ainsi que d'un gradient glycoprotéique provenant de la matrice extracellulaire. De plus, une fois la migration initiée, le contact entre PGCs joue aussi un rôle crucial dans leur migration. La colonisation des gonades en développement est une étape cruciale au développement des PGCs puisqu'elles apportent les facteurs nécessaires pour que ces dernières terminent leur différenciation, acquièrent leur morphologie (Donovan et al., 1986) ainsi qu'un profil d'expression spécifique, avec notamment l'expression des gènes suivants : *Gcna1* (*germ cell nuclear antigen 1*) (Enders and May, 1994), *Dazl* (*deleted in azoospermia like*) (Saunders et al., 2003), *Gcl* (*germ cell-less*) (Kimura et al., 1999), *ERR-β* (*estrogen related receptor β*) (Mitsunaga et al., 2004) et *Mvh* (*mouse vasa homologue*) (Noce et al., 2001; Toyooka et al., 2000).

La prolifération des PGCs se fait par mitose durant toute leur migration ainsi que un à deux jours après la colonisation des gonades (Schmahl et al., 2000). Leur nombre augmente de façon exponentielle, allant d'environ 100 cellules à E8,5 jusqu'à environ 25 000 cellules à E13,5 (Donovan et al., 1986; Tam and Snow, 1981).

En plus de la phase de prolifération, un programme de mort cellulaire se met en place dans les PGCs pendant leur développement, leur migration, leur colonisation des gonades et même après la naissance, afin de pallier aux tumeurs germinales (Göbel et al., 2000; Runyan



**Figure 2 : Schéma du profil de méthylation (5mC) chez les PGCs**

Dès la spécification, le génome des PGCs se déméthyle et ce, jusqu'au stade E13,5 où il est globalement déméthylé. A partir de E14,5, seul le génome des PGCs mâle se reméthyle tandis que chez la femelle, la reméthylation a lieu après la naissance dans les ovocytes. Certaines régions résistent à cette vague de déméthylation, comme certains éléments transposables tels que les IAPs, les LTRs et les LINE1. Ici, sont représentés les différents stades (E10,5, E11,5, E13,5 et E16,5) étudiés jusqu'à présent.

et al., 2008; Schneider et al., 2001) et d'assurer l'intégrité de la lignée germinale (Bristol-Gould et al., 2006). Chez les mâles, une vague initiale d'apoptose a lieu entre E13,5 et E17,5 et une seconde aux alentours de la naissance, où seulement 25% des cellules germinales survivent (Wang et al., 1998). Chez les femelles, une vague similaire d'apoptose a lieu à E13,5 et une seconde entre E15,5 et la naissance (Bakken and McClanahan, 1978; Borum, 1961; Coucouvanis et al., 1993; Ratts et al., 1995). Durant toute cette période, plus de 70% des cellules germinales femelles rentrent en apoptose (McClellan et al., 2003; Pepling and Spradling, 2001).

Chez les embryons femelles, les PGCs prolifèrent jusqu'à E13,5 et par la suite, entrent en prophase I de méiose (Hilscher et al., 1974; Speed, 1982) où elles s'arrêtent au stade diplotène. C'est après la naissance que les cellules germinales, appelées alors ovocytes, reprennent leur cycle méiotique mais jusqu'au stade de la métaphase II. Ce sera après fécondation que l'ovocyte terminera son cycle (Iwamatsu and Chang, 1972).

A l'inverse, les PGCs mâles entrent en arrêt mitotique après la colonisation des gonades, entre E12,5 et E14,5, et restent quiescentes en phase G0/G1 pendant toute l'embryogénèse (Hilscher et al., 1974; Western et al., 2008). Les cellules germinales mâles reprennent leur programme mitotique après la naissance et ce, jusqu'à la puberté, où commence la spermatogénèse, au travers de l'initiation de la phase méiotique.

## 2. Reprogrammation épigénétique des PGCs : que sait-on ?

Au cours du développement des cellules germinales, le génome de ces dernières subit des changements épigénétiques importants tant au niveau de la 5mC qu'au niveau de marques épigénétiques des histones. Les avancées technologiques faisant appel au séquençage haut débit (WGBS, ChIP-seq) ont permis l'étude de la dynamique de l'épigénome au cours de leur développement. Cependant, l'incidence physiologique de ces changements dynamiques est encore mal comprise. Dans ce paragraphe, je vous résumerai les connaissances actuelles sans rentrer dans les détails, puisque j'y reviendrai dans la partie Résultats I.C, lorsque je vous exposerai mes résultats.

L'ensemble des études déjà menées sur les PGCs murines démontrent que ces dernières rentrent dans une phase de déméthylation de l'ADN dès E9,5 jusqu'à E13,5, où le génome est globalement déméthylé (**Figure 2**) (Guibert et al., 2012; Kobayashi et al., 2013; Popp et al., 2010; Seisenberger et al., 2012). Il semblerait que cette vague de déméthylation soit la conséquence à la fois des mécanismes de déméthylation actifs et passifs.

Concernant les mécanismes de déméthylation actifs, plusieurs études ont été réalisées. L'une des premières a montré l'implication de la protéine AID qui a la capacité de déaminer la 5mC. L'utilisation de souris déficientes pour cette enzyme montre que le génome des PGCs est légèrement hyperméthylé en comparaison avec des animaux sauvages (Popp et al., 2010), suggérant son implication dans le processus de déméthylation active. Néanmoins, l'augmentation de méthylation n'est que modérée, suggérant l'implication d'autres enzymes. De ce fait, d'autres études ont été menées en particulier sur les protéines TET impliquées dans la déméthylation active. Dans les PGCs, seuls TET1 et TET2 sont exprimées (Hackett et al., 2013). Les expériences menées sur des PGCs déficientes pour ces deux protéines, de manière indépendante ou couplée, nous informent que ces enzymes prennent part au processus de déméthylation des PGCs. Cependant, elles n'affectent pas le niveau global de méthylation du génome des PGCs mais plutôt des loci spécifiques tels que les gènes spécifiques de la lignée germinale ou encore certains éléments transposables répétés (Vincent et al., 2013; Yamaguchi et al., 2012). De plus, les animaux déficients pour TET1 ou TET2 sont fertiles.

D'autres recherches sur la phase de déméthylation des PGCs ont porté sur la possibilité de l'implication d'un processus passif. Entre E8,5 et E13,5, les PGCs prolifèrent rapidement et des études démontrent que les deux ADN méthyltransférases *de novo*, DNMT3A et DNMT3B, sont exprimées faiblement dans les PGCs (Kagiyada et al., 2013). De plus, bien que DNMT1 soit exprimée, le recrutement de DNMT1 à la chromatine par le facteur UHRF1 ne peut se faire correctement puisque la protéine UHRF1 est faiblement exprimée et exclue du noyau, affectant ainsi l'activité de DNMT1 (Kagiyada et al., 2013; Seisenberger et al., 2012). De ce fait, l'idée qu'une déméthylation passive puisse contribuer principalement à l'hypométhylation du génome des PGCs est tout à fait plausible, sachant que la cinétique de déméthylation des régions soumises à empreintes corrèle avec cette hypothèse (Kagiyada et al., 2013).

Cette reprogrammation épigénétique est un évènement critique du développement des PGCs car elle est essentielle à l'activation des gènes spécifiques des PGCs, dont notamment *Mvh* et *Dazl*, dès E10,5 (Maatouk et al., 2006). Elle permet aussi la déméthylation des gènes soumis à empreinte, ce qui est nécessaire pour l'établissement du profil de méthylation sexe-spécifique des futurs gamètes (Brandeis et al., 1993; Ueda et al., 2000).

Cependant, la cinétique de déméthylation de l'ADN ne semble pas identique sur l'ensemble des régions génomiques. Ainsi la cinétique de déméthylation des gènes soumis à empreinte semble différente de celle de la plupart des autres gènes. Pour exemple, les gènes

maternels *Peg3*, *Lit1* et *Snrpn*, ainsi que le gène paternel *H19*, semblent se déméthyliser entre E11,5 et E12,5 (Hajkova et al., 2002), alors que des expériences d'immunohistochimie suggèrent que la déméthylation du génome des PGCs est initiée vers E9.0 (Seki 2005). Enfin, les éléments transposables tels que les rétrotransposons à LTR, les IAPs et les LINE1 se retrouvent aussi déméthylés à partir de E11,5 mais leur déméthylation reste incomplète à E13,5 (Guibert et al., 2012; Hajkova et al., 2002; Kobayashi et al., 2013; Lane et al., 2003; Seisenberger et al., 2012). Une hypothèse pour expliquer cette résistance à la déméthylation est qu'il serait nécessaire de garder un niveau résiduel de 5mC afin de verrouiller ces éléments pour protéger l'intégrité du génome.

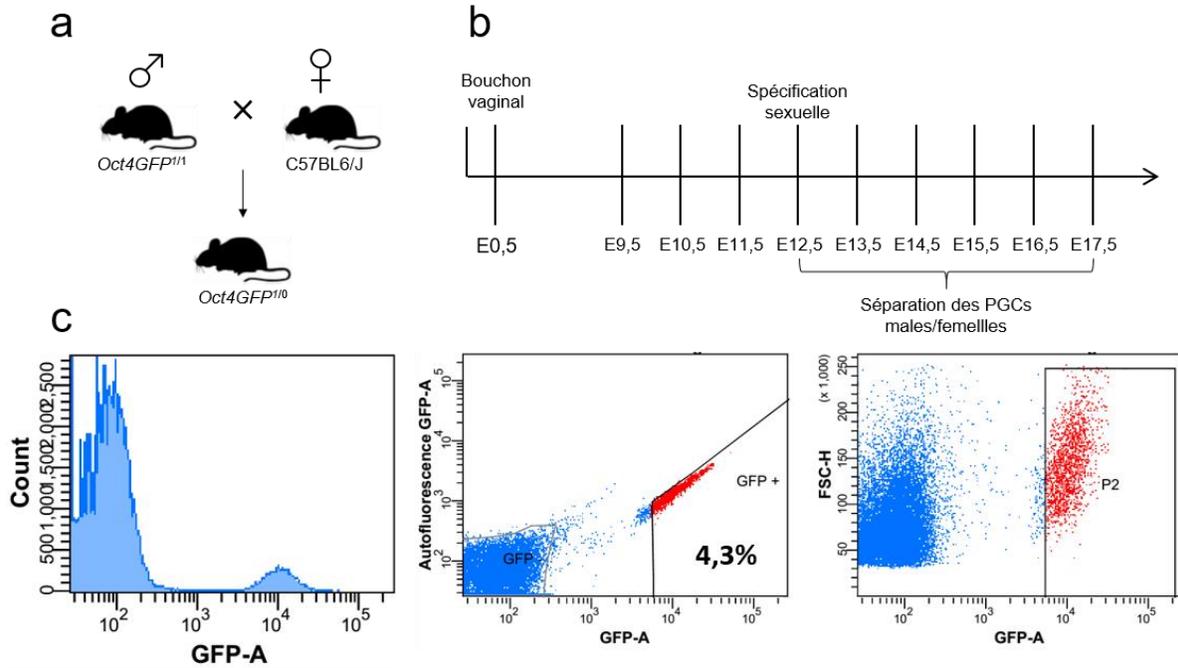
A E16,5, les PGCs mâles entrent dans une vague de reméthylation alors que les cellules germinales femelles entameront cette phase après la naissance (Figure 2) (Seisenberger et al., 2012). Selon la même étude, le génome des PGCs mâles se reméthylerait à hauteur de 31,8% à 50% à E16,5. Cependant, peu de choses sont encore connues sur les cibles et la cinétique de reméthylation des PGCs car à ce jour seuls les stades E13,5 et E16,5 ont été étudiés (Seisenberger et al., 2012).

Le génome des PGCs subit aussi une reprogrammation au niveau des histones (Hayashi et al., 2005; Mochizuki et al., 2012; Ng et al., 2013; Seki et al., 2005). Je n'aborderai ce sujet que brièvement. Ainsi, dès E8,5, le génome des PGCs perd la modification d'histone H3K9me2 (marque répressive) alors qu'il acquiert des marques d'histone permissives ou actives du type de H3K9ac, H3K4me2/me3, ainsi que des modifications répressives de type H3K27me3 (Seki et al., 2005). A E11,5, le linker d'histone H1 et les marques d'histone H3K27me3 et H3K9me3 sont perdus pour réapparaître à E12,5. De plus, le variant d'histone H2AZ ainsi que la marque H3K9ac sont perdus à E11,5. Récemment, il a été découvert que la marque d'histone H3K64me3 se localise au niveau de l'hétérochromatine péricentromérique, et qu'elle s'associe aux séquences satellites, ainsi qu'aux IAPs et LINE1. Cette modification est aussi transitoirement perdue entre E12,5 et E13,5 (Daujat et al., 2009). Les études menées par Ng et al., 2013, sur les PGCs à E13,5, révèlent la présence de la modification H3K27ac au niveau des gènes impliqués dans la différenciation et le maintien de la pluripotence. Ils observent aussi un enrichissement de H3K27me3 au niveau des gènes du développement et de la différenciation des PGCs. Enfin, de hauts niveaux de H3K9me3, catalysés en partie par l'enzyme SETDB1, marquent certains éléments transposables de type ERV et LINE1 dans les PGCs E13,5 et participent à la répression au niveau de ces sites au moment où la méthylation de l'ADN est réduite (Liu et al., 2014).

### 3. Objectifs du projet

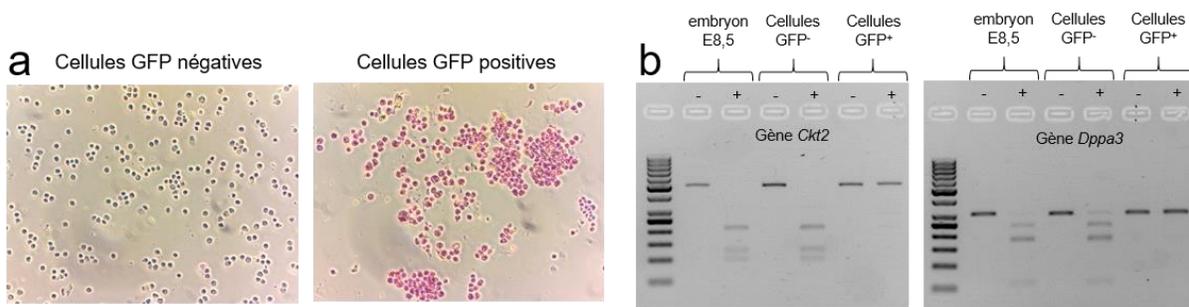
Les études publiées n'offrent qu'une vision partielle de la dynamique du méthylome dans les PGCs car les données de méthylome disponibles ne couvrent que quelques stades du développement. Par ailleurs, alors qu'il est bien établi que les PGCs subissent une déméthylation globale du génome et que la plupart des études passées ont essayé de comprendre les mécanismes de déméthylation, peu d'études se sont intéressées aux séquences qui maintiennent la méthylation dans les PGCs, à la fonction de cette méthylation et aux mécanismes moléculaires impliqués. J'ai donc initié une étude qui a trois objectifs :

- Décrire avec précision la dynamique du méthylome à tous les stades de développement des PGCs chez la souris.
- Etudier les séquences qui résistent à la déméthylation dans les PGCs
- Identifier les voies moléculaires responsables du maintien de la méthylation de l'ADN dans les PGCs



**Figure 3 : Stratégie de récupération des PGCs**

a. Croisement de souris mâles Oct4GFP1/1 avec des femelles sauvages dans le but d'obtenir des descendants hétérozygotes Oct4GFP1/0 qui expriment le marqueur GFP dans les PGCs. b. Stades d'étude de la cinétique de la 5mC dans les PGCs. Les croisements se déroulent sur la nuit et la vérification du bouchon vaginal est réalisé à 9h le matin. A chaque jour de la cinétique d'étude, les femelles gestantes sont euthanasiées à 9h. L'ensemble des gonades d'une même portée est regroupé, en faisant une distinction entre les sexes à partir de E12,5 en se basant sur les caractéristiques morphologiques. En effet, la gonade femelle est plus allongée et d'aspect duveteux, tandis que la gonade mâle est plus compacte et présente de stries visibles sous loupe binoculaire, correspondant à la mise en place des tubes séminifères. c. Exemples de résultats de cytométrie de flux permettant de purifier les cellules GFP positives. Le premier graphique représente l'intensité de la GFP en fonction du nombre total de cellules. Le second représente l'intensité de la GFP en fonction de l'auto-fluorescence dû à l'exposition des cellules au laser. Enfin, le dernier graphique représente l'intensité de la GFP en fonction de la taille des cellules. L'ensemble de ces paramètres m'ont permis de sélectionner les PGCs représentées ici par la fenêtre GFP + dans le deuxième graphe et la fenêtre P2 dans le troisième graphe.



**Figure 4 : Validation de la pureté des PGCs par marquage de la phosphatase alcaline et COBRA**

a. Après le tri par cytométrie de flux, les cellules positives et négatives pour le marqueur GFP ont été récupérées et soumises au test de la phosphatase alcaline. Ce test permet de visualiser, grâce à la coloration en rose des cellules, la présence d'un marqueur de surface spécifique des PGCs, la phosphatase alcaline. b. Analyse par COBRA des gènes *Ckt2* et *Dppa3* sur 3 échantillons : des cellules d'embryon E8,5, les cellules GFP<sup>-</sup> et les cellules GFP<sup>+</sup> obtenues lors du tri à E13,5. La présence de produits de digestion lorsque l'enzyme Taqα1 est présente (+) signifie que les séquences sont méthylées. Les résultats montrent que les gènes *Ckt2* et *Dppa3* sont méthylés dans les embryons E8.5 et les cellules GFP<sup>-</sup> mais pas dans les cellules GFP<sup>+</sup>, validant le fait que ces cellules sont des PGCs. (-) correspond au contrôle interne de l'expérience, c'est-à-dire l'absence de l'enzyme Taqα1.

## **B. Résultats**

### **1. Isolation des PGCs**

La première étape de mon projet a consisté à collecter les PGCs du stade E9,5 jusqu'au stade E17,5, tout en faisant la différence entre les gonades mâles et femelles à partir du stade embryonnaire E12,5 (stade d'engagement de la différenciation sexuelle) (**Figure 3-b**). Grâce à l'utilisation de souris transgéniques *Oct4*-GFP exprimant spécifiquement le marqueur GFP dans les PGCs (Kehler et al., 2004; Yoshimizu et al., 1999) (Matériels & méthodes ; **Annexe 1**), j'ai pu collecter les cellules germinales par tri cellulaire à tous les différents stades (**Figure 3a,c**). J'ai commencé l'étude par le stade E13,5 puisqu'il s'agit du stade où les PGCs sont les plus nombreux (environ 16 000 PGCs/embryons, selon mes expériences).

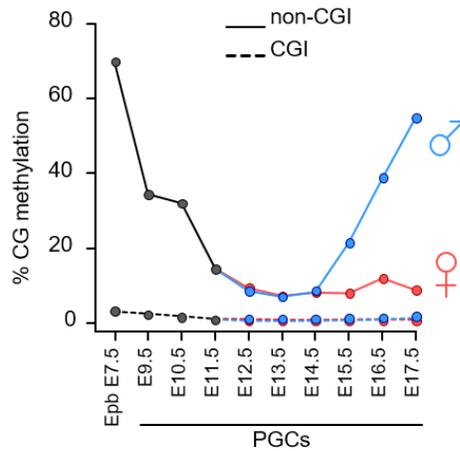
J'ai validé la pureté des PGCs triées par GFP de deux manières. Premièrement, j'ai réalisé un marquage de la phosphatase alcaline qui est un marqueur des cellules pluripotentes (**Figure 4a**). Dans cette expérience, seules les cellules GFP<sup>+</sup> présentent une coloration rose qui indique la présence du récepteur de la phosphatase alcaline au niveau de ces cellules, attestant de la robustesse du tri GFP (MacGregor et al., 1995). Deuxièmement, j'ai réalisé des validations par COBRA (**Figure 4b**). En effet, il est connu que les PGCs possèdent un génome globalement déméthylé à E13,5. J'ai mesuré par COBRA la méthylation au niveau des promoteurs des gènes *Ckt2* (*Casein Kinase 2*, un gène spécifique des testicules) et *Dppa3* (*Developmental Pluripotency Associated 3*, aussi appelé *Stella*) sur des cellules d'embryon E8,5, et sur les cellules GFP négatives et positives du tri. Les cellules embryonnaires et les cellules GFP-négatives présentent un profil de digestion indiquant que les deux gènes étudiés sont méthylés. Par contre, les cellules GFP-positives ne présentent pas de profil de digestion, ce qui indique que ces cellules ne sont pas méthylées au niveau des promoteurs de ces deux gènes et valide la pureté des PGCs.

La robustesse du tri GFP ayant été validée, j'ai collecté les PGCs aux différents stades embryonnaires. Leur ADNg (ADN génomique) a été extrait et a servi à la préparation de bibliothèques RRBS afin d'étudier la cinétique de la 5mC au cours du développement des PGCs.

### **2. Dynamique du méthylome au cours du développement des PGCs**

#### **a. Cinétique globale**

Pour l'analyse des niveaux globaux de méthylation par RRBS, nous avons séparé les CpG contenus dans les îlots CpG (CGIs) et les séquences non-CGI. En effet, les CGI, qui sont connus pour être hypométhylés, sont sur-représentés dans les données de RRBS et tendent à niveler le taux de méthylation vers le bas, ce qui ne serait pas représentatif de l'état de méthylation réel du



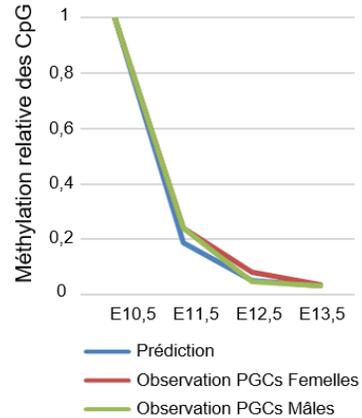
**Figure 5 : Cinétique de la méthylation (5mC) au cours du développement des PGCs**

Représentation du niveau de méthylation en pourcentage des CGIs ( — ) et des non-CGI ( --- ) au cours du développement des PGCs. A partir de la différenciation sexuelle, à E12,5, le niveau de 5mC est représenté en rouge pour les femelles et en bleu pour les mâles. Dès E9,5, le génome des PGCs est déméthylé en comparaison avec un épiblaste à E7,5. Par la suite, l'ensemble du génome des PGCs entre dans une vague de déméthylation, indépendamment des sexes, jusqu'à E13,5 où le génome se retrouve globalement déméthylé. Après E13,5, le génome des PGCs mâles se reméthyle tandis que celui des femelles reste globalement déméthylé. Epb = Epiblaste.

**a**

	E10.5	E11.5	E12.5 m	E12.5 f	E13.5 m	E13.5 f
Tri #1	147	1055	7675	4060	37750	29375
Tri #2	233	840	4571	2885	19925	12075
Tri #3	290	1238	5136	3955	8557	9333
Tri #4	151				16171	20900
Tri #5	298				13840	
Tri #6	146					
Tri #7	111					
<b>Moyenne</b>	<b>197</b>	<b>1044</b>	<b>5794</b>	<b>3633</b>	<b>19249</b>	<b>17921</b>

**b**



**Figure 6 : La vague de déméthylation des PGCs est compatible avec un mécanisme de déméthylation passif**

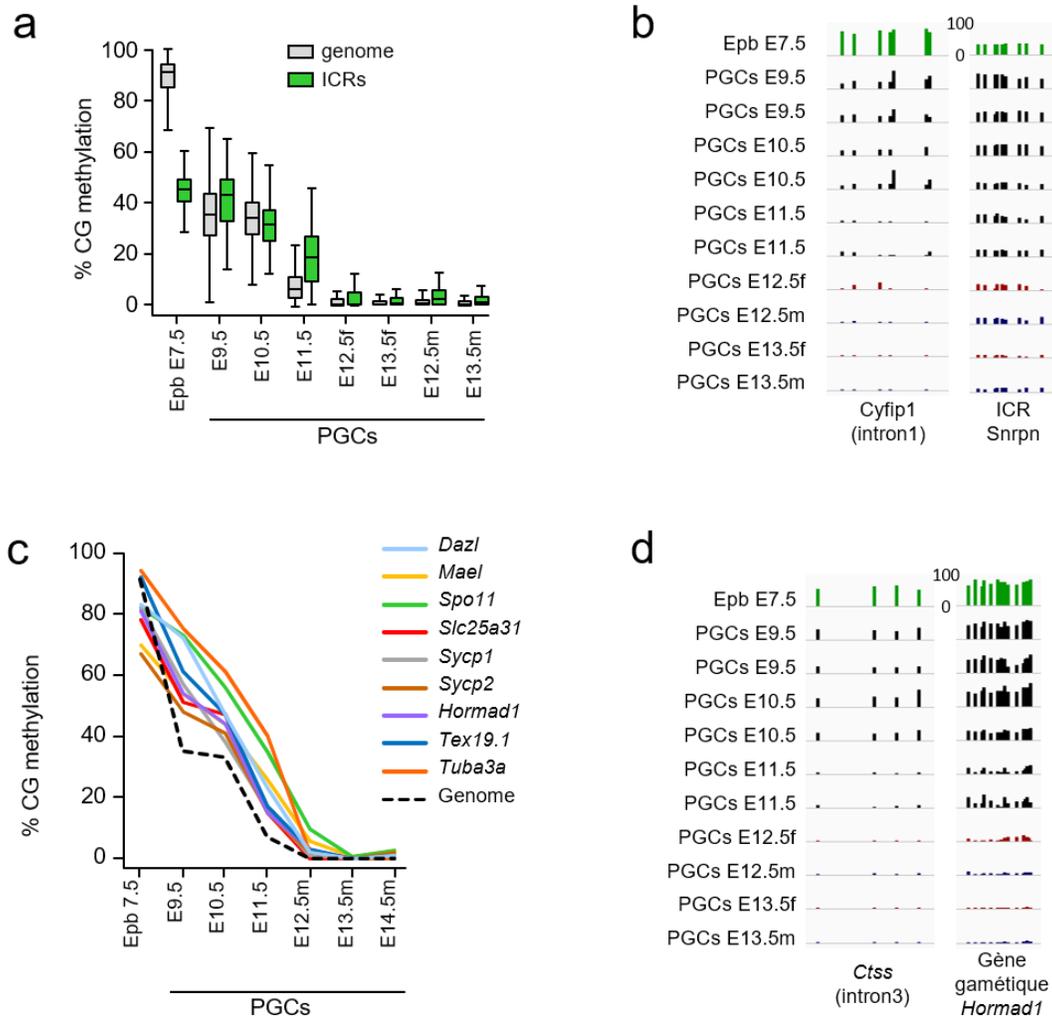
**a.** Tableau récapitulatif du nombre de PGCs obtenus par embryon au cours des différents tris cellulaires. A la vue du nombre de PGCs, on remarque qu'il y a environ deux divisions cellulaires entre chaque stade. **b.** Comparaison de la cinétique de déméthylation des PGCs avec un modèle théorique passif. Le graphique de prédiction dans le cas d'une déméthylation passive (basé sur une baisse de 50% de méthylation à chaque division cellulaire) est en bleu. Les données de méthylation des CpG obtenues par RRBS ont été normalisées par rapport à la valeur à E10.5. La comparaison révèle que la cinétique de déméthylation de PGCs est très proche de la cinétique théorique passive. f = femelles; m = mâles.

génomique. De ce fait, nous étudions la méthylation des CpG non-CGI afin d'estimer la méthylation globale du génome au cours du développement des PGCs.

Dans l'épiblaste E7,5, le niveau global de méthylation du génome est d'environ 70%. Dans les PGCs E9.5, le taux de 5mC est déjà diminué à environ 35%, indiquant que la déméthylation des PGCs démarre antérieurement à E9.5 (**Figure 5**). Ces résultats sont en accord avec les précédentes études ([Guibert et al., 2012](#); [Seisenberger et al., 2012](#); [Seki et al., 2005](#)). A partir de E9.5, le niveau de méthylation diminue à 31,5% à E10,5, puis diminue de façon plus drastique à E11,5 à 14%. Le génome des PGCs se déméthyle graduellement jusqu'à E13,5 où le génome des PGCs mâles ou femelles, est globalement déméthylé (les femelles sont à 7% et les mâles à 6,6%) (**Figure 5**). Par la suite, dans les PGCs femelles, le niveau de méthylation reste bas jusqu'à E17,5 (8%), alors que les PGCs mâles se reméthylent à partir de E14,5 jusqu'à E17,5 (~54,5%) (**Figure 5**), mettant en évidence une activité de méthylation *de novo* dans les cellules germinales mâles. Pour conclure, les génomes des PGCs semblent être soumis à deux phases de reprogrammation épigénétique. Jusqu'à E13,5, les génomes subissent une vague de déméthylation massive qui est suivie par une vague de reméthylation à partir de E14,5, uniquement chez les mâles. Ces résultats sont globalement en accord avec les études précédentes ([Guibert et al., 2012](#); [Seisenberger et al., 2012](#); [Wang et al., 2014b](#)).

#### **b. La cinétique de déméthylation des PGCs suggère une déméthylation passive**

Par la suite nous avons cherché à savoir si cette vague de déméthylation dans les PGCs entre E9.5 et E13,5 était liée à un processus de déméthylation actif ou passif. Dans cette optique, nous avons étudié le nombre de PGCs isolés à chaque tri aux stades embryonnaires E9,5, E10,5, E11,5, E12,5 et E13,5 (**Figure 6a**). Le comptage des PGCs aux différents stades suggère que ces cellules effectuent deux divisions cellulaires entre chaque stade. En prenant en compte ce résultat, nous avons réalisé une simulation de déméthylation passive du génome des PGCs en fonction du stade embryonnaire (**Figure 6b**). Pour réaliser cette simulation, nous avons normalisé à la valeur 1 le taux de méthylation des CpG dans les PGCs E10,5 puis estimé une courbe de déméthylation passive théorique en considérant que la méthylation perd 50% à chaque division cellulaire. Ainsi, à E11,5, le taux de méthylation théorique a été calculé de la façon suivante :  $(\text{Niveau de méthylation à E10,5} = 1) \times (\text{nombre de PGCs à E10,5}) / (\text{nombre de PGCs à E11,5})$ . Ce même calcul a été utilisé pour estimer le niveau de 5mC aux autres stades embryonnaires. Enfin, pour comparer la courbe de prédiction avec nos données, nous avons normalisé le niveau de méthylation des CpG obtenu par RRBS à chaque stade par rapport à la valeur obtenue à E10,5. La courbe normalisée de nos observations de méthylation est très proche de la courbe théorique passive (**Figure 6b**). Ainsi, il semblerait que la déméthylation du génome des PGCs



**Figure 7 : Cinétiques de déméthylation des ICRs et des gènes gamétiques dans les PGCs**

**a.** Graphique représentant la cinétique de déméthylation dans l'ensemble du génome (en sélectionnant les CpG méthylés >50% dans l'épiblaste E7.5, gris) et dans les régions de contrôle de l'empreinte [ICRs] (vert). Les ICRs se déméthylent de manière plus graduelle que le reste du génome, ce qui est visible à E11,5. **b.** Illustration à partir du logiciel IGV de la déméthylation de l'ICR *Snrpn* par rapport à une séquence contrôle (ici le premier intron du gène *Cyfip1* qui est un gène adjacent à *Snrpn*). Chaque barre représente le pourcentage de méthylation d'un CpG. **c.** Cinétique de déméthylation de certains gènes gamétiques dans les PGCs en comparaison avec l'ensemble du génome (en sélectionnant les CpG méthylés >50% dans l'épiblaste E7.5). **d.** Illustration à partir du logiciel IGV de la déméthylation du promoteur du gène gamétique *Hormad1* par rapport à une séquence contrôle (ici le troisième intron du gène *Ctss* qui est un gène adjacent à *Hormad1*). Epb = Epiblaste; f= femelle; m=mâle; ICR= Imprinting Control Region.

observée entre E10,5 et E13,5 soit globalement compatible avec un processus de déméthylation passif de l'ADN.

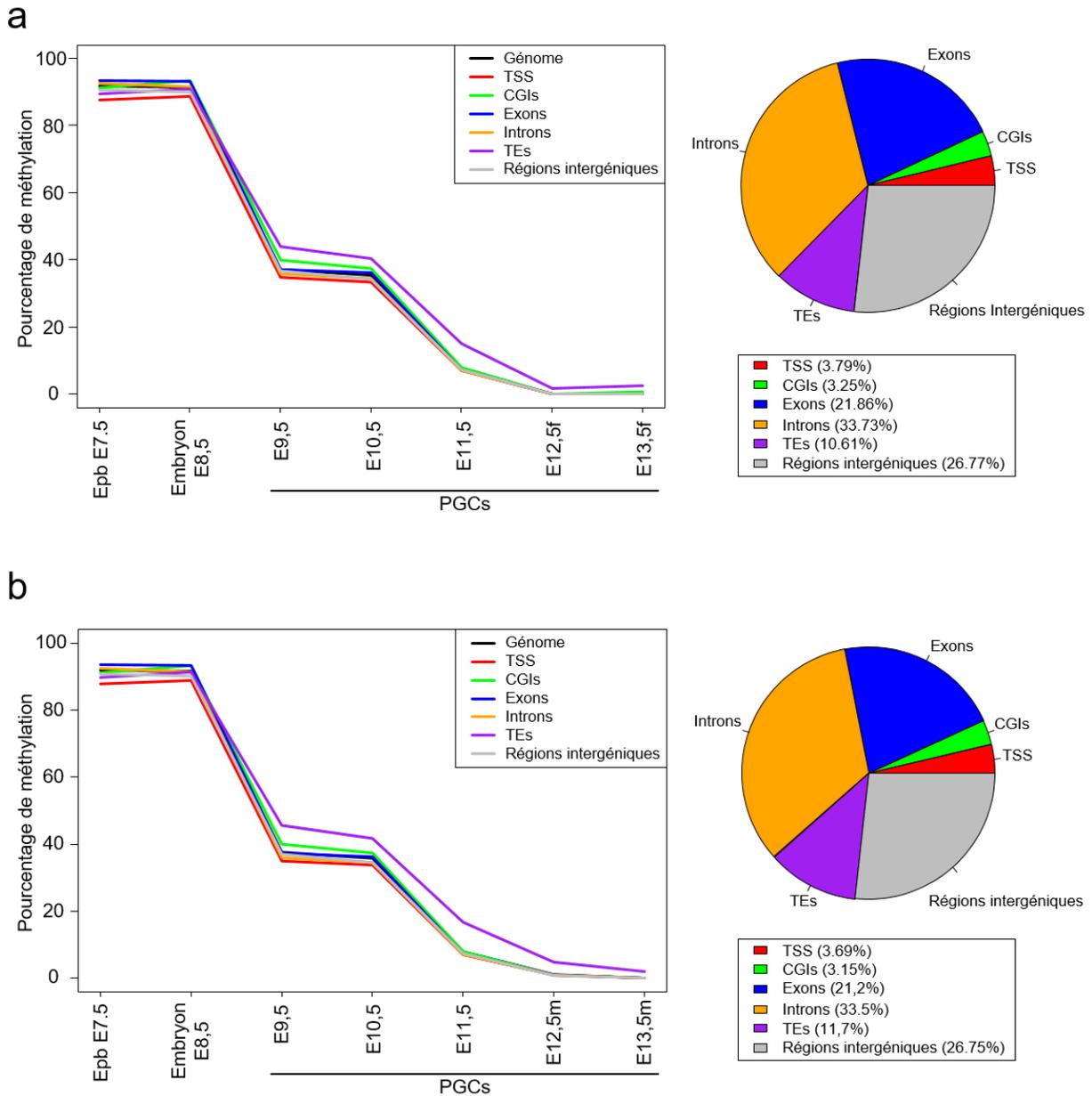
### c. Certaines régions se déméthylent plus tardivement que le génome

Bien que la déméthylation du génome des PGCs semble être globalement liée avec un système de déméthylation passif de l'ADN, de précédentes études ont suggéré que les gènes soumis à empreintes (ICRs) et certaines gènes gamétiques se déméthylent plus tardivement que le reste du génome des PGCs (Hajkova 2002; Guibert et al., 2012; Seisenberger et al., 2012). Ainsi, nous avons analysé nos données de RRBS afin de savoir si nos résultats corroborent les précédentes analyses menées par d'autres équipes.

Nous nous sommes tout d'abord intéressés aux régions de contrôle des gènes soumis à empreintes (ICRs) (Figure 7a-b). Pour servir de référence, nous avons sélectionné tous les CpGs dont le niveau de méthylation est supérieur à 50% dans l'épiblaste à E7,5. Au niveau de l'épiblaste E7,5, les ICRs sont normalement méthylés à 50% tandis que le niveau global de méthylation des séquences de référence dans le reste du génome est d'environ 90%. Dans les PGCs E9.5, le niveau global de méthylation du génome est déjà fortement diminué de plus de 40% par rapport à l'épiblaste, alors que le taux de 5mC des ICRs diminue très peu et est supérieur à celui du reste du génome (Figure 7a). Par la suite, le niveau de méthylation des ICRs est équivalent à celui du génome à E10,5, puis bien supérieur à E11,5 (~19%) en comparaison au reste du génome (~8%) (Figure 7a-b). Enfin, à partir de E12,5, que ce soit chez les femelles ou les mâles, les ICRs sont déméthylés tout comme le génome (Figure 7a-b). Par conséquent, les données indiquent que les ICRs présentent un retard de déméthylation par rapport au reste du génome (Figure 7a-b).

Nous avons ensuite étudié la cinétique de déméthylation au niveau des gènes gamétiques (Figure 7c-d). Tout comme les ICRs, certains gènes gamétiques présentent un profil de déméthylation retardé comparé à celui du génome, avec une déméthylation graduelle entre E9.5 et E12,5 (Figure 7c-d). Cette déméthylation tardive est particulièrement marquée pour les gènes *Dazl*, *Mael*, *Spo11*, *Slc25a31*, *Sycp1*, *Sycp2*, *Hormad1*, *Tex19.1* et *Tuba3a* (Figure 7c-d). Lorsqu'on y regarde de plus près, chacun de ces gènes gamétiques possède un profil de déméthylation qui lui est propre mais tous se retrouvent globalement déméthylés à E12,5 (Figure 7c-d).

Par la suite, nous nous sommes intéressés à l'ensemble des séquences du génome des PGCs afin de voir si certains types de séquences présentaient une cinétique de déméthylation différente (Figure 8a-b). Nous avons hiérarchisé le génome en comparant la cinétique de déméthylation



**Figure 8: Comparaison des cinétiques de déméthylation dans les PGCs embryonnaires.**

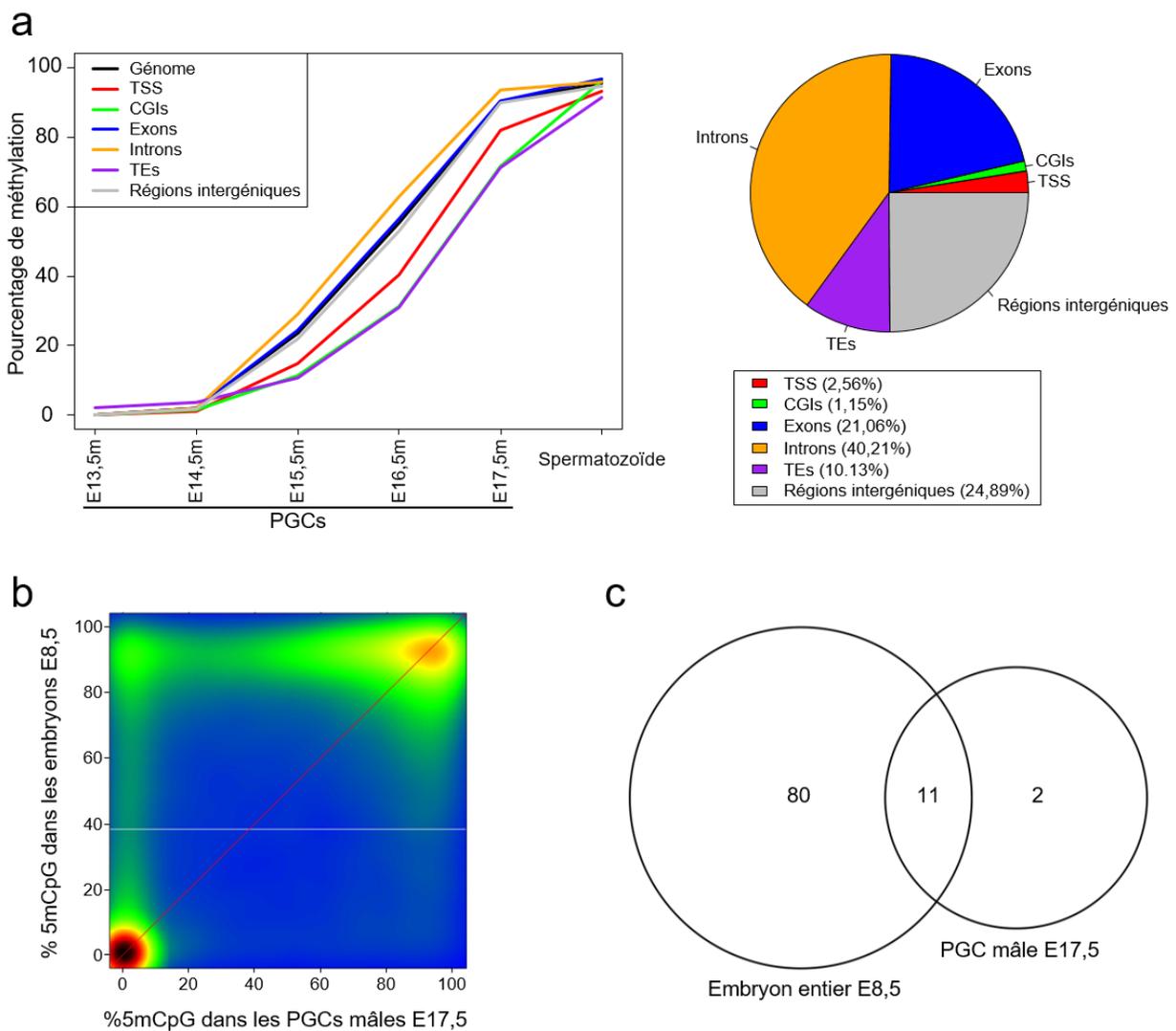
Pour ces graphiques, nous avons sélectionné les CpG méthylés >50% dans l'épiblaste E7,5 et <10% dans les PGCs E13,5. Le graphique **a** représente la cinétique de déméthylation dans les différents compartiments génomiques par rapport au génome entier chez les PGCs femelles et le graphique **b** représente la même chose chez les mâles. Les graphiques circulaires nous renseignent sur la proportion de chaque type de séquence dans les CpG analysés. Que ce soit chez les femelles ou les mâles, ce sont majoritairement les introns, les exons, les régions intergéniques ainsi que les TEs qui se déméthylent. De plus, on observe que le profil de déméthylation des introns, des exons, des séquences intergéniques, des TSS et des CGI sont globalement équivalents. Seuls les TEs présentent une cinétique de déméthylation légèrement décalée et ce, indépendamment du sexe. Epb = Epiblaste; f= femelle; m=mâle; TE = Eléments transposables ; TSS = promoteur de gènes (compris entre -1000 et +1000 pb) ; CGIs = îlots CpG.

des CpGs contenus dans les promoteurs (compris entre -1000 et 1000pb du TSS), les CGIs, les exons, les introns, les séquences intergéniques et les éléments transposables (TEs). Pour cette étude, nous avons filtré nos données afin de sélectionner les CpG avec un taux de méthylation supérieur à 50% dans l'épiblaste E7,5 et inférieur à 10% dans les PGCs E13,5. Comme l'attestent les graphiques circulaires (**Figure 8a-b**), les CpG ainsi sélectionnés sont contenus principalement dans les introns (33.73% chez les femelles et 33.5% chez les mâles), puis les régions intergéniques (26.77% chez les femelles et 26.75% chez les mâles), les exons (21.86% chez les femelles et 21.2% chez les mâles) et les TEs (10.61% chez les femelles et 11.7% chez les mâles). Les CGIs ne représentent qu'une faible proportion des CpG sélectionnés (respectivement 3.25% pour les femelles et 3.15% pour les mâles) car sur les 15 000 CGI du génome couverts par les données RRBS, seuls 500 sont méthylés dans l'épiblaste E7,5. Les résultats montrent que la cinétique de déméthylation de ces régions est, dans l'ensemble, semblable à celle du génome et ce, indépendamment du sexe (**Figure 8a-b**). Seuls les éléments transposables dont la méthylation est inférieure à 10% à E13,5 présentent un léger retard de déméthylation par rapport aux autres séquences. Je reviendrai plus tard sur le profil de méthylation particulier des éléments transposables dans le chapitre 3.

#### d. La phase de reméthylation

Après avoir analysé la phase de déméthylation des PGCs, nous nous sommes intéressés à la phase de reméthylation. Comme nous l'a montré la cinétique de méthylation des PGCs (**Figure 5**), cette vague de reméthylation touche uniquement le génome des PGCs mâles. De ce fait, nous avons regardé la cinétique de reméthylation dans les différents compartiments génomiques chez les PGCs mâles (**Figure 9a**). Pour cela, nous avons sélectionné les CpG méthylés à moins de 10% à E13,5 et plus de 30% à E17,5. Nous avons aussi ajouté le méthylome d'un spermatozoïde comme point final du développement des gamètes mâles.

Une fois encore ce sont les introns qui sont le plus représentés dans les séquences qui se méthylent à E17,5 (40.21% des CpG), suivis des régions intergéniques (24.89%), des exons (21.06%) et des transposons (10.13%). Viennent ensuite les TSS (2.56%) et les CGIs (1.15%) qui se méthylent très rarement dans les PGCs mâles (**Figure 9a**). Concernant la cinétique de reméthylation, le gain de 5mC est très faible (environ 2%) pour l'ensemble des différentes séquences génomiques entre E13,5 et E14,5. Ce n'est qu'à partir de E14,5 que le génome des PGCs se reméthyle de façon conséquente et continue jusqu'à E17,5 pour atteindre un niveau de méthylation à E17,5 proche de celui des spermatozoïdes. De manière intéressante, on observe que cette cinétique de reméthylation n'est pas la même pour les différents compartiments



**Figure 9 : Cinétiques de reméthylation dans les PGCs mâles.**

**a.** Le graphique représente la cinétique de reméthylation de différents compartiments génomiques chez les PGCs mâles. Pour ce graphique, nous avons sélectionné les CpG étant méthylés à moins de 10% dans le PGCs E13,5 et plus de 50% dans les PGCs E17,5. Le graphique en secteur résume la proportion de chaque type de séquences. Cette phase de reméthylation touche majoritairement les introns, les exons, les régions intergéniques et les TEs. Néanmoins, certains TSS et CGIs se reméthylent aussi dans les PGCs mâles. On observe que la cinétique de reméthylation des TSS, des CGIs et des TEs est plus lente que les autres types de séquence. **b.** Graphique de corrélation comparant le niveau de méthylation dans les embryons à E8,5 (en abscisse) et dans les PGCs mâle à E17,5 (en ordonnée). Majoritairement, les régions qui sont méthylées à plus de 80% dans l'embryon le sont aussi dans les PGCs mâle à E17,5. Néanmoins, il existe aussi de nombreuses régions qui sont méthylées dans l'embryon mais pas dans les PGCs (coin supérieur gauche). **c.** Graphique représentant le nombre de CGIs associés à des TSS méthylés à plus de 30%, soit dans l'embryon à E8,5, soit dans les PGCs mâles à E17,5, soit dans les deux.

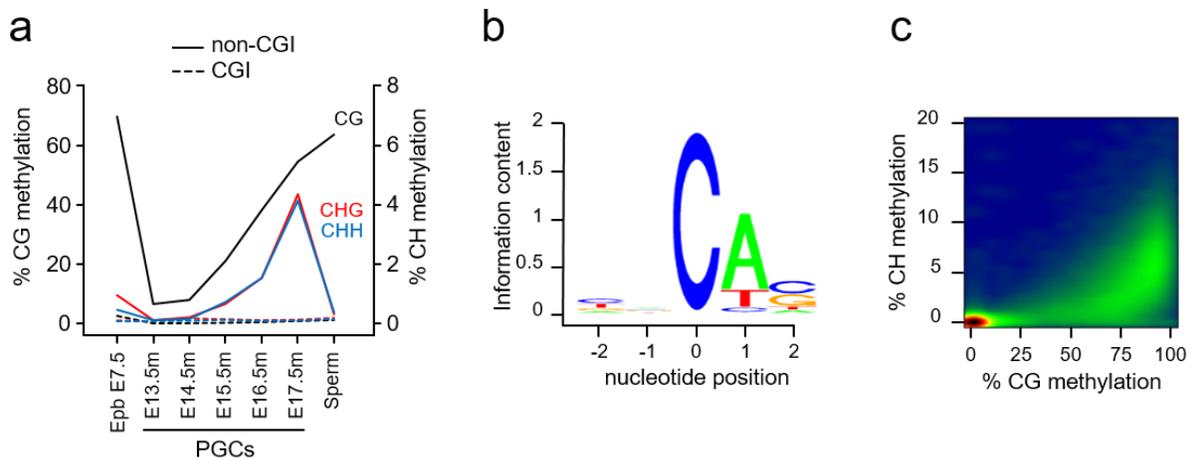
Epb = Epiblaste; m = mâle; TEs = Eléments transposables; TSS = Transcription Start Site; CGI = îlots CpG.

génomiques (**Figure 9a**). Les séquences qui se reméthylent le plus rapidement sont les introns qui passent d'un niveau moyen de méthylation de 3% à E14,5, à 31% à E15,5, puis à 60% à E16,5 et enfin à 88% à E17,5 où le taux de 5mC est relativement proche de celui d'un spermatozoïde (~90%). Les autres types de séquences accumulant rapidement de la méthylation sont les exons et les régions intergéniques, dont les courbes se chevauchent (**Figure 9a**) avec un pourcentage de méthylation de 3% à E14,5, 25% à E15,5, 53% à E16,5 et 83% à E17,5. Les compartiments génomiques qui se reméthylent le plus lentement sont les TSS, et surtout les CGIs et les éléments transposables. La méthylation de ces derniers augmente de façon beaucoup plus graduelle : 6% à E14,5, 14% à E15,5, 35% à E16,5 et 70% à E17,5 (**Figure 9a**).

Il existe deux grandes vagues de méthylation du génome chez la souris : dans l'embryon après l'implantation et dans les cellules germinales. Nous avons donc voulu savoir si les mêmes séquences sont ciblées par la méthylation dans l'embryon E8,5 et les PGCs mâles E17,5. Pour cela nous avons comparé le méthylome d'embryons E8,5 à celui des PGCs E17,5 (**Figure 9b**). L'analyse générale du graphe nous informe que, bien que la plupart des CpG soient méthylés de manière identique dans les embryons E8,5 et les PGCs mâles E17,5, il existe un grand nombre de séquences qui acquièrent la méthylation spécifiquement dans les embryons E8,5 et pas dans les PGCs mâles E17,5 (partie supérieure gauche du graphique, **Figure 9b**). Parmi ces séquences, nous nous sommes intéressés aux îlots CpG dans les promoteurs de gènes. Alors qu'une centaine de promoteurs à îlot CpG se méthylent dans l'embryon E8,5, il n'y en a presque aucun dans les PGCs. Ces promoteurs de gènes correspondent majoritairement aux gènes gamétiques qui se méthylent dans l'embryon. Ainsi, les PGCs mâles acquièrent un méthylome qui se distingue de l'embryon notamment par l'hypométhylation des promoteurs de gènes gamétiques (**Figure 9c**).

#### e. Les PGCs mâles accumulent de la méthylation non-CG

Par la suite nous avons regardé plus précisément la cinétique de méthylation des cytosines dans les contextes CG, CHG et CHH (H=A, T ou C), au cours de la vague de reméthylation des PGCs mâles (**Figure 10a**). Comme nous l'avons vu précédemment, le niveau de méthylation des cytosines dans un contexte CpG augmente dans les PGCs mâles entre E13,5 et E17,5. On constate aussi un gain de méthylation au niveau des cytosines présentes dans un contexte CHG



**Figure 10 : Accumulation de méthylation non-CG dans les PGCs mâles**

**a.** Cinétique de reméthylation des PGCs mâles dans les séquences CG et non-CG (CHH en bleu et CHG en rouge). On observe que les PGCs mâles accumulent un niveau non négligeable de méthylation CHG et CHH, même si les niveaux de méthylation ne sont pas comparables à celui des CpG. Par contre, cette méthylation non-CG disparaît dans une gamète mâle mature. **b.** Séquence logo définissant les motifs nucléotidiques préférentiellement méthylés dans le cadre de la méthylation non-CG dans les PGCs mâles E17,5. La méthylation non-CG a principalement lieu dans un contexte CA. **c.** Graphique de corrélation entre la méthylation CG et non-CG mesurée dans des fenêtres de 400pb dans les PGCs mâles E17,5. On constate que plus les CpG sont méthylés, plus on observe de la méthylation non-CG. Epb = Epiblaste; m=mâle.

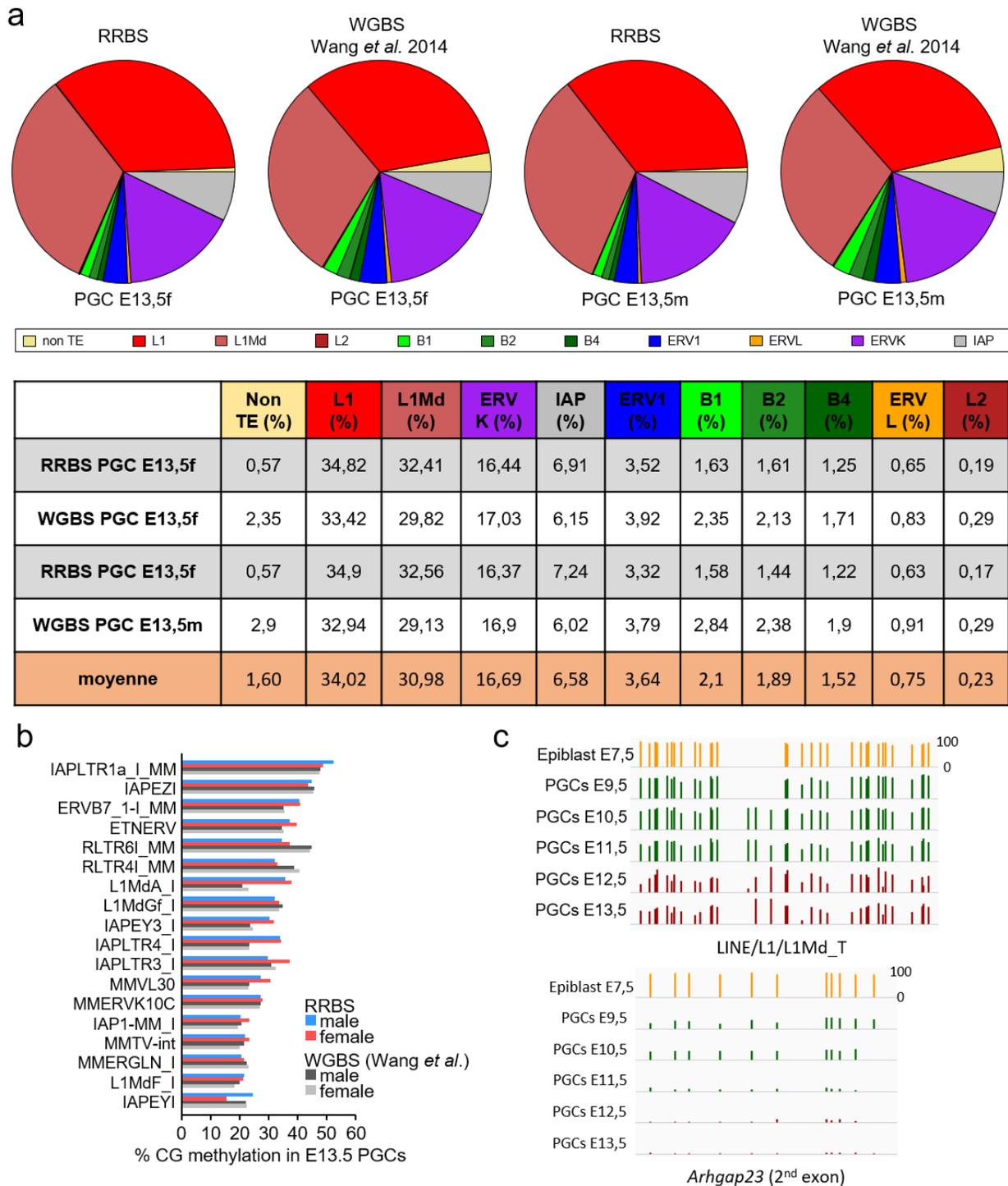
et CHH à partir de E15,5 jusqu'à E17,5 : alors que ces cytosines sont méthylées à E13,5 et E14,5, elles sont méthylées à environ 0.8% à E15,5, puis à environ 1.8% à E16,5, pour enfin atteindre un peu plus de 4% à E17,5 (**Figure 10a**). Par contre, cette méthylation non-CpG est absente au niveau des spermatozoïdes. Par la suite, nous avons voulu savoir dans quel contexte nucléotidique ces cytosines se méthylent préférentiellement. Nous avons réalisé une analyse LOGO qui révèle que ces cytosines hors contexte CpG sont préférentiellement méthylées dans un contexte CpA (**Figure 10b**). Enfin, nous avons cherché à savoir s'il existe une corrélation entre le pourcentage de méthylation des cytosines en contexte CG et CH (**Figure 10c**). Le graphique nous informe qu'il existe une corrélation positive entre le niveau de méthylation des cytosines dans un contexte CG et dans un contexte CH. En effet, plus les cytosines CG sont méthylées, plus les cytosines CH le sont elles aussi. En conclusion nous montrons que les cellules germinales mâles accumulent transitoirement de la méthylation CA au cours du développement, principalement au niveau des mêmes séquences que celles qui acquièrent la méthylation CG.

### **3. Analyse des régions résistantes à la vague de déméthylation**

#### **a. Quelles régions résistent à la vague de déméthylation des PGCs ?**

Afin d'étudier s'il existe des séquences génomiques qui résistent à la vague de déméthylation dans les PGCs, nous avons filtré nos données de RRBS afin de sélectionner les régions génomiques dont le niveau de méthylation est supérieur à 20% dans les PGCs E13,5. Etant donné que la technique du RRBS ne permet pas de couvrir tout le génome et afin d'analyser plus finement les régions résistantes à la vague de déméthylation, nous avons également comparé nos données de RRBS avec les données de WGBS publiées par (Wang et al., 2014b) dans les PGCs E13,5.

Le filtre à 20% nous indique qu'il existe de nombreuses régions résistantes à la phase de déméthylation dans les PGCs E13,5. De manière remarquable il s'agit principalement des éléments transposables (TEs) (**Figure 11a**). Ces derniers représentent plus de 99% de l'ensemble des régions résistantes dans les données de RRBS et plus de 97% des régions dans les résultats du WGBS. La majorité des éléments transposables résistants sont les L1 (34.02%), puis les L1Md (30.98%), les ERVK (16.69%) et les IAP (6.58%). Puis dans une plus faible proportion, on retrouve les ERV1 (3.64%), les B1 (2.10%), les B2 (1.89%), les B4 (1.52%), les ERVL (0.75%) et finalement les L2 (0.23%) (**Figure 11a**). Par la suite, nous avons analysé plus précisément le niveau de méthylation des différentes familles de transposons en alignant les données de séquençage sur les séquence consensus des familles de transposons classifiées dans la base de données RepBase (**Figure 11b-c**). Parmi les transposons les plus méthylés se trouvent majoritairement



**Figure 11 : Les régions résistantes à la vague de déméthylation des PGCs sont majoritairement des TE**

**a.** Pour ces graphiques en secteur, nous avons sélectionné les CpG méthylés à plus de 20% dans nos RRBS de PGCs femelles (f) et mâles (m) E13,5 ainsi que dans les WGBS de PGCs femelles et mâles réalisés par l'équipe de Liu (Wang *et al.* 2014). La majorité des régions résistantes correspondent aux TEs et notamment aux L1, L1Md et ERVK. **b.** Graphique en barres représentant le pourcentage de méthylation des différentes familles de TEs les plus méthylées dans les PGCs E13,5 dans nos RRBS et dans le WGBS de Wang. **c.** Illustration à partir du logiciel IGV de la méthylation d'un transposon LINE/L1/L1Md\_T et du gène *Arhgap23* qui lui est immédiatement adjacent. Chaque barre représente la méthylation d'un CpG.

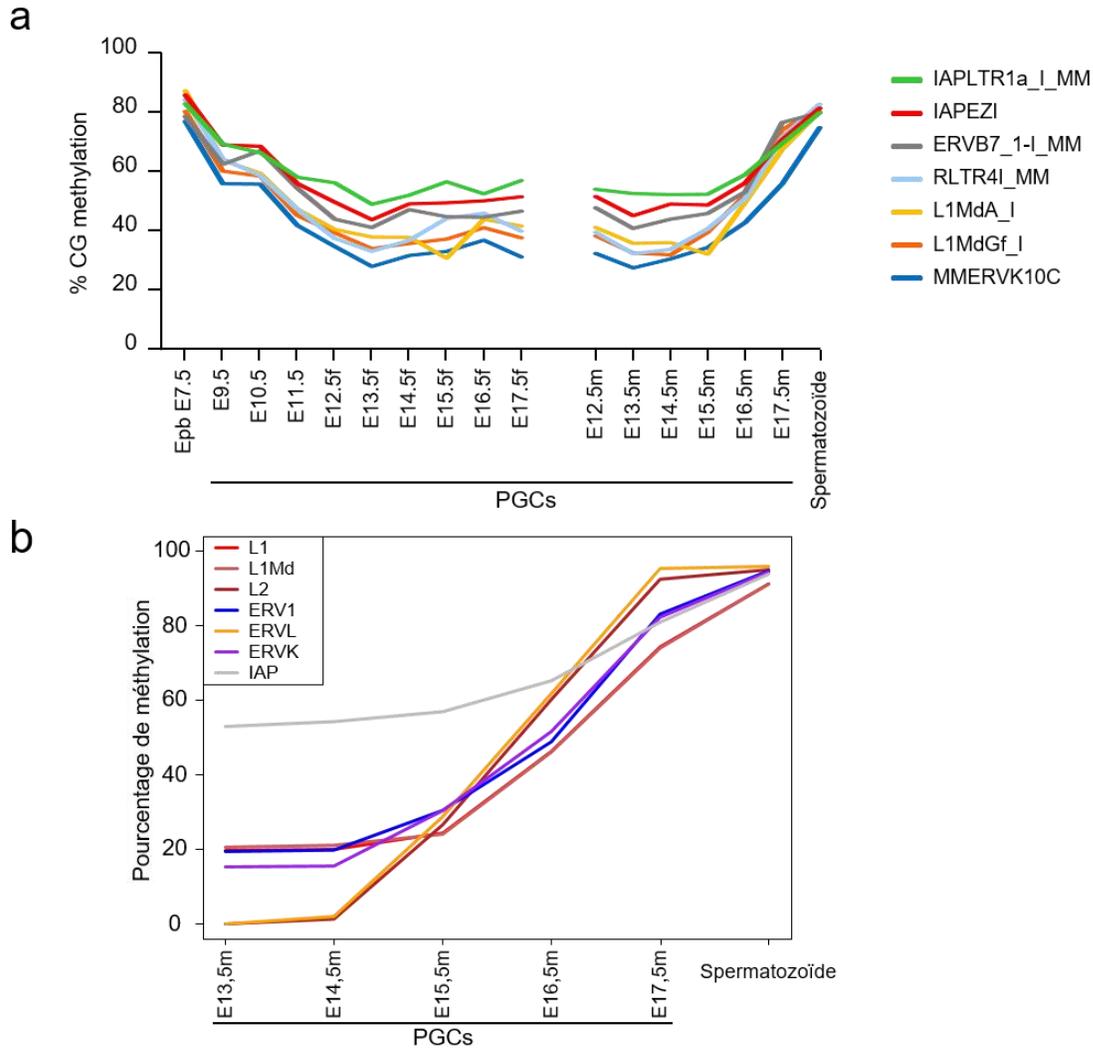
des transposons des familles IAP, ERVK, ERV1 et L1Md qui correspondent aux familles de transposons les plus jeunes et les plus actifs dans le génome de souris (**Figure 11b**), comme le montre l'exemple de la **Figure 11c**. Enfin, il n'y a pas de différence de méthylation en fonction des sexes et nos données de RRBS corrélient avec celles de WGBS de (Wang et al., 2014b) (**Figure 11b**).

Une très faible proportion des régions résistantes à la vague de déméthylation ne sont pas associés à des éléments transposables et représentent, selon nos données de RRBS, 0.57% des régions dans les PGCs femelles et 0.58% dans les PGCs mâles (**Figure 11a**). L'analyse WGBS réalisée par Wang et al., 2014 nous donne un peu plus de précision grâce à sa meilleure couverture du génome, notamment en dehors des CGIs, et nous révèle que les régions non-TEs résistantes à la vague de déméthylation représentent 2.35% des séquences résistantes chez les PGCs femelles et 2.90% chez les PGCs mâles (**Figure 11a**). Chez la femelle, nos analyses RRBS révèlent qu'il existe 18 régions résistantes dont deux au niveau des TSS des gènes *Zfp513* et *Cdr1*. L'analyse WGBS révèle l'existence de 79 régions dont cinq au niveau des TSS des gènes *Zfp513*, *Astl*, *Mir7037*, *Mir6922* et *Krt6a*. Chez les mâles, nos analyses RRBS indiquent l'existence de 17 régions résistantes dont deux associées à un TSS (*Inpp4a* et *Cdr1*). L'analyse WGBS révèle quant à elle la présence de 99 régions dont 16 au niveau de promoteurs. Compte tenu du faible nombre de gènes, l'analyse ontologique n'a pas permis de révéler des fonctions biologiques particulières associées à ces gènes.

#### b. Cinétique des éléments transposables au cours du développement des PGCs.

Ensuite, afin de savoir si la méthylation des transposons dans les PGCs E13,5 résulte d'une activité de maintien ou de méthylation *de novo*, nous avons sélectionné les familles de transposons les plus méthylées et étudié leur cinétique de méthylation au cours du développement des PGCs. Même si ceux-ci se déméthylent légèrement entre E9,5 et E13,5, ils maintiennent un haut niveau de méthylation à tous les stades du développement des PGCs, et ce indépendamment du sexe (**Figure 12a**). Cela suggère que la méthylation des transposons dans les PGCs résulte d'une activité de maintien sélective qui les protège de la vague de déméthylation. Après E13,5, ces transposons se reméthylent chez les PGCs mâles, jusqu'à atteindre un niveau d'environ 75% de 5mC à E17,5.

A la vue des résultats obtenus chez les PGCs mâles, nous avons regardé plus attentivement l'ensemble des éléments transposables au cours de la cinétique de reméthylation (**Figure 12b**).

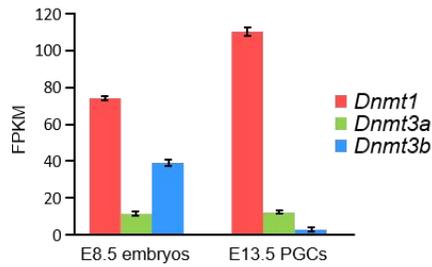


**Figure 12 : Cinétique de méthylation des TEs au cours du développement des PGCs**

**a.** Graphique représentant l'évolution du niveau de méthylation des principales familles de transposons résistantes à la phase de déméthylation. Chez la femelle, dès E14,5, l'ensemble des familles se reméthylent mais cela n'excède pas 60% à E17,5 alors que chez le mâle, l'ensemble des familles atteint ~70% de méthylation au stade E17,5. **b.** Ce graphique a été réalisé sans aucun filtre (235 872 CpG communs ont été analysé). Le graphique nous informe de la proportion de chacune des familles d'éléments transposables affectée par la phase de reméthylation. L'ensemble des familles de TEs se reméthylent dans les gamètes mâles afin de redevenir pleinement méthylés dans le sperme. Alors que la famille des ERVL et les L2 sont totalement déméthylées à E13,5, ces dernières présentent une courbe de reméthylation similaire. La famille des IAP représente les éléments les plus résistants à E13,5 (environ 54% de 5mC). A E17,5, seules les familles des ERVL et des L2 sont méthylés au même niveau qu'un spermatozoïde (~97%).

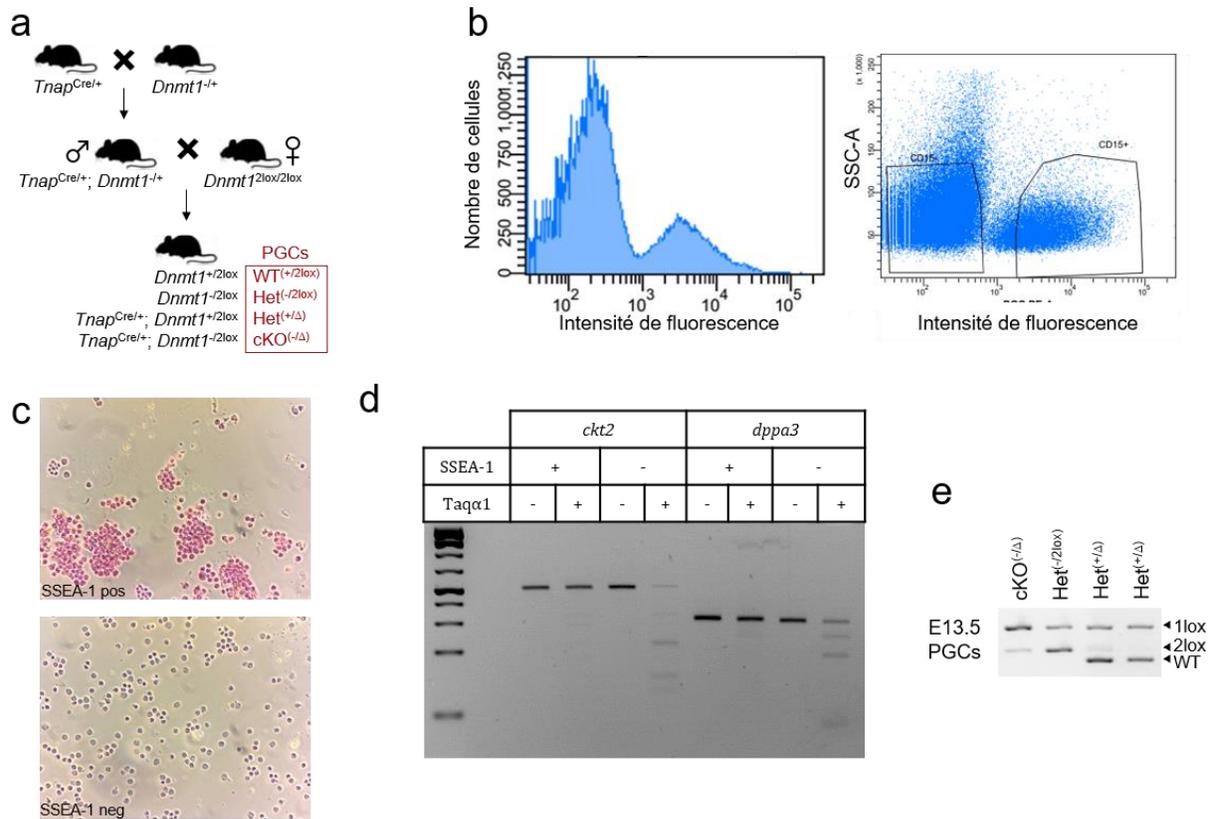
A E13,5, on remarque qu'à ce stade ce sont principalement les IAPs qui sont méthylés à hauteur de 55%. Les L1, les L1Md, les ERV1 et les ERVK possèdent un niveau de 5mC d'environ 20%. Seuls les ERVL et les L2 sont hypométhylés à E13,5 (**Figure 12b**). A partir de E14,5, l'ensemble des familles de transposons se reméthylent afin d'atteindre un niveau de méthylation élevé à E17,5. Cependant on constate que les IAPs, L1, L1Md, ERV1 et ERVK se reméthylent avec une cinétique retardée par rapport aux ERVL, L2 et au reste du génome. Ainsi les familles de transposons qui résistent à la déméthylation dans les PGCs avant E13,5 sont aussi celles qui se reméthylent tardivement dans les PGCs mâles après E14,5. (**Figure 12b**).

En résumé de cette partie, nous avons constaté que des régions résistent à la vague de déméthylation des PGCs et qu'il s'agit majoritairement d'éléments transposables. La suite de mon travail a consisté à étudier quels sont les acteurs moléculaires responsables du maintien de la méthylation de l'ADN au niveau de ces régions dans les PGCs.



**Figure 13 : Expression des gènes *Dnmt* dans les PGCs**

Ce graphique représente le niveau d'expression (en fpkm) des trois principaux gènes impliqués dans la méthylation de l'ADN (*Dnmt1*, *Dnmt3a* et *Dnmt3b*) dans l'embryon à E8,5 ainsi que dans les PGCs à E13,5. Ces données proviennent de RNA-seq que nous avons réalisés. *Dnmt1* est exprimé à haut niveau dans l'embryon et les PGCs, alors que le niveau d'expression de *Dnmt3a* et *Dnmt3b* est plus faible dans les PGCs E13,5. FPKM = fragments per kilobase of transcript per million mapped reads.



**Figure 14 : Isolation de PGCs E13,5 portant une inactivation conditionnelle de *Dnmt1***

**a.** Stratégie de croisement. Des souris *Tnap<sup>Cre/+</sup>* sont croisées avec des animaux *Dnmt1<sup>+/-</sup>* afin d'obtenir des mâles *Tnap<sup>Cre/+</sup> Dnmt1<sup>+/-</sup>*. Ces derniers sont croisés avec des femelles *Dnmt1<sup>2lox/2lox</sup>* afin d'obtenir des animaux nécessaires à notre étude. **b.** Exemple de tri cellulaire de PGCs avec l'anticorps SSEA1-PE (CD15-PE). Le premier graphique représente le nombre de cellules en fonction du niveau de fluorescence. Le graphique suivant représente l'intensité de fluorescence en fonction de la granulométrie des cellules (SSC-A). La fenêtre de tri pour les PGCs correspond aux CD15<sup>+</sup>, tandis que les cellules somatiques sont dans la fenêtre CD15<sup>-</sup>. **c.** Test de phosphatase alcaline réalisé sur les cellules triées par SSEA1. Les cellules SSEA1<sup>+</sup> se colorent en rose et correspondent aux PGCs tandis que les cellules SSEA1<sup>-</sup> ne se colorent pas. **d.** Résultats de COBRA mesurant la méthylation des gènes *ckt2* et *dppa3* dans les cellules SSEA-1 positives et négatives. Les cellules SSEA-1 positives ne présentent aucun signe de méthylation, validant le fait que ce soit des PGCs. Echelle: 1kb+ (ThermoFisher) **e.** Exemple de génotypage *Dnmt1*. La bande supérieure correspond à l'allèle délété (1lox), la seconde à celle de l'allèle 2lox et la plus basse à l'allèle WT. Dans le cKO et dans les Het<sup>+/-</sup>, on constate la présence résiduelle de la bande 2lox, dénotant que l'activité de la Cre n'est pas totale.

#### 4. DNMT1 : un acteur dans le maintien de la méthylation au cours du développement des PGCs

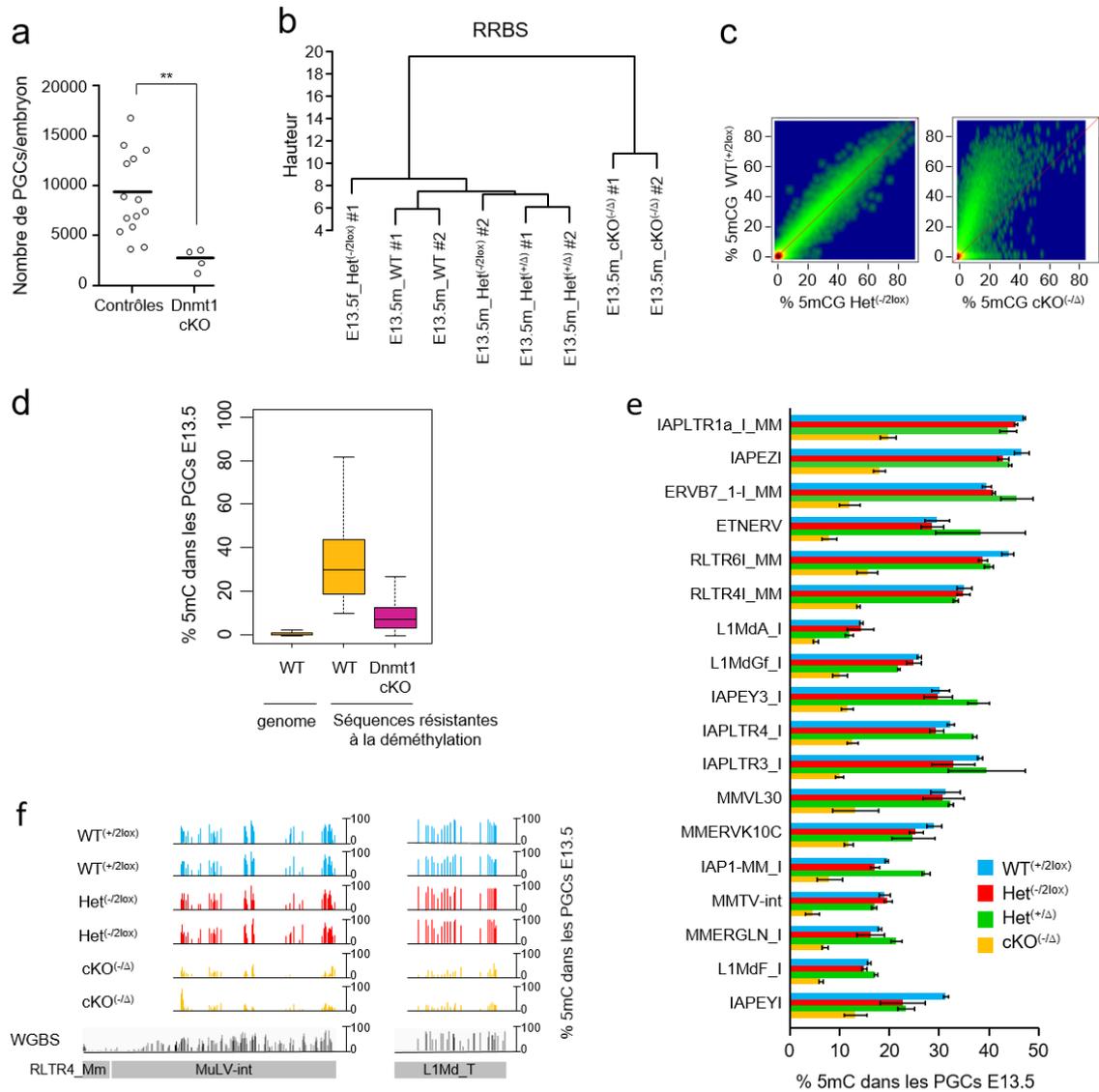
Les enzymes de méthylation *de novo*, c'est à dire DNMT3A et DNMT3B, sont connues pour être réprimées dans les PGCs, ce qui n'est pas le cas de DNMT1 (Kagiyada et al., 2013; Saitou et al., 2012; Seki et al., 2005). Grâce à nos expériences de RNA-seq, nous avons comparé l'expression des gènes *Dnmt3a*, *Dnmt3b* et *Dnmt1* dans les PGCs à E13,5 par rapport à des embryons E8,5 (Figure 13). Tout comme les études précédentes, nous avons constaté que l'expression des gènes *Dnmt3a* et *Dnmt3b* est réprimée dans les PGCs en comparaison avec un embryon E8,5 tandis que l'expression du gène *Dnmt1* se maintient à un haut niveau dans les PGCs. Ainsi, nous avons émis l'hypothèse que DNMT1 est l'enzyme responsable du maintien sélectif de la méthylation des CpG lors de la déméthylation des PGCs, protégeant ainsi de la déméthylation les éléments transposables et assurant l'intégrité du génome.

Pour tester cette hypothèse, j'ai créé une lignée de souris avec une délétion conditionnelle du gène *Dnmt1* (*Dnmt1* cKO) dans les PGCs avec la lignée *Tnap-Cre* qui exprime la CRE recombinase spécifiquement dans les PGCs. Pendant la réalisation de mon projet, un autre groupe a publié en 2016 la réalisation d'un KO conditionnel de *Dnmt1* dans les PGCs en utilisant une lignée *Blimp1-Cre* (Hargan-Calvopina et al., 2016). Même si nos deux délétions conditionnelles du gène *Dnmt1* n'ont pas été réalisées de la même manière, j'ai pu comparer certaines données de l'équipe d'Amander T. Clark avec les miennes lors de l'analyse de mes résultats.

##### a. Génération de souris avec une inactivation conditionnelle de *Dnmt1* dans les PGCs

Dans l'optique de générer une lignée de souris où *Dnmt1* est inactivé spécifiquement dans la lignée germinale, nous avons choisi d'utiliser le système Cre-Lox en croisant les lignées de souris *Dnmt1*<sup>2lox/2lox</sup> (Jackson-Grusby et al., 2001b) et *Tnap-Cre* qui exprime la recombinase CRE spécifiquement dans les PGCs à partir de E10,5 (Lomeli et al., 2000) (Matériels & méthodes ; Annexe 1). J'ai tout d'abord croisé des souris *Tnap-Cre* et *Dnmt1*<sup>-/+</sup> afin d'obtenir la lignée *Tnap*<sup>Cre/+</sup> *Dnmt1*<sup>-/+</sup>, puis j'ai croisé les souris *Tnap*<sup>Cre/+</sup> *Dnmt1*<sup>-/+</sup> et *Dnmt1*<sup>2lox/2lox</sup> entre elles afin d'avoir les génotypes suivants dans les cellules germinales : des cellules WT (*Dnmt1*<sup>2lox/+</sup>), des cellules hétérozygotes (soit *Dnmt1*<sup>2lox/-</sup>, soit *Tnap*<sup>Cre/+</sup> *Dnmt1*<sup>Δ/+</sup>) et des cellules cKO (*Tnap*<sup>Cre/+</sup> *Dnmt1*<sup>Δ/-</sup>) (Figure 14a).

Pour trier les PGCs, je n'ai pas utilisé la lignée *Oct4-GFP* comme précédemment car j'ai constaté que le croisement entre ces trois différentes lignées murines induisait une expression



**Figure 15 : *Dnmt1* a une incidence sur le nombre de PGCs et est impliqué dans le maintien de la méthylation des TEs**

**a.** Graphique représentant le nombre de PGCs obtenus lors des tris cellulaires à partir d'animaux sauvages (Contrôles,  $n = 11$ ) et de souris où le gène *dntm1* à été déléte de façon conditionnelle dans les PGCs (cKO,  $n=3$ ). La ligne correspond à la médiane. On constate que la délétion *Dnmt1* induit une diminution du nombre de PGCs **b.** Dendrogramme de nos données de RRBS. On constate que le méthylome des animaux WT et HeT sont regroupés sur la même branche tandis que les animaux cKO se situe au niveau d'une autre branche, démontrant une différence significative au sein de leur méthylome **c.** Graphique de corrélation représentant le niveau de méthylation des mêmes CpGs, soit dans des PGCs hétérozygotes pour *Dnmt1* (*Het*<sup>-2lox</sup>) par rapport aux PGCs WT, soit dans le KO conditionnel (cKO<sup>-Δ</sup>). Alors que le méthylome des PGCs WT et celui des *Het*<sup>-2lox</sup> corrént, ceci n'est pas le cas du cKO. En effet, on observe dans ce dernier, que les CpGs des PGCs cKO sont globalement déméthylés par rapport au WT. **d.** Diagramme en barres représentant le niveau global de 5mC dans un épiblaste à E7,5 puis dans les PGCs aux différents stades de développement. « E13,5\_Res » correspond aux régions du génome d'un PGC à E13,5 qui résistent à la phase de déméthylation et dont le niveau de 5mC est supérieur à 10%. « E13,5\_Res\_cKO<sup>-Δ</sup> » représente le pourcentage de méthylation de ces mêmes régions dans des PGCs où *dntm1* a été déléte. On remarque que l'absence de DNMT1 induit une diminution du niveau de 5mC **e.** Graphique en barres représentant le pourcentage de méthylation de différents TEs résistants à la vague de déméthylation à E13,5 dans les quatre génotypes. On constate que la délétion de *Dnmt1* induit une baisse significative du niveau de méthylation de l'ensemble des TEs. **f.** Illustration à partir du logiciel igv du niveau de 5mC au niveau de deux éléments transposables (MuLV-int et L1Md\_T). Chaque barre représente une cytosine. Plus la barre est basse, plus le niveau de méthylation est faible, et inversement. On observe que l'ensemble des cytosines des PGCs cKO sont déméthylés en comparaison avec les WT et Het.

ectopique de la GFP dans les cellules somatiques, rendant impossible le tri des PGCs. Pour isoler les PGCs, j'ai changé de stratégie et j'ai trié les cellules avec l'anticorps SSEA-1-PE, aussi appelé CD15-PE, qui marque spécifiquement les PGCs et permet de les trier en cytométrie de flux (**Figure 14b**). J'ai validé l'efficacité du tri grâce au test de la phosphatase alcaline (**Figure 14c**). On constate sur ces photos que seules les cellules SSEA-1 positives se colorent, validant ainsi la pureté de mon tri. En parallèle, j'ai extrait l'ADNg des cellules SSEA-1 positives et négatives afin de réaliser un COBRA sur les gènes *Ckt2* et *Dppa3* (**Figure 14d**). Les résultats du COBRA indiquent que les cellules SSEA-1 positives ne présentent pas de signes de méthylation sur les gènes *Ckt2* et *Dppa3*, ce qui valide une seconde fois la pureté du tri.

Enfin, après chaque tri cellulaire et extraction de l'ADNg des PGCs, j'ai réalisé un génotypage dans le but de vérifier l'efficacité de la délétion par le système Cre-Lox (**Figure 14e**). Ce gel d'électrophorèse est représentatif de l'ensemble des génotypages que j'ai réalisés et révèle une bonne délétion de l'allèle 2lox du gène *Dnmt1* par Tnap-Cre. Néanmoins, on peut constater la présence résiduelle de l'allèle 2lox, indiquant que la délétion n'est pas totale (**Figure 14e**).

#### **b. *Dnmt1* est nécessaire pour le maintien de la méthylation des transposons dans les PGCs**

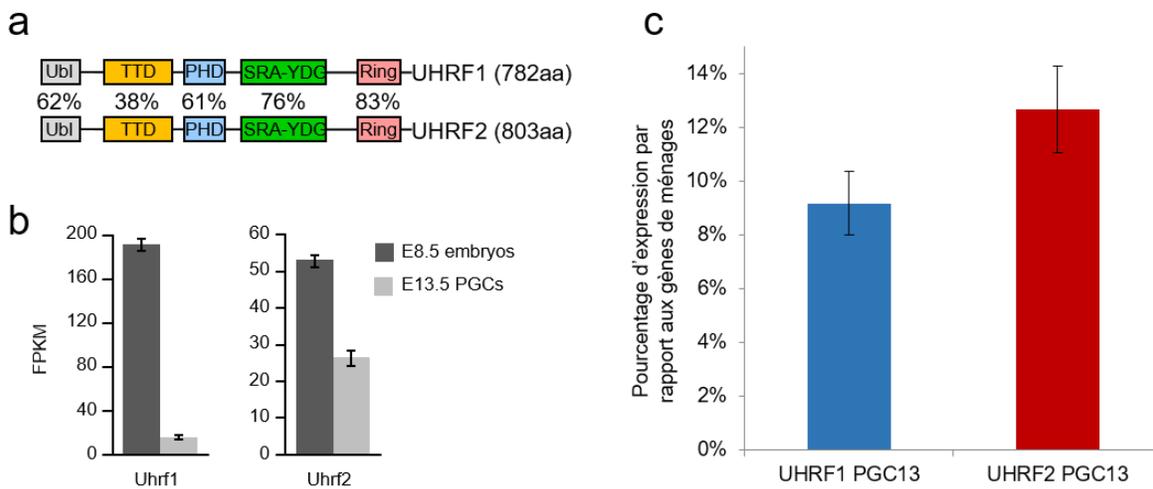
Par la suite, j'ai trié les PGCs au stade E13,5 afin de voir l'incidence de la délétion de *Dnmt1* sur les profils de méthylation de l'ADN des PGCs. Au fur et à mesure que je triais les PGCs, j'ai remarqué que le nombre de PGCs récoltés diminuait fortement lorsque *Dnmt1* est inactivé (**Figure 15a**). En effet, au lieu d'obtenir environ 10 000 PGCs/embryon, je n'ai trié qu'environ 2 000 PGCs dans les embryons cKO (**Figure 15a**). Mes résultats confirment ceux publiés par [Hargan-Calvopina et al., 2016](#) qui eux aussi observaient une diminution du nombre de PGCs de 10 000 PGCs/embryon à environ 2 000 PGCs/embryon dans les embryon cKO.

Malgré le faible nombre de PGCs, j'ai généré des bibliothèques RRBS en duplicatas biologiques pour les PGCs *Dnmt1* cKO et les trois génotypes contrôles (**Figure 14a**). Nous avons réalisé une classification hiérarchique des données et constaté que le dendrogramme regroupe dans une branche les PGCs *Dnmt1* cKO et dans une autres les PGCs WT et *Dnmt1* Het (**Figure 15b**). Ainsi, il semble que les méthylomes des PGCs *Dnmt1* cKO soient clairement différents de celui des autres PGCs. Parmi les génotypes hétérozygotes, l'échantillon E13,5f\_Het<sup>(-/2lox)</sup> #1 est séparé des autres. Ceci est peut-être dû au fait qu'il s'agit de l'unique échantillon femelle et qu'il existe une légère différence au niveau du méthylome des PGCs mâles et femelles.

Après cela, nous avons réalisé une analyse globale des méthylomes des PGCs de différents génotypes en réalisant un graphique de corrélation (**Figure 15c**). Aucune différence globale n'est observée lors de la comparaison du profil de méthylation des PGC *Dnmt1* Het avec les PGC WT.

Par contre, on constate que le génome des PGCs *Dnmt1* cKO est hypométhylé par rapport aux PGCs WT. Nous avons alors calculé la méthylation globale des régions résistantes à la déméthylation dans les PGCs WT et *Dnmt1* cKO (**Figure 15d**). On remarque que le niveau moyen de méthylation des régions résistantes passe de 30% dans les PGCs WT à 8% dans les PGCs *Dnmt1* cKO. La méthylation résiduelle dans les PGCs *Dnmt1* cKO est probablement due à la délétion incomplète de *Dnmt1* dans les PGCs, comme nous l'avons vu précédemment dans les profils de génotypage (**Figure 14e**). Dans un second temps, nous avons regardé plus précisément le comportement des familles de transposons dans les PGC E13,5 WT, Het et cKO (**Figure 15e**). La délétion hétérozygote d'un des deux allèles de *Dnmt1* n'a aucune influence sur le profil de méthylation des transposons (**Figure 15e**). Par contre, le niveau de méthylation de toutes les familles de transposons est fortement diminué dans les PGCs *Dnmt1* cKO. Par exemple, la méthylation des IAPEZI passe de 45% dans les PGCs contrôles à moins de 20% dans les PGCs *Dnmt1* cKO. La **Figure 15f** montre des exemples de cette déméthylation au niveau de la séquence d'un MuIV-int et un L1Md\_T.

En conclusion, la délétion conditionnelle de *Dnmt1* dans les PGCs induit une diminution conséquente du taux de 5mC à E13,5, malgré le fait que la délétion ne soit pas totale. Les régions normalement résistantes à la reprogrammation épigénétique dans les PGCs E13,5, notamment les éléments transposables, se retrouvent hypométhylées. Cela démontre que *Dnmt1* joue un rôle dans le maintien de la méthylation des transposons dans les PGCs. Les expériences pour savoir si l'expression des transposons est augmentée dans les PGCs *Dnmt1* cKO sont en cours au laboratoire. Cependant, nous avons effectué une analyse des données RNA-seq de l'équipe de Clark qui révèlent que malgré la perte de 5mC il y a peu d'incidence sur l'expression des transposons dans les PGCs *Dnmt1* cKO (**Hargan-Calvopina et al., 2016**).



**Figure 16 : UHRF2, homologue de UHRF1, est exprimée dans les PGCs E13,5**

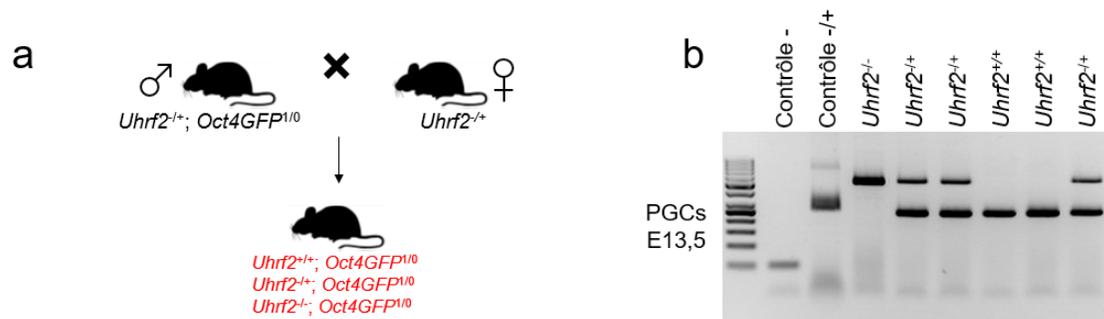
**a.** Comparaison schématique du taux de conservation des domaines entre UHRF1 et UHRF2 par blastP. Hormis le domaine TTD qui ne possède que 38% de similitude entre les deux protéines, l'ensemble des autres domaines sont homologues à plus de 60%. **b.** Graphique illustrant les niveaux d'expression des ARNm *Uhrf1* et *Uhrf2* mesurés par RNA-seq (en FPKM) dans les embryons E8,5 et les PGCs à E13,5. (Embryon E8,5 n=3 / PGCs E13,5 n=3). **c.** Graphique représentant le niveau d'expression de *Uhrf1* et *Uhrf2* dans des PGCs E13,5 mesuré par RT-qPCR (n=3). L'expression a été normalisé par rapport aux gènes de ménages *b-actin* et *Rpl13*. *Uhrf2* est plus exprimé que *Uhrf1* dans les PGCs. FPKM = fragments per kilobase of transcript per million mapped reads.

## 5. UHRF2 : un nouvel acteur de la méthylation de l'ADN dans les PGCs

Nous avons démontré dans le paragraphe précédent que *Dnmt1* joue un rôle dans le maintien sélectif de la 5mC au niveau des transposons au cours de la cinétique de déméthylation des PGCs. Cependant, une question demeure : comment la protéine DNMT1 est-elle recrutée au niveau des séquences des transposons dans les PGCs ? Il a été démontré que la protéine UHRF1, qui permet de recruter DNMT1 au niveau des fourches de réplication, est exclue du noyau des PGCs à E13,5 (Seisenberger et al., 2012). De plus, d'autres groupes ont démontré que l'expression du gène *Uhrf1* est diminuée dans les PGCs (Kagiwada et al., 2013; Saitou et al., 2012; Seisenberger et al., 2012; Seki et al., 2005). Si UHRF1 n'est pas active dans les PGCs, quels mécanismes recrutent DNMT1 au niveau des transposons ?

Il existe une protéine paralogue à UHRF1 : UHRF2 (Figure 16a). La comparaison par blastP montre que ces deux protéines partagent les mêmes domaines et ce, de manière très conservée (Figure 16a). Seul le domaine TTD présente un faible pourcentage d'homologie (38%), tandis que les autres possèdent au minimum 61% d'homologie. Des études antérieures démontrent que UHRF2 peut se lier à DNMT1, DNMT3A et DNMT3B, ainsi qu'à l'ADN méthylé, hémi-méthylé et hydroxyméthylé, et aux marques d'histone H3K9me2 et H3K9me3 (Pichler et al., 2011; Spruijt et al., 2013; Zhang et al., 2011; Zhou et al., 2014).

Pour démarrer cette étude, j'ai comparé le niveau d'expression des gènes *Uhrf1* et *Uhrf2* dans les PGCs E13,5 et les embryons E8,5 (Figure 16b-c). Nos données de RNA-seq démontrent que *Uhrf1* est considérablement plus exprimé dans les embryons E8,5 que dans les PGCs à E13,5, alors que l'expression de *Uhrf2* se maintient dans les PGCs (Figure 16b). On constate aussi que dans les PGCs E13,5, le niveau d'expression mesuré en FPKM (fragments per kilobase of transcript per million mapped reads) de *Uhrf2* (25.88) est plus élevé que *Uhrf1* (16.65). Dans le but de valider cette observation, j'ai réalisé l'expression des gènes par RT-qPCR dans les PGCs E13,5 (Figure 16c). J'ai normalisé l'expression des gènes *Uhrf1* et *Uhrf2* par rapport à l'expression des gènes de ménages *Rpl13* et  $\beta$ -*actin*. Ces résultats démontrent, une fois de plus, que le gène *Uhrf2* est plus exprimé que *Uhrf1* dans les PGCs. Ainsi, j'ai émis l'hypothèse qu'*Uhrf2* pourrait participer au maintien de la méthylation des éléments transposables lors du développement des PGCs, protégeant ainsi l'intégrité du génome.



**Figure 17** : Stratégie d'isolation de PGCs E13,5 invalidés pour *Uhrf2*

**a.** Stratégie de croisement. Des souris mâles *Uhrf2*<sup>-/-</sup>; *Oct4GFP*<sup>1/0</sup> sont croisées avec des femelles *Uhrf2*<sup>-/-</sup> dans le but d'obtenir les génotypes écrits en rouge. **b.** Exemple de résultat de génotypage des PGCs E13,5. La bande supérieure correspond à l'allèle KO (tm1b) et la plus basse à l'allèle WT.

### a. Stratégie d'obtention de PGCs *Uhrf2*<sup>-/-</sup>

Pour cette étude, j'ai utilisé la lignée murine *Uhrf2*<sup>-/+</sup> générée dans le cadre du projet « International Knockout Mouse Consortium » (Bradley et al., 2012) (Matériels & méthodes, souris Tm1b ; **Annexe 1**). Dans ces souris, l'exon 3 du gène *Uhrf2* a été excisé, générant un décalage de lecture et aboutissant à la production d'une protéine aberrante. Dans l'optique de générer un animal *Uhrf2*<sup>-/-</sup> où il me serait possible de trier les PGCs, j'ai croisé la lignée *Uhrf2*<sup>-/+</sup> avec la lignée *Oct4*-GFP (Yoshimizu et al., 1999) (Matériels & méthodes ; **Annexe 1**) afin de générer la lignée *Uhrf2*<sup>-/+</sup> *Oct4*-GFP<sup>1/0</sup> (**Figure 17a**). Après génération de cette lignée, j'ai croisé ces souris avec les souris *Uhrf2*<sup>+/+</sup> afin d'isoler les cellules germinales d'embryons WT (*Uhrf2*<sup>+/+</sup> *Oct4*-GFP<sup>1/0</sup>), hétérozygotes (*Uhrf2*<sup>-/+</sup> *Oct4*-GFP<sup>1/0</sup>) et KO (*Uhrf2*<sup>-/-</sup> *Oct4*-GFP<sup>1/0</sup>) (**Figure 17a**).

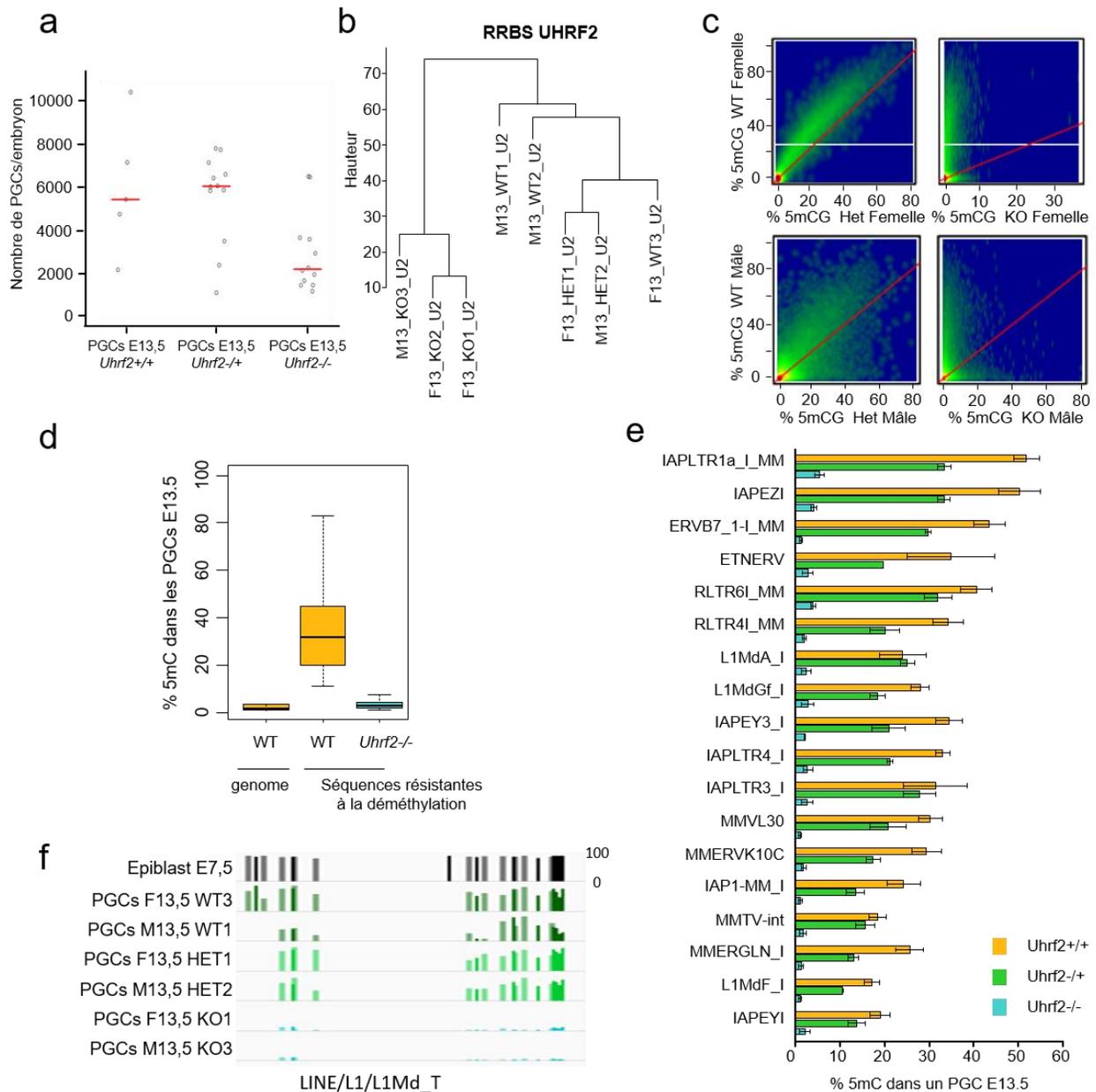
Dans le but de déterminer l'implication de *Uhrf2* dans le maintien de la méthylation au niveau des régions résistantes à la vague de déméthylation, j'ai trié les PGCs au stade E13,5 par cytométrie de flux grâce à l'expression du marqueur GFP. Après chaque tri cellulaire, j'ai extrait l'ADNg des cellules afin de vérifier le génotype par PCR (**Figure 17b**). Ce gel d'électrophorèse est représentatif de l'ensemble des génotypages que j'ai réalisés sur les PGCs. La bande supérieure correspond à l'allèle KO (506pb) et la bande inférieure à l'allèle WT (240pb).

### b. *Uhrf2* est nécessaire pour la méthylation des transposons dans les PGCs

Lors des différents tris cellulaires, j'ai remarqué que le nombre de PGCs récoltés était inférieur dans les embryons invalidés pour *Uhrf2*. En effet, par rapport aux 6 000 PGCs/embryon triés dans les embryons sauvages et hétérozygotes, je n'ai trié en moyenne qu'environ 2 000 PGCs dans les embryons *Uhrf2*<sup>-/-</sup> (**Figure 18a**).

J'ai généré des bibliothèques RRBS en triplicatas pour les PGCs *Uhrf2*<sup>-/-</sup> et *Uhrf2*<sup>+/+</sup>, ainsi qu'en duplicata pour les PGCs hétérozygotes. Nous avons réalisé une classification hiérarchique des données afin de valider les bibliothèques RRBS et voir si des groupes se forment. Le dendrogramme regroupe dans une branche les PGCs *Uhrf2*<sup>-/-</sup> et dans une autre les PGCs WT et *Uhrf2*<sup>+/+</sup> (**Figure 18b**). Ainsi, il semble que les PGCs *Uhrf2*<sup>-/-</sup> soient différents du point de vue de leur méthylome comparé aux autres PGCs.

Nous avons réalisé ensuite une analyse globale des méthylomes en réalisant un graphique de corrélation (**Figure 18c**). Pour cette analyse, nous avons analysé séparément les PGCs mâles et femelles. Les graphiques de corrélation entre le méthylome des PGCs WT vs hétérozygotes révèlent que la délétion d'un des deux allèles d'*Uhrf2* induit une diminution légère du niveau de 5mC au niveau des séquences méthylées dans les PGCs E13,5 (**Figure 18c**). De manière



**Figure 18 : *Uhrf2* est impliqué dans le maintien de la méthylation dans les PGCs**

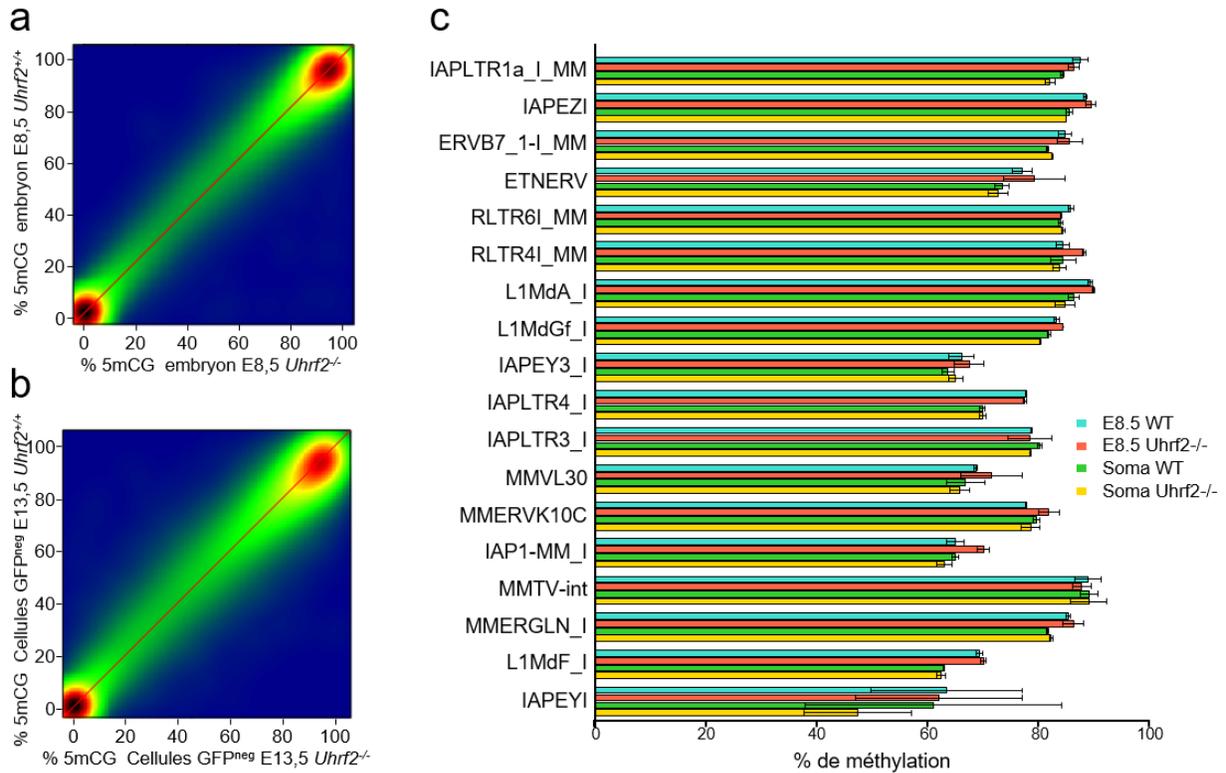
**a.** Graphique représentant le nombre de PGCs obtenus lors des tris cellulaires à partir d'animaux *Uhrf2*<sup>+/+</sup> (n =5), *Uhrf2*<sup>-/+</sup> (n=12) et *Uhrf2*<sup>-/-</sup> (n=12). La ligne rouge correspond à la médiane. On constate que la délétion d'*Uhrf2* induit une diminution du nombre de PGCs **b.** Dendrogramme des données de RRBS. On constate que le méthylome des PGCs *Uhrf2*<sup>-/-</sup> se distingue de celui des PGCs WT et *Uhrf2*<sup>-/+</sup> (HET). **c.** Graphiques de corrélation comparant le niveau de méthylation des CpGs des PGCs WT, *Uhrf2*<sup>-/+</sup> et *Uhrf2*<sup>-/-</sup>. La perte d'un allèle d'*Uhrf2* induit une légère baisse de méthylation comparée au WT. L'inactivation homozygote d'*Uhrf2* induit une déméthylation totale des PGCs. **d.** Diagramme représentant le niveau de 5mC dans les séquences qui résistent à la phase de déméthylation, dans les PGCs WT et *Uhrf2*<sup>-/-</sup>. **e.** Graphique représentant le pourcentage de méthylation des familles de transposons les plus résistantes à la vague de déméthylation, dans les PGCs WT, *Uhrf2*<sup>-/+</sup> et *Uhrf2*<sup>-/-</sup>. On constate que la délétion de *Uhrf2* induit une baisse significative du niveau de méthylation de l'ensemble des TEs. Même si l'effet est moins drastique, il y a aussi une perte de 5mC dans les PGCs *Uhrf2*<sup>-/+</sup>. **f.** Exemple de méthylation d'un transposon L1Md\_T dans les PGCs WT, *Uhrf2*<sup>-/+</sup> (HET) et *Uhrf2*<sup>-/-</sup> (KO).

remarquable, la comparaison des méthylomes des PGCs *Uhrf2*<sup>-/-</sup> avec ceux des PGCs WT montre que l'inactivation d'*Uhrf2* induit une hypométhylation totale des PGCs E13,5 (**Figure 18c**). Nous avons alors calculé la méthylation des régions résistantes à la reprogrammation épigénétique dans les PGCs WT et *Uhrf2*<sup>-/-</sup> (**Figure 18d**). Alors que ces séquences sont méthylées en moyenne aux alentours de 30% dans les PGCs WT, le niveau moyen de méthylation des régions résistantes chute drastiquement dans les PGCs *Uhrf2*<sup>-/-</sup> pour atteindre un niveau en dessous de 5% (**Figure 18d**). Dans un second temps, nous avons regardé le comportement des transposons dans les PGCs *Uhrf2*<sup>-/-</sup> (**Figure 18e**). La délétion hétérozygote d'un des deux allèles d'*Uhrf2* induit une diminution du niveau de méthylation de ces éléments transposables (**Figure 18e**). La délétion d'*Uhrf2* conduit à la perte quasi-totale de la méthylation de toutes les familles de transposons (**Figure 18e**). Pour exemple, dans la famille des IAPLTR1a\_I\_MM, le niveau de 5mC est de 48% dans les PGCs WT, puis chute à 28% dans les PGCs *Uhrf2*<sup>+/-</sup> et finalement à 7% dans les PGCs *Uhrf2*<sup>-/-</sup> (**Figure 18e**). Un exemple de transposon de la famille L1Md\_T totalement déméthylé dans les PGCs *Uhrf2*<sup>-/-</sup> est montré dans la **Figure 18f**.

### c. *Uhrf2* n'est pas impliqué dans la méthylation de l'ADN dans les cellules somatiques

Au vu des résultats décrits plus haut, nous avons cherché à savoir si *Uhrf2* régule également le profil de méthylation des TEs dans les cellules somatiques. Pour cela j'ai analysé la méthylation de l'ADN dans les embryons E8,5 et dans les cellules somatiques GFP-négatives des gonades triées en même temps que les PGCs à E13,5. Par conséquent, j'ai croisé les souris *Uhrf2*<sup>+/-</sup> entre elles afin de disséquer des embryons E8,5 *Uhrf2*<sup>+/+</sup> et *Uhrf2*<sup>-/-</sup>. J'ai réalisé et séquencé des bibliothèques RRBS en duplicatas pour les génotypes WT et KO.

L'analyse globale des méthylomes des embryons E8,5 et des cellules somatiques des gonades indique que la délétion d'*Uhrf2* n'a pas incidence sur le profil de méthylation global du génome dans les cellules somatiques. En effet, le méthylome de l'embryon sauvage corrèle parfaitement avec celui de l'embryon *Uhrf2*<sup>-/-</sup> (**Figure 19a**). On observe exactement la même chose lors de la comparaison des méthylomes des cellules somatiques E13,5 (**Figure 19b**). Dans un second temps, nous avons regardé plus précisément la méthylation des transposons qui sont méthylés sous l'influence d'*Uhrf2* dans les PGCs E13,5. Contrairement aux PGCs, la délétion d'*Uhrf2* n'a pas d'impact sur le niveau de méthylation des familles d'éléments transposables dans les cellules somatiques (**Figure 19c**). Nous concluons que le rôle d'*Uhrf2* dans le maintien de la méthylation des transposons est spécifique aux cellules germinales.



**Figure 19 : *Uhrf2* n'influence pas la méthylation de l'ADN dans les cellules somatiques**

**a.** Graphique de corrélation du niveau de méthylation des CpGs dans les embryons E8,5 *Uhrf2*<sup>+/+</sup> et *Uhrf2*<sup>-/-</sup>. Les deux méthylomes corrélient parfaitement. **b.** Graphique de corrélation du niveau de méthylation des CpGs dans les cellules somatiques (cellules GFP négatives triées à E13,5) *Uhrf2*<sup>+/+</sup> et *Uhrf2*<sup>-/-</sup>. Les deux méthylomes corrélient parfaitement. **c.** Graphique représentant le pourcentage de méthylation de différentes familles de transposons dans les embryons E8,5 ou les cellules somatiques GFP-négatives *Uhrf2*<sup>+/+</sup> (WT) et *Uhrf2*<sup>-/-</sup> (KO). On constate que la délétion d'*Uhrf2* n'a aucune incidence sur le niveau de méthylation des transposons dans ces cellules.

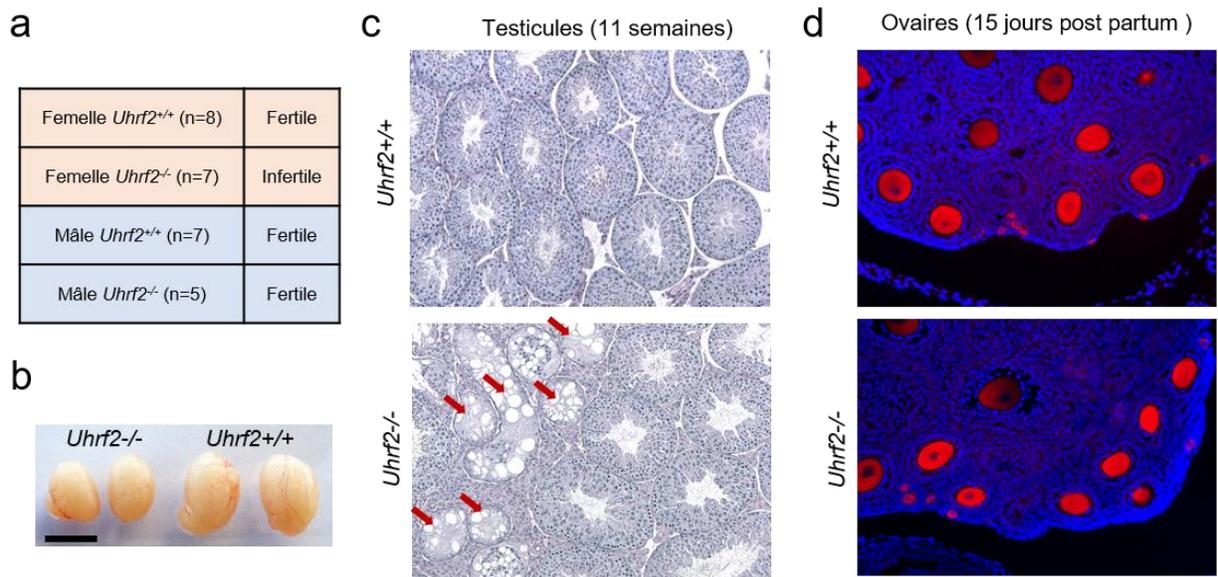
#### d. L'absence d'*Uhrf2* n'induit pas la réexpression des transposons dans les PGCs

A la vue des résultats obtenus dans les PGCs, nous avons testé si l'inactivation d'*Uhrf2* influence l'expression des transposons dans les PGCs. Pour cela j'ai réalisé des expériences de RNA-seq sur des PGCs E13,5 mâles *Uhrf2*<sup>-/-</sup> et *Uhrf2*<sup>+/+</sup> en triplicatas. De façon surprenante l'analyse des RNA-seq révèlent qu'aucune famille de transposons ne se réexprime dans les PGCs *Uhrf2*<sup>-/-</sup> (données non montrées). Néanmoins, 198 gènes sont sous exprimés dans les gamètes invalidés pour *Uhrf2*. L'analyse ontologique (valeur-P <0.05) par le logiciel DAVID (david.ncicrf.gov) nous informe qu'il s'agit de gènes impliqués dans les processus biologiques suivants : l'assemblage des nucléosomes, la méthylation des cytosines, la régulation positive épigénétique de l'expression génique, l'hétérotétramérisation des protéines, la régulation négative de la différenciation des mégacaryocytes, le silencing de l'ADN ribosomique, la formation des néphrons, l'initiation de la transcription et la régulation négative de l'adhésion cellulaire. De plus, 128 gènes sont significativement surexprimés dans les PGCs *Uhrf2*<sup>-/-</sup>. L'analyse ontologique (valeur-P <0.05) de ces gènes indique qu'il s'agit de gènes liés aux processus biologiques suivants : la régulation négative de la différenciation des myoblastes, la morphogénèse d'organes, le développement du système digestif, les mécanismes de réparation de l'ADN au cours de la méiose, la régulation de l'autophagie et la protéolyse.

#### e. Analyse préliminaire du phénotype reproductif des souris *Uhrf2*<sup>-/-</sup>

Enfin, nous avons voulu savoir si l'inactivation d'*Uhrf2* est associée à un phénotype de fertilité. Pour vérifier cela, j'ai tout d'abord généré des animaux *Uhrf2*<sup>-/-</sup> dans les deux sexes, puis j'ai croisé ces derniers avec des animaux C57Bl6J du sexe opposé dans le but de tester leur fertilité. Ces expériences nous ont appris que la délétion d'*Uhrf2* induit une infertilité chez les femelles mais pas chez les mâles (**Figure 20a**).

A la vue de ces résultats, j'ai disséqué des paires de testicules (*Uhrf2*<sup>-/-</sup> n=3, WT n=3) et des paires d'ovaires (*Uhrf2*<sup>-/-</sup> n=3, WT n=3) d'animaux âgés d'environ 11 semaines. Lors de la dissection des testicules, j'ai remarqué que ceux des animaux *Uhrf2*<sup>-/-</sup> sont plus petits que les WT (**Figure 20b**). J'ai donc pesé ces testicules afin de déterminer la différence de poids. Les testicules WT pèsent en moyenne 0.14g tandis que les testicules KO pèsent en moyenne 0.10g. Nous avons ensuite analysé l'histologie des testicules sur des coupes après fixation au Bouin et marquage avec l'hématoxyline (en collaboration avec Déborah Bourc'his). De manière intéressante, ces analyses préliminaires de coupes réalisées sur les testicules indiquent la présence de tubes séminifères vides sans spermatozoïdes chez les animaux *Uhrf2*<sup>-/-</sup> en comparaison avec les animaux WT (**Figure 20c**). J'ai également fixé au Bouin les ovaires de femelles *Uhrf2*<sup>-/-</sup> et WT



**Figure 20 : analyse du phénotype reproductif des souris *Uhrf2*<sup>-/-</sup>**

**a.** L'inactivation du gène *Uhrf2* induit une infertilité chez les femelles mais pas chez les mâles. **b.** Photographie de la taille des testicules d'animaux *Uhrf2*<sup>-/-</sup> et *Uhrf2*<sup>+/+</sup> âgés de 11 semaines. L'inactivation du gène *Uhrf2* est associée à une diminution de la taille des testicules. L'échelle (le trait noir) représente 1cm. **c.** Coupes histologiques de testicules d'animaux âgés de 11 semaines colorées à l'hématoxyline. On remarque la présence de tubes séminifères vides sans gamètes (flèches rouges) dans les testicules d'animaux *Uhrf2*<sup>-/-</sup>. **d.** Images d'immunohistochimie réalisées sur des coupes d'ovaires à 15 jours après la naissance avec un anticorps anti-DDX4 qui marque le cytoplasme des ovocytes. Il n'y a pas de différences de nombre d'ovocytes entre les animaux *Uhrf2*<sup>-/-</sup> et *Uhrf2*<sup>+/+</sup>.

âgés de 11 semaines et les ai envoyées au laboratoire du Dr. GHYSELINCK à l'IGBMC (Institut de génétique et de biologie moléculaire et cellulaire d'Illkirch) afin qu'ils réalisent des coupes et analyses histologiques des ovaires. Pour l'instant ces analyses n'ont pas révélé d'anomalies histologiques dans les ovaires d'animaux *Uhrf2*<sup>-/-</sup>, notamment en ce qui concerne le nombre et les stades de développement des ovocytes. Par ailleurs nous avons également réalisé des marquages par immunohistochimie avec un anticorps anti-DDX4 sur des coupes d'ovaires de femelles âgées de 15 jours, avant la puberté. Ces expériences n'ont pour l'instant pas non plus révélé d'anomalies concernant le nombre d'ovocytes dans les ovaires d'animaux *Uhrf2*<sup>-/-</sup> (**Figure 20d**).

En conclusion, la délétion d'*Uhrf2* induit une diminution du taux de 5mC uniquement dans les PGCs à E13,5. Les régions normalement résistantes à la reprogrammation épigénétique à E13,5, dont notamment les TEs, se retrouvent hypométhylées. Ainsi, il semblerait qu'*Uhrf2* joue un rôle dans le maintien de la méthylation des transposons dans les PGCs. Néanmoins, l'analyse transcriptomique n'a pour l'instant révélé aucune incidence sur l'expression des éléments transposables dans les PGCs. Enfin, l'inactivation d'*Uhrf2* induit une infertilité chez les femelles alors que l'ensemble des différents stades ovocytaires sont présents. Concernant les animaux mâles invalidés pour *Uhrf2*, ces derniers sont fertiles même s'ils présentent une proportion de tubes séminifères anormaux. De nombreux travaux complémentaires sont encore nécessaires pour comprendre le lien éventuel entre les défauts de méthylation observés dans les PGCs et les données phénotypiques préliminaires dans les animaux invalidés pour le gène *Uhrf2*.

## Discussion et Perspectives : PGCs

---

En utilisant des méthodes de cartographie à haute résolution, nous avons pu mettre en évidence la cinétique de méthylation des PGCs à partir de E9,5 jusqu'à E17,5. Ces données viennent grandement compléter les études déjà réalisées chez la souris qui se limitaient à un nombre restreint de stades embryonnaires (Guibert et al., 2012; Kobayashi et al., 2013; Popp et al., 2010; Seisenberger et al., 2012; Wang et al., 2014b).

Grâce à ces données, nous avons pu observer en détail les deux vagues de la reprogrammation épigénétique au cours du développement embryonnaire des PGCs. La première, qui a lieu entre 9,5 et 13,5 jours embryonnaires, est une phase durant laquelle l'ensemble du génome des PGCs se déméthyle, jusqu'à être presque totalement déméthylé à E13,5. La seconde, qui a lieu uniquement chez les mâles à partir E14,5, est une phase durant laquelle le génome des PGCs se reméthyle.

### **La vague de déméthylation du génome des PGCs**

Durant la première phase de reprogrammation épigénétique des PGCs, nous avons pu corréler, pour la première fois, la cinétique de déméthylation et de division cellulaire à l'échelle du génome. Nous avons constaté, comme d'autres études (Seisenberger et al., 2012), que la déméthylation globale du génome semble corréler avec un mécanisme passif de déméthylation de l'ADN, c'est-à-dire que DNMT1 n'est plus à même d'assurer le maintien de la méthylation au cours des divisions cellulaires. Ceci est d'autant plus plausible puisque son co-facteur UHRF1 est sous exprimé, comme nous l'avons vu par RNA-seq à E13,5 et RT-qPCR, et qu'il est exclu du noyau des PGCs durant cette période (Kurimoto et al., 2008; Seisenberger et al., 2012). Cependant d'autres études suggèrent que ce mécanisme de déméthylation passive fonctionne de pair avec le mécanisme de déméthylation active impliquant les enzymes TET et AID (Hackett et al., 2013; Popp et al., 2010; Vincent et al., 2013; Yamaguchi et al., 2013).

Nous avons aussi démontré qu'au cours de cette période, certaines régions se déméthylent plus tardivement, ce qui est notamment le cas des ICRs et de certains gènes gamétiques. En effet, ces séquences ne sont déméthylées qu'à partir de E12,5 alors que l'ensemble du génome est déjà majoritairement hypométhylé à ce stade. Mais est-ce que ces régions possèdent réellement un niveau de méthylation supérieur au reste du génome ? En effet, la technique de RRBS ne permet pas de différencier la 5mC de la 5hmC. Ainsi, une hypothèse serait que la « méthylation » que nous observons au niveau de ces régions à E10,5 et E11,5 serait le reflet de la présence de 5hmC et d'une déméthylation impliquant les protéines TET. L'étude sur l'activité des enzymes TET1 et TET2 réalisée par Vincent et al., 2013 alimente cette hypothèse puisqu'il a

été démontré que la déméthylation des promoteurs des gènes *Dazl*, *Sycp3* et *Mael* est lié à l'activité de ces deux enzymes. De plus, cette hypothèse est soutenue par les observations de l'équipe de Y. Zhang qui montre que l'inactivation de Tet1 modifie la méthylation des gènes méiotiques et des gènes soumis à empreinte dans les PGCs (Yamaguchi et al., 2012, 2013). Afin d'y voir plus clair, il serait nécessaire d'analyser la présence de 5hmC dans les PGCs en utilisant par exemple la technique de l'oxBS-seq (Booth et al., 2012), ou le TAB-seq (Yu et al., 2012) ou encore le RRHP (Reduced Representation 5-hydroxymethylcytosine Profiling) (Pettersson et al., 2014). Nous pourrions aussi regarder spécifiquement ces régions en utilisant des kits commerciaux basés sur l'utilisation d'enzymes de restriction comme celui proposé par NEB (EpiMark 5-hmC and 5-mC Analysis Kit).

### **Comment les transposons résistent à la déméthylation dans les PGCs ?**

Nos résultats démontrent que certaines régions résistent à la vague de déméthylation et parmi ces dernières, la majeure partie correspond aux éléments transposables. Parmi les éléments transposables les plus méthylés, on retrouve les IAP, les ERVK, les ERV1 et les L1Md. De plus, nous avons démontré que DNMT1 et UHRF2 sont elles aussi impliquées dans le maintien de la méthylation des éléments transposables lors de la vague de déméthylation des PGCs.

### **L'implication de DNMT1 dans le maintien de la méthylation au niveau des TES**

Grâce à l'utilisation de souris où *Dnmt1* est conditionnellement délété dans les PGCs, couplé à la technique de RRBS, nous avons mis en évidence l'implication de DNMT1 dans le maintien de la méthylation au niveau des éléments transposables. En effet, l'absence de cette enzyme au niveau des PGCs induit une baisse du taux de 5mC sur l'ensemble des familles de TEs.

Concernant l'étude de l'implication de DNMT1, d'autres expériences doivent être menées. D'une part, il serait intéressant de réaliser un RRBS des PGCs à E10,5 afin de savoir si l'absence de *Dnmt1* induit une hypométhylation précoce, tout comme le démontre une autre étude (Hargan-Calvopina et al., 2016). Deuxièmement, nous devrions réaliser des RRBS à des stades plus tardifs, par exemple dans les PGCs à E17,5, les spermatozoïdes et les ovocytes, afin de vérifier si les éléments transposables peuvent se remétyler en l'absence de DNMT1. Par la suite, il serait intéressant de laisser vivre ces animaux présentant une déplétion conditionnelle de *Dnmt1* au niveau de leurs cellules gamétiques et d'analyser la fertilité de ces derniers ainsi que l'incidence de cette déplétion sur la formation précoce des embryons.

## **L'implication d'UHRF2 dans le maintien de la méthylation au niveau des TEs**

Grâce à l'utilisation de souris *Uhrf2*<sup>-/-</sup>, couplé à la technique de RRBS, nous avons mis en évidence l'implication de cette protéine dans le maintien de la méthylation au niveau des éléments transposables. En effet, l'absence de cette enzyme induit une diminution drastique du taux de 5mC sur l'ensemble des régions résistantes à la vague de déméthylation, dont notamment les TEs. L'hypothèse selon laquelle cette protéine est impliquée dans le maintien de la méthylation des éléments transposable se base sur sa capacité à reconnaître l'hydroxyméthylation de l'ADN ainsi que les marques d'histones H3K9me2 et H3K9me3 (Pichler et al., 2011; Spruijt et al., 2013; Zhang et al., 2011; Zhou et al., 2014). Ainsi, dans les PGCs, cette protéine s'accrocherait spécifiquement au niveau des régions présentant ces marques d'histones, et de ce fait, ciblerait notamment les éléments transposables qui sont riches en marques d'histone H3K9me2 et H3K9me3. De plus, nous avons observé que ce rôle est spécifique des PGCs, puisque l'absence d'UHRF2 dans un embryon à E8,5, ainsi que dans des cellules somatiques à E13,5, n'induit aucun changement de 5mC, ni au niveau du génome, ni au niveau des éléments transposables. Néanmoins, deux études récentes ont démontré l'implication d'*Uhrf2* dans le maintien de la méthylation de l'ADN et dans la mise en place de l'hydroxyméthylation de l'ADN au niveau du cerveau (Chen et al., 2017b; Liu et al., 2017). Ainsi, il serait intéressant d'analyser l'implication d'UHRF2 au niveau d'autres types cellulaires, tel que par exemple, le lignage musculaire ou même hématopoïétique.

Par ailleurs, d'autres analyses complémentaires doivent être menées. D'une part, il serait intéressant de réaliser un RRBS des PGCs à E10,5 afin de savoir si l'absence d'*Uhrf2* induit une hypométhylation précoce. Deuxièmement, nous devrions réaliser des RRBS à des stades plus tardifs, par exemple dans les PGCs à E17,5, les spermatozoïdes et les ovocytes, afin de vérifier si les éléments transposables peuvent se reméthyler en l'absence d'UHRF2. Concernant les mâles *Uhrf2*<sup>-/-</sup>, alors que ces derniers sont fertiles, on observe néanmoins que certains tubes séminifères sont vides. Il serait donc intéressant de laisser vieillir ces animaux et de voir si ce phénomène s'aggrave et que par conséquent, ces souris deviennent infertiles.

D'autres travaux doivent être menés pour déterminer s'il existe un lien entre l'absence de méthylation des TEs dans les PGCs et la stérilité observée chez les femelles. En effet, nous avons constaté que la délétion d'*Uhrf2* induit un phénotype sur la fertilité des animaux femelles uniquement puisque ces dernières sont dans l'incapacité de donner naissance à des souriceaux. Malgré cela, les analyses immunocytochimiques préliminaires sur les ovaires n'ont pas révélé d'anomalies au niveau du développement des ovocytes. Ces études doivent être poursuivies pour savoir notamment s'il y a des problèmes de progression méiotique dans les ovocytes. Il serait

également intéressant d'étudier le méthylome des ovocytes matures d'animaux *Uhrf2*<sup>-/-</sup>. Par ailleurs, il ne peut pas être exclu que la stérilité des femelles ne soit pas liée à un problème de développement des ovocytes. En effet, le phénotype observé pourrait avoir plusieurs origines possibles, comme par exemple un problème hormonal, un problème d'environnement utérin ou de développement placentaire chez la femelle. On peut aussi imaginer que ces femelles arrivent à générer des embryons mais que ces derniers n'arrivent pas à maturité à cause d'un arrêt du développement embryonnaire. Afin de vérifier l'existence d'embryons, nous pourrions étudier le développement embryonnaire issu de ces femelles *Uhrf2*<sup>-/-</sup> aux stades blastocystes (E4,5) et à E8,5. Ce phénotype d'infertilité que nous observons auprès de nos animaux *Uhrf2*<sup>-/-</sup> n'a pas été observé lors des deux études portant sur l'implication d'UHRF2 dans le cerveau (Chen et al., 2017b; Liu et al., 2017). Ceci est peut-être lié au fait qu'il ne s'agit pas de la même construction génétique et du même fond génétique que ceux utilisés dans notre étude.

Enfin, afin de démontrer le lien entre la stérilité et l'action de UHRF2 dans les cellules germinales, il sera nécessaire de réaliser une lignée de souris avec une inactivation conditionnelle de *Uhrf2* dans les PGCs. Dans cette optique, j'ai commencé des croisements de souris *Tnap*<sup>Cre/+</sup> avec des animaux *Uhrf2*<sup>2lox/2lox</sup>, afin d'obtenir des animaux *Tnap*<sup>Cre/+</sup> ; *Uhrf2*<sup>2lox/2lox</sup>. J'ai aussi croisé des souris *Uhrf2*<sup>2lox/2lox</sup> avec des souris *Oct4GFP*<sup>1/0</sup> afin d'intégrer le marqueur des PGCs. A ce jour, il ne reste plus qu'à croiser ces deux lignées entre elles afin de récupérer les PGCs, analyser leur méthylome et étudier le phénotype reproductif des animaux.

### **Est-ce que DNMT1 et UHRF2 interagisse ensemble dans les PGCs ?**

Si on considère que DNMT1 est impliqué dans la maintenance de la 5mC au niveau des TEs, comment cette enzyme les cible-t-elle ? Etant donné que UHRF2 est elle aussi impliquée dans le maintien de la méthylation des TEs, est-ce que DNMT1 et UHRF2 interagissent ensemble ? Dans l'optique de découvrir le lien qu'il existe entre DNMT1 et UHRF2, j'avais dans l'intention de réaliser sur ces deux protéines une immunoprécipitation ainsi qu'une expérience de test de liaison par proximité - une technique d'immunomarquage permettant de démontrer la colocalisation de deux protéines à partir d'anticorps qui leur sont spécifiques - afin de prouver par une approche biochimique et par une approche d'imagerie cellulaire, qu'elles sont présentes au niveau d'un même complexe protéique. Cependant, je n'ai pas trouvé d'anticorps spécifique d'UHRF2 afin de mener ces analyses. En effet, j'ai testé l'ensemble des anticorps contre UHRF2 disponibles sur le marché en western blot, sur des échantillons protéiques *Uhrf2*<sup>-/-</sup>. Mais aucun n'a été concluant. Je pense que ceci est dû à la forte homologie d'UHRF2 avec UHRF1 et que, de ce fait, l'ensemble des anticorps commerciaux que j'ai testé reconnaissent UHRF1 dans mes échantillons *Uhrf2*<sup>-/-</sup>.

Pour répondre à cette question, il sera nécessaire d'essayer de développer un anticorps spécifique contre la protéine UHRF2, en faisant attention de choisir les peptides au niveau des régions les moins homologues à UHRF1. Nous avons actuellement des anticorps au laboratoire développé avec l'aide de la plateforme d'anticorps de l'IGBMC et il ne nous reste plus qu'à les tester.

### **Pourquoi les transposons résistent à la déméthylation dans les PGCs ?**

Ce qui est intéressant, c'est qu'il s'agit des familles de transposons les plus « jeunes » évolutivement parlant. Ces derniers représentent un potentiel danger pour l'intégrité du génome, puisque plusieurs études ont démontré qu'ils sont capables de se réactiver lorsqu'ils ne sont plus méthylés dans les cellules somatiques (Liu et al., 1994; Maksakova et al., 2006; Sharif et al., 2007; Walsh et al., 1998; Woodcock et al., 1997). Par ailleurs, ces éléments doivent être aussi fortement méthylés au niveau des gamètes matures: il a été démontré que leur méthylation dépend de deux enzymes spécifiques (DNMT3C et DNMT3L) et que leur méthylation est cruciale afin d'assurer leur répression et l'intégrité du génome au cours de la méiose (Barau et al., 2016; Bourc'his and Bestor, 2004). Ainsi, nous pouvons formuler l'hypothèse que le fait que ces éléments transposables soient méthylés alors que l'ensemble du génome des PGCs ne l'est plus, est nécessaire pour assurer leur répression et l'intégrité du génome au moment où le génome des PGCs est globalement reprogrammé.

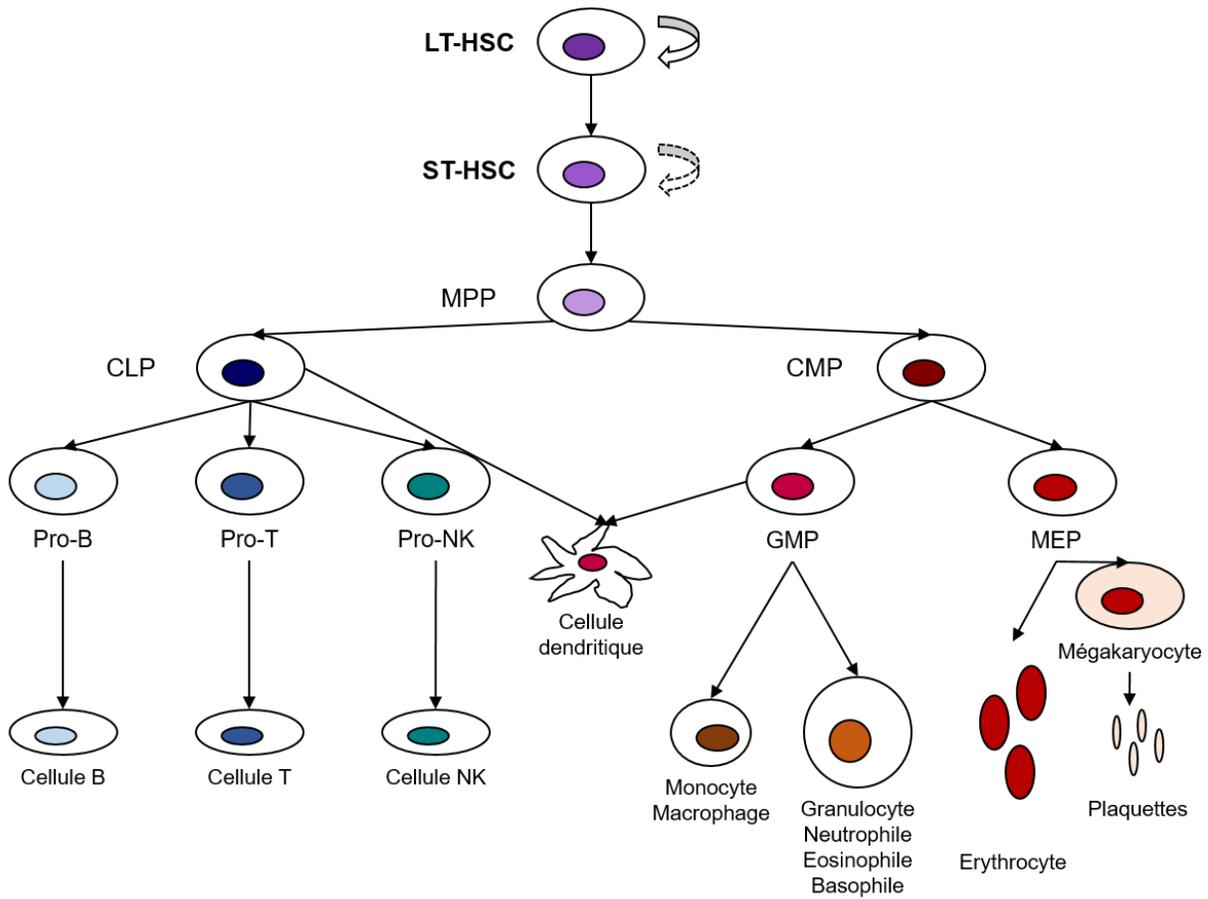
Cependant, de façon surprenante, notre analyse des données RNA-seq réalisées par Hargan-Calvopina et al., 2016 révèle que l'absence de DNMT1 dans les PGCs n'induit pas la réexpression des éléments transposables normalement méthylés à ce stade. De notre côté, nous devrions nous aussi réaliser des RNA-seq dans les PGCs E13,5 afin de vérifier si ces TEs réacquièrent la capacité de se réexprimer dans notre modèle de souris *Dnmt1* cKO. Nos données de RNA-seq réalisées sur des PGCs *Uhrf2*<sup>-/-</sup> à E13,5, confirme également que la déméthylation des TEs n'est pas corrélée avec une réexpression de ces derniers, tout comme nous l'avons précédemment vu pour le KO conditionnel de *Dnmt1* (Hargan-Calvopina et al., 2016). Cependant, il se peut que les TEs ne se réexpriment pas tout de suite et que cela ne soit visible qu'à des stades plus tardifs. Dans cette optique, il serait aussi intéressant d'étudier l'expression des transposons dans les PGCs à des stades plus tardifs comme E17,5 ainsi que dans les gamètes matures. Il est également possible que les transcrits issus des transposons soient très instables ou dégradés par d'autres mécanismes de surveillance. Une autre hypothèse serait que la méthylation des cytosines visible au niveau des éléments transposables ne soit qu'un mécanisme

secondaire de protection contre les transposons. En effet, les TEs sont couverts de marques répressives d'histones en plus de la méthylation de l'ADN : H3K9me2, H3K9me3 et H3K27me3. De plus, des études ont démontré que les cellules ES murines sont capables de maintenir réprimés les transposons sans passer par la méthylation de l'ADN, hormis pour les IAPs (Karimi et al., 2011; Matsui et al., 2010). De plus, une autre étude a démontré la réexpression de ces éléments transposables dans les PGCs à E13,5 lorsque SETDB1, une histone méthyltransférase spécifique de la triméthylation des histones H3K9, est conditionnellement inactivée dans ces cellules (Liu et al., 2014). Enfin, une autre étude dans les cellules ES murines a démontré que suite à l'induction de l'hypométhylation globale du génome, les transposons sont tout d'abord déréprimés lors de la déméthylation drastique de ces cellules, mais qu'ils sont très rapidement et efficacement re-réprimés, grâce à une reconfiguration de la chromatine (Walter et al., 2016). En outre, ils observent qu'au niveau des TEs, la marque H3K9me3 reste stable, tandis que la marque H3K9me2 disparaît pour laisser la place à une forte accumulation de H3K27me3 (Walter et al., 2016). Cela suggère que la perte de méthylation de l'ADN est compensée par des marques d'histones répressives au niveau des transposons. Ainsi, il serait intéressant, dans le cadre de ce projet, d'analyser les marques d'histones associées aux transposons dans les PGCs. Nous pourrions vérifier si l'absence de DNMT1 ou UHRF2 a un impact sur les marques d'histones H3K9me2, H3K9me3 et H3K27me3 au niveau des transposons.

### **La vague de reméthylation du génome des gamètes mâles**

Pendant cette phase, nous avons démontré que l'ensemble du génome des PGCs mâles se reméthyle jusqu'à atteindre un niveau de méthylation très proche de celui des spermatozoïdes. Certains types de séquences, tels que les introns, les exons et les régions intergéniques se méthylent beaucoup plus rapidement que le reste du génome (TSS, CGIs et TEs). Nous avons aussi remarqué que les familles de transposons qui résistent à la déméthylation dans les PGCs avant E13,5, sont aussi celles qui se reméthylent tardivement dans les gamètes mâles après E14,5. La non-réexpression des TEs lors de la phase de déméthylation ainsi que la reméthylation tardive observée au niveau des familles de TEs résistantes à la phase de déméthylation peut être liés à la voie moléculaire impliquant les piARN. En effet, nous savons que ces derniers sont impliqués dans la répression des TEs, notamment dans les gamètes (Siomi et al., 2011). Néanmoins, il serait intéressant de déterminer quelles enzymes sont responsables de la reméthylation des transposons lors de cette phase de reméthylation en réalisant des délétions conditionnelles des enzymes DNMT3A, DNMT3B, DNMT3C et DNMT3L.

Enfin, nous avons aussi montré qu'il existe de la méthylation non-CpG pendant cette période. Cette méthylation CHH et CHG, qui a principalement lieu dans un contexte CpA, corrèle avec l'augmentation de la méthylation des régions riches en CpG alentours. Ainsi, nous supposons que cette méthylation non-CpG est une résultante d'une propagation excessive de la 5mC générée par les ADN méthyltransférases *de novo*, puisque cette dernière n'est plus présente au niveau des gamètes matures.



**Figure 1 : Schéma de l'hématopoïèse**

Au sommet de la hiérarchie, il y a les LT-HSC (Long term HSC) qui sont les cellules à la base de la lignée hématopoïétique et qui possède une capacité d'auto-renouvellement dit de long terme (représentée ici par une flèche en trait plein). Les ST-HSC (Short term HSC) possèdent elles aussi une capacité d'auto-renouvellement mais de court terme (représentée ici par une flèche en pointillée). Les ST-HSCs se différencient en MPP (Multipotent progenitors) qui vont-elles mêmes se différencier soit en CLP (Common Lymphoid Progenitors), soit en CMP (Common Myeloid Progenitors). Tandis que les CLP pourvoiront à la lignée lymphoïde, les CMP donneront lieu à la lignée myéloïde. Parmi les cellules lymphoïdes, on retrouve les lymphocytes B et T ainsi que les cellules NK (natural killer). La lignée myéloïde recouvre deux autres types de cellules progénitrices, les GMP (Granulocyte-macrophage progenitors) et les MEP (Megakaryocyte-erythroid progenitor) qui donneront respectivement les monocytes, macrophages, granulocytes, neutrophiles, éosinophiles, basophiles, et les érythrocytes, mégakaryocytes ainsi que les plaquettes. Adapté de Luc *et al.*, 2007.

# Résultats

---

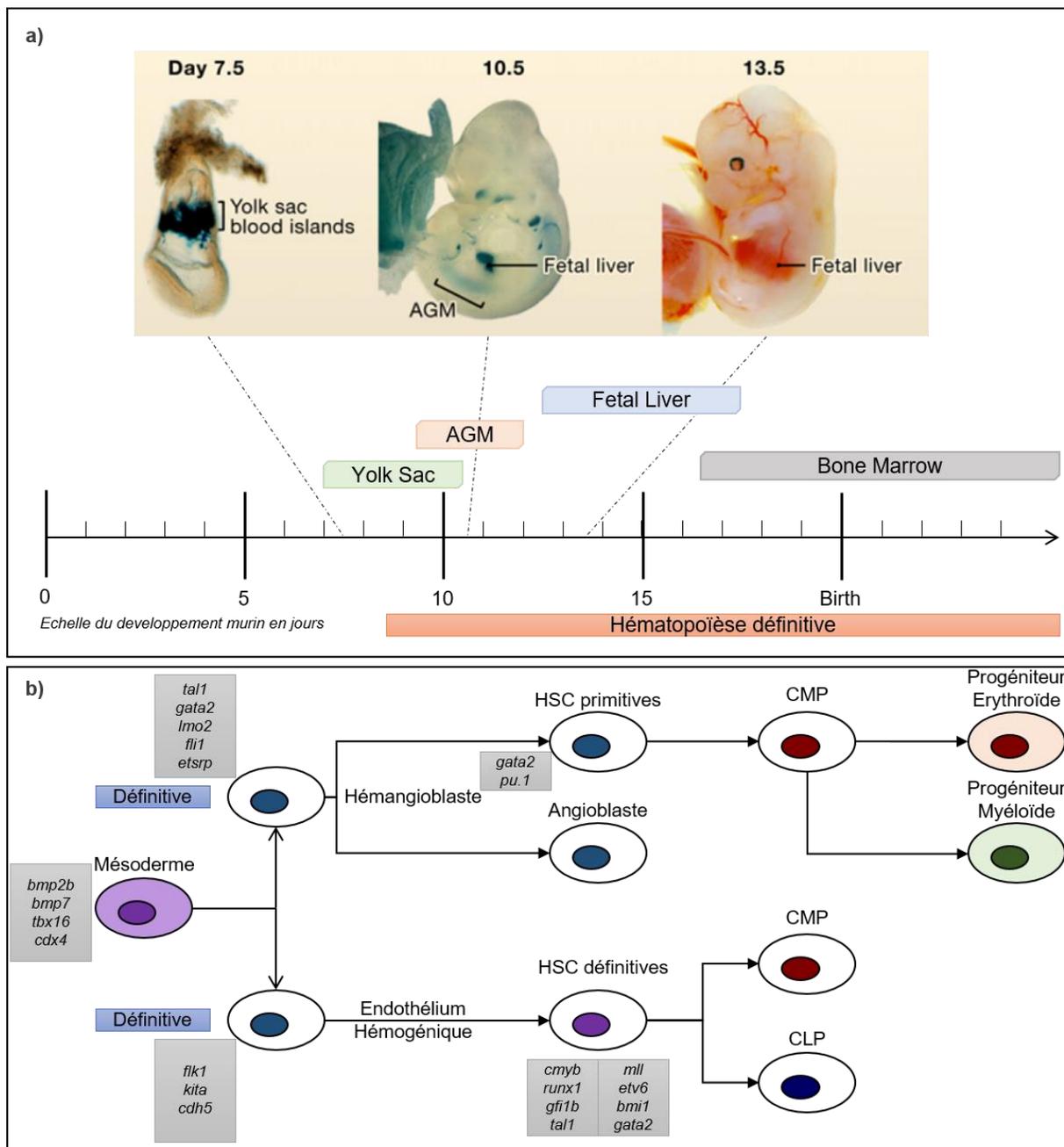
## II. La méthylation de l'ADN au cours du développement embryonnaire murin des cellules souches hématopoïétiques (HSCs)

### A. Introduction

#### 1. Le développement des HSCs

Les cellules souches hématopoïétiques sont les cellules à l'origine de l'ensemble des cellules du système hématopoïétique. Cette famille cellulaire est composée des cellules sanguines matures, des cellules de la moelle osseuse ainsi que des cellules lymphoïdes et myéloïdes. Les HSCs prolifèrent et se différencient continuellement afin d'entretenir et de renouveler l'ensemble du système hématopoïétique. C'est ce que l'on nomme l'hématopoïèse.

Les HSCs possèdent une hiérarchie qui a été basée sur l'isolement de ces dernières par cytométrie en flux (Akashi et al., 2000; Weissman, 2000). Cette hiérarchie permet de mieux appréhender leur potentiel de différenciation (**Figure 1**). Les cellules à la base du système hématopoïétique sont les LT-HSC (long term hematopoietic stem cells). Elles sont nommées ainsi car leur capacité d'auto-renouvellement est de long terme, c'est-à-dire tout au long de la vie de l'individu. Ces dernières se différencient en ST-HSC (short term hematopoietic stem cells), qui possèdent une capacité d'auto-renouvellement de seulement quelques mois. Par la suite, elles se différencient en MPP (multipotent progenitors), qui sont les progéniteurs les plus immatures pouvant se différencier soit en CLP (common lymphoid progenitors), qui donneront naissance à l'ensemble des cellules lymphoïdes (lymphocytes B et T, cellules NK, cellules dendritiques), soit en CMP (common myeloid progenitors), qui pourvoient à la lignée myéloïde (granulocytes, macrophages, cellules dendritiques d'une part et les érythrocytes et mégacaryocytes par l'intermédiaire des progéniteurs GMP (Granulocyte-macrophage progenitors) et MEP (Megakaryocyte-erythroid progenitors)). Les LT-HSC, les ST-HSC ainsi que les MPP peuvent être regroupées sous le terme de cellules LSK puisqu'elles expriment communément les marqueurs de surface Sca1 et cKit (Okada et al., 1992; Uchida and Weissman, 1992). De plus, elles n'expriment pas les marqueurs de différenciation des cellules matures (appelés Lin, lineage), tels que : CD3 (pour les cellules T), B220 (pour les cellules B), CD11b (pour les cellules myéloïdes), Ter119 (pour les érythrocytes), Gr1 (pour les granulocytes). Ainsi, ces cellules sont annotées LSK, c'est-à-dire Lin<sup>-</sup>Sca1<sup>+</sup>cKit<sup>+</sup> (Spangrude et al., 1988).



**Figure 2 : L'hématopoïèse au cours du développement murin**

**a)** L'hématopoïèse primitive démarre au niveau du sac vitellin (YS) à partir de E7,5. Par la suite, l'hématopoïèse, dite définitive se met en place dans le placenta et dans la région de l'AGM à partir de E8,5, puis les HSC migrent vers le foie fœtal (FL). Dans les photos ci-dessus qui proviennent de Orkin and Zon, 2008, on observe le YS grâce à un marquage LacZ lié à l'expression du gène hématopoïétique *Gata1*. Dans l'embryon à E10,5, le marquage LacZ est sous le contrôle d'un autre gène hématopoïétique, *Runx1*, qui permet de visualiser l'AGM ainsi que le foie fœtal. **b)** Hématopoïèse primitive et définitive pendant le développement embryonnaire avec les principaux gènes impliqués dans ces processus dans les boîtes grises. YS = Yolk Sac; AGM = Aorta-gonad-mesonephros; FL = Fetal liver; BM : Bone marrow. Adapté de Bogdan and Zon, 2013

Le développement hématopoïétique a lieu dès le développement embryonnaire et se fait en deux temps. Tout d'abord l'hématopoïèse dite primitive et transitoire (Galloway and Zon, 2003) démarre à E7,5 en impliquant un progéniteur érythroïde qui va se différencier en macrophages et érythrocyte afin d'oxygéner l'embryon (Orkin and Zon, 2008; Palis and Yoder, 2001). Ces progéniteurs érythroïdes qui sont nommés EMP (Erythroid-macrophage progenitors) et sont originaires de l'hémangioblaste (endothélium possédant des propriétés hémogéniques), apparaissent dans le sac vitellin (McGrath et al., 2011; Orkin and Zon, 2008). On dit d'eux qu'ils sont transitoires car ils ne possèdent pas de capacité d'auto-renouvellement. Par la suite, l'hématopoïèse dite définitive se met en place et implique les HSCs (Galloway and Zon, 2003) qui pourvoiront à l'ensemble des cellules sanguines de l'organisme adulte. Les HSC apparaissent dans la région mésonéphro-gonado-aortique, appelé AGM (aorta-gonad-mesonephros) à E8,5 et dans le placenta. Elles migrent ensuite dans le foie fœtal à E10,5, puis dans la moelle osseuse dans laquelle elles résident dès la naissance (Cumano and Godin, 2007; Orkin and Zon, 2008) (Figure 2-A). Le foie fœtal apparaît comme l'organe principal des HSC et de l'hématopoïèse lors de la seconde moitié de la gestation et jusqu'à la naissance.

L'hématopoïèse primitive est contrôlée par deux facteurs de transcription : GATA1 et PU.1 (appelé aussi SFPI1). Ces deux facteurs présentent une relation mutuellement inhibitrice pour l'engagement de la différenciation des EMP en cellules myéloïdes ou érythroïdes primitives. *Gata1* est le principal régulateur des érythrocytes (Cantor and Orkin, 2002). En effet, les souris déficientes pour *Gata1* meurent au cours de la gestation dû à un arrêt de la différenciation des proérythroblastes en érythrocytes matures. Ainsi, *Gata1* régulerait préférentiellement l'expression des gènes érythroïdes. A l'inverse, *Sfpi1* est le principal régulateur de la différenciation myéloïde (Scott et al., 1994). Il a été démontré que ces deux facteurs de transcription interagissent, suggérant une inhibition mutuelle afin de permettre une différenciation des progéniteurs en l'un ou l'autre type cellulaire (Cantor and Orkin, 2002).

L'hématopoïèse définitive est quant à elle régulée par différents facteurs de transcription (Figure 2-B) dont RUNX1 et MYB (Mucenski et al., 1991; Wang et al., 1996). Ainsi, il a été reporté que les souris déficientes pour *Runx1* n'ont plus la capacité de créer des cellules myéloïdes, lymphoïdes et érythroïdes définitives. Concernant les souris déficientes pour *Myb*, ces dernières meurent suite à un défaut de l'érythropoïèse au niveau du foie fœtal.

Les HSC fœtales et adultes diffèrent par de nombreux critères, tels que leur prolifération, leurs marqueurs de surface et leurs réponses aux cytokines. Par exemple, les HSCs du foie fœtal expriment les marqueurs de surface AA4.1 et CD11b alors que ce n'est pas le cas des cellules

adultes (Jordan et al., 1990; Morrison et al., 1995). Enfin, environ 30% des HSCs fœtales sont prolifératives tandis que les HSC adultes sont principalement quiescentes (75% des HSC) (Cheshier et al., 1999; Morrison et al., 1995).

## 2. HSCs et Méthylation de l'ADN, que sait-on ?

Depuis l'émergence des techniques de détection de la méthylation de l'ADN, de nombreuses études ont porté sur l'analyse des différents profils de 5mC au cours de l'hématopoïèse adulte. Ainsi, les différentes sous populations cellulaires de progéniteurs hématopoïétiques ont été analysées chez la souris et ont révélé de fortes variations des profils de méthylation de l'ADN en relation avec l'expression génique (Ji et al., 2010). Elles ont amené à l'identification de nouveaux gènes et voies moléculaires impliqués dans l'auto-renouvellement des HSCs ainsi que dans leur choix de destinée cellulaire au cours de leur différenciation (Ji et al., 2010). De plus, il a été démontré que la méthylation de l'ADN présente un fort dynamisme qui se traduit par un gain ou une perte progressive de la 5mC, spécifique à chacun des différents lignages (Cabezas-Wallscheid et al., 2014; Lipka et al., 2014). Par la suite, d'autres études ont mis en exergue la relation entre la dynamique de méthylation de l'ADN au niveau de certaines régions génomiques et l'association de facteurs de transcription dans les diverses lignées cellulaires hématopoïétiques (Bock et al., 2012). En effet, il a été suggéré que ces différents profils de 5mC soient le reflet ou peut-être même la cause des choix de destinée cellulaire des cellules hématopoïétiques, puisqu'ils pourraient influencer la liaison de facteurs de transcription hématopoïétiques clefs des différents lignages, comme par exemple GATA1, GATA2 ou encore RUNX1 (Cabezas-Wallscheid et al., 2014; Farlik et al., 2016; Lipka et al., 2014). Cette hypothèse est soutenue par une récente étude qui a mis en évidence que la méthylation de l'ADN d'un enhancer par DNMT1 contrôle l'activation du gène *Gata1* et l'initiation de l'érythropoïèse (Yu et al., 2017). Ainsi, il est suggéré que la méthylation de l'ADN fonctionne comme un gardien de la différenciation des HSCs et empêchant l'activation précoce ou au mauvais moment de facteurs de transcription au cours de leur développement. Pour exemple, il a été démontré que la lignée lymphoïde est associée à des gains de 5mC au niveau de loci contrôlant l'expression du programme de différenciation myéloïdes, garantissant ainsi la stabilité dans la lignée lymphoïde (Bock et al., 2012; Farlik et al., 2016; Ji et al., 2010). En accord avec cette hypothèse, les HSC de souris avec une inactivation conditionnelle du gène *Dnmt1* ne peuvent pas correctement réprimer certains facteurs myéloïdes, et donc peuvent se différencier en progéniteurs myéloïdes mais pas lymphoïdes (Broske et al., 2009; Trowbridge et al., 2009).

De plus, il a été observé qu'avec l'âge, la fonction des HSCs décline et que cela corrèle avec l'acquisition de profils épigénétiques aberrants (Beerman et al., 2013; Florian et al., 2012, 2013;

Sun et al., 2014). En effet, les HSCs âgées présentent une hyperméthylation au niveau des promoteurs de gènes impliqués dans leur différenciation, ainsi qu'une hypométhylation au niveau des gènes impliqués dans la maintenance des HSCs.

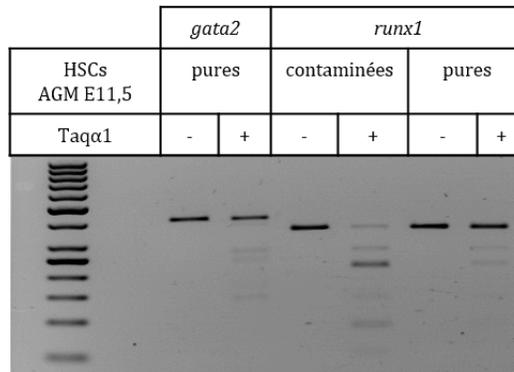
Le rôle de la méthylation de l'ADN est démontré par plusieurs études génétiques chez la souris. Comme décrit plus haut, il a notamment été démontré que les HSCs de souris présentant une réduction de l'expression de DNMT1 perdaient leur capacité d'auto-renouvellement ainsi que leur capacité de différenciation lymphoïde (Broske et al., 2009; Trowbridge et al., 2009). Par la suite, il a été démontré que *Dnmt3a* était, elle aussi, importante dans la capacité d'auto-renouvellement et de différenciation des HSCs puisque la délétion conditionnelle de cette enzyme dans les HSCs résultent en l'augmentation accrue des LT-HSCs au niveau de la moelle osseuse (Challen et al., 2012). En comparaison avec la délétion de *Dnmt3a*, celle de *Dnmt3b* n'est associée qu'à un phénotype moindre, suggérant que son absence est compensée par *Dnmt3a*. Cependant la double délétion conditionnelle de ces deux enzymes conduit à des effets synergiques, en bloquant les HSCs au stade d'auto-renouvellement constant et en bloquant leur différenciation (Challen et al., 2014).

Enfin, l'importance de ces profils épigénétiques pour les HSCs est renforcée par le fait que la méthylation de l'ADN joue un rôle prépondérant dans l'apparition des leucémies. Les gènes qui méthylent (*Dnmt3a*) ou déméthylent (*Tet2*) l'ADN sont fréquemment mutés dans les leucémies myéloïdes et lymphoïdes (syndromes myélodysplasiques, leucémies myéloïdes aiguës, leucémies myélomonocytaires chroniques, lymphomes, lymphomes T, lymphomes non hodgkiniens) (Couronne et al., 2012; Ley et al., 2010; Mayle et al., 2015). Chez la souris, la délétion de *Dnmt3a* n'induit de phénotype leucémique qu'au bout d'un certain nombre de transplantations suggérant que la perte de cette enzyme ne fait que prédisposer les cellules à une transformation leucémique. Une seconde étude portant sur l'inactivation de *Dnmt3a* arrive à la même conclusion (Cole et al., 2017). L'inactivation de *Tet2* chez la souris confirme un rôle dans la capacité d'auto-renouvellement et la différenciation des HSCs, puis que la délétion conditionnelle de cette enzyme dans les HSCs conduit à une augmentation de la capacité d'auto-renouvellement des HSCs, une expansion des différents compartiments des HSCs, une augmentation de la prolifération myéloïde et un dysfonctionnement de la différenciation myéloïde (Genovese et al., 2014; Jaiswal et al., 2017; Xie et al., 2014). On observe aussi des anomalies au niveau de la différenciation des cellules B et T dans ces souris (Quivoron et al., 2011). Néanmoins, tout comme la mutation au niveau de *Dnmt3a*, la délétion de *Tet2* ne confère qu'un phénotype pré-leucémique (Genovese et al., 2014; Jaiswal et al., 2017; Xie et al., 2014).

### 3. Objectifs du projet

L'ensemble des études génétiques réalisées à ce jour sont basées sur l'inactivation conditionnelle des gènes *Dnmt* dans les HSCs adultes avec un allèle MX1-Cre silencieux dans les souris saines et inductible par l'interféron. Ces études permettent d'étudier le rôle des *Dnmt* dans le maintien de la méthylation de l'ADN et de l'identité épigénétique des HSCs mais pas dans l'établissement de cette identité au cours du développement. Peu de choses sont connues sur la dynamique de méthylation de l'ADN au cours du développement dans les HSCs, et le rôle des *Dnmt* dans ce processus. C'est pourquoi j'ai initié une étude qui avait trois objectifs :

- Décrire la dynamique du méthylome au cours du développement embryonnaire des HSCs *in vivo* chez la souris
- Etudier les contributions de DNMT3A et DNMT3B dans la mise en place de l'identité de méthylation des HSCs au cours du développement
- Etudier les conséquences des perturbations de cette identité de méthylation dans les souris mutantes *Dnmt*. sur le transcriptome des HSCs et le phénotype des souris



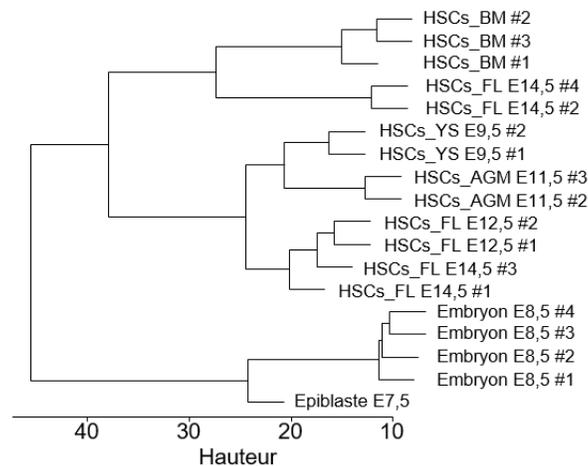
**Figure 3 :** Exemple de validation par COBRA des tris cellulaires de HSCs à E11,5

Visualisation sur gel d'agarose des profils de digestion de COBRA pour les gènes *gata2* et *runx1*, dans les échantillons d'HSCs triées à partir de l'AGM à E11,5. Les promoteurs des gènes *gata2* et *runx1* sont spécifiquement hypométhylés dans les cellules hématopoïétiques mais pas dans les autres lignages, permettant ainsi de valider l'efficacité du tri cellulaire. Le gel montre des exemples d'échantillons d'HSCs purs, dans lesquels les gènes *gata2* et *runx1* sont hypométhylés, et un exemple de cellules contaminées présentent un profil de digestion au niveau du gène *runx1*.

a)

	YS E9,5	AGM E11,5	FL E12,5	FL E14,5	BM
<b>Nombre d'HSCs/embryon</b>	1400	1000	2408	145	1706
<b>Nombre d'embryons disséqués</b>	65	70	49	85	11 adultes
<b>Nombre de librairies RRBS</b>	2	3	2	4	3

b)



**Figure 4 :** RRBS réalisés sur les échantillons de HSCs

**a) Tableau récapitulatif des dissections à chaque stade embryonnaire.** Au stade E9,5, environ 1400 HSCs sont récoltés par sac vitellin. 65 embryons ont été disséqués afin de réaliser des RRBS en duplicata. Au stade E11,5, environ 1000 HSCs sont récoltés par AGM. 70 embryons ont été disséqués afin de réaliser des RRBS en triplicata. Au stade E12,5, environ 2408 HSCs sont récoltés par foie fœtal. 49 embryons ont été disséqués afin de réaliser des RRBS en duplicata. Au stade E14,5, environ 145 HSCs sont récoltés par foie fœtal. 85 embryons ont été disséqués afin de réaliser des RRBS en quadruplicata. Au stade adulte dans la moelle osseuse, environ 1706 HSCs sont récoltés par individu. 11 adultes ont été disséqués afin de réaliser des RRBS en triplicata. **b) Dendrogramme des données de RRBS.** Les méthylomes des HSCs se distinguent fortement de celui des embryons. On remarque aussi que chacun des stades développementaux forment un groupe à part, à l'exception des HSCs des foies fœtaux à E14,5: sur les 4 échantillons, deux sont plus proches des HSCs provenant de foies fœtaux à E12,5, tandis que les deux autres sont plus proches de ceux de la moelle osseuse. YS= Sac Vitellin; AGM= Aorta-Gonad-Mesonephros; FL= Foie fœtal; BM= Moelle Osseuse.

## **B. Résultats**

### **1. Isolation des HSCs**

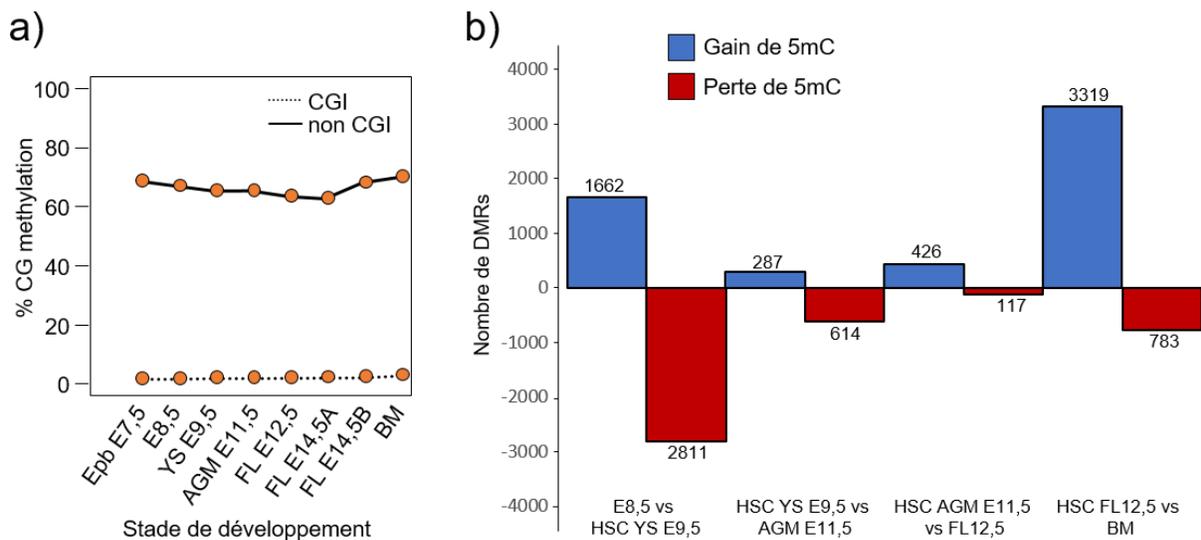
La première étape de mon projet a consisté à mettre au point des protocoles d'isolation des HSCs à partir des différents organes embryonnaires, c'est-à-dire au niveau du sac vitellin (YS) à E9,5, de l'axe gonado-mésonephrique (AGM) à E11,5, du foie fœtal (FL) à E12,5 et E14,5 et enfin, au niveau de la moelle osseuse (BM) de souris adultes âgés de 4-5 semaines. Pour ce faire, j'ai disséqué les différents organes de souris C57Bl/6J puis j'ai marqué les cellules issues de ces différents organes avec un cocktail spécifique d'anticorps couplés à des fluorochromes afin de pouvoir trier les HSCs par cytométrie en flux. Concernant les cocktails d'anticorps utilisés pour marquer les HSCs, je me suis basée sur la publication de [McKinney-Freeman et al., 2012](#) (Matériel & méthodes\_ **Figure B** ; **Annexe 2**). Lors de la mise au point de ces protocoles de tris, j'ai trié une seconde fois mes échantillons afin de vérifier leur pureté. Ces derniers révèlent une efficacité de plus de 75%, sachant que j'ai choisi d'être très strict au niveau de la sélection de ma fenêtre de tri, en préférant diminuer le nombre final de cellules triées afin d'assurer leur pureté. Lorsqu'on regarde les post-tris, l'ensemble des cellules triées présentent une appartenance claire à une population cellulaire distincte, comme l'atteste les graphes en contour (Matériel & méthodes\_ **Figure B**). A la vue du faible nombre de cellules trié par organe (**Figure 4a**), je n'ai plus réalisé ces post-tris lors des expériences d'analyse du méthylome par RRBS.

Pour valider les tris, j'ai vérifié par COBRA l'état de méthylation des promoteurs des gènes *Runx1* et *Gata2*. Nous avons choisi ces gènes car ils sont exprimés dans les HSCs et nos études préliminaires indiquaient que leur promoteur est spécifiquement hypométhylé dans les HSCs mais pas dans les cellules non-hématopoïétiques. La **Figure 3** montre des exemples de validations de tri cellulaire par COBRA. En effet, lors des tests COBRA, seuls les HSCs purs ne présentent pas de profil de digestion, tandis que les HSCs contaminées par d'autres types cellulaires présentent un profil de digestion qui atteste de la présence de méthylation au niveau de ces promoteurs provenant des cellules contaminantes (**Figure 3**).

### **2. Dynamique du méthylome des HSCs au cours du développement**

#### **a. Cinétique globale**

Grâce aux nombreuses dissections et tris cellulaires réalisés sur les différents organes hématopoïétiques, j'ai pu estimer le nombre d'HSCs par embryon en fonction des différents organes (**Figure 4a**). Ainsi, j'ai dénombré en moyenne 1400 HSCs par sac vitellin (YS) à E9,5, 1000 HSCs par axe gonado-mésonephrique (AGM) à E11,5, environ 2400 dans le foie fœtal



**Figure 5 : Deux vagues de reprogrammation du méthylome dans les HSCs**

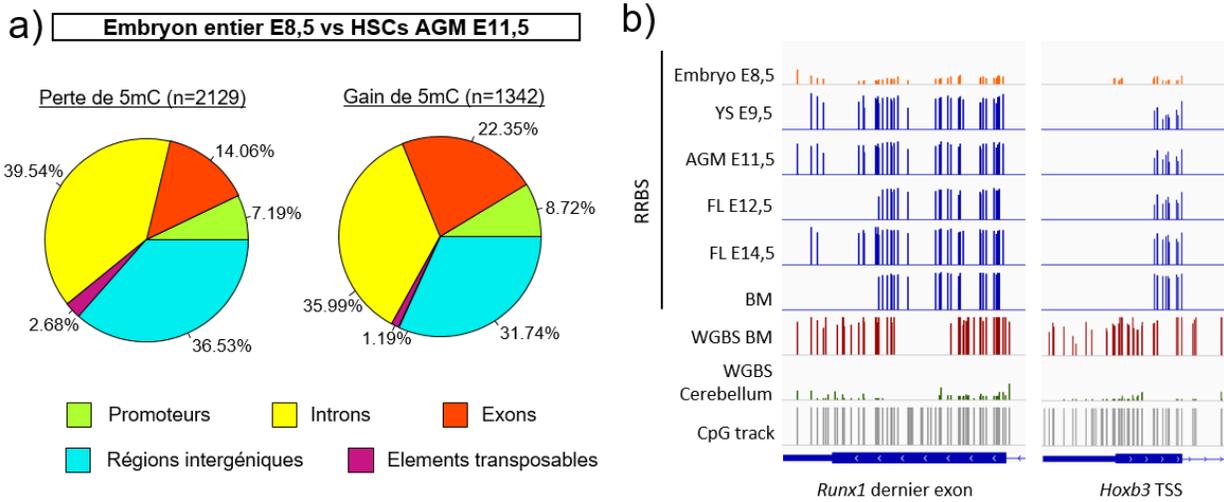
**a)** Variation du niveau global de méthylation des CpG dans les différents échantillons de HSCs. **b)** Ce graphique en barres représentent le nombre de régions différentiellement méthylés (DMR, >20% de différence de méthylation, en bleu: gain de méthylation, en rouge: perte de méthylation) entre les différents stades de développement des HSCs. Une première différence importante est observée entre le méthylome d'embryons E8,5 et celui des HSCs du sac vitellin à E9,5 (YS). Le méthylome des HSCs provenant du sac vitellin, de l'AGM et du foie fœtal ne présentent pas de grandes différences. Par contre, on constate un second changement drastique entre le méthylome d'HSCs provenant du foie fœtal et celui de la moelle osseuse. Ainsi, il semblerait qu'il existe une première vague de reprogrammation dès la mise en place des HSCs dans le sac vitellin et une seconde lorsque les HSCs colonisent la moelle osseuse.

CGI= îlots CpG; Epb E7,5 = Epiblastes E7,5; E8,5 = embryons E8,5 ; YS = HSCs de sacs vitellins à E9,5; AGM = HSCs d'AGM à E11,5; FL E12,5 =HSCs de foies fœtaux à E12,5; FL E14,5A : HSCs de foies fœtaux à E14,5 dont le méthylome est plus proche de ceux des foies fœtaux à E12,5; FL E14,5B : HSCs de foies fœtaux à E14,5 dont le méthylome est plus proche de ceux de la moelle osseuse; BM = LT-HSCs de la moelle osseuse.

(FL) à E12,5 et 145 dans le foie fœtal à E14,5. Cette diminution du nombre d'HSCs dans le foie à E14,5 peut être lié à la migration de ces cellules vers la moelle osseuse. Enfin, j'ai dénombré 1706 LT-HSCs au niveau de la moelle osseuse d'une souris adulte âgée de 4 à 5 semaines (**Figure 4a**).

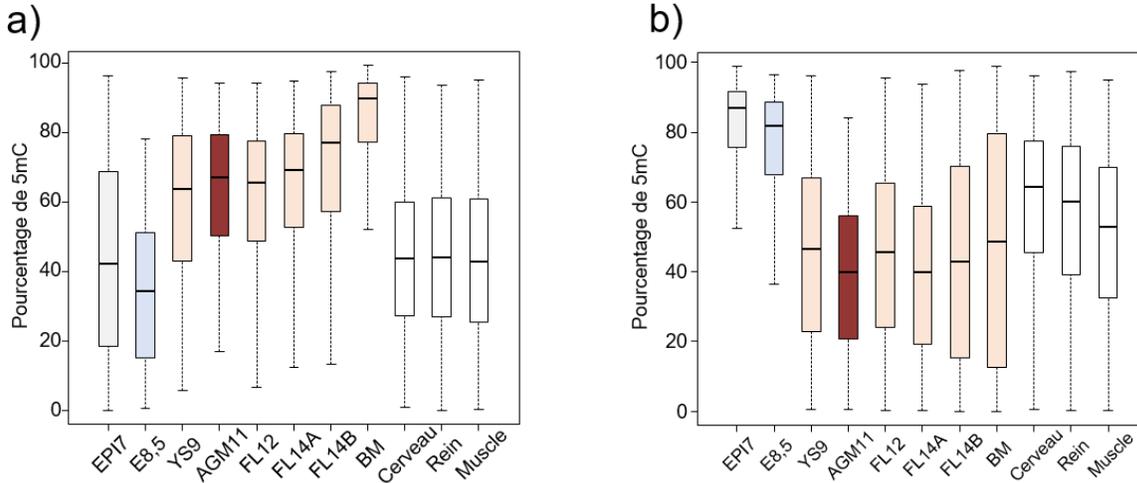
Une fois l'ensemble des échantillons de HSCs collectés, j'ai extrait l'ADN génomique et réalisé des expériences de préparation de bibliothèques RRBS afin d'étudier la dynamique de la 5mC au cours du développement dans les HSCs. J'ai réalisé des expériences de RRBS en duplicata pour le sac vitellin à E9,5 et le foie fœtal à E12,5, en triplicata pour l'AGM E11,5 et la BM, et en quadruplicata pour le foie fœtal à E14,5. Nous avons tout d'abord analysé les données en effectuant une classification hiérarchique des différents échantillons (**Figure 4b**). Au regard de ce dendrogramme, on constate que les méthylomes d'épiblastes E7,5 et d'embryons entiers à E8,5 sont regroupés dans une sous-branche différente de celles des HSCs, attestant que les HSCs acquièrent au cours du développement une signature de méthylation très différente de celle des cellules embryonnaires précoces. Concernant les méthylomes des HSCs, ces derniers se regroupent en fonction de leur origine développementale avec une branche qui regroupe les échantillons YS et AGM, une branche qui regroupe les échantillons FL E12,5 et une branche qui regroupe les échantillons de moelle osseuse. Il y a une exception pour les méthylomes des HSCs provenant des FL E14,5. En effet deux échantillons FL E14,5 se rapprochent des échantillons FL E12,5 et deux autres se rapprochent des échantillons de moelle osseuse (**Figure 4b**). Nous interprétons ce résultat comme le reflet d'une possible transition de méthylation à E14,5 qui corrèle avec le début de la migration des HSCs du foie fœtal vers la moelle osseuse.

Nous n'avons pas observé de changement important du niveau global de méthylation entre les stades de développement (**Figure 5a**). Par la suite, nous avons donc analysé les différences de méthylation entre les différents stades de développement des HSCs. Tout comme l'avait suggéré les études précédentes au niveau de l'hématopoïèse adulte, les méthylomes des HSCs se différencient les uns des autres par l'acquisition ou la perte de méthylation (**Figure 5b**). Pour l'ensemble des analyses suivantes, nous avons recherché les régions différentiellement méthylées (DMRs) ayant au minimum 3 CpG et 20% de différence de taux de 5mC. Nous avons tout d'abord comparé le méthylome des HSCs provenant du sac vitellin à E9,5 avec celui d'embryons entiers E8,5. Cette comparaison a révélé 1662 régions hyperméthylées (représentées en bleu) et 2811 régions hypométhylées (représentées en rouge) dans les HSCs provenant d'un sac vitellin par rapport à un embryon entier E8,5 (**Figure**



**Figure 6 : Analyse des changements de méthylation de l'ADN dans les HSCs de l'AGM E11,5**

**a)** Les graphiques circulaires représentent la répartition des catégories de séquences parmi les DMRs identifiées entre les embryons E8,5 et les HSCs provenant de l'AGM à E11,5. **b)** Exemples de profils de méthylation dans le dernier exon du gène *Runx1* et le promoteur du gène *Hoxb3*. Chaque barre verticale représente le pourcentage de méthylation d'un CpG avec une échelle de 0 à 100%.



**Figure 7 : Les HSCs acquièrent une signature de méthylation très précocement**

Ces graphiques représentent les niveaux de méthylation dans les DMRs identifiées entre les embryons E8,5 et les HSCs provenant de l'AGM E11,5. Le graphique **a** représente l'ensemble des régions (n=1342) gagnant de la méthylation dans les HSCs de l'AGM en comparaison avec les embryon E8,5 tandis que le graphique **b** représente l'ensemble des régions (n=2129) qui se déméthylent dans les HSCs de l'AGM en comparaison avec les embryon E8,5.

EPI7 = Epiblastes E7,5; E8,5 = embryons E8,5 ; YS9 = HSCs de sacs vitellins à E9,5; AGM11 = HSCs d'AGM à E11,5; FL12 =HSCs de foies fœtaux à E12,5; FL14A : HSCs de foies fœtaux à E14,5 dont le méthylome est plus proche de ceux des foies fœtaux à E12,5; FL14B : HSCs de foies fœtaux à E14,5 dont le méthylome est plus proche de ceux de la moelle osseuse; BM = LT-HSCs de la moelle osseuse.

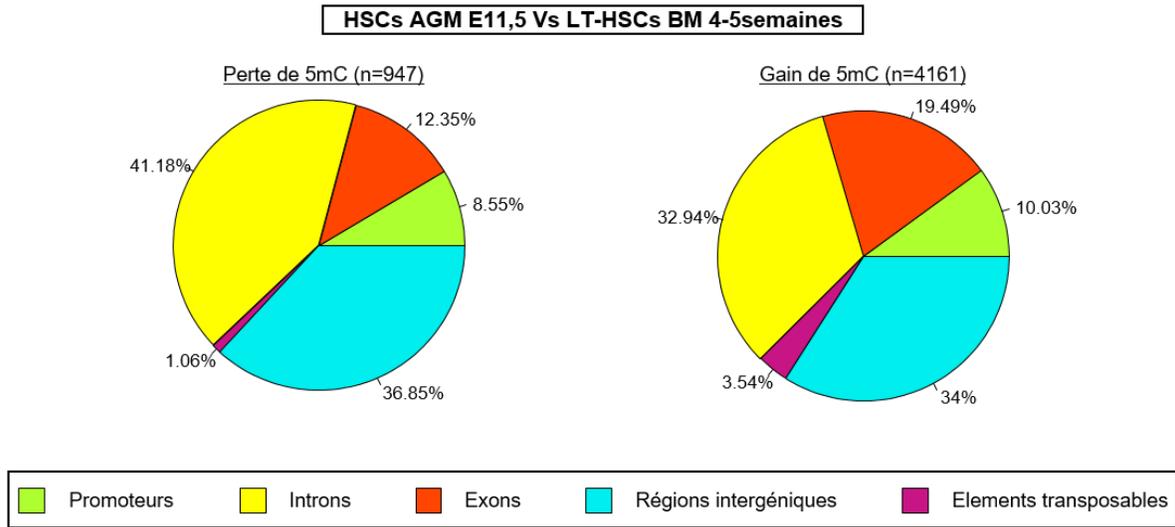
**5b).** Par la suite, on remarque que peu de régions sont différentiellement méthylées entre les HSCs provenant de sacs vitellins et d'AGM. En effet, nous avons pu identifier 287 régions hyperméthylées et 614 régions hypométhylées dans les AGM E11,5 par rapport aux YS E9,5. Il en est de même pour celles provenant des foies fœtaux à E12,5 avec 426 régions hyperméthylées et 117 régions hypométhylées dans les FL E12,5 par rapport aux AGM E11,5. Ayant vu que le méthylome des HSCs des FL E14,5 partageait à la fois des similitudes avec ceux des FL E12,5 et de la BM, nous avons choisi de ne pas les inclure dans cette étude. Enfin, les HSCs de la moelle osseuse acquièrent elles aussi un profil de méthylation particulier puisqu'on retrouve 3319 régions hyperméthylées et 783 régions hypométhylées dans les LT-HSCs de la BM par rapport aux HSCs des FL E12,5 (**Figure 5b**).

Ainsi, il semble qu'il existe deux vagues importantes d'acquisition des profils de 5mC dans les HSCs : une première entre l'embryon E8,5 et la mise en place des HSCs dans le YS et l'AGM, et une seconde lors de la sortie des HSCs du foie fœtal vers la moelle osseuse.

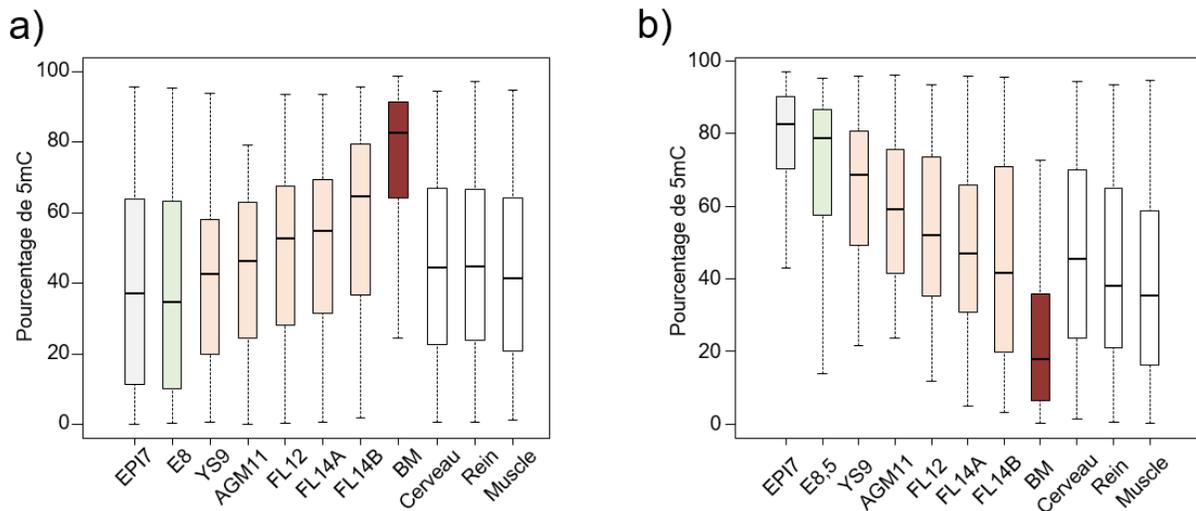
#### **b. La phase précoce d'acquisition de la signature hématopoïétique**

Nous nous sommes intéressés par la suite plus en détails aux changements précoces de méthylation de l'ADN observés entre l'embryon E8,5 et les HSCs provenant des AGM E11,5. Nous avons choisi l'AGM plutôt que le stade du sac vitellin pour deux raisons. Premièrement, à ces stades, les méthylomes des HSCs sont globalement identiques. De plus, les HSCs de l'AGM sont considérés comme étant définitives (**Galloway and Zon, 2003**) tandis que celles du sac vitellin qui sont dites primitives (**McGrath et al., 2011; Orkin and Zon, 2008**).

Nous avons voulu savoir quels types de séquence sont affectés par ces remaniements épigénétiques (**Figure 6a**). L'étude des DMRs entre les embryons E8,5 et les HSCs provenant de l'AGM à E11,5 révèle que le gain et la perte de 5mC concernent l'ensemble des différents compartiments génomiques. Sur les 2 129 régions hypométhylées dans les HSCs de l'AGM par rapport à l'embryon E8,5, ce sont majoritairement des introns (39.54%), des régions intergéniques (36.53%), des exons (14.06%) et des promoteurs (7.19%) qui sont affectés (**Figure 6a**). Le même profil est observé pour les régions hyperméthylées dans les HSCs de l'AGM E11,5 par rapport aux embryons E8,5 : elles sont principalement des introns (35.99%), des régions intergéniques (31.74%), des exons (22.35%) et des promoteurs (8.72%) (**Figure 6a**). Par exemple la **Figure 6b** montre des exemples de gain de méthylation importants dans les HSCs du sac vitellin et de l'AGM dans le dernier exon du gène *Runx1* et le promoteur du gène *Hoxb3*. Parmi les promoteurs affectés, on en dénombre exactement 164 hypométhylés et 123 hyperméthylés. Afin de savoir à quels processus biologiques ces derniers étaient associés, j'ai réalisé une analyse d'ontologie de gène à l'aide de l'application DAVID. Cette



**Figure 8 : Analyse des changements de méthylation de l'ADN dans les HSCs de la moelle osseuse**  
 Les graphiques circulaires représentent la répartition des catégories de séquences parmi les DMRs identifiées entre les HSCs de l'AGM à E11,5 et les HSCs adultes de la moelle osseuse.



**Figure 9 : Les HSCs semblent verrouiller leur signature de méthylation après colonisation de la BM**  
 Ces graphiques représentent les niveaux de méthylation dans les DMRs identifiées entre les HSCs de l'AGM E11,5 et celles de la moelle osseuse. Le graphique **a** représente l'ensemble des régions qui gagnent de la méthylation dans les HSCs de la moelle osseuse, tandis que le graphique **b** représente les régions qui se déméthylent dans les HSCs de la moelle osseuse.  
 EPI7 = Epiblastes E7,5; E8,5 = embryons E8,5 ; YS9 = HSCs de sacs vitellins à E9,5; AGM11 = HSCs d'AGM à E11,5; FL12 =HSCs de foies fœtaux à E12,5; FL14A : HSCs de foies fœtaux à E14,5 dont le méthylome est plus proche de ceux des foies fœtaux à E12,5; FL14B : HSCs de foies fœtaux à E14,5 dont le méthylome est plus proche de ceux de la moelle osseuse; BM = LT-HSCs de la moelle osseuse.

analyse révèle que les promoteurs hypométhylés dans des HSCs sont associés ( $p < 0.01$ ) principalement à des gènes impliqués dans la réponse immunitaire, tandis que les promoteurs hyperméthylés sont associés à des gènes impliqués dans le développement embryonnaire et celui du système nerveux, le transport et la régulation des ions, et enfin, la régulation positive de la migration cellulaire.

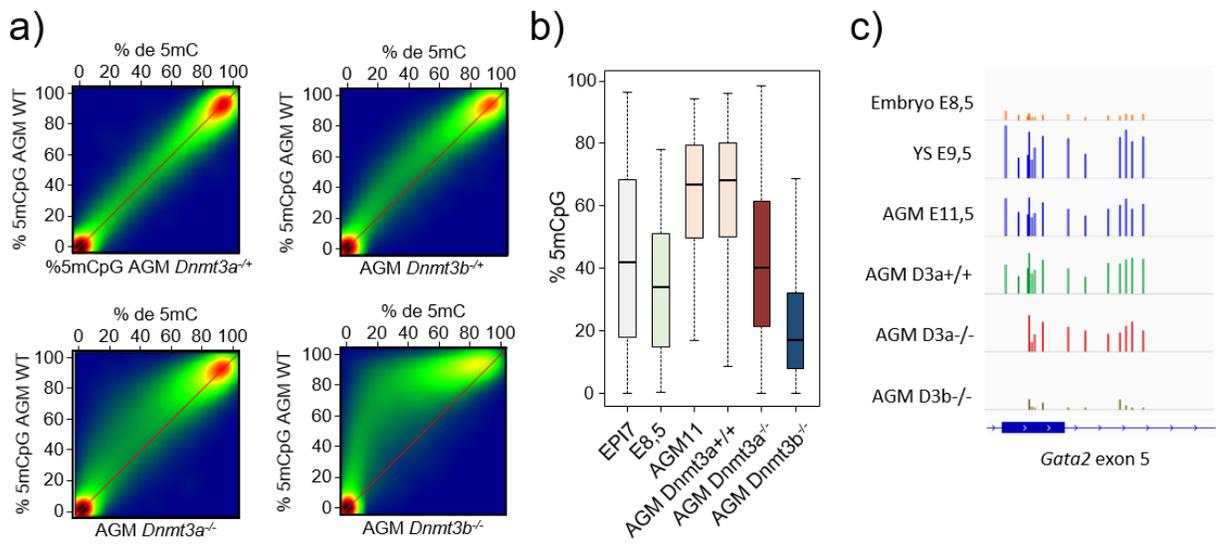
Ensuite nous avons étudié le devenir de la signature de méthylation acquise dans les HSCs de l'AGM E11,5. Concernant les DMRs hyperméthylées (**Figure 7a**), on constate que ces dernières continuent à se méthyliser au cours du développement des HSCs, jusqu'à atteindre un niveau moyen de 90% dans la moelle osseuse à l'âge adulte. En revanche, ces DMRs ne sont pas hyperméthylées dans d'autres tissus comme le cerveau, le rein et le muscle, ce qui montre que les HSCs acquièrent une signature de méthylome spécifique (**Figure 7a**). J'ai réalisé le même type d'analyse au niveau des DMRs hypométhylées entre les embryons E8,5 et les HSCs d'AGM E11,5 (**Figure 7b**). Au cours du développement, certaines de ces régions continuent de se déméthyliser tandis que certains réacquièrent de la méthylation au niveau de la moelle osseuse. Néanmoins, elles possèdent encore une fois un profil de méthylation spécifique en comparaison avec les méthylomes de cerveaux, de reins et de muscles.

Ainsi, il semble que les HSCs acquièrent très tôt, dans le sac vitellin et l'AGM, un profil de méthylation spécifique de leur devenir hématopoïétique.

### c. La phase tardive d'acquisition de la signature hématopoïétique

Ensuite, nous nous sommes intéressés aux changements tardifs de méthylation de l'ADN détectés dans les LT-HSCs provenant de la moelle osseuse, en étudiant les DMRs identifiées entre les HSCs provenant des AGM E11,5 et les LT-HSCs provenant de la moelle osseuse.

Nous avons cherché à savoir quels types de séquences sont affectés par ces remaniements épigénétiques (**Figure 8**). Sur les 947 régions hypométhylées, ce sont majoritairement des introns (41.18%), des régions intergéniques (36.85%), des exons (12.35%) et des promoteurs (8.55%) qui sont affectés (**Figure 8**). Au niveau des régions hyperméthylées ( $n=4161$ ), on constate qu'il s'agit principalement de régions intergéniques (34%), d'introns (32.94%), d'exons (19.49%), et de promoteurs (10.03%). Parmi les promoteurs affectés, on en dénombre exactement 86 hypométhylés et 607 hyperméthylés. L'analyse ontologique des promoteurs hypométhylés révèle qu'ils sont impliqués ( $p < 0.01$ ) dans les fonctions immunitaires. Les promoteurs hyperméthylés sont associés à des gènes impliqués dans le développement embryonnaire et la différenciation cellulaire mais sans enrichissement particulier dans une fonction biologique.



**Figure 10 : *Dnmt3b* a un rôle prépondérant pour la mise en place de la signature de méthylation des HSCs de l'AGM**

**a)** Comparaison des profils de méthylation des CpG dans les HSCs isolés à E11,5 d'AGM WT ou mutants pour *Dnmt3a* ou *Dnmt3b*. **b)** Graphique représentent le niveau de méthylation dans les DMRs identifiées hyperméthylées dans les HSCs de l'AGM E11,5 par rapport à l'embryon E8,5. **c)** Exemple de profil de méthylation dans le gène *Gata2*. Chaque barre verticale représente le pourcentage de méthylation d'un CpG avec une échelle de 0 à 100%. EPI7 = Epiblastes E7,5; E8,5 = embryons E8,5 ; YS= HSCs de sacs vitellins à E9,5.

Concernant les DMRs hyperméthylées (**Figure 9a**), on constate que ces dernières correspondent à des régions se méthylant au cours du développement murin dans les HSCs, jusqu'à atteindre un haut niveau de méthylation dans la moelle osseuse à l'âge adulte. En revanche, des régions ne sont pas méthylées dans d'autres tissus (**Figure 9a**), ce qui révèle encore une fois que la méthylation acquise dans les HSCs représente une signature de méthylation spécifique. J'ai réalisé le même type d'analyse au niveau des DMRs hypométhylées : ces régions se déméthylent et possèdent encore une fois un profil de méthylation spécifique en comparaison avec les méthylomes de cerveaux, de reins et de muscles (**Figure 9b**).

Ainsi, il semblerait que les HSCs finissent d'acquérir un profil de méthylation spécifique afin de verrouiller leur devenir hématopoïétique lors de leur passage dans la moelle osseuse.

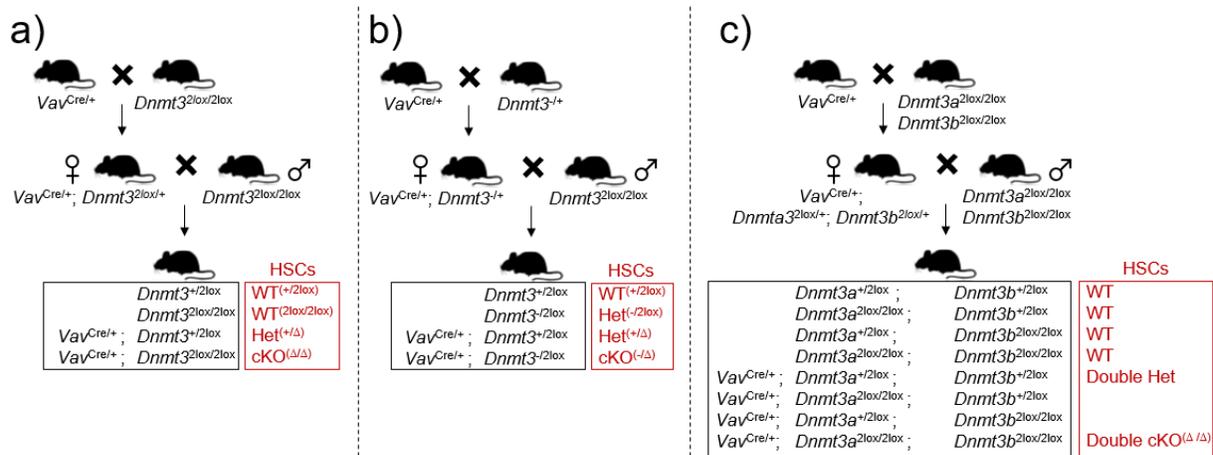
### **3. Dnmt3a et Dnmt3b sont acteurs dans la mise en place de la signature de méthylation des HSCs**

Lors de l'étude de la cinétique des méthylomes des HSCs au cours de leur développement, nous avons vu qu'il existait deux phases d'acquisition de leur signature hématopoïétique. Une première, très précoce, dès l'apparition des HSCs et une seconde, plus tardive, lors de la colonisation de la moelle osseuse à l'âge adulte. Par la suite, nous avons donc voulu savoir quelles enzymes étaient responsables de l'acquisition de ces profils de 5mC spécifiques.

#### **a. Dnmt3b est l'acteur principal de la méthylation précoce des HSCs**

Dans l'optique de savoir quelles enzymes étaient responsables de l'acquisition du profil de méthylation hématopoïétique précoce dans les HSCs de l'AGM, j'ai croisé entre elles des souris *Dnmt3a*<sup>-/-</sup> d'une part, et des souris *Dnmt3b*<sup>-/-</sup> d'une autre part, afin de récolter des HSCs d'AGM à E11,5 d'embryons *Dnmt3a*<sup>-/-</sup> et *Dnmt3b*<sup>-/-</sup>. L'inactivation de DNMT3A ou DNMT3B n'a pas d'effet sur le nombre d'HSCs récoltés à partir de l'AGM. Finalement, j'ai réalisé des RRBS en duplicata pour les génotypes suivants : *Dnmt3a*<sup>+/+</sup> (WT) ; *Dnmt3a*<sup>+/-</sup> (Het) ; *Dnmt3a*<sup>-/-</sup> (KO) ; *Dnmt3b*<sup>+/+</sup> (Het) ; *Dnmt3b*<sup>-/-</sup> (KO).

Nous avons réalisé une analyse globale des méthylomes des HSCs de l'AGM des différents génotypes en réalisant un graphique de corrélation (**Figure 10a**). Tandis que la délétion hétérozygote de *Dnmt3a* n'affecte pas le méthylome des HSCs en comparaison avec le WT, ce dernier est légèrement hypométhylé dans les HSCs *Dnmt3a*<sup>-/-</sup> (**Figure 10a**). En revanche, l'inactivation hétérozygote de *Dnmt3b* induit une hypométhylation du méthylome des HSCs par rapport à celui des WT, et cette perte de 5mC est beaucoup plus marquée dans les



**Figure 11 : Stratégie d'obtention de souris KO conditionnelles dans la lignée hématopoïétique**

**a) Création de souris KO conditionnelles.** Des souris  $Vav^{Cre/+}$  ont été croisées avec des souris  $Dnmt3^{2lox/2lox}$ . Les femelles  $Vav^{Cre/+}; Dnmt3^{2lox/+}$  ont par la suite été croisées avec des mâles  $Dnmt3^{2lox/2lox}$  afin de générer les différents génotypes contrôles et KO conditionnel (cKO) dans les HSCs. Ces expériences ont été menées sur le gène  $Dnmt3a$  puis sur le gène  $Dnmt3b$ . **b) Création de souris KO conditionnelles avec une efficacité de délétion supérieure.** Les souris  $Vav^{Cre/+}$  ont été croisées avec des souris  $Dnmt3^{-/+}$ . Les femelles  $Vav^{Cre/+}; Dnmt3^{-/+}$  ont par la suite été croisées avec des mâles  $Dnmt3^{2lox/2lox}$  afin de générer les différents génotypes contrôles et KO conditionnel (cKO) dans les HSCs. Dans ce cas, on estime que le KO conditionnel peut être plus efficace car la recombinase Cre n'a qu'un seul allèle 2lox à exciser. Ces expériences ont été mené sur le gène  $Dnmt3a$  puis sur le gène  $Dnmt3b$ . **c) Création de souris double KO conditionnelles.** Des souris  $Vav^{Cre/+}$  ont été croisé avec des souris  $Dnmt3a^{2lox/+}; Dnmt3b^{2lox/+}$ . Les femelles  $Vav^{Cre/+}; Dnmt3a^{2lox/+}; Dnmt3b^{2lox/+}$  ont par la suite été croisé avec des mâles  $Dnmt3a^{2lox/2lox}; Dnmt3b^{2lox/2lox}$  afin de générer les différents génotypes contrôles et double cKO.

HSCs *Dnmt3b*<sup>-/-</sup> (**Figure 10a**). Ces résultats corroborent parfaitement ceux obtenus précédemment au laboratoire dans les embryons *Dnmt3a*<sup>-/-</sup> et *Dnmt3b*<sup>-/-</sup> (Auclair et al., 2014).

Pour étudier précisément le rôle de ces enzymes dans la mise en place de la signature de méthylation spécifique des HSCs de l'AGM, nous avons analysé la méthylation des DMRs que nous avons identifiées entre les embryons E8,5 et les HSCs de l'AGM. Les résultats montrent que la méthylation des DMRs est réduite à la fois dans les HSCs des AGM *Dnmt3a*<sup>-/-</sup> et *Dnmt3b*<sup>-/-</sup>, ce qui suggère que les deux enzymes coopèrent. Cependant la baisse de méthylation est plus drastique dans les HSCs des AGM *Dnmt3b*<sup>-/-</sup>, indiquant que DNMT3B est l'enzyme majoritairement responsable de la mise en place des profils de méthylation spécifiques des HSCs de l'AGM (**Figure 10b**). Ceci peut être visualisé par exemple dans le gène *Gata2* (**Figure 10c**).

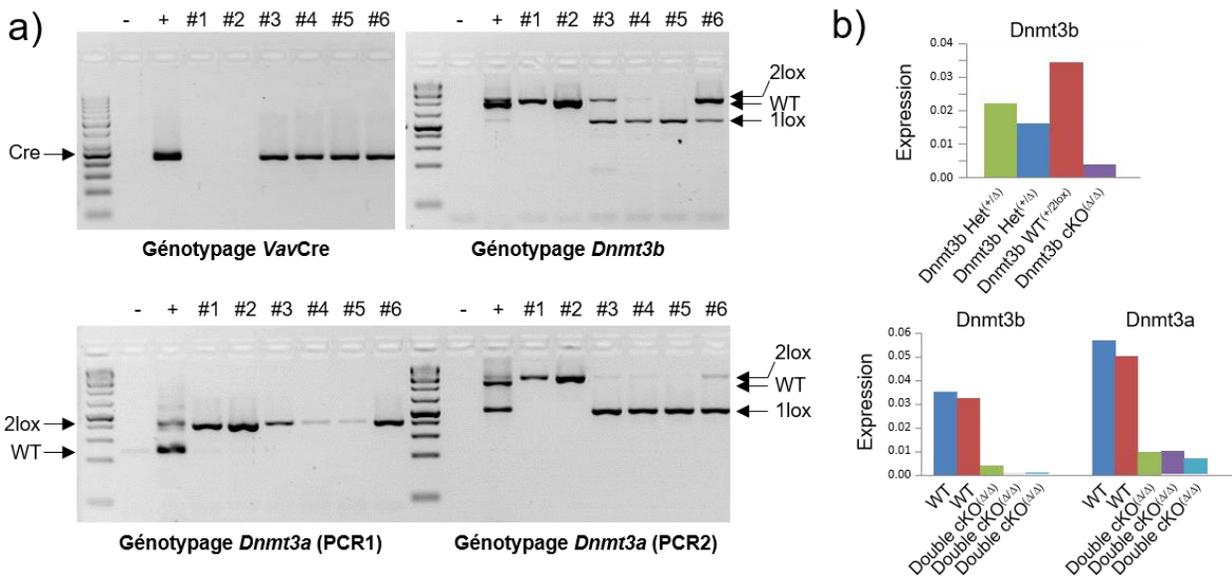
#### b. Génération de modèles de souris avec une inactivation conditionnelle des gènes *Dnmt3a/b* dans les HSCs au cours du développement

Dans un second temps, j'ai cherché à savoir quelles enzymes étaient responsables de l'acquisition du profil de méthylation hématopoïétique au cours du développement tardif.

J'ai réalisé une partie de mon étude sur des animaux présentant une délétion totale de *Dnmt3a* en récoltant les HSCs provenant de la moelle osseuse à 3 semaines car ces souris meurent entre 4 et 5 semaines (Okano et al., 1999). Je n'ai pas fait cela pour *Dnmt3b* car la délétion totale de ce gène induit une mort embryonnaire à mi-gestation (Okano et al., 1999).

Afin de mieux appréhender le rôle de ces deux enzymes dans l'hématopoïèse, j'ai donc généré des souris portant une inactivation conditionnelle de *Dnmt3a* ou *Dnmt3b*, ou des deux gènes, spécifiquement dans les HSCs embryonnaires (**Figure 11**). Pour ce faire, j'ai utilisé des souris *Vav*<sup>Cre/+</sup> (Georgiades et al., 2002) exprimant spécifiquement la CRE recombinase au niveau des cellules souches hématopoïétiques à partir de E11,5, ainsi que dans le lignage endothélial. D'une part, j'ai généré des souris conditionnelles en croisant les animaux *Vav*<sup>Cre/+</sup> avec des animaux *Dnmt3a*<sup>2lox/2lox</sup> ou *Dnmt3b*<sup>2lox/2lox</sup>, selon le schéma de croisements décrit dans la **Figure 11a**. J'ai ainsi pu obtenir des animaux *Vav*<sup>Cre/+</sup> *Dnmt3*<sup>2lox/2lox</sup> où le gène *Dnmt3a* ou *Dnmt3b* est inactivé conditionnellement dans les HSCs au cours des stades précoces de développement embryonnaire (**Figure 11a**).

En parallèle, j'ai aussi effectué d'autres séries de croisements entre des animaux *Vav*<sup>Cre/+</sup> et des souris *Dnmt3a*<sup>-/+</sup> ou *Dnmt3b*<sup>-/+</sup>, dans le but d'obtenir des animaux *Vav*<sup>Cre/+</sup> *Dnmt3*<sup>-/+</sup>. Puis ces animaux ont été croisés avec des mâles *Dnmt3a*<sup>2lox/2lox</sup> ou *Dnmt3b*<sup>2lox/2lox</sup>, selon le schéma de croisement décrit dans la **Figure 11b**, afin d'obtenir des animaux *Vav*<sup>Cre/+</sup> *Dnmt3*<sup>2lox</sup> dans



**Figure 12 : Exemples de génotypes des souris KO conditionnelles**

**a)** Exemples de gels d'agarose de génotypage de HSCs isolées à partir de souris de génotypes *VavCre*<sup>0/0</sup>; *Dnmt3a*<sup>2lox/2lox</sup>; *Dnmt3b*<sup>2lox/2lox</sup> (#1-2) et *VavCre*<sup>1/0</sup>; *Dnmt3a*<sup>2lox/2lox</sup>; *Dnmt3b*<sup>2lox/2lox</sup> (double cKO, #3-6). Le contrôle négatif (-) correspond à de l'eau tandis que le contrôle positif (+) correspond à l'ADNg d'un animal *VavCre*<sup>1/0</sup>; *Dnmt3a*<sup>2lox/+</sup>; *Dnmt3b*<sup>2lox/+</sup>. Concernant le génotypage de *Dnmt3a*, deux PCRs sont faites: une pour différencier l'allèle 2lox (400pb) de l'allèle WT (815pb) et une seconde pour différencier l'allèle 1lox (550pb) de l'allèle WT (815pb) et 2lox (951pb). Dans ce génotypage, on constate que seules les souris #4 et #5 présentent une très bonne efficacité de délétion à la fois des gènes *Dnmt3a* et *Dnmt3b* dans les HSCs. **b)** Exemples de validations par RT-qPCR de l'efficacité de délétion des exons floxés dans les HSCs. Les graphiques montrent le niveau d'expression mesuré par RT-qPCR dans les exons floxés, normalisé par rapport à l'expression de deux gènes de ménages (*Rpl13* et *Actb*).

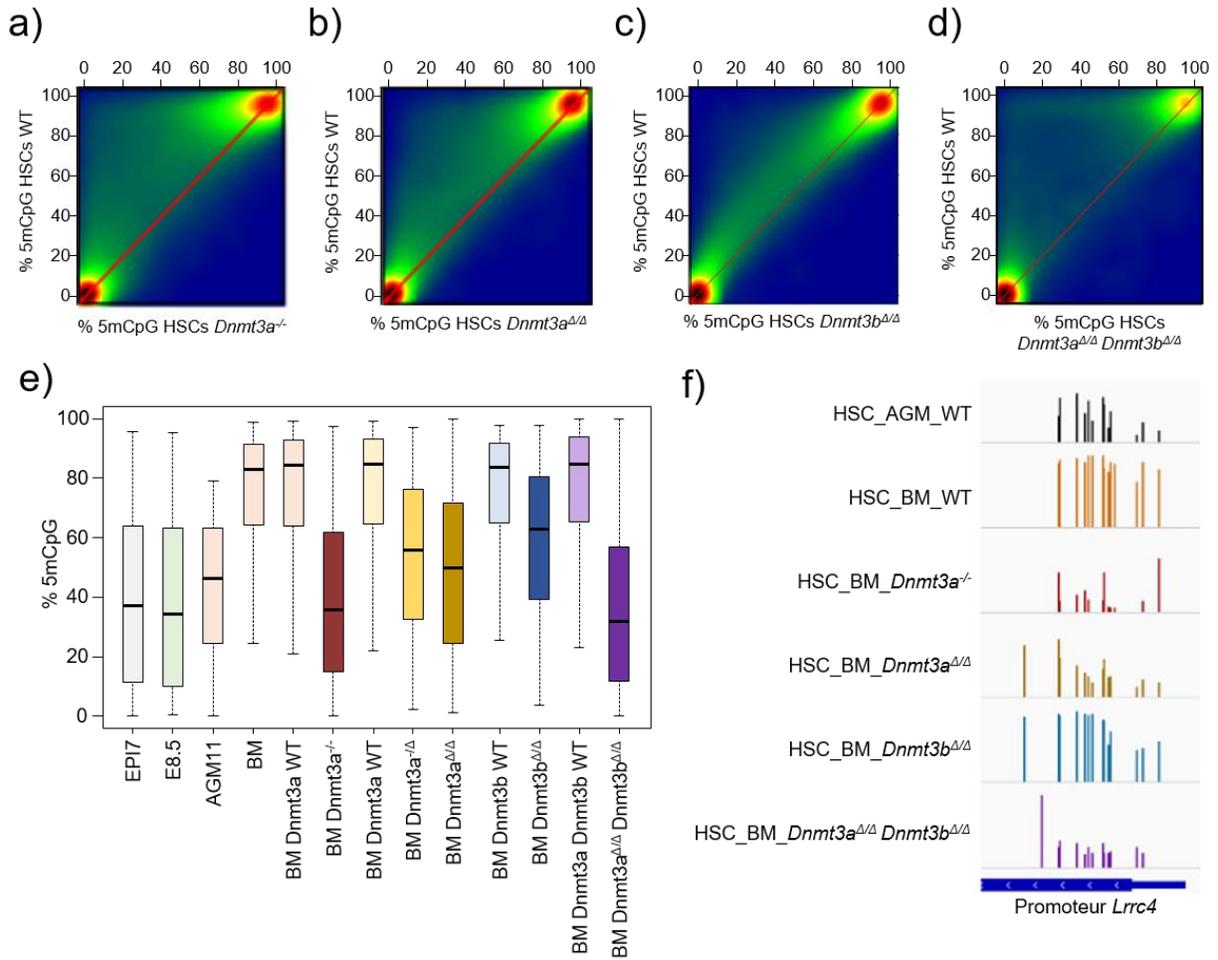
lesquels l'efficacité de recombinaison des sites Lox est potentiellement plus efficace. En effet, dans ces lignées, il n'y a qu'un seul allèle à déléter (**Figure 11b**).

Enfin, j'ai réalisé des croisements permettant d'étudier les conséquences de la double inactivation conditionnelle de *Dnmt3a* et *Dnmt3b* dans les HSCs. Pour cela j'ai croisé les animaux *Vav<sup>Cre/+</sup>* avec des animaux *Dnmt3a<sup>2lox/2lox</sup>* ; *Dnmt3b<sup>2lox/2lox</sup>*, selon le schéma de croisements décrit dans la **Figure 11c**. J'ai ainsi obtenu des animaux *Vav<sup>Cre/+</sup>* ; *Dnmt3a<sup>2lox/2lox</sup>* ; *Dnmt3b<sup>2lox/2lox</sup>* où les deux enzymes sont simultanément inactivées dans les HSCs (**Figure 11c**).

Pour chaque série de croisement, j'ai isolé les HSCs de la moelle osseuse dans les animaux cKO, ainsi que les animaux contrôles (WT ou Het) issus des mêmes portées. Lors de ces expériences, j'ai cette fois-ci trié une population mixte de ST-HSC et de MPP car le nombre de LT-HSC est trop insuffisant dans les animaux KO conditionnels pour réaliser un RRBS. Néanmoins, j'ai comparé le méthylome de cette population mixte triée à partir d'animaux sauvages avec ceux des LT-HSCs que j'avais réalisés lors de l'étude de la cinétique au cours du développement, et j'ai observé que les méthylomes des LT-HSCs et ST-HSCs sont extrêmement similaires.

Dans le but de vérifier l'efficacité de recombinaison des allèles *Dnmt3a-2lox* et *Dnmt3b-2lox* dans les HSCs, j'ai systématiquement génotypé les échantillons de HSCs en réalisant quatre PCRs : une pour valider la présence de la *VavCre*, une pour valider la recombinaison du site 2lox pour le gène *Dnmt3b* et deux autres pour valider la recombinaison du site 2lox pour le gènes *Dnmt3a* (**Figure 12a**). Par ailleurs, j'ai également quantifié l'efficacité de délétion des exons floxés par RT-qPCR sur des échantillons d'ARN isolés à partir des HSCs (**Figure 12b**). Ces validations montrent que l'efficacité de délétion des allèles floxés des gènes *Dnmt3a* et *Dnmt3b* peut être variable selon les animaux mais qu'elle est souvent supérieure à 80% (**Figure 12a-b**).

Une fois les échantillons de HSCs récoltés, j'ai préparé des bibliothèques RRBS pour un grand nombre d'échantillons, dont le détail est donné dans l'**annexe 5**. En résumé, j'ai pu préparer et séquencer des bibliothèques RRBS pour deux animaux *Dnmt3a<sup>-/-</sup>*, trois animaux *Dnmt3a* cKO, cinq animaux *Dnmt3b* cKO, trois animaux double cKO, et un grand nombre d'animaux de génotypes contrôles.



**Figure 13 : Rôles de *Dnmt3a* et *Dnmt3b* dans l'acquisition des profils de méthylation des HSCs de la moelle osseuse**

**a-d)** Comparaisons des niveaux de méthylation des CpG dans les HSCs de moelle osseuse d'animaux WT par rapport aux animaux *Dnmt3a*<sup>-/-</sup>, *Dnmt3a* cKO, *Dnmt3b* cKO et double cKO. **e)** Graphique représentant le niveau de méthylation, dans les génotypes indiqués, des DMRs qui acquièrent de la méthylation dans les HSCs de moelle osseuse par rapport aux HSCs de l'AGN E11,5. **f)** Exemples de profils de méthylation dans le promoteur du gène *Lrrc4* qui acquièrent la méthylation dans la moelle osseuse. Chaque barre verticale représente le pourcentage de méthylation d'un CpG avec une échelle de 0 à 100%.

EPI7 = Epiblastes E7,5; E8,5 = embryons E8,5 ; AGM11 = HSCs d'AGM à E11,5; BM = HSCs de la moelle osseuse.

c. *Dnmt3a* et *Dnmt3b* participent à la mise en place des profils de méthylation des HSCs de la moelle osseuse

Une fois les bibliothèques RRBS séquencées, nous avons réalisé une analyse globale des méthylomes des HSCs provenant de la moelle osseuse dans les souris de différents génotypes (**Figure 13**).

Les HSCs de moelle osseuse isolées à partir d'animaux *Dnmt3a*<sup>-/-</sup> (c'est-à-dire présentant une délétion totale de *Dnmt3a* présentent une hypométhylation d'un grand nombre de séquences du génome comparé au génotype sauvage (**Figure 13a**). La délétion conditionnelle de *Dnmt3a* dans les HSCs (*Dnmt3a*<sup>Δ/Δ</sup>) récapitule en grande partie cette hypométhylation par rapport aux animaux contrôles (**Figure 13b**). Au contraire, la délétion conditionnelle de *Dnmt3b* dans les HSCs (*Dnmt3b*<sup>Δ/Δ</sup>) induit un effet plus limité sur le méthylome des HSCs, même si on note quand même une hypométhylation par rapport aux animaux contrôles (**Figure 13c**). Enfin, la double inactivation conditionnelle de *Dnmt3a* et *Dnmt3b* (*Dnmt3a*<sup>Δ/Δ</sup> *Dnmt3b*<sup>Δ/Δ</sup>) semble avoir un effet synergique (**Figure 13d**).

Pour quantifier plus précisément la contribution de *Dnmt3a* et *Dnmt3b*, nous avons étudié la méthylation des DMRs qui normalement gagnent de la méthylation entre les HSCs de l'AGM E11,5 et les HSCs de la moelle osseuse (**Figure 13e**). Cette analyse nous informe que les régions se méthylant normalement au niveau des HSCs de la moelle osseuse, ne sont plus à même de se méthyler dans les animaux *Dnmt3a*<sup>-/-</sup> (**Figure 13e-f**). Dans ces animaux, la méthylation de ces DMRs dans les HSCs de la moelle osseuse est proche du niveau observé dans les HSCs de l'AGM à E11,5, ou dans l'épiblaste à E7,5 et l'embryon E8,5. La méthylation de ces DMRs est également fortement compromise dans les HSCs d'animaux *Dnmt3a* cKO (**Figure 13e-f**), ce qui valide le rôle prépondérant de *Dnmt3a* dans la mise en place de la signature de méthylation des HSCs de la moelle osseuse. De manière surprenante, l'effet est moins marqué dans les animaux *Dnmt3a*<sup>-Δ</sup> par rapport aux animaux *Dnmt3a*<sup>Δ/Δ</sup>, ce qui pourrait s'expliquer par une efficacité de la Cre recombinase variable entre animaux. La méthylation des DMRs est également réduite dans les HSCs d'animaux *Dnmt3b* cKO (*Dnmt3b*<sup>Δ/Δ</sup>) (**Figure 13e-f**), mais l'impact de l'inactivation de *Dnmt3b* est moindre que celui de *Dnmt3a*. Enfin on note un effet synergique de la double inactivation conditionnelle de *Dnmt3a* et *Dnmt3b* (**Figure 13e-f**). En effet dans les HSCs isolées d'animaux double cKO (*Dnmt3a*<sup>Δ/Δ</sup> *Dnmt3b*<sup>Δ/Δ</sup>), la méthylation des DMRs est à un niveau identique de celui des embryons E8,5.

En conclusion, nos résultats indiquent que les deux enzymes DNMT3A et DNMT3B collaborent pour mettre en place la signature de méthylation des HSCs dans la moelle osseuse, mais que DNMT3A a un rôle plus important que DNMT3B.

d. L'inactivation conditionnelle de *Dnmt3a* dans les HSCs semble influencer la survie des animaux

La suite logique de mes résultats consistait à étudier quelles sont les conséquences sur le transcriptome et l'homéostasie hématopoïétique lorsque la mise en place des profils de méthylation spécifiques des HSCs de la moelle osseuse est perturbée dans les souris *Dnmt3b* cKO.

Mon absence du laboratoire m'a empêchée pour l'instant de répondre à ces questions. Cependant j'ai pu faire des observations préliminaires intéressantes sur le phénotype dans animaux *Dnmt3* cKO. Dans la lignée où le gène *Dnmt3a* est délété de façon conditionnelle amélioré (-/ $\Delta$ ), on observe que l'ensemble des animaux cKO (n=3) sont morts entre 7 mois et un an, et ce, indépendamment du sexe de l'animal. Par contre, aucun animal délété conditionnellement pour *Dnmt3b*, qu'il soit cKO (-/ $\Delta$ ) (n=2) ou cKO ( $\Delta$ / $\Delta$ ) (n=5) n'est mort jusqu'à l'âge de 1 an et demi. Ces observations corrént avec le fait que nos données montrent que *Dnmt3a* a un rôle plus important que *Dnmt3b* dans la mise en place de la signature de méthylation des HSCs de la moelle osseuse (**Figure 13e-f**). Elles ouvrent des perspectives intéressantes suggérant que la mise en place de profils de méthylation spécifiques dans les HSCs de la moelle osseuse est importante pour maintenir la stabilité du transcriptome et l'identité des HSCs, et garantir l'homéostasie hématopoïétique.

## Discussion et Perspectives : HSCs

---

En utilisant des méthodes de cartographie à haute résolution, nous avons pu mettre en évidence, pour la première fois, la cinétique de méthylation de l'ADN des cellules souches hématopoïétiques (HSCs) au cours de leur développement chez a souris. Grâce aux données de méthylomes d'HSCs à différents stades embryonnaires et à l'âge adulte, nous avons mis en évidence l'existence de deux phases d'acquisition de la signature de méthylation des HSCs.

### **Les HSCs acquièrent de façon précoce une signature de méthylation qui leur est propre.**

La première chose que nous avons constaté est que le méthylome des HSCs (YS E9,5) dites « primitives » (McGrath et al., 2011; Orkin and Zon, 2008) est très semblable à celles provenant de l'AGM à E11,5, alors que ces dernières sont considérées comme des HSCs « définitives » par la communauté scientifique (Galloway and Zon, 2003). Entre ces deux populations d'HSCs, il existe 51 promoteurs différenciellement méthylés pouvant avoir une incidence potentielle sur l'expression génique. Parmi ces derniers, on en retrouve 21 qui gagnent en méthylation dans les HSCs de l'AGM et 30 qui se déméthylent. Cependant, les différences de méthylation observées entre les HSCs provenant du sac vitellin à E9,5 et de l'AGM E11,5 sont faibles, de l'ordre de 20% seulement. Ainsi, malgré un transcriptome différent, comme en conclue l'étude de McKinney-Freeman et al., 2012, les cellules hématopoïétiques « primitives » sont très semblables aux cellules hématopoïétiques « définitives » de l'AGM, du point de vue de leur méthylome. Cela suggère que leur différence au niveau transcriptomique n'est pas uniquement liée à la méthylation de l'ADN mais peut-être liée à leur état chromatinien (Hu and Zhao, 2014). Une analyse par WGBS nous permettrait de couvrir l'ensemble du génome et nous permettrait de vérifier si les résultats que nous observons en RRBS sont le reflet du génome dans sa globalité.

Les HSCs provenant des AGM étant considérées comme les premières HSCs « définitives », ce sont leur méthylome que nous avons comparé à celui d'embryons E8,5. Lors de cette analyse, nous avons mis en exergue que le profil de méthylation des HSCs leur est propre et qu'elles l'acquièrent très tôt dès leur apparition dans l'embryon précoce. Cette signature épigénétique précoce se caractérise davantage par des pertes de méthylation que des gains de méthylation. Ainsi parmi les DMRs, nous avons constaté qu'il existe 164 promoteurs hypométhylés et 123 promoteurs hyperméthylés dans les HSCs d'AGM E11,5. L'analyse ontologique nous apprend que les promoteurs hypométhylés appartiennent à des gènes impliqués dans les fonctions immunitaires, tandis que les promoteurs hyperméthylés sont associés à des gènes impliqués dans des fonctions diverses du développement embryonnaire et celui du système nerveux. Ainsi, il

semblerait que lors de la mise en place des HSCs, ces dernières déméthylent un certain nombre de promoteurs qui sont nécessaires à l'expression de leur transcriptome hématopoïétique et qu'à l'inverse, elles méthylent les promoteurs de certains gènes qui ne leur sont pas nécessaires.

Pour l'instant, nous n'avons pas réalisé de comparatif entre nos données de RRBS et les données transcriptomiques réalisées par [McKinney-Freeman et al., 2012](#), mais il serait intéressant de réaliser ce type d'analyse afin de voir s'il existe une corrélation entre la méthylation de ces promoteurs et leur expression. Il serait également très intéressant d'élargir nos données à l'échelle du génome entier par la technique du WGBS afin d'avoir une vue plus complète des changements de méthylation de l'ADN dans les HSCs de l'AGM. Grâce au protocole de WGBS mis au point dans le laboratoire pour de faibles quantités de cellules, cette technologie est désormais compatible avec le faible nombre de cellules HSCs isolées d'embryons précoces.

### **Les HSCs verrouillent leur signature de méthylation lors de la colonisation de la moelle osseuse.**

L'étude de la cinétique de méthylation des HSCs au cours du développement murin a aussi révélé une forte reprogrammation du méthylome lorsque les HSCs quittent le foie fœtal pour coloniser la moelle osseuse. En effet, lors du passage dans la moelle osseuse, les HSCs méthylent un grand nombre de régions dont notamment 417 promoteurs. Nous faisons l'hypothèse que ces gains de méthylation de l'ADN au cours du développement des HSCs servent à stabiliser le transcriptome des HSCs, comme pour verrouiller leur identité hématopoïétique. Cette hypothèse est soutenue par le fait que les promoteurs de gènes qui se méthylent au cours du développement des HSCs sont impliqués dans le développement embryonnaire et le développement d'autres lignages cellulaires non hématopoïétiques. Ainsi on peut imaginer que la méthylation de l'ADN participe à réprimer de façon stable certains gènes qui n'ont plus de fonction dans le lignage hématopoïétique, ce qui favorise la stabilité du lignage hématopoïétique. Là aussi, il serait très intéressant de compléter nos données de RRBS par des données de WGBS sur le génome entier, ce qui permettrait d'avoir une vue exhaustive des changements de méthylation de l'ADN notamment au niveau des régions « enhancers » qui sont très peu couvertes par le RRBS. Nous pourrions également analyser le transcriptome des mêmes échantillons par RNA-seq afin de corrélérer dans le temps les changements de méthylation de l'ADN et d'expression de gènes.

### **Rôle de DNMT3A et DNMT3B dans les HSCs**

Nous avons aussi mis en évidence, grâce à l'utilisation de souris KO conditionnelles, la contribution des enzymes DNMT3A et DNMT3B dans l'acquisition des profils de méthylation spécifiques des HSCs. Cela a été réalisé par le système Cre-Lox en utilisant une lignée de souris *Vav-Cre* qui induit une délétion dans progéniteurs hématopoïétiques à partir de E10,5. Néanmoins cette lignée présente également une expression de la Cre dans les cellules endothéliales et les cellules germinales, ce qui pourrait affecter notre étude. Une alternative aurait été d'utiliser le système Cre-Lox avec la lignée *Runx1-CreER* (Samokhvalov et al., 2007).

Concernant l'acquisition des profils de méthylation au niveau des HSCs de l'AGM, mes résultats démontrent que DNMT3B est l'enzyme principalement responsable des gains de méthylation, puisque la délétion de ce gène induit une hypométhylation plus importante que *Dnmt3a*. En effet, 117 promoteurs dépendent spécifiquement de *Dnmt3a*, tandis que 667 dépendent de *Dnmt3b*. La perte de méthylation au niveau de ces promoteurs pourrait potentiellement induire la réexpression de ces derniers qui sont principalement impliqués dans la mise en place de lignages cellulaires non-hématopoïétiques. A ce jour, aucune analyse transcriptomique n'a été réalisé afin de regarder la corrélation avec le taux de méthylation de ces promoteurs. Dans cette optique, nous allons réaliser des RNA-seq ou des RT-qPCR au niveau des deux génotypes.

Mes données montrent aussi, que DNMT3A est l'enzyme majoritairement responsable des gains de méthylation dans les HSCs de la moelle osseuse. L'inactivation conditionnelle de *Dnmt3a* induit une hypométhylation globale du génome beaucoup plus drastique que celle induite par l'inactivation de *Dnmt3b*. Elle touche aussi beaucoup plus de régions que *Dnmt3b* (4195 contre 2256). Par ailleurs, il semblerait que ces enzymes méthylent spécifiquement certaines régions du génome. En effet, 550 promoteurs dépendent spécifiquement de *Dnmt3a*, tandis que 69 dépendent de *Dnmt3b*. L'étude du double KO conditionnel nous apprend aussi que ces deux enzymes possèdent une action synergique puisque cette délétion induit une hypométhylation plus forte du génome que celles causées par l'inactivation simple de *Dnmt3a* ou *Dnmt3b*. L'ensemble de ces résultats sont globalement en accord avec les études menées dans le laboratoire du Dr Goodell (Challen et al., 2012, 2014). Cependant notre approche est fondamentalement différente puisque nous inactivons les enzymes *Dnmt* au cours du développement embryonnaire avant ou pendant l'acquisition de la signature de méthylation spécifique des HSCs. Au contraire, les études du Dr Goodell sont réalisées par inactivation conditionnelle chez l'adulte et mettent en jeu le rôle des enzymes DNMT3 dans le maintien du profil épigénétique des HSCs, ce qui pourrait expliquer

le fait qu'ils doivent réaliser plusieurs transplantations successives afin de révéler des perturbations épigénétiques.

De façon surprenante, nous avons aussi remarqué qu'un ICR, celui du gène *Mcts2*, qui est normalement méthylé à hauteur de 50% dans les cellules somatiques et embryonnaires, se méthyle à hauteur de 90% dans les HSCs provenant de la moelle osseuse. C'est l'enzyme DNMT3B qui est à l'origine de cette méthylation puisqu'en son absence le niveau de méthylation repasse à 50%, ce qui est le cas aussi dans les HSCs des animaux double cKO *Dnmt3a/Dnmt3b*. A ma connaissance, c'est un cas rare de plasticité de méthylation de l'ADN au niveau d'un ICR dans le développement somatique. Il sera intéressant d'étudier l'expression du gène *Mcts2* dans les différents mutants afin de comprendre l'utilité de l'hyperméthylation de cette ICR au niveau des HSCs de la moelle osseuse.

### **Importance de la signature épigénétique des HSCs**

La question la plus importante est de savoir quelle est l'importance des changements de méthylation de l'ADN pour le transcriptome et le fonctionnement des HSCs. L'étude plus approfondie des modèles de souris cKO nous permettra de répondre à ces questions en étudiant les conséquences des perturbations de la signature de méthylation de l'ADN des HSCs sur le transcriptome des HSCs et le phénotype des souris.

Concernant l'étude transcriptomique, il sera très intéressant de réaliser des expériences de RNA-seq sur les HSCs provenant d'animaux cKO. Ces expériences ont été planifiées durant la thèse mais je n'ai pas pu les mener à terme. Nous prévoyons de réaliser ces expériences en priorité sur les animaux *Dnmt3a* cKO et *Dnmt3a/3b* cKO car ils sont les plus perturbés au niveau de la méthylation de l'ADN des HSCs. Il sera particulièrement intéressant de tester si la baisse de méthylation corrèle avec l'activation ectopique de certains gènes embryonnaires ou non-hématopoïétiques, ce qui validerait l'hypothèse que la méthylation de l'ADN participe à stabiliser l'identité des HSCs en contribuant à maintenir certains gènes dans un état réprimé. Pour aller plus loin, nous pourrions également valider le rôle de la méthylation de l'ADN sur certains gènes cibles avec les nouvelles techniques de modification ciblée de la méthylation de l'ADN par dCas9-TET (Morita et al., 2016).

Enfin, il sera intéressant d'étudier les conséquences de l'inactivation des *Dnmt* dans les HSCs sur le phénotype des animaux. Pour l'instant, nous n'avons observé qu'un phénotype visible auprès des animaux *Dnmt3a* cKO (n=3). Ces derniers meurent entre 7 mois et un an. Par contre,

aucun phénotype n'est visible au niveau des animaux cKO *Dnmt3b* (n=7). Concernant les doubles cKO, je n'ai pas encore eu l'opportunité de suivre des cohortes d'animaux et d'observer leur phénotype. Le laboratoire est en train de régénérer la lignée double cKO afin d'étudier plus en détail son phénotype. Il serait intéressant d'étudier plusieurs paramètres sur ces souris, comme par exemple le nombre de HSCs récoltées dans les moelles osseuses, ainsi que l'homéostasie des différentes populations de cellules hématopoïétiques par des expériences de tri cellulaire afin de détecter des problèmes de prolifération et différenciation cellulaire des HSCs. En parallèle, nous allons mettre en place une collaboration avec l'Institut Clinique de la souris afin qu'il nous vienne en aide pour réaliser un phénotypage plus approfondi de ces trois lignées.

## Bibliographie

---

Aapola, U., Kawasaki, K., Scott, H.S., Ollila, J., Vihinen, M., Heino, M., Shintani, A., Kawasaki, K., Minoshima, S., Krohn, K., et al. (2000). Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. *Genomics* 65, 293–298.

Aapola, U., Lyle, R., Krohn, K., Antonarakis, S.E., and Peterson, P. (2001). Isolation and initial characterization of the mouse Dnmt3l gene. *Cytogenet. Cell Genet.* 92, 122–126.

Aapola, U., Liiv, I., and Peterson, P. (2002). Imprinting regulator DNMT3L is a transcriptional repressor associated with histone deacetylase activity. *Nucleic Acids Res.* 30, 3602–3608.

Achour, M., Jacq, X., Rondé, P., Alhosin, M., Charlot, C., Chataigneau, T., Jeanblanc, M., Macaluso, M., Giordano, A., Hughes, A.D., et al. (2008). The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of VEGF gene expression. *Oncogene* 27, 2187–2197.

Akashi, K., Traver, D., Miyamoto, T., and Weissman, I.L. (2000). A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* 404, 193–197.

Alvarez-Nuñez, F., Bussaglia, E., Mauricio, D., Ybarra, J., Vilar, M., Lerma, E., de Leiva, A., Matias-Guiu, X., and Thyroid Neoplasia Study Group (2006). PTEN promoter methylation in sporadic thyroid carcinomas. *Thyroid Off. J. Am. Thyroid Assoc.* 16, 17–23.

Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* 23, 185–188.

Anderson, R., Copeland, T.K., Scholer, H., Heasman, J., and Wylie, C. (2000). The onset of germ cell migration in the mouse embryo. *Mech Dev* 91, 61–68.

Antequera, F. (2003). Structure, function and evolution of CpG island promoters. *Cell. Mol. Life Sci. CMLS* 60, 1647–1658.

Aoki, A., Suetake, I., Miyagawa, J., Fujio, T., Chijiwa, T., Sasaki, H., and Tajima, S. (2001). Enzymatic properties of de novo-type mouse DNA (cytosine-5) methyltransferases. *Nucleic Acids Res.* 29, 3506–3512.

Aran, D., and Hellman, A. (2013). DNA methylation of transcriptional enhancers and cancer predisposition. *Cell* 154, 11–13.

Arand, J., Spieler, D., Karius, T., Branco, M.R., Meilinger, D., Meissner, A., Jenuwein, T., Xu, G., Leonhardt, H., Wolf, V., et al. (2012). In Vivo Control of CpG and Non-CpG DNA Methylation by DNA Methyltransferases. *PLoS Genet* 8, e1002750.

Arnaud, P. (2010). Genomic imprinting in germ cells: imprints are under control. *Reprod. Camb. Engl.* 140, 411–423.

- Arthur, E.I., Zlotogora, J., Lerer, I., Dagan, J., Marks, K., and Abeliovich, D. (1997). Transient neonatal diabetes mellitus in a child with invdup(6)(q22q23) of paternal origin. *Eur. J. Hum. Genet. EJHG* 5, 417–419.
- Auclair, G., and Weber, M. (2012). Mechanisms of DNA methylation and demethylation in mammals. *Biochimie* 94, 2202–2211.
- Auclair, G., Guibert, S., Bender, A., and Weber, M. (2014). Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. *Genome Biol* 15, 545.
- Augui, S., Nora, E.P., and Heard, E. (2011). Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat. Rev. Genet.* 12, 429–442.
- Babenko, A.P., Polak, M., Cavé, H., Busiah, K., Czernichow, P., Scharfmann, R., Bryan, J., Aguilar-Bryan, L., Vaxillaire, M., and Froguel, P. (2006). Activating mutations in the ABCC8 gene in neonatal diabetes mellitus. *N. Engl. J. Med.* 355, 456–466.
- Bachman, K.E., Rountree, M.R., and Baylin, S.B. (2001). Dnmt3a and Dnmt3b are transcriptional repressors that exhibit unique localization properties to heterochromatin. *J. Biol. Chem.* 276, 32282–32287.
- Bakken, A.H., and McClanahan, M. (1978). Patterns of RNA synthesis in early meiotic prophase oocytes from fetal mouse ovaries. *Chromosoma* 67, 21–40.
- Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., Guillou, F., and Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science* 354, 909–912.
- Barlow, D.P., Stöger, R., Herrmann, B.G., Saito, K., and Schweifer, N. (1991). The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature* 349, 84–87.
- Barrès, R., Osler, M.E., Yan, J., Rune, A., Fritz, T., Caidahl, K., Krook, A., and Zierath, J.R. (2009). Non-CpG methylation of the PGC-1 $\alpha$  promoter through DNMT3B controls mitochondrial density. *Cell Metab.* 10, 189–198.
- Bartolomei, M.S., Zemel, S., and Tilghman, S.M. (1991). Parental imprinting of the mouse H19 gene. *Nature* 351, 153–155.
- Barton, S.C., Surani, M.A., and Norris, M.L. (1984). Role of paternal and maternal genomes in mouse development. *Nature* 311, 374–376.
- Bashtrykov, P., Jankevicius, G., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2014). The UHRF1 protein stimulates the activity and specificity of the maintenance DNA methyltransferase DNMT1 by an allosteric mechanism. *J. Biol. Chem.* 289, 4106–4115.
- Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schubeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature*.

- Beerman, I., Bock, C., Garrison, B.S., Smith, Z.D., Gu, H., Meissner, A., and Rossi, D.J. (2013). Proliferation-Dependent Alterations of the DNA Methylation Landscape Underlie Hematopoietic Stem Cell Aging. *Cell Stem Cell* 12, 413–425.
- Berger, S.L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. *Genes Dev.* 23, 781–783.
- Berkyurek, A.C., Suetake, I., Arita, K., Takeshita, K., Nakagawa, A., Shirakawa, M., and Tajima, S. (2014). The DNA Methyltransferase Dnmt1 Directly Interacts with the SET and RING Finger-associated (SRA) Domain of the Multifunctional Protein Uhrf1 to Facilitate Accession of the Catalytic Center to Hemimethylated DNA. *J. Biol. Chem.* 289, 379–386.
- Berman, B.P., Weisenberger, D.J., Aman, J.F., Hinoue, T., Ramjan, Z., Liu, Y., Noushmehr, H., Lange, C.P.E., van Dijk, C.M., Tollenaar, R.A.E.M., et al. (2011). Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat. Genet.* 44, 40–46.
- Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J. Mol. Biol.* 203, 971–983.
- Bhatnagar, S., Zhu, X., Ou, J., Lin, L., Chamberlain, L., Zhu, L.J., Wajapeyee, N., and Green, M.R. (2014). Genetic and pharmacological reactivation of the mammalian inactive X chromosome. *Proc. Natl. Acad. Sci. U. S. A.* 111, 12591–12598.
- Biniszkiwicz, D., Gribnau, J., Ramsahoye, B., Gaudet, F., Eggan, K., Humpherys, D., Mastrangelo, M.-A., Jun, Z., Walter, J., and Jaenisch, R. (2002). Dnmt1 overexpression causes genomic hypermethylation, loss of imprinting, and embryonic lethality. *Mol. Cell. Biol.* 22, 2124–2135.
- Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev* 16, 6–21.
- Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. *Nature* 321, 209–213.
- Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* 40, 91–99.
- Birling, M.-C., Dierich, A., Jacquot, S., Héroult, Y., and Pavlovic, G. (2012). Highly-efficient, fluorescent, locus directed cre and FlpO deleter mice on a pure C57BL/6N genetic background. *Genesis* 50, 482–489.
- Bock, C. (2012). Analysing and interpreting DNA methylation data. *Nat. Rev. Genet.* 13, 705–719.
- Bock, C., Beerman, I., Lien, W.-H., Smith, Z.D., Gu, H., Boyle, P., Gnirke, A., Fuchs, E., Rossi, D.J., and Meissner, A. (2012). DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. *Mol. Cell* 47, 633–647.
- Bogdanović, O., and Lister, R. (2017). DNA methylation and the preservation of cell identity. *Curr. Opin. Genet. Dev.* 46, 9–14.

- Bogdanović, O., Smits, A.H., de la Calle Mustienes, E., Tena, J.J., Ford, E., Williams, R., Senanayake, U., Schultz, M.D., Hontelez, S., van Kruijsbergen, I., et al. (2016). Active DNA demethylation at enhancers during the vertebrate phylotypic period. *Nat. Genet.* *48*, 417–426.
- Booth, M.J., Branco, M.R., Ficz, G., Oxley, D., Krueger, F., Reik, W., and Balasubramanian, S. (2012). Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* *336*, 934–937.
- Borgel, J., Guibert, S., Li, Y., Chiba, H., Schubeler, D., Sasaki, H., Forne, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet* *42*, 1093–1100.
- Borum, K. (1961). Oogenesis in the mouse. A study of the meiotic prophase. *Exp. Cell Res.* *24*, 495–507.
- Bostick, M., Kim, J.K., Estève, P.-O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* *317*, 1760–1764.
- Boulard, M., Edwards, J.R., and Bestor, T.H. (2015). FBXL10 protects Polycomb-bound genes from hypermethylation. *Nat. Genet.* *47*, 479–485.
- Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* *431*, 96–99.
- Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science* *294*, 2536–2539.
- Bradley, A., Anastassiadis, K., Ayadi, A., Battey, J.F., Bell, C., Birling, M.-C., Bottomley, J., Brown, S.D., Bürger, A., Bult, C.J., et al. (2012). The mammalian gene function resource: the International Knockout Mouse Consortium. *Mamm. Genome Off. J. Int. Mamm. Genome Soc.* *23*, 580–586.
- Brandeis, M., Kafri, T., Ariel, M., Chaillet, J.R., McCarrey, J., Razin, A., and Cedar, H. (1993). The ontogeny of allele-specific methylation associated with imprinted genes in the mouse. *EMBO J.* *12*, 3669–3677.
- Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Names, A., Temper, V., Razin, A., and Cedar, H. (1994). Spl elements protect a CpG island from de novo methylation. *Nature* *371*, 435–438.
- Bristol-Gould, S.K., Kreeger, P.K., Selkirk, C.G., Kilen, S.M., Mayo, K.E., Shea, L.D., and Woodruff, T.K. (2006). Fate of the initial follicle pool: empirical and mathematical evidence supporting its sufficiency for adult fertility. *Dev. Biol.* *298*, 149–154.
- Bronner, C., Achour, M., Arima, Y., Chataigneau, T., Saya, H., and Schini-Kerth, V.B. (2007). The UHRF family: oncogenes that are drugable targets for cancer therapy in the near future? *Pharmacol. Ther.* *115*, 419–434.
- Broske, A.M., Vockentanz, L., Kharazi, S., Huska, M.R., Mancini, E., Scheller, M., Kuhl, C., Enns, A., Prinz, M., Jaenisch, R., et al. (2009). DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. *Nat Genet* *41*, 1207–1215.

Brunetti, L., Gundry, M.C., and Goodell, M.A. (2017). DNMT3A in Leukemia. *Cold Spring Harb. Perspect. Med.* 7, a030320.

Brzezinski, J., Shuman, C., Choufani, S., Ray, P., Stavropoulos, D.J., Basran, R., Steele, L., Parkinson, N., Grant, R., Thorner, P., et al. (2017). Wilms tumour in Beckwith-Wiedemann Syndrome and loss of methylation at imprinting centre 2: revisiting tumour surveillance guidelines. *Eur. J. Hum. Genet. EJHG.*

Butler, M.G. (2009). Genomic imprinting disorders in humans: a mini-review. *J. Assist. Reprod. Genet.* 26, 477–486.

Cabezas-Wallscheid, N., Klimmeck, D., Hansson, J., Lipka, D.B., Reyes, A., Wang, Q., Weichenhan, D., Lier, A., von Paleske, L., Renders, S., et al. (2014). Identification of Regulatory Networks in HSCs and Their Immediate Progeny via Integrated Proteome, Transcriptome, and DNA Methylome Analysis. *Cell Stem Cell* 15, 507–522.

Cantor, A.B., and Orkin, S.H. (2002). Transcriptional regulation of erythropoiesis: an affair involving multiple partners. *Oncogene* 21, 3368–3376.

Challen, G.A., Sun, D., Jeong, M., Luo, M., Jelinek, J., Berg, J.S., Bock, C., Vasanthakumar, A., Gu, H., Xi, Y., et al. (2012). Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat Genet* 44, 23–31.

Challen, G.A., Sun, D., Mayle, A., Jeong, M., Luo, M., Rodriguez, B., Mallaney, C., Celik, H., Yang, L., Xia, Z., et al. (2014). Dnmt3a and dnmt3b have overlapping and distinct functions in hematopoietic stem cells. *Cell Stem Cell* 15, 350–364.

Chedin, F., Lieber, M.R., and Hsieh, C.-L. (2002). The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proc. Natl. Acad. Sci. U. S. A.* 99, 16916–16921.

Chen, R., Zhang, Q., Duan, X., York, P., Chen, G.-D., Yin, P., Zhu, H., Xu, M., Chen, P., Wu, Q., et al. (2017a). The 5-Hydroxymethylcytosine (5hmC) Reader UHRF2 Is Required for Normal Levels of 5hmC in Mouse Adult Brain and Spatial Learning and Memory. *J. Biol. Chem.* 292, 4533–4543.

Chen, R., Zhang, Q., Duan, X., York, P., Chen, G.-D., Yin, P., Zhu, H., Xu, M., Chen, P., Wu, Q., et al. (2017b). The 5-Hydroxymethylcytosine (5hmC) Reader UHRF2 Is Required for Normal Levels of 5hmC in Mouse Adult Brain and Spatial Learning and Memory. *J. Biol. Chem.* 292, 4533–4543.

Chen, T., Ueda, Y., Xie, S., and Li, E. (2002). A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. *J. Biol. Chem.* 277, 38746–38754.

Chen, T., Ueda, Y., Dodge, J.E., Wang, Z., and Li, E. (2003). Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* 23, 5594–5605.

Chen, T., Tsujimoto, N., and Li, E. (2004). The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. *Mol. Cell. Biol.* 24, 9048–9058.

- Chen, Z.-X., Mann, J.R., Hsieh, C.-L., Riggs, A.D., and Chédin, F. (2005). Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family. *J. Cell. Biochem.* *95*, 902–917.
- Cheng, X., and Roberts, R.J. (2001). AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Res.* *29*, 3784–3795.
- Cheshier, S.H., Morrison, S.J., Liao, X., and Weissman, I.L. (1999). In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* *96*, 3120–3125.
- Chiang, J.W., Karlan, B.Y., Cass, L., and Baldwin, R.L. (2006). BRCA1 promoter methylation predicts adverse ovarian cancer prognosis. *Gynecol. Oncol.* *101*, 403–410.
- Chiquoine, A.D. (1954). The identification, origin, and migration of the primordial germ cells in the mouse embryo. *Anat. Rec.* *118*, 135–146.
- Chotalia, M., Smallwood, S.A., Ruf, N., Dawson, C., Lucifero, D., Frontera, M., James, K., Dean, W., and Kelsey, G. (2009). Transcription is required for establishment of germline methylation marks at imprinted genes. *Genes Dev* *23*, 105–117.
- Citterio, E., Papait, R., Nicassio, F., Vecchi, M., Gomiero, P., Mantovani, R., Di Fiore, P.P., and Bonapace, I.M. (2004). Np95 is a histone-binding protein endowed with ubiquitin ligase activity. *Mol. Cell. Biol.* *24*, 2526–2535.
- Cohen, N.M., Kenigsberg, E., and Tanay, A. (2011). Primate CpG islands are maintained by heterogeneous evolutionary regimes involving minimal selection. *Cell* *145*, 773–786.
- Cole, C.B., Russler-Germain, D.A., Ketkar, S., Verdoni, A.M., Smith, A.M., Bangert, C.V., Helton, N.M., Guo, M., Klco, J.M., O’Laughlin, S., et al. (2017). Haploinsufficiency for DNA methyltransferase 3A predisposes hematopoietic cells to myeloid malignancies. *J. Clin. Invest.* *127*, 3657–3674.
- Compere, S.J., and Palmiter, R.D. (1981). DNA methylation controls the inducibility of the mouse metallothionein-I gene lymphoid cells. *Cell* *25*, 233–240.
- Cooper, D.N., Taggart, M.H., and Bird, A.P. (1983). Unmethylated domains in vertebrate DNA. *Nucleic Acids Res.* *11*, 647–658.
- Cortázar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F., et al. (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* *470*, 419–423.
- Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., et al. (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* *146*, 67–79.
- Costello, J.F., Berger, M.S., Huang, H.S., and Cavenee, W.K. (1996). Silencing of p16/CDKN2 expression in human gliomas by methylation and chromatin condensation. *Cancer Res.* *56*, 2405–2410.

- Coucouvani, E.C., Sherwood, S.W., Carswell-Crumpton, C., Spack, E.G., and Jones, P.P. (1993). Evidence that the mechanism of prenatal germ cell death in the mouse is apoptosis. *Exp. Cell Res.* *209*, 238–247.
- Couronne, L., Bastard, C., and Bernard, O.A. (2012). TET2 and DNMT3A mutations in human T-cell lymphoma. *N Engl J Med* *366*, 95–96.
- Csankovszki, G., Nagy, A., and Jaenisch, R. (2001). Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. *J. Cell Biol.* *153*, 773–784.
- Cumano, A., and Godin, I. (2007). Ontogeny of the Hematopoietic System. *Annu. Rev. Immunol.* *25*, 745–785.
- Daujat, S., Weiss, T., Mohn, F., Lange, U.C., Ziegler-Birling, C., Zeissler, U., Lappe, M., Schübeler, D., Torres-Padilla, M.-E., and Schneider, R. (2009). H3K64 trimethylation marks heterochromatin and is dynamically remodeled during developmental reprogramming. *Nat. Struct. Mol. Biol.* *16*, 777–781.
- Dawlaty, M.M., Breiling, A., Le, T., Raddatz, G., Barrasa, M.I., Cheng, A.W., Gao, Q., Powell, B.E., Li, Z., Xu, M., et al. (2013). Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev Cell* *24*, 310–323.
- Dean, W., Santos, F., Stojkovic, M., Zakhartchenko, V., Walter, J., Wolf, E., and Reik, W. (2001). Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos. *Proc. Natl. Acad. Sci. U. S. A.* *98*, 13734–13738.
- DeChiara, T.M., Robertson, E.J., and Efstratiadis, A. (1991). Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* *64*, 849–859.
- Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2010). The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J. Biol. Chem.* *285*, 26114–26120.
- Di Francesco, A., Arosio, B., Falconi, A., Micioni Di Bonaventura, M.V., Karimi, M., Mari, D., Casati, M., Maccarrone, M., and D’Addario, C. (2015). Global changes in DNA methylation in Alzheimer’s disease peripheral blood mononuclear cells. *Brain. Behav. Immun.* *45*, 139–144.
- Dickson, J., Gowher, H., Strogantsev, R., Gaszner, M., Hair, A., Felsenfeld, G., and West, A.G. (2010). VEZF1 Elements Mediate Protection from DNA Methylation. *PLoS Genet.* *6*.
- Dodge, J.E., Okano, M., Dick, F., Tsujimoto, N., Chen, T., Wang, S., Ueda, Y., Dyson, N., and Li, E. (2005). Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. *J. Biol. Chem.* *280*, 17986–17991.
- Dong, A., Yoder, J.A., Zhang, X., Zhou, L., Bestor, T.H., and Cheng, X. (2001). Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA. *Nucleic Acids Res.* *29*, 439–448.
- Donovan, P.J., Stott, D., Cairns, L.A., Heasman, J., and Wylie, C.C. (1986). Migratory and postmigratory mouse primordial germ cells behave differently in culture. *Cell* *44*, 831–838.

- Down, T.A., Rakyan, V.K., Turner, D.J., Flicek, P., Li, H., Kulesha, E., Graf, S., Johnson, N., Herrero, J., Tomazou, E.M., et al. (2008). A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat Biotechnol* 26, 779–785.
- Eguizabal, C., Herrera, L., De Oñate, L., Montserrat, N., Hajkova, P., and Izpisua Belmonte, J.C. (2016). Characterization of the Epigenetic Changes During Human Gonadal Primordial Germ Cells Reprogramming. *Stem Cells Dayt. Ohio* 34, 2418–2428.
- Ehrenhofer-Murray, A.E. (2017). Cross-Talk between Dnmt2-Dependent tRNA Methylation and Queuosine Modification. *Biomol.* 2218-273X 7, 1–21.
- Ehrlich, M. (2003). The ICF syndrome, a DNA methyltransferase 3B deficiency and immunodeficiency disease. *Clin. Immunol. Orlando Fla* 109, 17–28.
- Elhamamsy, A.R. (2017). Role of DNA methylation in imprinting disorders: an updated review. *J. Assist. Reprod. Genet.* 34, 549–562.
- Enders, G.C., and May, J.J. (1994). Developmentally regulated expression of a mouse germ cell nuclear antigen examined from embryonic day 11 to adult in male and female mice. *Dev. Biol.* 163, 331–340.
- Esteller, M. (2011). Non-coding RNAs in human disease. *Nat. Rev. Genet.* 12, 861–874.
- Ewen, K.A., and Koopman, P. (2010). Mouse germ cell development: from specification to sex determination. *Mol. Cell. Endocrinol.* 323, 76–93.
- Fadloun, A., Le Gras, S., Jost, B., Ziegler-Birling, C., Takahashi, H., Gorab, E., Carninci, P., and Torres-Padilla, M.-E. (2013). Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat. Struct. Mol. Biol.* 20, 332–338.
- Farlik, M., Halbritter, F., Müller, F., Choudry, F.A., Ebert, P., Klughammer, J., Farrow, S., Santoro, A., Ciaurro, V., Mathur, A., et al. (2016). DNA Methylation Dynamics of Human Hematopoietic Stem Cell Differentiation. *Cell Stem Cell* 19, 808–822.
- Farthing, C.R., Ficiz, G., Ng, R.K., Chan, C.F., Andrews, S., Dean, W., Hemberger, M., and Reik, W. (2008). Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLoS Genet* 4, e1000116.
- Fasolino, M., and Zhou, Z. (2017). The Crucial Role of DNA Methylation and MeCP2 in Neuronal Function. *Genes* 8.
- Fatemi, M., Hermann, A., Gowher, H., and Jeltsch, A. (2002). Dnmt3a and Dnmt1 functionally cooperate during de novo methylation of DNA. *Eur. J. Biochem.* 269, 4981–4984.
- Feinberg, A.P., and Vogelstein, B. (1983). Hypomethylation of ras oncogenes in primary human cancers. *Biochem. Biophys. Res. Commun.* 111, 47–54.
- Felle, M., Joppien, S., Németh, A., Diermeier, S., Thalhammer, V., Dobner, T., Kremmer, E., Kappler, R., and Längst, G. (2011). The USP7/Dnmt1 complex stimulates the DNA methylation activity of Dnmt1 and regulates the stability of UHRF1. *Nucleic Acids Res.* 39, 8355–8365.

- Feng, S., Jacobsen, S.E., and Reik, W. (2010). Epigenetic reprogramming in plant and animal development. *Science* 330, 622–627.
- Florian, M.C., Dörr, K., Niebel, A., Daria, D., Schrezenmeier, H., Rojewski, M., Filippi, M.-D., Hasenberg, A., Gunzer, M., Scharffetter-Kochanek, K., et al. (2012). Cdc42 Activity Regulates Hematopoietic Stem Cell Aging and Rejuvenation. *Cell Stem Cell* 10, 520–530.
- Florian, M.C., Nattamai, K.J., Dörr, K., Marka, G., Uberle, B., Vas, V., Eckl, C., Andrä, I., Schiemann, M., Oostendorp, R.A.J., et al. (2013). A canonical to non-canonical Wnt signalling switch in haematopoietic stem-cell ageing. *Nature* 503, 392–396.
- Formosa, A., Lena, A.M., Markert, E.K., Cortelli, S., Miano, R., Mauriello, A., Croce, N., Vandesompele, J., Mestdagh, P., Finazzi-Agrò, E., et al. (2013). DNA methylation silences miR-132 in prostate cancer. *Oncogene* 32, 127–134.
- Frauer, C., Hoffmann, T., Bultmann, S., Casa, V., Cardoso, M.C., Antes, I., and Leonhardt, H. (2011). Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS One* 6, e21306.
- Fröhlich, L.F., Mrakovcic, M., Steinborn, R., Chung, U.-I., Bastepe, M., and Jüppner, H. (2010). Targeted deletion of the Nesp55 DMR defines another Gnas imprinting control region and provides a mouse model of autosomal dominant PHP-1b. *Proc. Natl. Acad. Sci. U. S. A.* 107, 9275–9280.
- Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., and Paul, C.L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. U. S. A.* 89, 1827–1831.
- Fujimori, A., Matsuda, Y., Takemoto, Y., Hashimoto, Y., Kubo, E., Araki, R., Fukumura, R., Mita, K., Tatsumi, K., and Muto, M. (1998). Cloning and mapping of Np95 gene which encodes a novel nuclear protein associated with cell proliferation. *Mamm. Genome Off. J. Int. Mamm. Genome Soc.* 9, 1032–1035.
- Fuks, F., Hurd, P.J., Deplus, R., and Kouzarides, T. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res.* 31, 2305–2312.
- Galloway, J.L., and Zon, L.I. (2003). Ontogeny of hematopoiesis: examining the emergence of hematopoietic cells in the vertebrate embryo. *Curr. Top. Dev. Biol.* 53, 139–158.
- Gama-Sosa, M.A., Slagel, V.A., Trewyn, R.W., Oxenhandler, R., Kuo, K.C., Gehrke, C.W., and Ehrlich, M. (1983). The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Res.* 11, 6883–6894.
- Gardiner-Garden, M., and Frommer, M. (1987). CpG Islands in vertebrate genomes. *J. Mol. Biol.* 196, 261–282.
- Gaudet, F., Rideout, W.M., Meissner, A., Dausman, J., Leonhardt, H., and Jaenisch, R. (2004). Dnmt1 Expression in Pre- and Postimplantation Embryogenesis and the Maintenance of IAP Silencing. *Mol. Cell. Biol.* 24, 1640–1648.
- Ge, Y.-Z., Pu, M.-T., Gowher, H., Wu, H.-P., Ding, J.-P., Jeltsch, A., and Xu, G.-L. (2004). Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. *J. Biol. Chem.* 279, 25447–25454.

- Gendrel, A.-V., Apedaile, A., Coker, H., Termanis, A., Zvetkova, I., Godwin, J., Tang, Y.A., Huntley, D., Montana, G., Taylor, S., et al. (2012). Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Dev. Cell* **23**, 265–279.
- Genovese, G., Kähler, A.K., Handsaker, R.E., Lindberg, J., Rose, S.A., Bakhoum, S.F., Chambert, K., Mick, E., Neale, B.M., Fromer, M., et al. (2014). Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487.
- Georgiades, P., Ogilvy, S., Duval, H., Licence, D.R., Charnock-Jones, D.S., Smith, S.K., and Print, C.G. (2002). VavCre transgenic mice: a tool for mutagenesis in hematopoietic and endothelial lineages. *Genes*. N. Y. N **2000** **34**, 251–256.
- Gibbons, R.J., and Higgs, D.R. (2000). Molecular-clinical spectrum of the ATR-X syndrome. *Am. J. Med. Genet.* **97**, 204–212.
- Ginsburg, M., Snow, M.H., and McLaren, A. (1990). Primordial germ cells in the mouse embryo during gastrulation. *Development* **110**, 521–528.
- Göbel, U., Schneider, D.T., Calaminus, G., Haas, R.J., Schmidt, P., and Harms, D. (2000). Germ-cell tumors in childhood and adolescence. GPOH MAKEI and the MAHO study groups. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **11**, 263–271.
- Goldstone, A.P. (2004). Prader-Willi syndrome: advances in genetics, pathophysiology and treatment. *Trends Endocrinol. Metab. TEM* **15**, 12–20.
- Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.-L., Zhang, X., Golic, K.G., Jacobsen, S.E., and Bestor, T.H. (2006). Methylation of tRNA<sup>Asp</sup> by the DNA methyltransferase homolog Dnmt2. *Science* **311**, 395–398.
- Greer, E.L., Blanco, M.A., Gu, L., Sendinc, E., Liu, J., Aristizábal-Corrales, D., Hsu, C.-H., Aravind, L., He, C., and Shi, Y. (2015). DNA Methylation on N6-Adenine in *C. elegans*. *Cell* **161**, 868–878.
- Greiner, V.J., Kovalenko, L., Humbert, N., Richert, L., Birck, C., Ruff, M., Zaporozhets, O.A., Dhe-Paganon, S., Bronner, C., and Mély, Y. (2015). Site-Selective Monitoring of the Interaction of the SRA Domain of UHRF1 with Target DNA Sequences Labeled with 2-Aminopurine. *Biochemistry (Mosc.)* **54**, 6012–6020.
- Gu, H., Smith, Z.D., Bock, C., Boyle, P., Gnirke, A., and Meissner, A. (2011a). Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat Protoc* **6**, 468–481.
- Gu, T.-P., Guo, F., Yang, H., Wu, H.-P., Xu, G.-F., Liu, W., Xie, Z.-G., Shi, L., He, X., Jin, S., et al. (2011b). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606–610.
- Guenatri, M., Duffié, R., Iranzo, J., Fauque, P., and Bourc'his, D. (2013). Plasticity in Dnmt3L-dependent and -independent modes of de novo methylation in the developing mouse embryo. *Dev. Camb. Engl.* **140**, 562–572.

- Guibert, S., Forne, T., and Weber, M. (2012). Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Res* 22, 633–641.
- Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., Wu, X., Wen, L., Gu, T.-P., Hu, B., et al. (2014a). Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell* 15, 447–458.
- Guo, F., Yan, L., Guo, H., Li, L., Hu, B., Zhao, Y., Yong, J., Hu, Y., Wang, X., Wei, Y., et al. (2015a). The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells. *Cell* 161, 1437–1452.
- Guo, H., Zhu, P., Wu, X., Li, X., Wen, L., and Tang, F. (2013). Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res.* 23, 2126–2135.
- Guo, J.U., Su, Y., Zhong, C., Ming, G., and Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* 145, 423–434.
- Guo, J.U., Su, Y., Shin, J.H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., et al. (2014b). Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat. Neurosci.* 17, 215–222.
- Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015b). Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* 517, 640–644.
- Hackett, J.A., Sengupta, R., Zyllicz, J.J., Murakami, K., Lee, C., Down, T.A., and Surani, M.A. (2013). Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* 339, 448–452.
- Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M.A. (2002). Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* 117, 15–23.
- Hajkova, P., Jeffries, S.J., Lee, C., Miller, N., Jackson, S.P., and Surani, M.A. (2010). Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* 329, 78–82.
- Hansen, R.S., Wijmenga, C., Luo, P., Stanek, A.M., Canfield, T.K., Weemaes, C.M., and Gartler, S.M. (1999). The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc. Natl. Acad. Sci. U. S. A.* 96, 14412–14417.
- Hargan-Calvopina, J., Taylor, S., Cook, H., Hu, Z., Lee, S.A., Yen, M.-R., Chiang, Y.-S., Chen, P.-Y., and Clark, A.T. (2016). Stage-Specific Demethylation in Primordial Germ Cells Safeguards against Precocious Differentiation. *Dev. Cell* 39, 75–86.
- Hashimoto, H., Vertino, P.M., and Cheng, X. (2010). Molecular coupling of DNA methylation and histone methylation. *Epigenomics* 2, 657–669.
- Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* 40, 4841–4849.

- Hata, K., Okano, M., Lei, H., and Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* 129, 1983–1993.
- Hayashi, K., Yoshida, K., and Matsui, Y. (2005). A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature* 438, 374–378.
- He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303–1307.
- Heard, E., and Martienssen, R.A. (2014). Transgenerational Epigenetic Inheritance: Myths and Mechanisms. *Cell* 157, 95–109.
- Hemberger, M., Dean, W., and Reik, W. (2009). Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. *Nat. Rev. Mol. Cell Biol.* 10, 526–537.
- Hilscher, B., Hilscher, W., Bülthoff-Ohnolz, B., Krämer, U., Birke, A., Pelzer, H., and Gauss, G. (1974). Kinetics of gametogenesis. I. Comparative histological and autoradiographic studies of oocytes and transitional prospermatogonia during oogenesis and prespermatogenesis. *Cell Tissue Res.* 154, 443–470.
- Hirasawa, R., Chiba, H., Kaneda, M., Tajima, S., Li, E., Jaenisch, R., and Sasaki, H. (2008). Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes Dev.* 22, 1607–1616.
- Hitt, M.M., Wu, T.L., Cohen, G., and Linn, S. (1988). De novo and maintenance DNA methylation by a mouse plasmacytoma cell DNA methyltransferase. *J. Biol. Chem.* 263, 4392–4399.
- Holliday, R., and Pugh, J.E. (1975). DNA modification mechanisms and gene activity during development. *Science* 187, 226–232.
- Hon, G.C., Hawkins, R.D., Caballero, O.L., Lo, C., Lister, R., Pelizzola, M., Valsesia, A., Ye, Z., Kuan, S., Edsall, L.E., et al. (2012). Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.* 22, 246–258.
- Hon, G.C., Rajagopal, N., Shen, Y., McCleary, D.F., Yue, F., Dang, M.D., and Ren, B. (2013). Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat. Genet.* 45, 1198–1206.
- Hopfner, R., Mousli, M., Jeltsch, J.M., Voulgaris, A., Lutz, Y., Marin, C., Bellocq, J.P., Oudet, P., and Bronner, C. (2000). ICBP90, a novel human CCAAT binding protein, involved in the regulation of topoisomerase IIalpha expression. *Cancer Res.* 60, 121–128.
- Howell, C.Y., Bestor, T.H., Ding, F., Latham, K.E., Mertineit, C., Trasler, J.M., and Chaillet, J.R. (2001). Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. *Cell* 104, 829–838.
- Howlett, S.K., and Reik, W. (1991). Methylation levels of maternal and paternal genomes during preimplantation development. *Dev. Camb. Engl.* 113, 119–127.

- Hsieh, C.-L. (1999). In Vivo Activity of Murine De Novo Methyltransferases, Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* *19*, 8211–8218.
- Hu, G., and Zhao, K. (2014). Correlating histone modification patterns with gene expression data during hematopoiesis. *Methods Mol. Biol. Clifton NJ* *1150*, 175–187.
- Huang, H., Jiang, X., Li, Z., Li, Y., Song, C.-X., He, C., Sun, M., Chen, P., Gurbuxani, S., Wang, J., et al. (2013). TET1 plays an essential oncogenic role in MLL-rearranged leukemia. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 11994–11999.
- Illingworth, R., Kerr, A., Desousa, D., Jorgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J., et al. (2008). A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol* *6*, e22.
- Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet* *6*.
- Iwamatsu, T., and Chang, M.C. (1972). Sperm penetration in vitro of mouse oocytes at various times during maturation. *J. Reprod. Fertil.* *31*, 237–247.
- Iyer, L.M., Abhiman, S., and Aravind, L. (2011). Natural history of eukaryotic DNA methylation systems. *Prog. Mol. Biol. Transl. Sci.* *101*, 25–104.
- Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E., et al. (2001a). Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nat Genet* *27*, 31–39.
- Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E., et al. (2001b). Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nat. Genet.* *27*, 31–39.
- Jain, D., Meydan, C., Lange, J., Claeys Bouuaert, C., Mason, C.E., Anderson, K.V., and Keeney, S. (2017). rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline. *PLoS Genet.* *13*, e1006964.
- Jaiswal, S., Natarajan, P., Silver, A.J., Gibson, C.J., Bick, A.G., Shvartz, E., McConkey, M., Gupta, N., Gabriel, S., Ardissino, D., et al. (2017). Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N. Engl. J. Med.* *377*, 111–121.
- Janulaitis, A., Klimasauskas, S., Petrusyte, M., and Butkus, V. (1983). Cytosine modification in DNA by BcnI methylase yields N4-methylcytosine. *FEBS Lett.* *161*, 131–134.
- Jeanblanc, M., Mousli, M., Hopfner, R., Bathami, K., Martinet, N., Abbady, A.-Q., Siffert, J.-C., Mathieu, E., Muller, C.D., and Bronner, C. (2005). The retinoblastoma gene and its product are targeted by ICBP90: a key mechanism in the G1/S transition during the cell cycle. *Oncogene* *24*, 7337–7345.
- Jelinic, P., and Shaw, P. (2007). Loss of imprinting and cancer. *J. Pathol.* *211*, 261–268.

- Jeltsch, A., and Jurkowska, R.Z. (2014). New concepts in DNA methylation. *Trends Biochem. Sci.* *39*, 310–318.
- Ji, H., Ehrlich, L.I., Seita, J., Murakami, P., Doi, A., Lindau, P., Lee, H., Aryee, M.J., Irizarry, R.A., Kim, K., et al. (2010). Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* *467*, 338–342.
- Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., and Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* *449*, 248–251.
- Jiang, L., Zhang, J., Wang, J.-J., Wang, L., Zhang, L., Li, G., Yang, X., Ma, X., Sun, X., Cai, J., et al. (2013). Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Cell* *153*, 773–784.
- Jin, S.-G., Jiang, Y., Qiu, R., Rauch, T.A., Wang, Y., Schackert, G., Krex, D., Lu, Q., and Pfeifer, G.P. (2011). 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer Res.* *71*, 7360–7365.
- Jones, P.A., and Laird, P.W. (1999). Cancer epigenetics comes of age. *Nat. Genet.* *21*, 163–167.
- Jones, P.A., and Liang, G. (2009). Rethinking how DNA methylation patterns are maintained. *Nat. Rev. Genet.* *10*, 805–811.
- Jordan, C.T., McKearn, J.P., and Lemischka, I.R. (1990). Cellular and developmental properties of fetal hematopoietic stem cells. *Cell* *61*, 953–963.
- Kagiyada, S., Kurimoto, K., Hirota, T., Yamaji, M., and Saitou, M. (2013). Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *EMBO J* *32*, 340–353.
- Kaneda, A., and Feinberg, A.P. (2005). Loss of imprinting of IGF2: a common epigenetic modifier of intestinal tumor risk. *Cancer Res.* *65*, 11236–11240.
- Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* *429*, 900–903.
- Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L., Jones, S., Hirst, M., et al. (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell Stem Cell* *8*, 676–687.
- Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., Okano, M., Li, E., Nozaki, M., and Sasaki, H. (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum Mol Genet* *16*, 2272–2280.
- Kehler, J., Tolkunova, E., Koschorz, B., Pesce, M., Gentile, L., Boiani, M., Lomeli, H., Nagy, A., McLaughlin, K.J., Scholer, H.R., et al. (2004). Oct4 is required for primordial germ cell survival. *EMBO Rep* *5*, 1078–1083.
- Kigami, D., Minami, N., Takayama, H., and Imai, H. (2003). MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol. Reprod.* *68*, 651–654.

- Kim, G.-D., Ni, J., Kelesoglu, N., Roberts, R.J., and Pradhan, S. (2002). Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. *EMBO J.* *21*, 4183–4195.
- Kim, J.K., Estève, P.-O., Jacobsen, S.E., and Pradhan, S. (2009). UHRF1 binds G9a and participates in p21 transcriptional regulation in mammalian cells. *Nucleic Acids Res.* *37*, 493–505.
- Kim, S.J., Zhao, H., Hardikar, S., Singh, A.K., Goodell, M.A., and Chen, T. (2013). A DNMT3A mutation common in AML exhibits dominant-negative effects in murine ES cells. *Blood* *122*, 4086–4089.
- Kimura, T., Yomogida, K., Iwai, N., Kato, Y., and Nakano, T. (1999). Molecular cloning and genomic organization of mouse homologue of *Drosophila* germ cell-less and its expression in germ lineage cells. *Biochem. Biophys. Res. Commun.* *262*, 223–230.
- Kinoshita, H., Shi, Y., Sandefur, C., Meisner, L.F., Chang, C., Choon, A., Reznikoff, C.R., Bova, G.S., Friedl, A., and Jarrard, D.F. (2000). Methylation of the androgen receptor minimal promoter silences transcription in human prostate cancer. *Cancer Res.* *60*, 3623–3630.
- Klein, C.J., Botuyan, M.-V., Wu, Y., Ward, C.J., Nicholson, G.A., Hammans, S., Hojo, K., Yamanishi, H., Karpf, A.R., Wallace, D.C., et al. (2011). Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss. *Nat. Genet.* *43*, 595–600.
- Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R., et al. (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* *468*, 839–843.
- Kobayashi, H., Sakurai, T., Imai, M., Takahashi, N., Fukuda, A., Yayoi, O., Sato, S., Nakabayashi, K., Hata, K., Sotomaru, Y., et al. (2012). Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet.* *8*, e1002440.
- Kobayashi, H., Sakurai, T., Miura, F., Imai, M., Mochiduki, K., Yanagisawa, E., Sakashita, A., Wakai, T., Suzuki, Y., Ito, T., et al. (2013). High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Res* *23*, 616–627.
- Kohli, R.M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* *502*, 472–479.
- Kozioł, M.J., Bradshaw, C.R., Allen, G.E., Costa, A.S.H., Frezza, C., and Gurdon, J.B. (2016). Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nat. Struct. Mol. Biol.* *23*, 24–30.
- Krebs, A.R., Dessus-Babus, S., Burger, L., and Schübeler, D. (2014). High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions. *eLife* *3*.
- Kriaucionis, S., and Heintz, N. (2009). The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science* *324*, 929–930.

- Kulis, M., Heath, S., Bibikova, M., Queirós, A.C., Navarro, A., Clot, G., Martínez-Trillos, A., Castellano, G., Brun-Heath, I., Pinyol, M., et al. (2012). Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat. Genet.* *44*, 1236–1242.
- Kulis, M., Queirós, A.C., Beekman, R., and Martín-Subero, J.I. (2013). Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer. *Biochim. Biophys. Acta* *1829*, 1161–1174.
- Kunarso, G., Chia, N.-Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.-S., Ng, H.-H., and Bourque, G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* *42*, 631–634.
- Kurdyukov, S., and Bullock, M. (2016). DNA Methylation Analysis: Choosing the Right Method. *Biology* *5*.
- Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K., and Saitou, M. (2008). Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev* *22*, 1617–1635.
- Laker, R.C., and Ryall, J.G. (2016). DNA Methylation in Skeletal Muscle Stem Cell Specification, Proliferation, and Differentiation. *Stem Cells Int.* *2016*, 5725927.
- Lallous, N., Legrand, P., McEwen, A.G., Ramón-Maiques, S., Samama, J.-P., and Birck, C. (2011). The PHD finger of human UHRF1 reveals a new subgroup of unmethylated histone H3 tail readers. *PLoS One* *6*, e27599.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* *409*, 860–921.
- Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., Walter, J., and Reik, W. (2003). Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis* *35*, 88–93.
- Lawson, K.A., and Hage, W.J. (1994). Clonal analysis of the origin of primordial germ cells in the mouse. *Ciba Found. Symp.* *182*, 68-84; discussion 84-91.
- Le Fevre, A., Beygo, J., Silveira, C., Kamien, B., Clayton-Smith, J., Colley, A., Buiting, K., and Dudding-Byth, T. (2017). Atypical Angelman syndrome due to a mosaic imprinting defect: Case reports and review of the literature. *Am. J. Med. Genet. A.* *173*, 753–757.
- Lei, H., Oh, S.P., Okano, M., Juttermann, R., Goss, K.A., Jaenisch, R., and Li, E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* *122*, 3195–3205.
- Leonhardt, H., Page, A.W., Weier, H.-U., and Bestor, T.H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* *71*, 865–873.
- Ley, T.J., Ding, L., Walter, M.J., McLellan, M.D., Lamprecht, T., Larson, D.E., Kandoth, C., Payton, J.E., Baty, J., Welch, J., et al. (2010). DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* *363*, 2424–2433.

- Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* *69*, 915–926.
- Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. *Nature* *366*, 362–365.
- Li, H., Rauch, T., Chen, Z.-X., Szabó, P.E., Riggs, A.D., and Pfeifer, G.P. (2006). The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. *J. Biol. Chem.* *281*, 19489–19500.
- Li, J., Huang, Q., Zeng, F., Li, W., He, Z., Chen, W., Zhu, W., and Zhang, B. (2014). The prognostic value of global DNA hypomethylation in cancer: a meta-analysis. *PLoS One* *9*, e106290.
- Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., and Ferguson-Smith, A.C. (2008). A maternal-zygotic effect gene, *Zfp57*, maintains both maternal and paternal imprints. *Dev. Cell* *15*, 547–557.
- Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F., and Schubeler, D. (2011). Identification of genetic elements that autonomously determine DNA methylation states. *Nat Genet* *43*, 1091–1097.
- Lipka, D.B., Wang, Q., Cabezas-Wallscheid, N., Klimmeck, D., Weichenhan, D., Herrmann, C., Lier, A., Brocks, D., von Paleske, L., Renders, S., et al. (2014). Identification of DNA methylation changes at cis-regulatory elements during early steps of HSC differentiation using tagmentation-based whole genome bisulfite sequencing. *Cell Cycle* *13*, 3476–3487.
- Lister, R., O’Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. *Cell* *133*, 523.
- Lister, R., Pelizzola, M., Downen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* *462*, 315–322.
- Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., Antosiewicz-Bourget, J., O’Malley, R., Castanon, R., Klugman, S., et al. (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* *471*, 68–73.
- Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., et al. (2013). Global epigenomic reconfiguration during mammalian brain development. *Science* *341*, 1237905.
- Liu, S., Brind’Amour, J., Karimi, M.M., Shirane, K., Bogutz, A., Lefebvre, L., Sasaki, H., Shinkai, Y., and Lorincz, M.C. (2014). *Setdb1* is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes Dev* *28*, 2041–2055.
- Liu, W.M., Maraia, R.J., Rubin, C.M., and Schmid, C.W. (1994). Alu transcripts: cytoplasmic localisation and regulation by DNA methylation. *Nucleic Acids Res.* *22*, 1087–1095.
- Liu, Y., Zhang, B., Meng, X., Korn, M.J., Parent, J.M., Lu, L.-Y., and Yu, X. (2017). UHRF2 regulates local 5-methylcytosine and suppresses spontaneous seizures. *Epigenetics* *0*.

- Lomeli, H., Ramos-Mejia, V., Gertsenstein, M., Lobe, C.G., and Nagy, A. (2000). Targeted insertion of Cre recombinase into the TNAP gene: excision in primordial germ cells. *Genesis* 26, 116–117.
- Long, H.K., Sims, D., Heger, A., Blackledge, N.P., Kutter, C., Wright, M.L., Grützner, F., Odom, D.T., Patient, R., Ponting, C.P., et al. (2013a). Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. *eLife* 2.
- Long, H.K., Blackledge, N.P., and Klose, R.J. (2013b). ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochem. Soc. Trans.* 41, 727–740.
- Long, H.K., King, H.W., Patient, R.K., Odom, D.T., and Klose, R.J. (2016). Protection of CpG islands from DNA methylation is DNA-encoded and evolutionarily conserved. *Nucleic Acids Res.* 44, 6693–6706.
- Lorincz, M.C., Schübeler, D., Hutchinson, S.R., Dickerson, D.R., and Groudine, M. (2002). DNA methylation density influences the stability of an epigenetic imprint and Dnmt3a/b-independent de novo methylation. *Mol. Cell. Biol.* 22, 7572–7580.
- Lorsbach, R.B., Moore, J., Mathew, S., Raimondi, S.C., Mukatira, S.T., and Downing, J.R. (2003). TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia* 17, 637–641.
- Lou, S., Lee, H.-M., Qin, H., Li, J.-W., Gao, Z., Liu, X., Chan, L.L., Kl Lam, V., So, W.-Y., Wang, Y., et al. (2014). Whole-genome bisulfite sequencing of multiple individuals reveals complementary roles of promoter and gene body methylation in transcriptional regulation. *Genome Biol.* 15, 408.
- Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260.
- Lujambio, A., Ropero, S., Ballestar, E., Fraga, M.F., Cerrato, C., Setién, F., Casado, S., Suarez-Gauthier, A., Sanchez-Cespedes, M., Git, A., et al. (2007). Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. *Cancer Res.* 67, 1424–1429.
- Lyko, F., Ramsahoye, B.H., Kashevsky, H., Tudor, M., Mastrangelo, M.-A., Orr-Weaver, T.L., and Jaenisch, R. (1999). Mammalian (cytosine-5) methyltransferases cause genomic DNA methylation and lethality in *Drosophila*. *Nat. Genet.* 23, 363–366.
- Lyon, M.F. (1961). Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* 190, 372–373.
- Ma, H., Morey, R., O’Neil, R.C., He, Y., Daughtry, B., Schultz, M.D., Hariharan, M., Nery, J.R., Castanon, R., Sabatini, K., et al. (2014). Abnormalities in human pluripotent cells due to reprogramming mechanisms. *Nature* 511, 177–183.
- Maatouk, D.M., Kellam, L.D., Mann, M.R.W., Lei, H., Li, E., Bartolomei, M.S., and Resnick, J.L. (2006). DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. *Dev. Camb. Engl.* 133, 3411–3418.

- Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* *487*, 57–63.
- MacGregor, G.R., Zambrowicz, B.P., and Soriano, P. (1995). Tissue non-specific alkaline phosphatase is expressed in both embryonic and extraembryonic lineages during mouse embryogenesis but is not required for migration of primordial germ cells. *Development* *121*, 1487–1496.
- Macleod, D., Charlton, J., Mullins, J., and Bird, A.P. (1994). Sp1 sites in the mouse *aprt* gene promoter are required to prevent methylation of the CpG island. *Genes Dev.* *8*, 2282–2292.
- Maher, E.R., and Reik, W. (2000). Beckwith-Wiedemann syndrome: imprinting in clusters revisited. *J. Clin. Invest.* *105*, 247–252.
- Maksakova, I.A., Romanish, M.T., Gagnier, L., Dunn, C.A., Lagemaat, L.N. van de, and Mager, D.L. (2006). Retroviral Elements and Their Hosts: Insertional Mutagenesis in the Mouse Germ Line. *PLOS Genet.* *2*, e2.
- Maksakova, I.A., Mager, D.L., and Reiss, D. (2008). Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. *Cell. Mol. Life Sci. CMLS* *65*, 3329–3347.
- Maloisel, L., and Rossignol, J.L. (1998). Suppression of crossing-over by DNA methylation in *Ascobolus*. *Genes Dev.* *12*, 1381–1389.
- Manakov, S.A., Pezic, D., Marinov, G.K., Pastor, W.A., Sachidanandam, R., and Aravin, A.A. (2015). MIWI2 and MILI Have Differential Effects on piRNA Biogenesis and DNA Methylation. *Cell Rep.* *12*, 1234–1243.
- Manzardo, A.M., and Butler, M.G. (2016). Examination of Global Methylation and Targeted Imprinted Genes in Prader-Willi Syndrome. *J. Clin. Epigenetics* *2*.
- Martienssen, R.A., and Colot, V. (2001). DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science* *293*, 1070–1074.
- Mastroeni, D., Chouliaras, L., Grover, A., Liang, W.S., Hauns, K., Rogers, J., and Coleman, P.D. (2013). Reduced RAN expression and disrupted transport between cytoplasm and nucleus; a key event in Alzheimer’s disease pathophysiology. *PloS One* *8*, e53349.
- Matsui, T., Leung, D., Miyashita, H., Maksakova, I.A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M.C., and Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* *464*, 927–931.
- Mayle, A., Yang, L., Rodriguez, B., Zhou, T., Chang, E., Curry, C.V., Challen, G.A., Li, W., Wheeler, D., Rebel, V.I., et al. (2015). *Dnmt3a* loss predisposes murine hematopoietic stem cells to malignant transformation. *Blood* *125*, 629–638.
- McClellan, K.A., Gosden, R., and Taketo, T. (2003). Continuous loss of oocytes throughout meiotic prophase in the normal mouse ovary. *Dev. Biol.* *258*, 334–348.

- McGrath, J., and Solter, D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* 37, 179–183.
- McGrath, K.E., Frame, J.M., Fromm, G.J., Koniski, A.D., Kingsley, P.D., Little, J., Bulger, M., and Palis, J. (2011). A transient definitive erythroid lineage with unique regulation of the  $\beta$ -globin locus in the mammalian embryo. *Blood* 117, 4600–4608.
- McGraw, S., Oakes, C.C., Martel, J., Cirio, M.C., Zeeuw, P. de, Mak, W., Plass, C., Bartolomei, M.S., Chaillet, J.R., and Trasler, J.M. (2013). Loss of DNMT1o Disrupts Imprinted X Chromosome Inactivation and Accentuates Placental Defects in Females. *PLOS Genet.* 9, e1003873.
- McKinney-Freeman, S., Cahan, P., Li, H., Lacadie, S.A., Huang, H.T., Curran, M., Loewer, S., Naveiras, O., Kathrein, K.L., Konantz, M., et al. (2012). The transcriptional landscape of hematopoietic stem cell ontogeny. *Cell Stem Cell* 11, 701–714.
- Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454, 766–770.
- Mendenhall, E.M., Koche, R.P., Truong, T., Zhou, V.W., Issac, B., Chi, A.S., Ku, M., and Bernstein, B.E. (2010). GC-Rich Sequence Elements Recruit PRC2 in Mammalian ES Cells. *PLoS Genet.* 6.
- Mertineit, C., Yoder, J.A., Taketo, T., Laird, D.W., Trasler, J.M., and Bestor, T.H. (1998). Sex-specific exons control DNA methyltransferase in mammalian germ cells. *Dev. Camb. Engl.* 125, 889–897.
- Mitsunaga, K., Araki, K., Mizusaki, H., Morohashi, K.-I., Haruna, K., Nakagata, N., Giguère, V., Yamamura, K.-I., and Abe, K. (2004). Loss of PGC-specific expression of the orphan nuclear receptor ERR-beta results in reduction of germ cell number in mouse embryos. *Mech. Dev.* 121, 237–246.
- Miura, F., Enomoto, Y., Dairiki, R., and Ito, T. (2012). Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res.* 40, e136.
- Moarefi, A.H., and Chédin, F. (2011). ICF syndrome mutations cause a broad spectrum of biochemical defects in DNMT3B-mediated de novo DNA methylation. *J. Mol. Biol.* 409, 758–772.
- Mochizuki, K., Tachibana, M., Saitou, M., Tokitake, Y., and Matsui, Y. (2012). Implication of DNA demethylation and bivalent histone modification for selective gene regulation in mouse primordial germ cells. *PloS One* 7, e46036.
- Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schübeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* 30, 755–766.
- Molaro, A., Falciatori, I., Hodges, E., Aravin, A.A., Marran, K., Rafii, S., McCombie, W.R., Smith, A.D., and Hannon, G.J. (2014). Two waves of de novo methylation during mouse germ cell development. *Genes Dev* 28, 1544–1549.
- Molyneaux, K.A., Stallock, J., Schaible, K., and Wylie, C. (2001). Time-Lapse Analysis of Living Mouse Germ Cell Migration. *Dev. Biol.* 240, 488–498.

- Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M.E., Vasanthakumar, A., Patel, J., Zhao, X., et al. (2011). Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* 20, 11–24.
- Mori, T., Ikeda, D.D., Yamaguchi, Y., and Unoki, M. (2012). NIRF/UHRF2 occupies a central position in the cell cycle network and allows coupling with the epigenetic landscape. *FEBS Lett* 586, 1570–1583.
- Morita, S., Noguchi, H., Horii, T., Nakabayashi, K., Kimura, M., Okamura, K., Sakai, A., Nakashima, H., Hata, K., Nakashima, K., et al. (2016). Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nat. Biotechnol.* 34, 1060–1065.
- Morrison, S.J., Hemmati, H.D., Wandycz, A.M., and Weissman, I.L. (1995). The purification and characterization of fetal liver hematopoietic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 92, 10302–10306.
- Mousli, M., Hopfner, R., Abbady, A.-Q., Monté, D., Jeanblanc, M., Oudet, P., Louis, B., and Bronner, C. (2003). ICBP90 belongs to a new family of proteins with an expression that is deregulated in cancer cells. *Br. J. Cancer* 89, 120–127.
- Mucenski, M.L., McLain, K., Kier, A.B., Swerdlow, S.H., Schreiner, C.M., Miller, T.A., Pietryga, D.W., Scott, W.J., and Potter, S.S. (1991). A functional c-myc gene is required for normal murine fetal hepatic hematopoiesis. *Cell* 65, 677–689.
- Müller, M., Hartmann, M., Schuster, I., Bender, S., Thüring, K.L., Helm, M., Katze, J.R., Nellen, W., Lyko, F., and Ehrenhofer-Murray, A.E. (2015). Dynamic modulation of Dnmt2-dependent tRNA methylation by the micronutrient queuine. *Nucleic Acids Res.* 43, 10952–10962.
- Müller, T., Gessi, M., Waha, A., Isselstein, L.J., Luxen, D., Freihoff, D., Freihoff, J., Becker, A., Simon, M., Hammes, J., et al. (2012). Nuclear exclusion of TET1 is associated with loss of 5-hydroxymethylcytosine in IDH1 wild-type gliomas. *Am. J. Pathol.* 181, 675–683.
- Nakamura, T., Arai, Y., Umehara, H., Masuhara, M., Kimura, T., Taniguchi, H., Sekimoto, T., Ikawa, M., Yoneda, Y., Okabe, M., et al. (2007). PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nat. Cell Biol.* 9, 64–71.
- Nakamura, T., Liu, Y.-J., Nakashima, H., Umehara, H., Inoue, K., Matoba, S., Tachibana, M., Ogura, A., Shinkai, Y., and Nakano, T. (2012). PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* 486, 415–419.
- Ng, J.-H., Kumar, V., Muratani, M., Kraus, P., Yeo, J.-C., Yaw, L.-P., Xue, K., Lufkin, T., Prabhakar, S., and Ng, H.-H. (2013). In vivo epigenomic profiling of germ cells reveals germ cell molecular signatures. *Dev. Cell* 24, 324–333.
- Nicholls, R.D., Saitoh, S., and Horsthemke, B. (1998). Imprinting in Prader-Willi and Angelman syndromes. *Trends Genet.* TIG 14, 194–200.
- Nielsen, N.H., Roos, G., Emdin, S.O., and Landberg, G. (2001). Methylation of the p16(Ink4a) tumor suppressor gene 5'-CpG island in breast cancer. *Cancer Lett.* 163, 59–69.

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

Noce, T., Okamoto-Ito, S., and Tsunekawa, N. (2001). Vasa homolog genes in mammalian germ cell development. *Cell Struct. Funct.* *26*, 131–136.

Okada, S., Nakauchi, H., Nagayoshi, K., Nishikawa, S., Miura, Y., and Suda, T. (1992). In vivo and in vitro stem cell function of c-kit- and Sca-1-positive murine hematopoietic cells. *Blood* *80*, 3044–3050.

Okano, M., Xie, S., and Li, E. (1998a). Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Res.* *26*, 2536–2540.

Okano, M., Xie, S., and Li, E. (1998b). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat. Genet.* *19*, 219–220.

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

Olkhov-Mitsel, E., and Bapat, B. (2012). Strategies for discovery and validation of methylated and hydroxymethylated DNA biomarkers. *Cancer Med.* *1*, 237–260.

Ono, R., Taki, T., Taketani, T., Taniwaki, M., Kobayashi, H., and Hayashi, Y. (2002). LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer Res.* *62*, 4075–4080.

Ooi, S.K., Wolf, D., Hartung, O., Agarwal, S., Daley, G.Q., Goff, S.P., and Bestor, T.H. (2010). Dynamic instability of genomic methylation patterns in pluripotent stem cells. *Epigenetics Chromatin* *3*, 17.

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

Oostra, B.A., and Willemsen, R. (2002). The X chromosome and fragile X mental retardation. *Cytogenet. Genome Res.* *99*, 257–264.

Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. *Cell* *132*, 631–644.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. *Curr Biol* *10*, 475–478.

Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Rep.* *10*, 1235–1241.

Palis, J., and Yoder, M.C. (2001). Yolk-sac hematopoiesis: The first blood cells of mouse and man. *Exp. Hematol.* *29*, 927–936.

- Peaston, A.E., Evsikov, A.V., Graber, J.H., de Vries, W.N., Holbrook, A.E., Solter, D., and Knowles, B.B. (2004). Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* 7, 597–606.
- Peat, J.R., Dean, W., Clark, S.J., Krueger, F., Smallwood, S.A., Ficz, G., Kim, J.K., Marioni, J.C., Hore, T.A., and Reik, W. (2014). Genome-wide bisulfite sequencing in zygotes identifies demethylation targets and maps the contribution of TET3 oxidation. *Cell Rep.* 9, 1990–2000.
- Pepling, M.E., and Spradling, A.C. (2001). Mouse ovarian germ cell cysts undergo programmed breakdown to form primordial follicles. *Dev. Biol.* 234, 339–351.
- Petterson, A., Chung, T.H., Tan, D., Sun, X., and Jia, X.-Y. (2014). RRHP: a tag-based approach for 5-hydroxymethylcytosine mapping at single-site resolution. *Genome Biol.* 15, 456.
- Pichler, G., Wolf, P., Schmidt, C.S., Meilinger, D., Schneider, K., Frauer, C., Fellingner, K., Rottach, A., and Leonhardt, H. (2011). Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *J Cell Biochem* 112, 2585–2593.
- Pinarbasi, E., Elliott, J., and Hornby, D.P. (1996). Activation of a yeast pseudo DNA methyltransferase by deletion of a single amino acid. *J. Mol. Biol.* 257, 804–813.
- Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* 463, 1101–1105.
- Potok, M.E., Nix, D.A., Parnell, T.J., and Cairns, B.R. (2013). Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. *Cell* 153, 759–772.
- Pradhan, S., Bacolla, A., Wells, R.D., and Roberts, R.J. (1999). Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. *J. Biol. Chem.* 274, 33002–33010.
- Prisette, M., El-Maarri, O., Arnaud, D., Walter, J., and Avner, P. (2001). Methylation profiles of DXPas34 during the onset of X-inactivation. *Hum. Mol. Genet.* 10, 31–38.
- Qian, C., Li, S., Jakoncic, J., Zeng, L., Walsh, M.J., and Zhou, M.-M. (2008). Structure and hemimethylated CpG binding of the SRA domain from human UHRF1. *J. Biol. Chem.* 283, 34490–34494.
- Qin, W., Leonhardt, H., and Spada, F. (2011). Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1. *J. Cell. Biochem.* 112, 439–444.
- Qin, W., Wolf, P., Liu, N., Link, S., Smets, M., La Mastra, F., Forne, I., Pichler, G., Horl, D., Fellingner, K., et al. (2015). DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. *Cell Res* 25, 911–929.
- Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P.V., Grimaldi, G., Riccio, A., et al. (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol. Cell* 44, 361–372.

- Quivoron, C., Couronne, L., Della Valle, V., Lopez, C.K., Plo, I., Wagner-Ballon, O., Do Cruzeiro, M., Delhommeau, F., Arnulf, B., Stern, M.H., et al. (2011). TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer Cell* 20, 25–38.
- Raiber, E.-A., Beraldi, D., Ficz, G., Burgess, H.E., Branco, M.R., Murat, P., Oxley, D., Booth, M.J., Reik, W., and Balasubramanian, S. (2012). Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biol.* 13, R69.
- Rajakumara, E., Wang, Z., Ma, H., Hu, L., Chen, H., Lin, Y., Guo, R., Wu, F., Li, H., Lan, F., et al. (2011). PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression. *Mol. Cell* 43, 275–284.
- Ramesh, V., Bayam, E., Cernilogar, F.M., Bonapace, I.M., Schulze, M., Riemenschneider, M.J., Schotta, G., and Götz, M. (2016). Loss of Uhrf1 in neural stem cells leads to activation of retroviral elements and delayed neurodegeneration. *Genes Dev.* 30, 2199–2212.
- Rasmussen, K.D., and Helin, K. (2016). Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev.* 30, 733–750.
- Ratts, V.S., Flaws, J.A., Kolp, R., Sorenson, C.M., and Tilly, J.L. (1995). Ablation of bcl-2 gene expression decreases the numbers of oocytes and primordial follicles established in the post-natal female mouse gonad. *Endocrinology* 136, 3665–3668.
- Rauch, T.A., Zhong, X., Wu, X., Wang, M., Kernstine, K.H., Wang, Z., Riggs, A.D., and Pfeifer, G.P. (2008). High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. *Proc. Natl. Acad. Sci. U. S. A.* 105, 252–257.
- Ravenel, J.D., Broman, K.W., Perlman, E.J., Niemitz, E.L., Jayawardena, T.M., Bell, D.W., Haber, D.A., Uejima, H., and Feinberg, A.P. (2001). Loss of imprinting of insulin-like growth factor-II (IGF2) gene in distinguishing specific biologic subtypes of Wilms tumor. *J. Natl. Cancer Inst.* 93, 1698–1703.
- Richardson, B.E., and Lehmann, R. (2010). Mechanisms guiding primordial germ cell migration: strategies from different organisms. *Nat. Rev. Mol. Cell Biol.* 11, 37–49.
- Riggs, A.D. (1975). X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet* 14, 9–25.
- Rothbart, S.B., Krajewski, K., Nady, N., Tempel, W., Xue, S., Badeaux, A.I., Barsyte-Lovejoy, D., Martinez, J.Y., Bedford, M.T., Fuchs, S.M., et al. (2012). Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nat. Struct. Mol. Biol.* 19, 1155–1160.
- Rottach, A., Frauer, C., Pichler, G., Bonapace, I.M., Spada, F., and Leonhardt, H. (2010). The multi-domain protein Np95 connects DNA methylation and histone modification. *Nucleic Acids Res.* 38, 1796–1804.
- Rouleau, J., Tanigawa, G., and Szyf, M. (1992). The mouse DNA methyltransferase 5'-region. A unique housekeeping gene promoter. *J. Biol. Chem.* 267, 7368–7377.
- Ruchirawat, M., Noshari, J., and Lapeyre, J.N. (1987). Kinetic mechanisms and interaction of rat liver DNA methyltransferase with defined DNA substrates. *Mol. Cell. Biochem.* 76, 45–54.

Runyan, C., Gu, Y., Shoemaker, A., Looijenga, L., and Wylie, C. (2008). The distribution and behavior of extragonadal primordial germ cells in Bax mutant mice suggest a novel origin for sacrococcygeal germ cell tumors. *Int. J. Dev. Biol.* 52, 333–344.

Russler-Germain, D.A., Spencer, D.H., Young, M.A., Lamprecht, T.L., Miller, C.A., Fulton, R., Meyer, M.R., Erdmann-Gilmore, P., Townsend, R.R., Wilson, R.K., et al. (2014). The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers. *Cancer Cell* 25, 442–454.

Sado, T., Fenner, M.H., Tan, S.-S., Tam, P., Shioda, T., and Li, E. (2000). X Inactivation in the Mouse Embryo Deficient for *Dnmt1*: Distinct Effect of Hypomethylation on Imprinted and Random X Inactivation. *Dev. Biol.* 225, 294–303.

Sado, T., Okano, M., Li, E., and Sasaki, H. (2004). De novo DNA methylation is dispensable for the initiation and propagation of X chromosome inactivation. *Dev. Camb. Engl.* 131, 975–982.

Saied, M.H., Marzec, J., Khalid, S., Smith, P., Down, T.A., Rakyan, V.K., Molloy, G., Raghavan, M., Debernardi, S., and Young, B.D. (2012). Genome wide analysis of acute myeloid leukemia reveal leukemia specific methylome and subtype specific hypomethylation of repeats. *PLoS One* 7, e33213.

Saito, Y., Kanai, Y., Sakamoto, M., Saito, H., Ishii, H., and Hirohashi, S. (2002). Overexpression of a splice variant of DNA methyltransferase 3b, DNMT3b4, associated with DNA hypomethylation on pericentromeric satellite regions during human hepatocarcinogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 99, 10060–10065.

Saitou, M., Kagiwada, S., and Kurimoto, K. (2012). Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. *Development* 139, 15–31.

Sakai, T., Toguchida, J., Ohtani, N., Yandell, D.W., Rapaport, J.M., and Dryja, T.P. (1991). Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. *Am. J. Hum. Genet.* 48, 880–888.

Samokhvalov, I.M., Samokhvalova, N.I., and Nishikawa, S. (2007). Cell tracing shows the contribution of the yolk sac to adult haematopoiesis. *Nature* 446, 1056–1061.

Santos, F., Hendrich, B., Reik, W., and Dean, W. (2002). Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev. Biol.* 241, 172–182.

Santos, F., Peat, J., Burgess, H., Rada, C., Reik, W., and Dean, W. (2013). Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. *Epigenetics Chromatin* 6, 39.

Sasaki, H., and Matsui, Y. (2008). Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat Rev Genet* 9, 129–140.

Saunders, P.T.K., Turner, J.M.A., Ruggiu, M., Taggart, M., Burgoyne, P.S., Elliott, D., and Cooke, H.J. (2003). Absence of *mDazl* produces a final block on germ cell development at meiosis. *Reprod. Camb. Engl.* 126, 589–597.

Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 1412–1417.

Schaefer, M., Pollex, T., Hanna, K., Tuorto, F., Meusburger, M., Helm, M., and Lyko, F. (2010). RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev.* *24*, 1590–1595.

Schermelleh, L., Haemmer, A., Spada, F., Rösing, N., Meilinger, D., Rothbauer, U., Cardoso, M.C., and Leonhardt, H. (2007). Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res.* *35*, 4301–4312.

Schiffers, S., Ebert, C., Rahimoff, R., Kosmatchev, O., Steinbacher, J., Bohne, A.-V., Spada, F., Michalakis, S., Nickelsen, J., Müller, M., et al. (2017). Quantitative LC-MS Provides No Evidence for m(6) dA or m(4) dC in the Genome of Mouse Embryonic Stem Cells and Tissues. *Angew. Chem. Int. Ed Engl.* *56*, 11268–11271.

Schmahl, J., Eicher, E.M., Washburn, L.L., and Capel, B. (2000). Sry induces cell proliferation in the mouse gonad. *Dev. Camb. Engl.* *127*, 65–73.

Schneider, D.T., Schuster, A.E., Fritsch, M.K., Hu, J., Olson, T., Lauer, S., Göbel, U., and Perlman, E.J. (2001). Multipoint imprinting analysis indicates a common precursor cell for gonadal and nongonadal pediatric germ cell tumors. *Cancer Res.* *61*, 7268–7276.

Schomacher, L., Han, D., Musheev, M.U., Arab, K., Kienhöfer, S., von Seggern, A., and Niehrs, C. (2016). Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation. *Nat. Struct. Mol. Biol.* *23*, 116–124.

Scott, E.W., Simon, M.C., Anastasi, J., and Singh, H. (1994). Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science* *265*, 1573–1577.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol Cell* *48*, 849–862.

Seisenberger, S., Peat, J.R., Hore, T.A., Santos, F., Dean, W., and Reik, W. (2013). Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. *Philos Trans R Soc Lond B Biol Sci* *368*, 20110330.

Seki, Y., Hayashi, K., Itoh, K., Mizugaki, M., Saitou, M., and Matsui, Y. (2005). Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev Biol* *278*, 440–458.

Seki, Y., Yamaji, M., Yabuta, Y., Sano, M., Shigeta, M., Matsui, Y., Saga, Y., Tachibana, M., Shinkai, Y., and Saitou, M. (2007). Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. *Development* *134*, 2627–2638.

Selker, E.U. (2002). Repeat-induced gene silencing in fungi. *Adv. Genet.* *46*, 439–450.

- Sen, G.L., Reuter, J.A., Webster, D.E., Zhu, L., and Khavari, P.A. (2010). DNMT1 maintains progenitor function in self-renewing somatic tissue. *Nature* *463*, 563–567.
- Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., et al. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* *450*, 908–912.
- Shen, L., Wu, H., Diep, D., Yamaguchi, S., D’Alessio, A.C., Fung, H.-L., Zhang, K., and Zhang, Y. (2013). Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* *153*, 692–706.
- Shen, L., Inoue, A., He, J., Liu, Y., Lu, F., and Zhang, Y. (2014). Tet3 and DNA replication mediate demethylation of both the maternal and paternal genomes in mouse zygotes. *Cell Stem Cell* *15*, 459–470.
- Shield, J.P., Gardner, R.J., Wadsworth, E.J., Whiteford, M.L., James, R.S., Robinson, D.O., Baum, J.D., and Temple, I.K. (1997). Aetiopathology and genetic basis of neonatal diabetes. *Arch. Dis. Child. Fetal Neonatal Ed.* *76*, F39–42.
- Shih, A.H., Abdel-Wahab, O., Patel, J.P., and Levine, R.L. (2012). The role of mutations in epigenetic regulators in myeloid malignancies. *Nat. Rev. Cancer* *12*, 599–612.
- Shirane, K., Toh, H., Kobayashi, H., Miura, F., Chiba, H., Ito, T., Kono, T., and Sasaki, H. (2013). Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. *PLoS Genet.* *9*, e1003439.
- Shlush, L.I., Zandi, S., Mitchell, A., Chen, W.C., Brandwein, J.M., Gupta, V., Kennedy, J.A., Schimmer, A.D., Schuh, A.C., Yee, K.W., et al. (2014). Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* *506*, 328–333.
- Silber, J.R., Bobola, M.S., Blank, A., and Chamberlain, M.C. (2012). O(6)-methylguanine-DNA methyltransferase in glioma therapy: promise and problems. *Biochim. Biophys. Acta* *1826*, 71–82.
- Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* *12*, 246–258.
- Smallwood, S.A., Tomizawa, S., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S.R., and Kelsey, G. (2011). Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat Genet* *43*, 811–814.
- Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* *11*, 817–820.
- Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nat Rev Genet* *14*, 204–220.
- Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature*.

- Spangrude, G.J., Heimfeld, S., and Weissman, I.L. (1988). Purification and characterization of mouse hematopoietic stem cells. *Science* 241, 58–62.
- Speed, R.M. (1982). Meiosis in the foetal mouse ovary. I. An analysis at the light microscope level using surface-spreading. *Chromosoma* 85, 427–437.
- Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W., Bauer, C., Munzel, M., Wagner, M., Muller, M., Khan, F., et al. (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* 152, 1146–1159.
- Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480, 490–495.
- Stancheva, I., and Meehan, R.R. (2000). Transient depletion of xDnmt1 leads to premature gene activation in *Xenopus* embryos. *Genes Dev.* 14, 313–327.
- Stewart, K.R., Veselovska, L., Kim, J., Huang, J., Saadeh, H., Tomizawa, S., Smallwood, S.A., Chen, T., and Kelsey, G. (2015). Dynamic changes in histone modifications precede de novo DNA methylation in oocytes. *Genes Dev.* 29, 2449–2462.
- Sun, D., Luo, M., Jeong, M., Rodriguez, B., Xia, Z., Hannah, R., Wang, H., Le, T., Faull, K.F., Chen, R., et al. (2014). Epigenomic Profiling of Young and Aged HSCs Reveals Concerted Changes during Aging that Reinforce Self-Renewal. *Cell Stem Cell* 14, 673–688.
- Surani, M.A., Barton, S.C., and Norris, M.L. (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* 308, 548–550.
- Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* 9, 465–476.
- Swartz, M.N., Trautner, T.A., and Kornberg, A. (1962). Enzymatic synthesis of deoxyribonucleic acid. XI. Further studies on nearest neighbor base sequences in deoxyribonucleic acids. *J. Biol. Chem.* 237, 1961–1967.
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935.
- Takagi, N., and Abe, K. (1990). Detrimental effects of two active X chromosomes on early mouse development. *Dev. Camb. Engl.* 109, 189–201.
- Takagi, H., Tajima, S., and Asano, A. (1995). Overexpression of DNA methyltransferase in myoblast cells accelerates myotube formation. *Eur. J. Biochem.* 231, 282–291.
- Tam, P.P.L., and Snow, M.H.L. (1981). Proliferation and migration of primordial germ cells during compensatory growth in mouse embryos. *Development* 64, 133–147.

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

Temple, I.K., and Shield, J.P.H. (2002). Transient neonatal diabetes, a disorder of imprinting. *J. Med. Genet.* 39, 872–875.

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R., Deaton, A., Andrews, R., James, K.D., et al. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* 464, 1082–1086.

Tomizawa, S., Kobayashi, H., Watanabe, T., Andrews, S., Hata, K., Kelsey, G., and Sasaki, H. (2011). Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Dev. Camb. Engl.* 138, 811–820.

Toyooka, Y., Tsunekawa, N., Takahashi, Y., Matsui, Y., Satoh, M., and Noce, T. (2000). Expression and intracellular localization of mouse Vasa-homologue protein during germ cell development. *Mech. Dev.* 93, 139–149.

Trotzler, M.-A., Bronner, C., Bathami, K., Mathieu, E., Abbady, A.-Q., Jeanblanc, M., Muller, C.D., Rochette-Egly, C., and Mousli, M. (2004). Phosphorylation of ICBP90 by protein kinase A enhances topoisomerase IIalpha expression. *Biochem. Biophys. Res. Commun.* 319, 590–595.

Trowbridge, J.J., Snow, J.W., Kim, J., and Orkin, S.H. (2009). DNA Methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells. *Cell Stem Cell* 5, 442–449.

Tucker, K.L., Talbot, D., Lee, M.A., Leonhardt, H., and Jaenisch, R. (1996). Complementation of methylation deficiency in embryonic stem cells by a DNA methyltransferase minigene. *Proc. Natl. Acad. Sci. U. S. A.* 93, 12920–12925.

Tuorto, F., Liebers, R., Musch, T., Schaefer, M., Hofmann, S., Kellner, S., Frye, M., Helm, M., Stoecklin, G., and Lyko, F. (2012). RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nat. Struct. Mol. Biol.* 19, 900–905.

Tuorto, F., Herbst, F., Alerasool, N., Bender, S., Popp, O., Federico, G., Reitter, S., Liebers, R., Stoecklin, G., Gröne, H.-J., et al. (2015). The tRNA methyltransferase Dnmt2 is required for accurate polypeptide synthesis during haematopoiesis. *EMBO J.* 34, 2350–2362.

Uchida, N., and Weissman, I.L. (1992). Searching for hematopoietic stem cells: evidence that Thy-1.1lo Lin- Sca-1+ cells are the only stem cells in C57BL/Ka-Thy-1.1 bone marrow. *J. Exp. Med.* 175, 175–184.

Ueda, T., Abe, K., Miura, A., Yuzuriha, M., Zubair, M., Noguchi, M., Niwa, K., Kawase, Y., Kono, T., Matsuda, Y., et al. (2000). The paternal methylation imprint of the mouse H19 locus is acquired in the gonocyte stage during foetal testis development. *Genes Cells* 5, 649–659.

Unoki, M., Nishidate, T., and Nakamura, Y. (2004). ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain. *Oncogene* 23, 7601–7610.

- Van den Wyngaert, I., Sprengel, J., Kass, S.U., and Luyten, W.H. (1998). Cloning and analysis of a novel human putative DNA methyltransferase. *FEBS Lett.* *426*, 283–289.
- Van Emburgh, B.O., and Robertson, K.D. (2011). Modulation of Dnmt3b function in vitro by interactions with Dnmt3L, Dnmt3a and Dnmt3b splice variants. *Nucleic Acids Res.* *39*, 4984–5002.
- Vanyushin, B.F., Belozersky, A.N., Kokurina, N.A., and Kadirova, D.X. (1968). 5-Methylcytosine and 6-Methylaminopurine in Bacterial DNA. *Nature* *218*, 1066–1067.
- Vertino, P.M., Yen, R.W., Gao, J., and Baylin, S.B. (1996). De novo methylation of CpG island sequences in human fibroblasts overexpressing DNA (cytosine-5)-methyltransferase. *Mol. Cell. Biol.* *16*, 4555–4565.
- Veselovska, L., Smallwood, S.A., Saadeh, H., Stewart, K.R., Krueger, F., Maupetit-Méhouas, S., Arnaud, P., Tomizawa, S.-I., Andrews, S., and Kelsey, G. (2015). Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. *Genome Biol.* *16*, 209.
- Vilkaitis, G., Suetake, I., Klimasauskas, S., and Tajima, S. (2005). Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. *J. Biol. Chem.* *280*, 64–72.
- Vincent, J.J., Huang, Y., Chen, P.Y., Feng, S., Calvopina, J.H., Nee, K., Lee, S.A., Le, T., Yoon, A.J., Faull, K., et al. (2013). Stage-specific roles for tet1 and tet2 in DNA demethylation in primordial germ cells. *Cell Stem Cell* *12*, 470–478.
- Viré, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.-M., et al. (2006). The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* *439*, 871–874.
- Vlachogiannis, G., Niederhuth, C.E., Tuna, S., Stathopoulou, A., Viiri, K., de Rooij, D.G., Jenner, R.G., Schmitz, R.J., and Ooi, S.K.T. (2015). The Dnmt3L ADD Domain Controls Cytosine Methylation Establishment during Spermatogenesis. *Cell Rep.*
- Voo, K.S., Carlone, D.L., Jacobsen, B.M., Flodin, A., and Skalnik, D.G. (2000). Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Mol. Cell. Biol.* *20*, 2108–2121.
- Wachter, E., Quante, T., Merusi, C., Arczewska, A., Stewart, F., Webb, S., and Bird, A. (2014). Synthetic CpG islands reveal DNA sequence determinants of chromatin structure. *eLife* *3*.
- Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* *20*, 116–117.
- Walter, M., Teissandier, A., Pérez-Palacios, R., and Bourc'his, D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *eLife* *5*.
- Wang, J., Xie, G., Singh, M., Ghanbarian, A.T., Raskó, T., Szvetnik, A., Cai, H., Besser, D., Prigione, A., Fuchs, N.V., et al. (2014a). Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* *516*, 405–409.

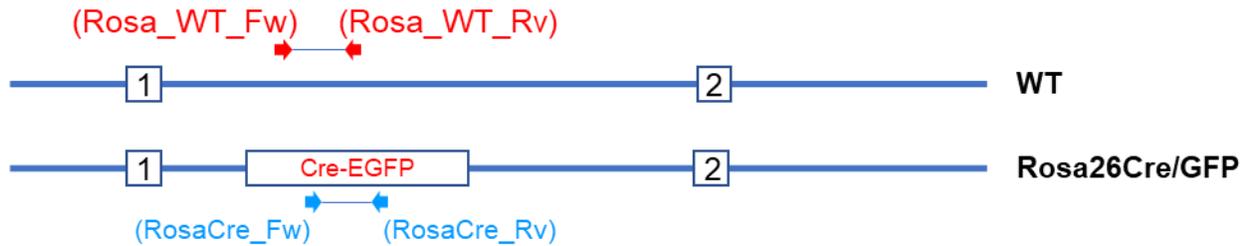
- Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W., et al. (2014b). Programming and inheritance of parental DNA methylomes in mammals. *Cell* *157*, 979–991.
- Wang, Q., Stacy, T., Binder, M., Marin-Padilla, M., Sharpe, A.H., and Speck, N.A. (1996). Disruption of the *Cbfa2* gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. *Proc. Natl. Acad. Sci. U. S. A.* *93*, 3444–3449.
- Wang, R.A., Nakane, P.K., and Koji, T. (1998). Autonomous cell death of mouse male germ cells during fetal and postnatal period. *Biol. Reprod.* *58*, 1250–1256.
- Watanabe, D., Suetake, I., Tada, T., and Tajima, S. (2002). Stage- and cell-specific expression of *Dnmt3a* and *Dnmt3b* during embryogenesis. *Mech Dev* *118*, 187–190.
- Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schubeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* *37*, 853–862.
- Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* *39*, 457–466.
- Weissman, I.L. (2000). Stem Cells. *Cell* *100*, 157–168.
- Western, P.S., Miles, D.C., van den Bergen, J.A., Burton, M., and Sinclair, A.H. (2008). Dynamic Regulation of Mitotic Arrest in Fetal Male Germ Cells. *STEM CELLS* *26*, 339–347.
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* *8*, 973–982.
- Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A.C., Rappaport, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* *473*, 343–348.
- Williamson, C.M., Ball, S.T., Dawson, C., Mehta, S., Beechey, C.V., Fray, M., Teboul, L., Dear, T.N., Kelsey, G., and Peters, J. (2011). Uncoupling antisense-mediated silencing and DNA methylation in the imprinted *Gnas* cluster. *PLoS Genet.* *7*, e1001347.
- Winkelmann, J., Lin, L., Schormair, B., Kornum, B.R., Faraco, J., Plazzi, G., Melberg, A., Cornelio, F., Urban, A.E., Pizza, F., et al. (2012). Mutations in *DNMT1* cause autosomal dominant cerebellar ataxia, deafness and narcolepsy. *Hum. Mol. Genet.* *21*, 2205–2210.
- Wolff, E.M., Byun, H.-M., Han, H.F., Sharma, S., Nichols, P.W., Siegmund, K.D., Yang, A.S., Jones, P.A., and Liang, G. (2010). Hypomethylation of a *LINE-1* promoter activates an alternate transcript of the *MET* oncogene in bladders with cancer. *PLoS Genet.* *6*, e1000917.
- Woodcock, C.L., and Ghosh, R.P. (2010). Chromatin Higher-order Structure and Dynamics. *Cold Spring Harb. Perspect. Biol.* *2*, a000596.

- Woodcock, D.M., Lawler, C.B., Linsenmeyer, M.E., Doherty, J.P., and Warren, W.D. (1997). Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. *J. Biol. Chem.* **272**, 7810–7816.
- Woodcock, D.M., Linsenmeyer, M.E., Doherty, J.P., and Warren, W.D. (1999). DNA methylation in the promoter region of the p16 (CDKN2/MTS-1/INK4A) gene in human breast tumours. *Br. J. Cancer* **79**, 251–256.
- Wu, H., D’Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473**, 389–393.
- Wu, H., Wu, X., Shen, L., and Zhang, Y. (2014). Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. *Nat. Biotechnol.* **32**, 1231–1240.
- Wu, T.P., Wang, T., Seetin, M.G., Lai, Y., Zhu, S., Lin, K., Liu, Y., Byrum, S.D., Mackintosh, S.G., Zhong, M., et al. (2016). DNA methylation on N6-adenine in mammalian embryonic stem cells. *Nature* **532**, 329–333.
- Wylie, C. (1999). Germ Cells. *Cell* **96**, 165–174.
- Xie, M., Lu, C., Wang, J., McLellan, M.D., Johnson, K.J., Wendl, M.C., McMichael, J.F., Schmidt, H.K., Yellapantula, V., Miller, C.A., et al. (2014). Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* **20**, 1472–1478.
- Xie, S., Jakoncic, J., and Qian, C. (2012a). UHRF1 double tudor domain and the adjacent PHD finger act together to recognize K9me3-containing histone H3 tail. *J. Mol. Biol.* **415**, 318–328.
- Xie, W., Barr, C.L., Kim, A., Yue, F., Lee, A.Y., Eubanks, J., Dempster, E.L., and Ren, B. (2012b). Base-resolution Analyses of Sequence and Parent-of-Origin Dependent DNA Methylation in the Mouse Genome. *Cell* **148**, 816–831.
- Xiong, Z., and Laird, P.W. (1997). COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res* **25**, 2532–2534.
- Xu, G.L., Bestor, T.H., Bourc’his, D., Hsieh, C.L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J., and Viegas-Pequignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402**, 187–191.
- Xu, J., Wang, Y.-Y., Dai, Y.-J., Zhang, W., Zhang, W.-N., Xiong, S.-M., Gu, Z.-H., Wang, K.-K., Zeng, R., Chen, Z., et al. (2014). DNMT3A Arg882 mutation drives chronic myelomonocytic leukemia through disturbing gene expression/DNA methylation in hematopoietic cells. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2620–2625.
- Xue, J.-H., Xu, G.-F., Gu, T.-P., Chen, G.-D., Han, B.-B., Xu, Z.-M., Bjørås, M., Krokan, H.E., Xu, G.-L., and Du, Y.-R. (2016). Uracil-DNA Glycosylase UNG Promotes Tet-mediated DNA Demethylation. *J. Biol. Chem.* **291**, 731–738.
- Yamaguchi, S., Hong, K., Liu, R., Shen, L., Inoue, A., Diep, D., Zhang, K., and Zhang, Y. (2012). Tet1 controls meiosis by regulating meiotic gene expression. *Nature* **492**, 443–447.

- Yamaguchi, S., Shen, L., Liu, Y., Sendler, D., and Zhang, Y. (2013). Role of Tet1 in erasure of genomic imprinting. *Nature* *504*, 460–464.
- Yamashita, Y., Yuan, J., Suetake, I., Suzuki, H., Ishikawa, Y., Choi, Y.L., Ueno, T., Soda, M., Hamada, T., Haruta, H., et al. (2010). Array-based genomic resequencing of human leukemia. *Oncogene* *29*, 3723–3731.
- Yan, X.-J., Xu, J., Gu, Z.-H., Pan, C.-M., Lu, G., Shen, Y., Shi, J.-Y., Zhu, Y.-M., Tang, L., Zhang, X.-W., et al. (2011). Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat. Genet.* *43*, 309–315.
- Yoder, J.A., and Bestor, T.H. (1998). A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. *Hum. Mol. Genet.* *7*, 279–284.
- Yoder, J.A., Yen, R.W., Vertino, P.M., Bestor, T.H., and Baylin, S.B. (1996). New 5' regions of the murine and human genes for DNA (cytosine-5)-methyltransferase. *J. Biol. Chem.* *271*, 31092–31097.
- Yoder, J.A., Soman, N.S., Verdine, G.L., and Bestor, T.H. (1997a). DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. *J. Mol. Biol.* *270*, 385–395.
- Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997b). Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet. TIG* *13*, 335–340.
- Yoshimizu, T., Sugiyama, N., De Felice, M., Yeom, Y.I., Ohbo, K., Masuko, K., Obinata, M., Abe, K., Scholer, H.R., and Matsui, Y. (1999). Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. *Dev Growth Differ* *41*, 675–684.
- You, J.S., and Jones, P.A. (2012). Cancer genetics and epigenetics: two sides of the same coin? *Cancer Cell* *22*, 9–20.
- Yu, M., Hon, G.C., Szulwach, K.E., Song, C.-X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., et al. (2012). Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* *149*, 1368–1380.
- Yu, X., Wu, C., Bhavanasi, D., Wang, H., Gregory, B.D., and Huang, J. (2017). Chromatin dynamics during the differentiation of long-term hematopoietic stem cells to multipotent progenitors. *Blood Adv.* *1*, 887–898.
- Zamudio, N., Barau, J., Teissandier, A., Walter, M., Borsos, M., Servant, N., and Bourc'his, D. (2015). DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. *Genes Dev.* *29*, 1256–1270.
- Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. *Science* *328*, 916–919.
- Zhang, G., Huang, H., Liu, D., Cheng, Y., Liu, X., Zhang, W., Yin, R., Zhang, D., Zhang, P., Liu, J., et al. (2015). N6-Methyladenine DNA Modification in *Drosophila*. *Cell* *161*, 893–906.

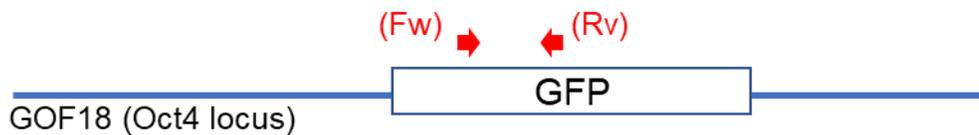
- Zhang, J., Gao, Q., Li, P., Liu, X., Jia, Y., Wu, W., Li, J., Dong, S., Koseki, H., and Wong, J. (2011). S phase-dependent interaction with DNMT1 dictates the role of UHRF1 but not UHRF2 in DNA methylation maintenance. *Cell Res* 21, 1723–1739.
- Zhang, S.J., Endo, S., Ichikawa, T., Washiyama, K., and Kumanishi, T. (1998). Frequent deletion and 5' CpG island methylation of the p16 gene in primary malignant lymphoma of the brain. *Cancer Res.* 58, 1231–1237.
- Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W., et al. (2010). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Res.* 38, 4246–4253.
- Zhao, H., and Chen, T. (2013). Tet family of 5-methylcytosine dioxygenases in mammalian development. *J. Hum. Genet.* 58, 421–427.
- Zhou, T., Xiong, J., Wang, M., Yang, N., Wong, J., Zhu, B., and Xu, R.M. (2014). Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. *Mol Cell* 54, 879–886.
- Zhu, C., Gao, Y., Guo, H., Xia, B., Song, J., Wu, X., Zeng, H., Kee, K., Tang, F., and Yi, C. (2017). Single-Cell 5-Formylcytosine Landscapes of Mammalian Early Embryos and ESCs at Single-Base Resolution. *Cell Stem Cell* 20, 720–731.e5.
- Ziller, M.J., Gu, H., Müller, F., Donaghey, J., Tsai, L.T.-Y., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500, 477–481.

## Annexe 1 : Lignées murines utilisées



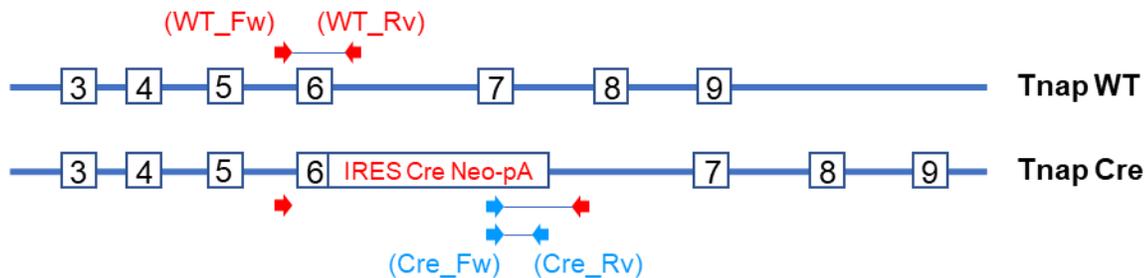
### Lignée Rosa26Cre/GFP (Birling et al. 2012)

Dans cette lignée, la séquence de la Cre fusionnée à la GFP, a été insérée au niveau du locus de *Rosa26* (gène exprimé de façon ubiquitaire dans l'embryon), permettant une expression ubiquitaire de la Cre. Cette lignée transgénique nous a permis de créer les différentes lignées hétérozygotes à partir de lignée *fllox* (*Dnmt1<sup>2lox/+</sup>*, *Dnmt3a<sup>2lox/+</sup>*, *Dnmt3b<sup>2lox/+</sup>*). Cette souris transgénique peut être identifiée par PCR à l'aide des primers Rosa\_WT\_Fw/Rosa\_WT\_Rv (représentés ici en rouge) pour l'allèle sauvage et les primers RosaCre\_Fw/RosaCre\_Rv (représentés ici en bleu) pour l'allèle contenant la Cre. (Voir Annexe de la liste des primers)



### Lignée Oct4GFP (Yoshimizu et al. 1999)

Cette lignée a été générée par transgénèse en insérant un transgène contenant la séquence codante de la GFP et 18kb de régions flanquantes du gène *Oct4* qui permettent de récapituler l'expression endogène du gène *Oct4* (Yeom et al. 1996). Cette souris transgénique peut être identifiée par PCR à l'aide de primer reconnaissant la séquence de la GFP, Fw/Rv, représentés ici en rouge. (Voir Annexe de la liste des primers)



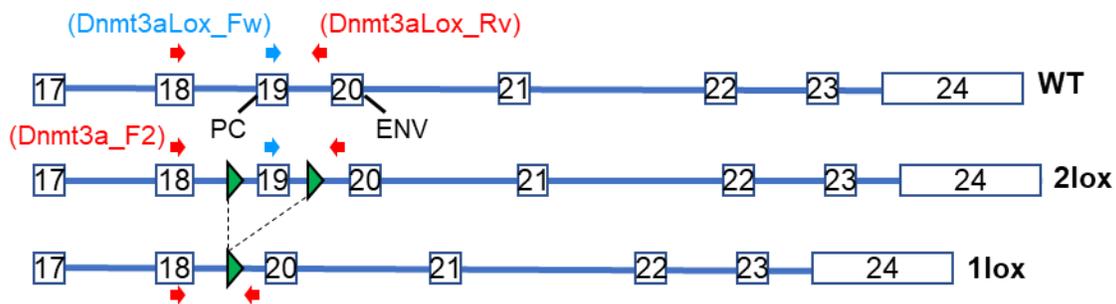
### Lignée TnapCre (Lomeli et al. 2000)

Dans cette lignée, la séquence de la Cre a été insérée dans le locus du gène *Tnap* (Tissue Non-Specific Alakaline Phosphatase), permettant une expression spécifique au niveau de la lignée germinale. Le fait que le gène *Tnap* soit hétérozygote n'impacte pas le lignage germinale, ce qui corrèle avec les études de la souris *Tnap<sup>-/-</sup>* (MacGregor et al. 1995). L'allèle Tnap-Cre doit provenir du père sinon son expression est ectopique dans l'embryon précoce. La Cre commence à s'exprimer entre E9,5 et E10,5, avec une activité ectopique dans les intestins, le placenta et le tube neuronal, et à E13,5, son expression est spécifique des PGCs. Cette souris transgénique peut être identifiée par PCR à l'aide des primers WT\_Fw/WT\_Rv (représentés ici en rouge) pour l'allèle sauvage et les primers Cre\_Fw/Cre\_Rv (représentés ici en bleu) pour l'allèle contenant la Cre. (Voir Annexe de la liste des primers)



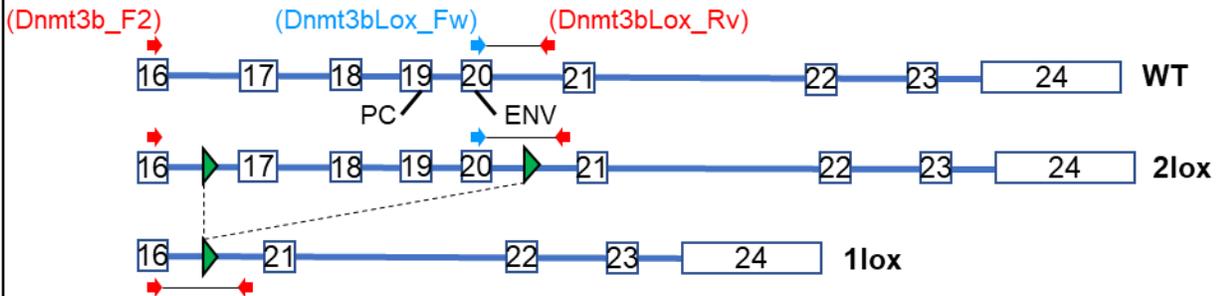
**Lignée VavCre** (Georgiades et al. 2002)

Dans cette lignée transgénique, la Cre est exprimée sous le contrôle des séquences régulatrices du gène *Vav1*. La Cre est exprimée majoritairement dans la lignée hématopoïétique, néanmoins elle présente une expression ectopique dans l'endothélium vasculaire et les cellules germinales mâles. Au niveau embryonnaire, l'expression du gène *Vav* débute à E11,5. Cette souris transgénique peut être identifiée par PCR à l'aide des primers TgVav-Cre\_For/ TgVav-Cre\_Rev (représentés ici en rouge) pour l'allèle contenant la Cre. (Voir Annexe de la liste des primers)



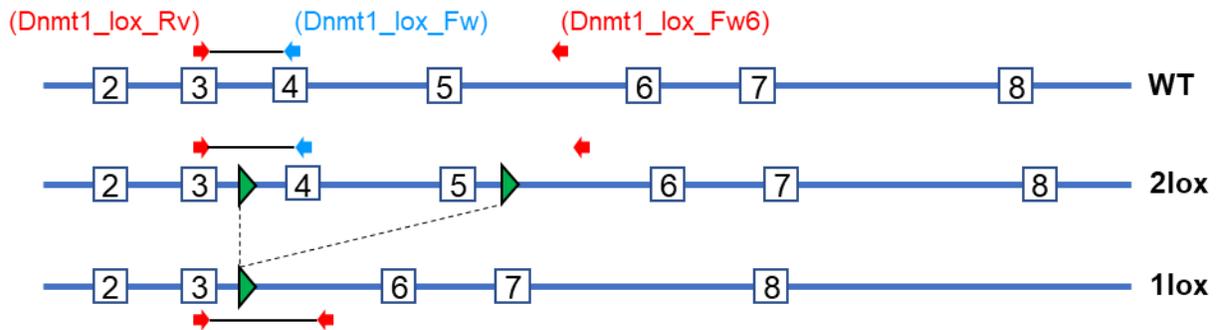
**Lignée Dnmt3a<sup>2lox</sup>** (Kaneda et al. 2004)

Dans cette lignée, des séquences loxP (triangles vert) ont été insérées entre les exons 18/19 et 19/20. La recombinaison induite par un système Cre induit la perte de l'exon 19 qui code pour le motif PC du domaine catalytique, ainsi qu'un décalage de lecture. Cette souris transgénique peut être identifié par PCR à l'aide des primers Dnmt3aLox\_Fw/Dnmt3aLox\_Rv (représentés ici en bleu et rouge) pour l'allèle sauvage et les primers Dnmt3a\_F2/Dnmt3aLox\_Rv (représentés en rouge) pour l'allèle possédant le ou les sites loxP . (Voir Annexe de la liste des primers)



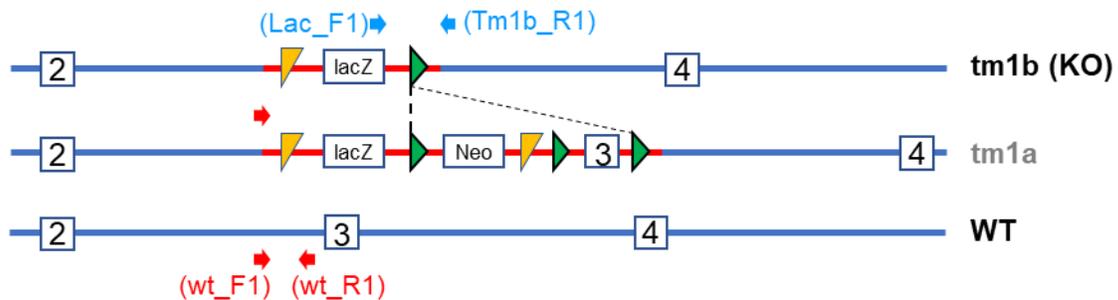
**Lignée Dnmt3b<sup>2lox</sup>** (Dodge et al. 2005)

Dans cette lignée, des séquences loxP (triangles vert) ont été insérées entre les exons 16/17 et 20/21. La recombinaison induite par un système Cre induit la perte des exons 17 à 20 qui encode pour les motifs PC et ENV du domaine catalytique, ainsi qu'un décalage de lecture. Cette souris transgénique peut être identifié par PCR à l'aide des primers Dnmt3bLox\_Fw/Dnmt3bLox\_Rv (représentés ici en bleu et rouge) pour l'allèle sauvage et les primers Dnmt3b\_F2/Dnmt3bLox\_Rv (représentés en rouge) pour l'allèle possédant le ou les sites loxP . (Voir Annexe de la liste des primers)



**Lignée *Dnmt1*<sup>2lox</sup>** (Jackson-Grusby et al. 2001)

Dans cette lignée, des séquences loxP ont été insérées entre les exons 3/4 et 5/6 du gène *Dnmt1*. La délétion induite par Cre induit la perte des exons 4 et 5 et un décalage de lecture, produisant une protéine tronquée ne possédant plus les motifs de localisation ni le domaine catalytique. Cette souris transgénique peut être identifiée par PCR à l'aide des primers *Dnmt1\_lox\_Fw*/*Dnmt1\_lox\_Rv* (représentés ici en bleu et rouge) pour l'allèle sauvage et les primers *Dnmt1\_Fw6*/*Dnmt1\_Rv* (représentés ici en rouge) pour l'allèle excisé. (Voir Annexe de la liste des primers)



**Lignée *Uhrf2* KO *tm1b*** (Bradley et al. 2012)

Cette lignée a été générée dans le cadre du projet « International Knockout Mouse Consortium ». Cette lignée provient de la souche *tm1a* (que nous n'avons pas en possession), dans laquelle les sites loxP (triangles vert) ont été recombines, excisant l'exon 3 et générant un décalage de lecture conduisant à une protéine tronquée. Cette souris transgénique peut être identifiée par PCR à l'aide des primers *Uhrf2\_WT\_F1*/*Uhrf2\_WT\_R1* (représentés ici en rouge) pour l'allèle sauvage et les primers *Uhrf2\_Lac\_F1*/*Uhrf2\_Tm1b\_R1* (représentés ici en bleu) pour l'allèle KO. (Voir Annexe de la liste des primers)

## Annexe 2 : Liste des anticorps utilisés en cytométrie de flux

<i>Projet HSC</i>			
<b>Référence</b>	<b>Désignation</b>	<b>Clone</b>	<b>Fournisseur</b>
17-1171-82	Anti-Mouse CD117 (c-Kit) APC	2B8	ebioscience
11-0341-82	Anti-Mouse CD34 FITC	RAM34	ebioscience
12-0411-81	Anti-Mouse CD41 PE	MWReg30	ebioscience
12-1441-80	Anti-Mouse CD144 (VE-Cadherin)PE	BV13	ebioscience
11-0451-81	Anti-Mouse CD45 FITC	30-F11	ebioscience
25-5981-82	Anti-Mouse Ly-6A/E (Sca-1) PE-Cy7	D7	ebioscience
45-0112-80	Anti-Mouse CD11b PerCP-Cy5.5 (Mac-1)	M1/70	ebioscience
12-1501-80	Anti-Mouse CD150 PE	9D1	ebioscience
46-0481-80	Anti-Mouse CD48 PerCP-eFluor® 710	HM48-1	ebioscience
Lineage			
13-5921-82	Anti-Mouse TER-119 Biotin	TER-119	ebioscience
13-0031-82	Anti-Mouse CD3e Biotin	145-2C11	ebioscience
13-0041-82	Anti-Mouse CD4 Biotin	GK1.5	ebioscience
13-0081-82	Anti-Mouse CD8a Biotin	53-6.7	ebioscience
13-5790-82	Anti-Mouse IgM Biotin	II/41	ebioscience
13-0193-82	Anti-Mouse CD19 Biotin	1D3	ebioscience
13-5931-82	Anti-Mouse Ly-6G (Gr-1) Biotin	RB6-8C5	ebioscience
48-4317-82	Streptavidin eFluor® 450		ebioscience
<i>Projet PGC</i>			
<b>Référence</b>	<b>Désignation</b>	<b>Clone</b>	<b>Fournisseur</b>
130-104-936	Anti-SSEA-1-PE, human and mouse	REA321	Moltenyi Biotec

## Annexe 3 : Primers de qPCR et de PCR bisulfite pour COBRA

qPCR		
Désignation	séquence	Tm
Rpl13a	GAGTCCGTTGGTCTTGAGGA	60°C
	GGTTACGGAAACAGGCAGAA	
Actb	TGATCCACATCTGCTGGAAG	60°C
	GCGCAAGTACTCTGTGTGGA	
Dnmt1	CAAAAGAGGAGGCTGCTACCA	60°C
	AAAAACATCAGTGCACGTGGG	
Dnmt3b	CACCTCACCTGTCCCCTTTTT	64°C
	CACTGAGCAGGAGAAAGCTGA	
Dnmt3b_Floxed exon	CCTCCGAAGTCTGTGCAGAG	60°C
	TGGTGATTTTCCGGACGTCA	
Dnmt3a	AGACGTCTCCAACATGAGCC	64°C
	GAAGAGGTGGCGGATGACC	
Dnmt3a_Floxed exon	GCCATTTCGACCTGGTGATT	60°C
	GGCAGGGTTGACAATGGAGA	
Uhrf1	CCAGCCTGTCGCTTTGAG	60°C
	ATAGCCAGGGAAGACCTGGT	
Uhrf2	TCCTTTAAGGCCAGGTTTT	60°C
	GCTGTAGCCAGGGAAGAAGA	

PCR bisulfite		
Désignation	séquence	Tm
Dppa3	TAGATTGATTGTTAATTGGGTTTTG	58°C
	AAAACCTCCCTAATCCTCCACTAAATC	
Ckt2	CCCTAAATATCATCTAAACTACACCTTCCT	58°C
	GAGGTTTTTGTGGTTATTGAGTGAAGTAG	
Runx1	GTTTATTATGGAGAATTGGTAGGAAT	60°C
	CAAAAACACCCTAAACAATCCTATATATC	
Gata2	GGGGATTTTAGTTTTTGTGTTGTTT	60°C
	CTATACCTTCAACCCAAACCCAATATA	

## Annexe 4 : Primers de géotypage

Nom	Séquence	Amplicon(s)	Gène	Application
EGFP_Fw	AACCACTACCTGAGCACCC	180bp	GFP	Géotypage de la lignée Oct4-GFP
EGFP_Rv	ACCTCTACAAATGTGGTATG			
Tnap_WT_Fw	GAGGTACGTGTGCAGGGACT	388bp	Tnap (Apl)	Géotypage de la lignée Tnap-Cre
Tnap_WT_Rev	CTTACAAGGGCTGCCTGTTC			
Tnap_Cre_Fw	CACGTCCGATGGCCGCTCTA	262bp	Cre-neo cassette	
Tnap_Cre_Rv	TAAGGGCCAGCTCATTCTCC			
Rosa_Cre_Fw	AAAGTCGCTCTGAGTTGTTAT	154bp	Cre	Géotypage de la lignée Rosa26Cre/GFP
Rosa_Cre_Rv	CCTTAAGCCTGCCAGAAG			
Rosa_WT_Fw	GCAAGCTGGTGGCTGGACCAATG	238bp	Rosa26	
Rosa_WT_Rv	AGGGTTGGACTCGACGTCTCC			
Dnmt1_lox_Fw	CTTGGGCCTGGATCTTGGGGA	335bp (WT) 368bp (2lox) 460bp (1lox)	Dnmt1	Géotypage de la lignée Dnmt1 WT/2lox/1lox
Dnmt1_lox_Rv	AAGCCAGTTGTGTGACTTGG			
Dnmt1_1lox_Fw6	GCAGGCAAGGCTCTAATGT			
Uhrf2_wt_F1	GCCAAGCAAGTTACAATCCATGGCAG	240bp	Uhrf2	Géotypage de la lignée Uhrf2 tm1b
Uhrf2_wt_R1	CGCAGGAATCATGTTAAGAATGCCGCC			
Uhrf2-Lac_F1	CCATTACCAGTTGGTCTGGTGTC	506bp	Uhrf2 tm1b	
Uhrf2_tm1b_R1	TCACTCTCAGTCATGCCAATCCTGG			
TgVav-Cre-For	CTGACCGTACACAAAATTTGCCTG	211bp	Cre	Géotypage de la lignée Vav-Cre
TgVav-Cre-Rev	GATAATCGCGAACATCTTCAGGTTC			
Dnmt3aLox_Fw	CTGGTGATTGGAGGCAGTCCATGCA	232bp (WT) 400bp (2lox)	Dnmt3a	Géotypage de la lignée Dnmt3a WT/2lox/1lox
Dnmt3aLox_Rv	TAGCTGAGGCTGTCTGCATCGGACA			
Dnmt3a_F2	GCACCAGGGAAAGATCATGT	500bp (1lox) 997bp (WT) 1400bp (2lox)		
Dnmt3aLox_Rv	TAGCTGAGGCTGTCTGCATCGGACA			
Dnmt3bLox_Fw	GGATGTTTCGAGAATGTTGTGGCC	815bp (WT) 951bp (2lox) 550bp (1lox)	Dnmt3b	Géotypage de la lignée Dnmt3b WT/2lox/1lox
Dnmt3b_F2	CGCAGGAAAGATTGGAACAT			
Dnmt3bLox_Rv	CAGGTCAGACCTCTCTGGTGACAAG			

## Annexe 5 : Tableau des statistiques de RRBS du projet PGCs

Sample (f=female, m=male)	Paired-end reads	Reads with unique hits (%)	Conversion rate (%)	Sequencing depth (median)	CpGs 8x
Epiblast E7,5 WT	30 735 607	75.9	99.51	69	1 303 989
PGC E9,5 WT	37 476 510	58.82	97.86	59	1 311 022
PGC E9,5 WT	21 392 551	59.83	99.85	39	1 225 318
PGC E10,5 WT	24 330 565	65.44	99.69	50	1 021 015
PGC E10,5 WT	27 730 384	69.25	99.85	46	1 212 401
PGC E11,5 WT	17 478 136	70.32	99.73	38	1 119 055
PGC E11,5 WT	22 010 125	65.68	99.8	36	1 115 487
PGC E12,5f WT	17 262 811	62.54	99.67	31	1 159 563
PGC E13,5f WT	26 719 446	72.56	99.83	54	1 208 873
PGC E14,5f WT	21 689 530	70.33	99.85	45	1 159 725
PGC E15,5f WT	22 105 243	69.03	99.65	45	1 108 914
PGC E16,5f WT	20 122 771	69.84	99.88	42	1 115 093
PGC E17,5f WT	44 002 897	65.71	99.72	93	1 096 064
PGC E12,5m WT	19 814 437	73.68	99.85	45	1 148 352
PGC E13,5m WT	23 925 345	75.21	99.89	49	1 122 405
PGC E14,5m WT	32 290 159	69.87	99.83	63	1 192 841
PGC E15,5m WT	35 823 418	67.36	99.43	59	1 295 767
PGC E16,5m WT	23 842 226	68.18	98.81	42	1 265 995
PGC E17,5m WT	34 654 500	68.46	97.01	76	1 059 686
PGC E13,5m TnapCre0/0 Dnmt1+/2lox	42 548 061	74.1	99.92	93	1 287 603
PGC E13,5m TnapCre0/0 Dnmt1+/2lox	32 805 444	72.3	99.84	72	1 241 211
PGC E13,5f TnapCre0/0 Dnmt1-/2lox	81 554 956	68.61	99.43	165	1 186 121
PGC E13,5m TnapCre0/0 Dnmt1-/2lox	30 723 575	61.28	99.5	57	1 182 551
PGC E13,5m TnapCre1/0 Dnmt1+/2lox	63 420 689	72.32	99.89	115	765 025
PGC E13,5m TnapCre1/0 Dnmt1+/2lox	56 386 711	72.55	99.91	190	1 014 757
PGC E13,5m TnapCre1/0 Dnmt1+/2lox	35 901 036	69.63	99.92	133	937 052
PGC E13,5m TnapCre1/0 Dnmt1-/2lox (cKO)	30 753 654	60.82	99.89	61	1 089 621
PGC E13,5m TnapCre1/0 Dnmt1-/2lox (cKO)	33 676 774	56.37	99.86	61	1 016 532
PGC E13,5m Uhrf2+/+	31 771 619	61.69	99.84	41	1 039 873
PGC E13,5m Uhrf2+/+	65 181 662	31.36	99.81	64	1 080 568
PGC E13,5f Uhrf2+/+	25 523 178	66.97	99.87	60	1 186 265
PGC E13,5f Uhrf2-/+	36 893 171	72.74	99.84	84	1 281 804
PGC E13,5m Uhrf2-/+	24 649 253	69.14	99.84	54	1 202 808
PGC E13,5m Uhrf2-/- (KO)	33 328 166	44.45	99.83	41	1 055 472
PGC E13,5f Uhrf2-/- (KO)	26 611 464	72.27	99.83	62	1 234 238
PGC E13,5f Uhrf2-/- (KO)	38 232 635	52.82	99.9	70	1 113 033
Soma WT	35 567 613	73.21	99.93	79	1 299 671
Soma WT	27 192 440	72.4	99.93	62	1 299 921
Soma Uhrf2-/- (KO)	53 838 620	70.65	99.91	112	1 312 443
Soma Uhrf2-/- (KO)	41 924 022	71.23	99.89	90	1 389 198
E8.5 WT	32 754 559	74.12	99.78	85	1 220 038
E8.5 WT	28 696 676	73.64	99.79	73	1 263 562
E8.5 Uhrf2-/- (KO)	34 081 556	74.17	99.79	96	1 111 734
E8.5 Uhrf2-/- (KO)	36 698 031	68.15	99.62	65	1 197 440

## Annexe 5 : Tableau des statistiques de RRBS du projet HSCs

Sample (YS=Yolk Sack , FL=Fetal Liver, BM=Bone Marrow)	Paired-end reads	Reads with unique hits (%)	Conversion rate (%)	Sequencing depth (median)	CpGs 8x
HSC YS E9,5 WT	30 223 470	71.68	99.76	61	1 274 593
HSC YS E9,5 WT	29 288 053	72.77	99.81	60	1 314 039
HSC AGM E11,5 WT	26 503 831	73.52	99.85	56	1 262 388
HSC AGM E11,5 WT	47 407 302	47.65	99.92	60	1 252 919
HSC FL E12,5 WT	32 420 330	62.01	99.75	61	1 086 185
HSC FL E12,5 WT	39 577 556	60.91	99.79	67	1 211 264
HSC FL E14,5 WT	21 124 250	71.19	99.77	39	1 214 036
HSC FL E14,5 WT	23 787 334	67.12	99.84	37	1 208 033
HSC FL E14,5 WT	27 731 458	72.04	99.82	52	1 312 593
HSC FL E14,5 WT	34 793 247	74.47	99.87	65	1 356 220
HSC BM Adult WT	20 963 990	34.37	99.77	22	1 078 392
HSC BM Adult WT	55 459 071	65.16	99.8	89	1 288 366
HSC BM Adult WT	39 850 387	62.94	99.81	67	1 231 761
HSC AGM E11,5 Dnmt3a+/+	41 325 719	35.13	99.82	39	1 138 197
HSC AGM E11,5 Dnmt3a+/+	33 614 358	68.84	99.92	63	1 262 017
HSC AGM E11,5 Dnmt3a+/-	30 989 351	53.88	99.77	53	1 280 631
HSC AGM E11,5 Dnmt3a+/-	50 002 942	24.88	99.79	40	1 207 966
HSC AGM E11,5 Dnmt3a-/- (KO)	35 345 561	42.01	99.92	46	1 127 681
HSC AGM E11,5 Dnmt3a-/- (KO)	27 170 656	66.06	99.93	47	1 224 426
HSC AGM E11,5 Dnmt3b+/-	26 591 499	66.76	99.72	42	1 244 433
HSC AGM E11,5 Dnmt3b+/-	38 717 878	68.3	99.72	84	1 320 514
HSC AGM E11,5 Dnmt3b-/- (KO)	22 811 393	64.28	99.93	47	1 105 930
HSC AGM E11,5 Dnmt3b-/- (KO)	23 682 079	27.98	99.93	24	899 948
HSC BM Dnmt3a+/+	30 658 284	71.99	99.76	58	1 275 988
HSC BM Dnmt3a-/- (KO)	28 104 900	69.58	99.4	59	1 116 941
HSC BM Dnmt3a-/- (KO)	21 421 297	67.86	99.92	42	1 149 825
HSC BM VavCre0/0 Dnmt3a +/2lox	35 613 412	72.73	99.82	98	1 081 410
HSC BM VavCre0/0 Dnmt3a +/2lox	32 462 757	72.48	99.87	78	1 245 659
HSC BM VavCre0/0 Dnmt3a -/2lox	49 502 218	72.13	99.84	134	1 180 188
HSC BM VavCre0/0 Dnmt3a -/2lox	33 780 580	71.32	99.85	76	1 180 378
HSC BM VavCre1/0 Dnmt3a -/2lox (cKO)	24 450 955	73.01	99.85	66	1 155 088
HSC BM VavCre1/0 Dnmt3a -/2lox (cKO)	30 085 500	72.44	99.88	78	1 164 589
HSC BM VavCre1/0 Dnmt3a +/2lox	36 953 033	74.16	99.89	92	1 280 074
HSC BM VavCre1/0 Dnmt3a +/2lox	28 566 472	73.78	99.89	70	1 240 889
HSC BM VavCre1/0 Dnmt3a 2lox/2lox (cKO)	28 121 924	74.21	99.89	78	1 201 891
HSC BM VavCre0/0 Dnmt3b +/2lox	20 522 339	70.72	99.64	38	1 180 677
HSC BM VavCre0/0 Dnmt3b +/2lox	33 985 836	73.32	99.46	66	1 259 601
HSC BM VavCre0/0 Dnmt3b +/2lox	29 087 477	73.35	99.83	65	1 170 396
HSC BM VavCre1/0 Dnmt3b -/2lox (cKO)	26 608 955	68.48	99.92	51	1 263 819
HSC BM VavCre1/0 Dnmt3b -/2lox (cKO)	36 203 427	74.16	99.94	76	1 307 015
HSC BM VavCre1/0 Dnmt3b -/2lox (cKO)	19 571 647	73.63	99.93	42	1 170 935
HSC BM VavCre0/0 Dnmt3b +/2lox	53 206 488	73.03	99.84	124	1 421 060
HSC BM VavCre0/0 Dnmt3b 2lox/2lox	51 215 870	63.12	99.81	95	1 265 450
HSC BM VavCre1/0 Dnmt3b 2lox/2lox (cKO)	43 173 186	72.08	98.38	85	1 294 283
HSC BM VavCre1/0 Dnmt3b 2lox/2lox (cKO)	39 662 574	68.54	99.76	76	1 358 197
HSC BM VavCre0/0 D3a2lox/2lox D3b2lox/2lox	44 288 356	60.61	99.68	116	1 034 260
HSC BM VavCre0/0 D3a2lox/2lox D3b2lox/2lox	47 938 370	70.44	99.52	109	1 321 917
HSC BM VavCre1/0 D3a2lox/2lox D3b2lox/2lox (cKO)	20 687 224	49.79	99.42	31	1 253 725
HSC BM VavCre1/0 D3a2lox/2lox D3b2lox/2lox (cKO)	51 655 954	48.75	99.72	26	783 591
HSC BM VavCre1/0 D3a2lox/2lox D3b2lox/2lox (cKO)	53 026 259	67.11	99.57	55	755 462

## **Annexe 6 : Protocoles RRBS**

---

**Revue: DNA methylation: an identity card for brain cells**

---

# Preparation of Reduced Representation Bisulfite Sequencing (RRBS) libraries



**Ambre Bender and Michael Weber**

CNRS  
University of Strasbourg  
UMR7242 Biotechnology and Cell signalling  
300, Bd Sébastien Brant  
67412 Illkirch, France

Email: [michael.weber@unistra.fr](mailto:michael.weber@unistra.fr)

**Keywords:** DNA methylation, bisulfite sequencing, RRBS, CpG islands.

## Introduction

DNA methylation is a well-known epigenetic mark consisting in the addition of a methyl group to the cytosine producing the 5-methylcytosine (5mC). 5mC is abundant in mammalian genomes and occurs almost exclusively at CpG dinucleotides. Globally, mammalian genomes are CpG-poor but some regions, termed CpG islands, have a high CpG frequency and are found frequently in promoters. Only 2 to 4% of CpG islands are methylated in somatic cells. This process is essential to maintain stable gene repression of developmental, imprinted and X-linked genes. Several methods are available for mapping DNA methylation. Compared to the Whole Genome Bisulfite Sequencing, the Reduced Representation Bisulfite Sequencing (RRBS) allows a good coverage of gene promoters and CpG islands for a lower sequencing cost (Meissner et al. 2008). Here we describe an optimized RRBS protocol, derived from the original protocol described by (Gu et al. 2011), which allows the use of small amount of starting DNA (2ng) with good efficiency.

## Procedure

### *Preparation of methylated adapters*

Methylated adapter oligos, in which all Cs are replaced by 5mCs (**Table 1**), are purchased from IDT with double high-performance liquid chromatography. The 5'-terminal phosphate in the bottom oligo facilitates the ligation of annealed adapters to DNA inserts; the phosphorothioate linkage in the top oligo prevents nuclease cleavage of the T overhang, which is required for ligation with A-tailed DNA fragments.

To produce 200  $\mu$ l of 15  $\mu$ M methylated adapters:

- Dissolve oligos in TE buffer to generate a 100  $\mu$ M stock for each oligo.
- Mix 30  $\mu$ l of the top oligo and 30  $\mu$ l of the bottom oligo in a 1.5 ml microtube.
- Add 2  $\mu$ l of 5 M NaCl and 138  $\mu$ l of TE buffer.
- Incubate in a 98°C water bath for 5 min, then turn off the water bath and let the water cool down to room temperature.
- Split the annealed adapter solution into 50  $\mu$ l aliquots. Aliquots can be stored at -80°C for 2 years.

### *Preparation on genomic DNA*

The extraction of genomic DNA from cells or tissues is a critical step, especially for very small amounts of starting material. A low purity of DNA will interfere with the optimal generation of RRBS libraries. For high amount of cells, standard genomic DNA extraction kits can be used. For low amounts of cells (<30,000), we recommend a proteinase K /phenol/chloroform extraction as described below:

- On the dry cell pellet, add 100  $\mu$ L of Lysis Buffer 2X, 100  $\mu$ L of H<sub>2</sub>O and 0.5  $\mu$ g/ $\mu$ L of proteinase K.
- Incubate at 55°C for at least 3 h.
- Add 0.5  $\mu$ L of RNase A/T1 and incubate 20 min at 37°C.
- Add 1 volume of UltraPure Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v) and mix vigorously.
- Spin 10 min at maximum speed.
- Transfer the supernatant to a DNA LoBind 1.5 ml microtube.
- Add 10  $\mu$ g of Glycogen and vortex.
- Add 1/10 volume of NaOAC 3M and vortex.
- Add 2.5 volumes of EtOH 100% (-20°C) and inverse the tube several times.
- Incubate at -20°C overnight.
- Spin 1 h at 4°C at maximum speed.
- Rinse the pellet with 500  $\mu$ L EtOH 70% (-20°C).
- Spin 15 min at 4°C at maximum speed.
- Dry the pellet.
- Resuspend the pellet in water with an adequate volume.
- Measure DNA concentration with a Qubit fluorometer per the manufacturer's instructions (**Note 1**).

Visit <http://www.epigenesys.eu> for other epigenetics and systems biology protocols

### *MspI* digestion

-Prepare the following reaction in a PCR tube:

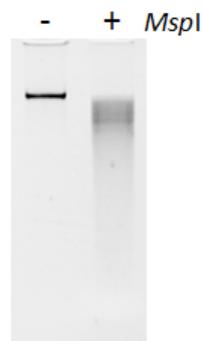
? $\mu\text{L}$	Genomic DNA (2-200 ng, max 15 $\mu\text{L}$ )
1.8 $\mu\text{L}$	Tango 10X Buffer
1.2 $\mu\text{L}$	<i>MspI</i> (10 U/ $\mu\text{L}$ )
qsp 18 $\mu\text{L}$	H <sub>2</sub> O

Final volume: 18  $\mu\text{L}$

-Incubate at 37°C for 5 h in a thermocycler.

-If possible, verify the digestion by loading 10 ng of the digested DNA alongside a non-digested control on a 1% Agarose Gel (**Figure 1**) ([Note 2](#)).

-Inactivate *MspI* by incubating at 80°C for 20 min.



**Figure 1:** Example of *MspI*-digested DNA visualized by electrophoresis on a 1% agarose gel.

### *End-Repair and A-Tailing*

-Prepare the following reaction in the same PCR tube:

18 $\mu\text{L}$	Digested DNA from the previous step
1 $\mu\text{L}$	Klenow Fragment exo- (5 U/ $\mu\text{L}$ )
0.8 $\mu\text{L}$	dNTP (10 mM dATP, 1 mM dCTP, 1 mM dGTP)
0.2 $\mu\text{L}$	Tango 10X Buffer

Final volume: 20  $\mu\text{L}$

-Incubate at 37°C for 40 min in a thermocycler.

-Inactivate the Klenow enzyme by incubating at 75°C for 15 min.

### *Adapter Ligation*

-Prepare the following reaction in the same PCR tube:

20 $\mu\text{L}$	End-Repaired DNA from the previous step
1 $\mu\text{L}$	T4 DNA ligase (30 U/ $\mu\text{L}$ )
1 $\mu\text{L}$	Methylated adapters (15 $\mu\text{M}$ )
0.5 $\mu\text{L}$	Tango 10X Buffer
0.25 $\mu\text{L}$	ATP (100 mM)
2.25 $\mu\text{L}$	H <sub>2</sub> O

Final volume: 25  $\mu\text{L}$

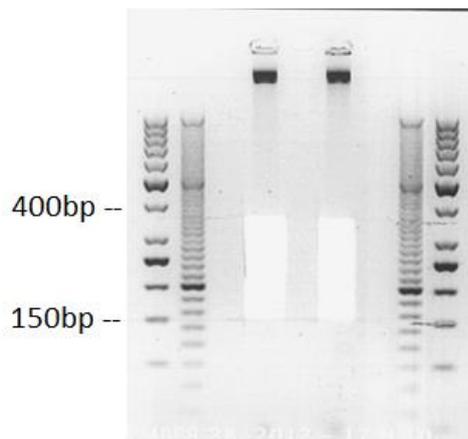
-Incubate overnight at 16°C.

-Inactivate the DNA ligase enzyme by incubating at 65°C for 20 min.

Visit <http://www.epigenesys.eu> for other epigenetics and systems biology protocols

### Size selection of the ligated DNA

- Prepare a 3% agarose TBE 0.5X gel, with additional wells at the bottom of the gel to facilitate gel excision.
- Load the samples carefully, leaving at least one empty well between samples to avoid cross contaminations.
- Load 2  $\mu\text{L}$  of 20 bp DNA Ladder and 4  $\mu\text{L}$  of 50 bp DNA Ladder (**Note 3**).
- Run the gel for 2 h at 120V until the bromophenol blue dye has migrated 7,5 cm.
- Excise the gel between 150 and 400 bp, using a different razor blade for each sample to avoid cross contaminations (**Figure 2**) (**Note 4**). This corresponds to a selection of genomic fragments between 40 and 300 bp without counting the adapters.
- Collect the gel slice in 15 mL falcon tubes.



**Figure 2:** Examples of size selection of adapter-ligated DNA by electrophoresis on a 3% agarose gel. 20 bp and 50 bp DNA Ladders are loaded on each side of the gel.

- Extract the DNA from the gel slice with the Qiagen MinElute Gel Extraction Kit following the kit instructions. Each sample is loaded on 2 MinElute columns and DNA from both columns is eluted with 12  $\mu\text{L}$  of EB Buffer in the same DNA LoBind 1.5 mL microtube. The total volume of recovered DNA is around 20  $\mu\text{L}$ .

### Verification of the size selection and ligation efficiency (optional)

- Prepare the following PCR reaction (as well as a negative PCR control):

2.5 $\mu\text{L}$	PfUTurbo Cx hotstart DNA polymerase buffer (10X)
0.6 $\mu\text{L}$	dNTP (10 mM each)
3 $\mu\text{L}$	Paired-end Illumina Primers (2.5 $\mu\text{M}$ each) ( <b>Table 1</b> )
0.5 $\mu\text{L}$	PfUTurbo Cx hotstart DNA polymerase (2.5 U/ $\mu\text{L}$ )
0.5 $\mu\text{L}$	Gel-extracted DNA from the previous step OR H <sub>2</sub> O for the negative control
17.9 $\mu\text{L}$	H <sub>2</sub> O

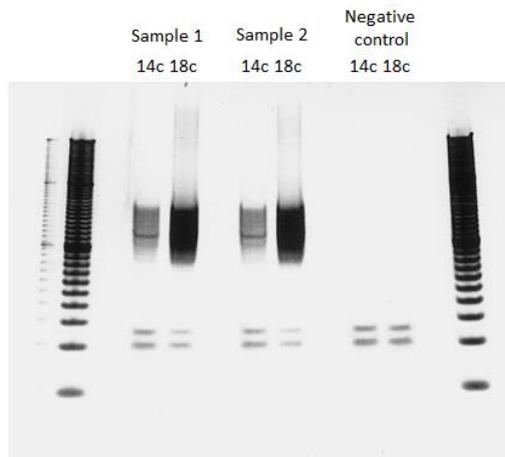
Final volume: 25  $\mu\text{L}$

- Dispense 10  $\mu\text{L}$  of the PCR reaction in two PCR tubes and perform two PCR reactions with varying number of PCR cycles (14 and 18 cycles) (**Note 5**) with the following program:

- 2 min at 95°C
- 14 or 18 cycles: 30 s at 95°C, 30 s at 65°C, 45 s at 72°C
- 7 min at 72°C

Visit <http://www.epigenesys.eu> for other epigenetics and systems biology protocols

- Load the entire PCR products on a 4-20% Criterion TBE polyacrylamide gel with 2  $\mu$ L of 20 bp DNA Ladder.
- Run at 120V for 1 h (migration of 4.5 cm).
- Stain the gel in a SYBR Green bath (1/10 000, TBE 0.5X) for 30 min on an orbital shaker (**Figure 3**).



**Figure 3:** Example of PCR validation of the size selection and adapter ligation by 14 and 18 cycles of PCR.

### *Bisulfite conversion of DNA*

Perform two consecutive rounds of bisulfite conversion using the Qiagen EpiTect Bisulfite kit, following the manufacturer's protocol designated for DNA extracted from FFPE tissues. Two rounds are necessary to achieve a C- to U conversion rate of >99.5% (**Note 6**). After the first conversion, elute the DNA twice with 11  $\mu$ L of preheated (65°C) EB Buffer in a DNA LoBind 1.5 mL microtube, and use 20  $\mu$ L to perform the second conversion. After the second conversion, elute the DNA twice with 22  $\mu$ L of preheated (65°C) EB Buffer in a DNA LoBind 1.5 mL microtube. The volume of recovered DNA is around 40  $\mu$ L.

### *Preparation of RRBS libraries*

We recommend performing a test PCR to determine the minimal number of PCR cycles for the final library.

- Prepare the following test PCR reaction (as well as a negative PCR control):

4 $\mu$ L	PfUTurbo Cx hotstart DNA polymerase buffer (10X)
1 $\mu$ L	dNTP (10 mM each)
4.8 $\mu$ L	Paired-end Illumina Primers (2.5 $\mu$ M each) ( <b>Table 1</b> )
0.8 $\mu$ L	PfUTurbo Cx hotstart DNA polymerase (2.5 U/ $\mu$ L)
0.8 $\mu$ L	Converted DNA from the previous step OR H <sub>2</sub> O for the negative control
28.6 $\mu$ L	H <sub>2</sub> O

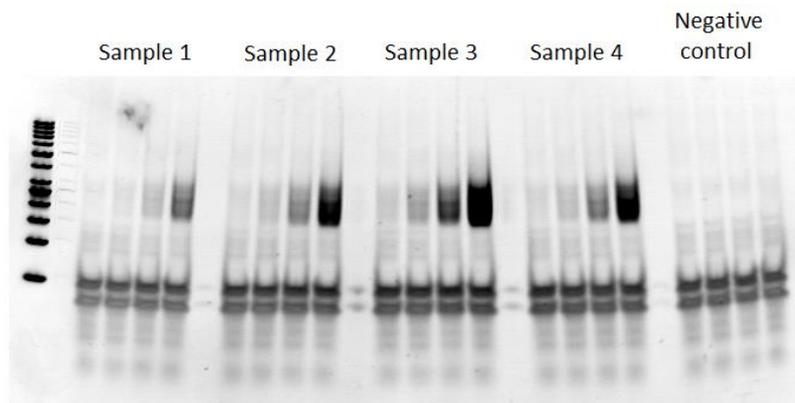
Final volume: 40  $\mu$ L

- Dispense 10  $\mu$ L of the PCR reaction in four PCR tubes and perform four PCR reactions with varying number of cycles (12, 14, 16 and 18 cycles) with the following program:

- 2 min at 95°C
- 12, 14, 16 or 18 cycles: 30 s at 95°C, 30 s at 65°C, 45 s at 72°C
- 7 min at 72°C

Visit <http://www.epigenesys.eu> for other epigenetics and systems biology protocols

- Load the entire PCR products on a 4-20% Criterion TBE polyacrylamide gel with a DNA Ladder.
- Run at 120V for 1 h (Migration of 4.5 cm).
- Stain the gel in a SYBR Green bath (1/10 000, TBE 0.5X) for 30 min on an orbital shaker (**Figure 4**) (**Note 7**).
- Determine from the gel the lowest number of cycles for the final library preparation. In our conditions, we choose two cycles below the first cycle that generates a detectable signal on the gel (see examples shown in the **Figure 4**) (**Note 8**).



**Figure 4:** Examples of test PCRs for RRBS libraries. For each sample we performed PCR with 12, 14, 16 and 18 cycles (from left to right). In these examples, we choose 14 cycles for the final PCR of sample 1, and 12 cycles for the final PCR of the samples 2, 3 and 4.

- Set up a large-scale PCR reaction for the final library preparation (as well as a negative PCR control) as follows:

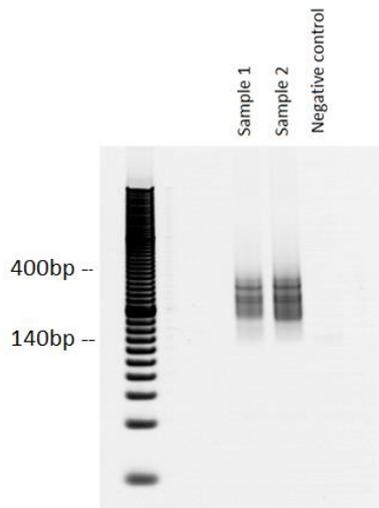
20 $\mu$ L	PfUTurbo Cx hotstart DNA polymerase buffer (10X)
5 $\mu$ L	dNTP (10 mM each)
24 $\mu$ L	Paired-end Illumina Primers (2.5 $\mu$ M each) ( <b>Table 1</b> )
4 $\mu$ L	PfUTurbo Cx hotstart DNA polymerase (2.5 U/ $\mu$ L)
24 $\mu$ L ( <b>Note 9</b> )	Converted DNA from the previous step OR H <sub>2</sub> O for the negative control
123 $\mu$ L	H <sub>2</sub> O

Final volume: 200  $\mu$ L

- Split the PCR reaction in 8x25 $\mu$ L in 8 PCR tubes.
- Perform the PCR amplification as previously described using the lowest number of cycles:
  - 2 min at 95°C
  - ?? cycles: 30 s at 95°C, 30 s at 65°C, 45 s at 72°C
  - 7 min at 72°C
- Pool the PCR products in a DNA LoBind 1.5 mL microtube.
- Add 360  $\mu$ L of AMPure magnetic beads and mix well by pipetting up and down.
- Leave at room temperature for 15 min.
- Place the tubes on a magnetic rack for 5 min until the solution becomes clear.
- Carefully discard the aqueous phase.
- Add 1 mL of fresh 70% EtOH and incubate for 5 min.
- Remove the ethanol and repeat this step.
- Let the magnetic beads air-dry completely.
- Remove the tubes from the magnetic rack and add 40  $\mu$ L of EB Buffer. Vortex.
- Place the tubes on a magnetic rack for 5 min until the solution becomes clear.
- Transfer the DNA solution to a DNA LoBind 1.5 mL microtube.

Visit <http://www.epigenesys.eu> for other epigenetics and systems biology protocols

- Quantify 1  $\mu$ L of the library with a Qubit fluorometer (Quant-iT ds DNA HS Assay Kit). Our protocol typically yields between 30 and 100 ng of final library, which is sufficient for several sequencing reactions on an Illumina HiSeq.
- Verify the size of the library and the removal of primers by loading 5-10 ng of the library on a 4-20% Criterion TBE polyacrylamide gel with a DNA Ladder.
- Run at 120V for 1 h.
- Stain the gel in a SYBR Green bath (1/10 000, TBE 0.5X) for 30 min on an orbital shaker (40 rpm) (**Figure 5**).



**Figure 5:** Examples of final RRBS libraries.

- The RRBS library is ready for sequencing with an Illumina HiSeq.

## Authors Notes

- Note 1:** We do not recommend quantifying DNA with a spectrophotometer (such as the NanoDrop) because it tends to overestimate DNA quantities.
- Note 2:** If the DNA sample is in limiting amount, one alternative is to digest another DNA in a separate tube as a parallel digestion control.
- Note 3:** We recommend diluting the DNA Ladders in 1X Tango Buffer to ensure a uniform migration.
- Note 4:** Protect the samples from UV exposure as much as possible to avoid degradation.
- Note 5:** The number of cycles can be adjusted depending on the amount of starting DNA. We use 10-14 cycles when starting from >30ng DNA and 14-18 cycles when starting from <30ng DNA.
- Note 6:** Bisulfite conversion kits from other manufacturers (e.g. Zymo Research) can be used and might require only one round of conversion. In our tests, two rounds of conversion with the Qiagen Epiect kit give the best results in terms of yield and conversion rate.
- Note 7:** Compared to the pre-bisulfite PCR, you should expect a strong decrease of the PCR signal because of DNA degradation during the bisulfite conversion.
- Note 8:** We recommend not using more than 18 cycles for the final PCR.
- Note 9:** When starting with large amounts of DNA (>10 ng), we use 24  $\mu$ L out of the 40  $\mu$ L of converted DNA for the final PCR and keep the rest as a backup. In the case of very small amounts of starting DNA (2-10 ng), it is recommended to use the 40  $\mu$ L in the final PCR.

## Material and Reagents

- TE buffer: 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA.
- Lysis Buffer 2X: 40 mM Tris-HCl pH 8.0, 10 mM EDTA, 20 mM NaCl, 2% SDS.
- Proteinase K [ThermoFisher, ref. 25530-015].
- RNase A/T1 [ThermoFisher, ref. EN0551].
- UltraPure Phenol:Chloroform:Isoamyl Alcohol (25:24:1 v/v) [ThermoFisher, ref. 15593-031].
- Glycogen 5 mg/ml [ThermoFisher, ref. AM9510].
- Tango 10X Buffer [ThermoFisher, ref. BY5].
- MspI* enzyme (10 U/μL) [ThermoFisher, ref. ER0541].
- Klenow Fragment exo- (5 U/μL) [ThermoFisher, ref. EP0421].
- T4 DNA ligase (30 U/μL) [ThermoFisher, ref. EL0013].
- 20 bp DNA Ladder [Sigma, ref. P1598].
- GeneRuler 50 bp DNA Ladder [ThermoFisher, ref. SM0371].
- Qiagen MinElute Gel Extraction Kit [Qiagen, ref. 28604].
- PfuTurbo Cx hotstart DNA polymerase (2.5 U/μL) [Agilent, ref. 600412].
- Precast 4–20% Criterion TBE Polyacrylamide Gel [Biorad, ref. 3450061].
- SYBR Green I Nucleic Acid Gel Stain 10,000X [ThermoFisher, ref. S-7563].
- EpiTect Bisulfite kit [Qiagen, ref. 59104]
- Beckman Agencourt AMPure XP 5 mL [Beckman Coulter, ref. A63880]
- Qubit fluorometer [ThermoFisher].
- DNA LoBind 1.5 ml tubes [Eppendorf]

## Tables

Name	Sequence (5' to 3')	Modifications
Methylated paired-end Adapter (top)	AcAcTcTTTcccTAcAcGAcGcTcTTccGATc*T	*: Phosphorothioate bond p: 5'-phosphate c: 5-methyl C in red: Illumina index
Methylated paired-end Adapter (bottom)	pGATcGGAAGAGcGGTTcAGcAGGAATGccGAG	
PCR primer PE 1.0	AATGATACGGCGACCACCGAGATCTCACTCTTCCCTACACGACGCTCTTCCGATC*T	
PCR primer PE 2.0	CAAGCAGAAGACGGCATAcGAGATCTCACGCATTCTGCTGAACCGCTCTTCCGATC*T	

**Table 1:** Sequences and specific modifications of oligos used in the protocol.

## References:

- Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A. 2011. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nature protocols* **6**: 468-481.
- Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB et al. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**: 766-770.

Visit <http://www.epigenesys.eu> for other epigenetics and systems biology protocols

RESEARCH HIGHLIGHT

# DNA methylation: an identity card for brain cells

Ambre Bender<sup>1</sup> and Michael Weber<sup>1\*</sup>

## Abstract

A recently published study has revealed the genome-wide dynamics of DNA methylation and hydroxymethylation patterns at single-base resolution in the human and mouse developing brain.

**Keywords:** Brain, DNA methylation, glial cells, hydroxymethylation, neurons

Epigenetic modifications are chemical changes to DNA and histones that modulate gene expression without altering the DNA sequence. A well-known epigenetic mark in mammals is the addition of a methyl group to cytosine, one of the four bases of DNA, which produces 5-methylcytosine (5mC). 5mC is found almost exclusively as a symmetrical mark in CG dinucleotides, and enzymes of the DNA (cytosine-5)-methyltransferase (DNMT) family catalyze its formation. In mammals, 5mC abounds throughout the genome and is crucial to complete embryonic development. Over the last few decades, several functions have been assigned to this epigenetic mark, such as the long-lasting silencing of genes and of parasitic mobile elements. As development proceeds, many loci undergo cell-type-specific changes in DNA methylation that ultimately contribute to the establishment and stabilization of a chosen gene expression program.

Another layer of complexity comes from the discovery that 5mC can be further oxidized by ten-eleven translocation (TET) proteins into 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). These novel modified forms of cytosine are primarily viewed as intermediates of DNA demethylation reactions.

Intriguingly, DNA methylation behaves uniquely in the brain compared with other tissues. Recent studies revealed that genomic DNA from the adult brain contains high levels of 5mC in a non-CG context [1]. In addition,

the levels of 5hmC in the brain dramatically exceed those observed in other tissues [2]. Genetic studies further indicated that changes in DNA methylation are important for brain development and learning. The loss of Dnmt1, Dnmt3A or Tet1 in the adult brain leads to cognitive deficits in mice [3,4]. In humans, mutations in DNMT1 are associated with a form of neurodegenerative disease [5], and mutations in methyl-CpG-binding protein 2 (MECP2), a protein that binds methylated DNA, are observed in several neurodevelopmental disorders [6]. Despite these indications that DNA epigenetic marks are crucial for cognitive functions, we still know very little about the underlying regulatory mechanisms involved.

A recent article by Lister *et al.* [7] has taken a step toward the better understanding of the dynamics and cell-type specificity of DNA methylation in the brain. The authors carried out a genome-wide mapping of cytosine methylation (by methylated cytosine sequencing (MethylC-seq)) and hydroxymethylation (by Tet-assisted bisulfite sequencing (TAB-seq)) at single-base resolution during the postnatal development of the mouse and human frontal cortex. In the mammalian nervous system, neural stem cells differentiate not only into neurons but also into glial cells (such as astrocytes and oligodendrocytes) that support neuronal activity. To account for possible cell-type-specific variations, the authors also generated epigenetic maps in separate populations of neurons and glial cells isolated by fluorescence-activated cell sorting. This technical *tour de force* provides an integrated view of the dynamic epigenome during mammalian brain development at an unprecedented resolution. These data confirm some previous observations by others and also reveal novel and unexpected findings that provide insights into potential pivotal functions.

## Accumulation of non-CG methylation and CG hydroxymethylation in adult neurons

Lister *et al.* [7] first showed that while global CG methylation (mCG) levels are stable during brain development, the transition from the fetal to the adult cortex is characterized by a striking accumulation of DNA methylation in a non-CG context (that is, mCH, where H = A, C or T). mCH occurs mostly in CA sequences and rapidly accumulates during childhood concomitantly

\*Correspondence: michael.weber@unistra.fr

<sup>1</sup>UMR 7242 Biotechnology and Cell Signaling, University of Strasbourg, CNRS, 300 Bd Sébastien Brant, BP 10413, 67412 Illkirch Cedex, France

with synaptogenesis. In mice, this coincides with the transient upregulation of Dnmt3A, suggesting a role of this methyltransferase in establishing non-CG methylation.

In keeping with previous reports [8], Lister *et al.* [7] also demonstrated that the cortex acquires 5hmC during postnatal development; 5hmC occurs exclusively in the CG context, indicating that mCH is not susceptible to TET-mediated oxidation in the brain. They identified a small number of megabase-sized regions that are curiously refractory to the acquisition of mCH and 5hmC in the adult cortex, even though they have high levels of mCG. These regions, termed 'mCH deserts', are enriched for large gene clusters and show signs of lower chromatin accessibility. By profiling isolated neurons and glial cells, Lister *et al.* then showed that mCH accumulates specifically in mature neurons, whereas glial cells retain low levels that resemble those of the fetal brain. In the cortical neurons of adult humans, mCH reaches levels that have never been observed in other mammalian cells and becomes the dominant form of methylation, accounting for approximately 53% of methylated cytosine residues, whereas mCG represents approximately 47%. In addition, the mCH positions throughout the genome are highly conserved between unrelated individuals, suggesting that mCH is the product of a controlled process with possible biological roles. Alternatively, the precise positioning of mCH might reflect the physical constraints of chromatin rather than a conserved, regulated function.

### **CG and non-CG methylation patterns discriminate functional gene categories**

Are patterns of CG and CH methylation associated with specific gene functions? To address this question, Lister *et al.* [7] performed an in-depth analysis of methylation levels in 1 kb bins within and around each gene. One remarkable feature is that cell-type-specific variations in methylation are precisely localized to the bodies of genes. Highly expressed genes involved in neuronal and synaptic function do not acquire intragenic mCH in mature neurons, and are marked by reduced intragenic mCG in neurons compared with glial cells. Conversely, genes associated with glial functions gain high intragenic mCH in mature neurons and are marked by reduced intragenic mCG in glial cells compared with neurons.

The inverse relationship observed between intragenic mCH and mRNA abundance suggests a possible role in transcriptional repression. This is surprising because studies in other cell types have uncovered a positive correlation between intragenic DNA methylation and transcription, which suggests that distinct mechanisms of epigenetic regulation may operate in the brain compared with other tissues. Notable differences in gene body methylation were also evident on the X chromosome;

Lister *et al.* [7] identified a subset of genes, in both humans and mice, with greater intragenic mCH levels (but not mCG levels) in female neurons compared with males. Remarkably, these genes correspond to those that escape X-inactivation in females, suggesting that their epigenetic signature could serve as a predictive feature. Indeed, on this basis, seven novel genes that escape X-inactivation were predicted.

### **CG hydroxymethylation marks sites of activation during brain development**

Lister *et al.* [7] asked if 5hmC appears at specific genomic sites during postnatal brain development and if it is implicated in developmentally regulated demethylation. The base resolution analysis by TAB-seq revealed that 5hmC in the adult murine frontal cortex is depleted in transcription start sites but accumulates throughout the body of highly transcribed genes, where it inversely correlates with intragenic mCG levels. This confirms recent results obtained by other groups that used different mapping technologies to investigate mouse neuronal cell types [8,9]. 5hmC is also present at many DNase I hypersensitive sites and enhancer regions in the mouse brain.

Recently, the first single-base resolution map of 5hmC in embryonic stem cells revealed that this mark is enriched at active enhancers that undergo activation-coupled CG demethylation [10]. The work by Lister *et al.* [7] found that this is likely also to be the case in the brain *in vivo*: a set of enhancers were identified that undergo CG demethylation in the adult brain and are marked by 5hmC in the fetal brain, suggesting that 5hmC marks these poised enhancers for demethylation and activation later in development. To validate this model, Lister *et al.* [7] investigated CG methylation in the frontal cortex of Tet2-deficient mice and observed a small increase in DNA methylation at enhancers that normally undergo demethylation. The tiny magnitude of this hypermethylation could be explained by a functional redundancy with the other Tet enzymes.

In summary, Lister *et al.* [7] have shown that neurons undergo an extensive epigenetic reconfiguration during postnatal brain development and acquire a peculiar epigenetic marking of DNA characterized by modified cytosine residues in both CG and non-CG contexts. In contrast, glial cells maintain epigenetic profiles that closely resemble those of the fetal brain. The fact that many aspects of this epigenetic pattern are conserved between mouse and humans denotes functional importance, which opens stimulating perspectives for future research. Why do neurons acquire an epigenetic profile of DNA that is so different from other cell types? Is it related to the fact that neurons stop dividing? How do these epigenetic profiles influence gene expression states?

And is there additional epigenetic heterogeneity between neuron subtypes?

We can now expect new answers from the study of mouse genetic models, as well as from the identification of proteins that bind to methylated and hydroxymethylated cytosine residues [9]. This will help to tackle the fascinating challenge of understanding how epigenetic marks regulate brain maturation, which in turn has major clinical applications for neurological disorders.

#### Abbreviations

5mC, 5-methylcytosine; 5hmC, 5-hydroxymethylcytosine; DNMT, DNA (cytosine-5) methyltransferase; MethylC-seq, methylated cytosine sequencing; MECP2, methyl-CpG-binding protein 2; TAB-seq, Tet-assisted bisulfite sequencing; TET, Ten-eleven translocation.

#### Competing interests

The authors declare that they have no competing interests.

#### Acknowledgements

The work carried out in our laboratory is supported by the Ligue Contre le Cancer, the Fondation ARC pour la Recherche sur le Cancer, MEDDTL (11-MRES-PNRPE-9-CVS-072) France, Aviesan's Atip-Avenir program and the EpiGeneSys research initiative.

Published: 27 August 2013

#### References

- Xie W, Barr CL, Kim A, Yue F, Lee AY, Eubanks J, Dempster EL, Ren B: **Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome.** *Cell* 2012, **148**:816-831.
- Nestor CE, Ottaviano R, Reddington J, Sproul D, Reinhardt D, Dunican D, Katz E, Dixon JM, Harrison DJ, Meehan RR: **Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes.** *Genome Res* 2012, **22**:467-477.
- Feng J, Zhou Y, Campbell SL, Le T, Li E, Sweatt JD, Silva AJ, Fan G: **Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons.** *Nat Neurosci* 2010, **13**:423-430.
- Zhang RR, Cui QY, Murai K, Lim YC, Smith ZD, Jin S, Ye P, Rosa L, Lee YK, Wu HP, Liu W, Xu ZM, Yang L, Ding YQ, Tang F, Meissner A, Ding C, Shi Y, Xu GL: **Tet1 regulates adult hippocampal neurogenesis and cognition.** *Cell Stem Cell* 2013, **13**:237-245.
- Klein CJ, Botuyan MV, Wu Y, Ward CJ, Nicholson GA, Hammans S, Hojo K, Yamanishi H, Karpf AR, Wallace DC, Simon M, Lander C, Boardman LA, Cunningham JM, Smith GE, Litchy WJ, Boes B, Atkinson EJ, Middha S, B Dyck PJ, Parisi JE, Mer G, Smith DJ, Dyck PJ: **Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss.** *Nat Genet* 2011, **43**:595-600.
- Gonzales ML, LaSalle JM: **The role of MeCP2 in brain development and neurodevelopmental disorders.** *Curr Psychiatry Rep* 2010, **12**:127-134.
- Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, Lucero J, Huang Y, Dwork AJ, Schultz MD, Yu M, Tonti-Filippini J, Heyn H, Hu S, Wu JC, Rao A, Esteller M, He C, Haghghi FG, Sejnowski TJ, Behrens MM, Ecker JR: **Global epigenomic reconfiguration during mammalian brain development.** *Science* 2013. doi: 10.1126/science.1237905.
- Song CX, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen CH, Zhang W, Jian X, Wang J, Zhang L, Looney TJ, Zhang B, Godley LA, Hicks LM, Lahn BT, Jin P, He C: **Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine.** *Nat Biotechnol* 2011, **29**:68-72.
- Mellén M, Ayata P, Dewell S, Kriaucionis S, Heintz N: **MeCP2 Binds to 5hmC enriched within active genes and accessible chromatin in the nervous system.** *Cell* 2012, **151**:1417-1430.
- Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B, Min JH, Jin P, Ren B, He C: **Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome.** *Cell* 2012, **149**:1368-1380.

doi:10.1186/gb-2013-14-8-131

Cite this article as: Bender A, Weber M: DNA methylation: an identity card for brain cells. *Genome Biology* 2013, **14**:131.

## Méthylation de l'ADN & Identité Cellulaire : Fonctions de la méthylation de l'ADN dans les lignages gamétiques et hématopoïétiques chez la souris.

### Résumé

La méthylation de l'ADN est la marque épigénétique la plus connue. Elle consiste en l'ajout d'un groupement méthyle au niveau de la cytosine, produisant la 5-méthyl-cytosine (5mC). Cette réaction chimique est catalysée par des ADN méthyltransférases : DNMT1, DNMT3A et DNMT3B. Peu de choses sont connues concernant les changements de 5mC au cours des lignages cellulaires dans l'embryon et comment cette marque contribue à l'établissement ou au maintien de l'identité cellulaire. Nous avons cherché à mieux comprendre ces mécanismes en étudiant la 5mC dans deux lignages cellulaires : les cellules primordiales germinales (PGCs) et les cellules souches hématopoïétiques (HSCs). Nous avons généré les premiers méthylomes de ces cellules au cours de leur développement chez la souris. Chez les PGCs, nous avons mis en évidence l'existence de deux phases de reprogrammation de la 5mC. Une première phase entre E9,5 et E13,5, où le génome des PGCs se déméthyle et une phase de reméthylation entre E14,5 et E17,5, chez les gamètes mâles uniquement. Néanmoins, certaines régions, dont notamment les éléments transposables sont résistants à la vague de déméthylation. L'utilisation de souris conditionnellement, nous a permis de mettre en évidence l'implication des protéines DNMT1 et UHRF2 dans le maintien de la 5mC au niveau de ces séquences. Concernant les HSCs, nous avons mis en évidence qu'elles acquièrent un profil de 5mC qui leur est propre lors de deux phases. La première a lieu dès l'apparition des HSCs dans l'organisme tandis que l'acquisition de la signature hématopoïétique définitive se déroule à l'âge adulte dans la moelle osseuse. L'utilisation de souris conditionnelles, nous a permis de mettre en exergue l'implication de DNMT3A et DNMT3B dans la mise en place de ces profils, avec un rôle prépondérant de DNMT3B lors de la phase d'acquisition précoce et de DNMT3A lors du verrouillage de leur profil de 5mC.

**Mots clés :** Méthylation de l'ADN, PGCs, HSCs, signature épigénétique

### Abstract

The methylation of DNA is a well-known epigenetic mark. It consists in adding a methyl group to a cytosine producing the 5-methylcytosine (5mC). This is catalysed by the DNA methyltransferase (DNMT) family: DNMT1, DNMT3A and DNMT3B. Little is known about the changes in DNA methylation that follow lineage decisions in the embryo and how these contribute, establish or maintain cellular identities. We are addressing these questions using as a model the specification of mouse primordial germ cells (PGCs) and mouse hematopoietic stem cells (HSCs) in the mouse embryo. We generate the first genome-wide maps of 5mC during their development. These maps highlight two waves of DNA methylation in PGCs. The first one takes place between E9,5 and E13,5, where the genome demethylates while the second one corresponds to a remethylation phase only in male PGCs between E14,5 and E17,5. Nevertheless, some regions, notably the transposable elements, are resistant to this demethylation wave. We demonstrate the implication of DNMT1 and UHRF2 in maintaining the 5mC on these regions using transgenic mice presenting specific deletion in PGCs. In HSCs, the 5mC maps highlight two wave of DNA methylation. The first one correlates with the first appearance of the HSCs in early embryos while the second one corresponds to their migration to the bone marrow and seems to act as a definitive lock for their hematopoietic identity. Using transgenic mice presenting specific deletions in HSCs, we prove the implication of DNMT3A and DNMT3B in hematopoietic stem cells, with a major role in locking HSC identity of DNMT3B during the first wave and a DNMT3A during the second one respectively.

**Key words :** DNA methylation, PGCs, HSCs, epigenetic signature