



HAL
open science

Identification de cibles et régulateurs de la méthylation de l'ADN chez la souris

Ghislain Auclair

► **To cite this version:**

Ghislain Auclair. Identification de cibles et régulateurs de la méthylation de l'ADN chez la souris. Biologie du développement. Université de Strasbourg, 2015. Français. NNT : 2015STRAJ061 . tel-03270788

HAL Id: tel-03270788

<https://theses.hal.science/tel-03270788v1>

Submitted on 25 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ÉCOLE DOCTORALE VIE ET SANTE

UMR 7242

THÈSE présentée par :

Ghislain AUCLAIR

Soutenue le : 22 octobre 2015

Pour obtenir le grade de : **Docteur de l'université de Strasbourg**
Discipline/ Spécialité : Aspects moléculaires et cellulaires de la biologie

**Identification de cibles et régulateurs de
la méthylation de l'ADN chez la souris**

THÈSE DIRIGÉE PAR :

Dr WEBER Michael
IREBS, Illkirch-graffenstaden

RAPPORTEUR INTERNE :

Dr TORRES PADILLA Maria Elena
IGBMC, Illkirch-graffenstaden

RAPPORTEURS EXTERNES:

Dr CAMMAS Florence
IRCM, Montpellier

Dr ARNAUD Philippe
GReD, Clermont-Ferrand

REMERCIEMENTS

Voici donc venu pour moi le temps de terminer cette thèse et de remercier tout le monde...

Je remercie donc tout d'abord Maria Elena Torres Padilla, Philippe Arnaud et Florence Cammas qui ont accepté de juger mon travail de thèse.

Et puis je vais remercier la personne la plus importante à remercier pour cette thèse, c'est bien sûr Michael Weber, alias « le chef ». Qu'il, m'excuse tout de suite, si mes virgules sont, mal placées. Je, te promets de faire, un effort de ce côté, là. Pendant que j'écris ces mots, tu es derrière le mur derrière mon écran, en train de corriger ce manuscrit. Je tiens à te remercier pour beaucoup de choses, je vais commencer par le début.

Je te remercie de m'avoir accueilli dans ton laboratoire de recherche. Quand j'ai su qu'un labo d'épigénétique se montait sur Illkirch, j'ai sauté sur l'occasion. C'était pour moi le seul endroit où je voulais passer ma thèse. On a eu un RdV et tu m'as choisi. Pour moi c'était fantastique, j'allais commencer l'aboutissement de mon rêve, passer un doctorat en biologie, qui plus est, en Epigénétique, mon domaine favoris.

Au début, j'ai eu très peur, le labo était vide, pas une pipette, pas un frigo. Juste de la poussière et un Sylvain ronchon qui ne cessait de s'inquiéter pour le moindre petit bruit. Mais très rapidement, je me suis rendu compte que tu avais de grands projets et petit à petit, j'ai vu grandir un superbe labo et vous avez su m'apprendre les rudiments de la recherche en méthylation de l'ADN. Tu as su me remettre sur les rails quand je m'égarais dans des hypothèses scientifiques douteuses, me montrer comment m'y prendre avec les résultats, être plus autonome, plus rigoureux, plus efficace. Je t'en remercie, tu as su maintenir la barre haute et grâce à toi, je sais que j'ai pu grandement évoluer depuis mon arrivée.

Le 3 décembre 2012 naquit « 1 », la première souris à l'animalerie. Et ce fut le départ d'une grande aventure... Jamais je n'oublierai cette inondation où nous étions presque toute l'équipe à éponger les raz de marée envahissant les locaux. Merci

pour tout Michael, pour ma formation, et pour ton aide dévouée et plus que précieuse à l'écriture de mon manuscrit.

Sylvain Guibert... Aaah Msieur Guibert, on en passé du temps ensemble... J'étais tellement sur que tu resterais dans mon dos jusqu'à la fin de ma thèse... Quand j'ai fait ta rencontre en octobre 2011, j'avoue que je ne savais vraiment pas comment m'y prendre avec toi. Râleur, impitoyable, maniaque,... Merci ! Parce que quand on apprend à te connaître Sylvain Guibert (et il faut un certain temps), on se rend compte que tu es une bonne personne qui en veut plus au système qu'à ma chaise qui grince. Une personne qui pose un couteau sur son bureau uniquement pour taillader une orange en milieu d'après-midi. Une personne qui répond d'un rapide « va te faire foutre » pour mieux prendre le temps d'expliquer. Et ouè mon gars, je t'ai percé ! ^^

Malheureusement, le maître incontesté de Bourrino a fini par partir... Plus personne n'y croyait au labo, on avait tous cet espoir que tu nous enterrerais... Et pourtant... En plus tu as fait un pot de départ, chose totalement inattendue ! En tout cas après ça, tu nous as manqué Sylvain, et pas seulement pour tes grandes analyses bioinformatiques (là encore, un grand merci !), mais pour ta personne.

Ça me manque moi ces moments passés à regarder par la fenêtre ensemble... Au hasard vers les étudiants de l'ESBS où sur le toboggan mouillé glissent et glissent... Tu sais, on doit avoir encore un tunnel ou deux d'inachevé, alors si un jour tu veux taper du Wither, hésite pas à m'envoyer un wagon !

Et puis un jour. Frais, droit, juste (et prude)... Elouan est arrivé. Aaah mon vieil Elouan, toutes ces années passées ensemble. L'animalerie, le labo, les biberons, les colis, le RU !

Je te remercie d'avoir toujours été là, à l'écoute, ouvert, que ça soit professionnellement (merci pour les géno !) ou personnellement. Pour ces moments où l'on a pu bien rigoler, en remplissant des biberons ou en allant au RU. Pour ces soirées jeux, ces sorties badminton. Tu es quelqu'un que j'apprécie beaucoup, et je te souhaite le meilleur à toi et Barbara pour la suite de votre vie.

Et puis une bourrasque est arrivée, emportant tout sur son passage, que dis-je une bourrasque, une tornade, un cyclone !!! Ambre !

Quand tu es arrivé Ambre, j'ai eu du mal à m'adapter, quelqu'un qui veut tout contrôler, avoir la main sur chaque chose vivante ou morte ! Un véritable ouragan ! Et puis peu à peu, j'ai découvert le personnage. Une fille parfois Barbie, parfois camionneur. Une fille à cent à l'heure, une fille géniale avec laquelle on tisse rapidement des liens d'amitié. Comme pour Elouan, que de bons moments passés à l'animalerie, au labo, aux bibs, aux colis... Mais aussi à l'extérieur, nos soirées bouffe et jeux... Nos folles déconnades (pas discrètes) au labo. The Snail, les monstruosités sous la loupe, ton humour noir, nos clin d'œil, nos galères à l'animalerie, nos stress de thésards, et tout ce qu'on a partagé avec Elouan, les parties de frisbee, les paillasses bien décorées et j'en passe !!!

Merci pour tout Ambre, pour m'avoir écouté et pour m'avoir filé un coup de main dès que j'en avais besoin. On a formé une super équipe !

Dans quelques jours, je serai là à assister à ton mariage avec ce cher Nael. Je suis sûr que vous serez heureux tous les deux et je vous souhaite de vivre des moments inoubliables le plus souvent que possible. Bonne chance pour ta thèse, c'est toi la prochaine !

Sans vous quatre Michael, Sylvain, Elouan et Ambre, m'a thèse n'aurait jamais pu être ce qu'elle a été, et pour tout ça, je vous remercie encore !

Le labo de Michael Weber, c'est un labo en constante évolution, et bon nombres de nouveaux arrivants ont suivi...

Andrea, l'espagnol ! Que d'éclats de rires nous avons pu avoir au labo ! Et je ne parle même pas de ce congrès à Heidelberg avec Ambre, où j'ai pu voir des choses bien plus personnelles de vous deux... ^^ C'était génial en tout cas, continue de t'amuser comme ça Andrea, bonne chance pour ta thèse et profite bien !

Puis plus tard, sont arrivés Judith, Claire, Thomas et Michael(2)...

Judith, un grand merci pour ton aide pour le CHIP G9a, l'aide que tu m'as apporté m'a été plus que précieuse, pour le reste aussi malgré les petits couacs ! ^^ Reste motivée, je suis sûr que tu es capable de faire de grandes choses. Et merci encore pour ton soutien, et nos superbes duo en chambre froide ! ^^

Claire, un énorme merci pour ta joie et ta bonne humeur ! Ton arrivée, bien que tardive, a apporté une nouvelle vague de sourires et de rires au labo. Reste

comme tu es, tu es géniale ! Bon courage pour la suite, je te souhaite que du bonheur.

Michael(2), on n'a pas eu le temps de beaucoup se côtoyer, toi le nouveau cerveau de Bourrino, et moi dans mes dernières manips de paillasses ! Merci pour ton aide bioinformatique, tu seras un digne successeur de Sylvain ! Finalement, on a appris à se connaître un peu plus en dehors du labo, autour d'une table et dans une bonne randonnée. Je te souhaite une bonne continuation dans le bureau D420a, et nous on se retrouvera à l'extérieur !

Thomas. Thomas, Thomas, Thomas... Mon stagiaire ! Tu es un mec super Thomas, et 6 mois ont été trop courts pour apprendre à bien se connaître. D'après toi, j'ai été excellent. Je t'en remercie, pour moi ça été super enrichissant de t'encadrer. D'excellentes discussions scientifiques, de bonnes rigolades, j'ai pris plaisir à te passer la main et à te montrer ce que je sais faire. Merci pour l'aide que tu m'as apporté sur le projet E2F6, et je te souhaite bonne chance pour ta thèse et tout le meilleur dans ta vie perso.

Voilà merci au labo Weber et à toute son équipe, ce fut pour moi un grand plaisir d'apporter une pierre dans sa construction et j'ai vraiment pris beaucoup de plaisir à y faire des recherches sur le pourquoi du comment et surtout à la grande question, mais à quoi sert la méthylation de l'ADN ?

Vous allez me manquer.

Merci aussi à tous les autres membres de l'ESBS que j'ai pu côtoyer, au comité des fêtes, à Mintah, à Jacqui. Un remerciement tout particulier à Mike, Christian, Carole, Kathline, Eléa et Thomas (K), avec lesquels j'ai passé d'excellents moments au RU, GN et à la cafèt, à me détendre et à bien rire en votre compagnie. J'espère qu'on restera tous en contact et je vous souhaite bonne chance pour vos thèses Carole, Kathline et Eléa.

En dehors du labo, j'ai beaucoup de monde à remercier, mais il y en a une que je veux par-dessus tout remercier en premier. Cette personne a toujours été là quand j'avais besoin, cette personne m'a accompagné tous les matins et tous les soirs à cette thèse, m'a soutenue dans les moments difficiles, m'a apporté son aide en cas de coup dur, était compréhensive quand je devais venir tard le soir ou les

weekends... Cette personne c'est ma fiancée, Rose-Marie. J'aurais tellement de chose à dire qu'il me faudrait plus de temps que j'en ai. Alors je vais juste te dire Merci Rose-Marie, merci d'être toujours là. Si je t'ai demandé ta main récemment, c'est parce que je sais que je ne pourrai plus vivre sans toi. Je t'aime.

Merci à mes parents. Papa, Maman, vous avez toujours été là pour me soutenir, pour me rappeler qu'il faut tenir la barre, et que vous êtes fiers de moi.

Merci à toute ma famille, pour leur soutien, à tous mes amis (ça sera plus simple que de tous vous citer, vous étiez si nombreux à me soutenir).

Un merci particulier aux nains... Où pendant quelques années, j'ai pris un énorme plaisir à jouer ce barbare vert dans nos quêtes les plus folles ! Vous avez occupé une grande place dans ma thèse, je savais que le jeudi soir, nous nous retrouverions pour nous asseoir autour d'une bière, d'une grande feuille quadrillée, de quelques pions et de chocobons. Merci pour ces grands moments à la fois philosophiques, intelligents et si absurdes à la fois. Mais bon, c'est Magique, ta g... Parmi eux, merci à Remy, mon meilleur ami qui a toujours été prêt pour une soirée bière et discussion.

Merci à Foflo, pour ton soutien avant et pendant ma thèse. Les paroles d'une grande chercheuse qui ont eu un poids en or sur certains de mes choix.

Un grand merci aux MTG players, pour votre soutien durant cette thèse, pour nos retrouvailles toujours plus folles, pourvu qu'elles continuent encore pour les années et les années à venir. Un jour nous vaincrons les slivoïdes !

Je remercie aussi l'ADDAL, l'Association des Docteurs et Doctorants d'Alsace. Pour les sorties organisées, les opportunités de rencontre et le travail associatif que j'ai pu entreprendre en tant que trésorier. Puisse-t-elle continuer à prospérer pour tous les futurs doctorants de Strasbourg.

Merci à *Two Steps* pour me maintenir motivé quel qu'en soit la situation, pendant ma thèse et surtout pendant son écriture.

Voilà voilà, je suis sûr d'oublier bien des gens. J'en suis désolé, qu'ils sachent que je leur adresse un grand merci pour l'aide ou le soutien qu'ils ont pu m'apporter.

Je profite de ce paragraphe libre pour ne pas remercier le système d'édition des articles scientifiques. Voir sans cesse que l'accès à un très grand nombre de papiers est payant m'a rendu dingue. La recherche devrait être publique. Les chercheurs ne devraient pas avoir à payer deux fois le fruit de leur travail. Payer pour

publier, payer pour lire son propre article... J'ai même reçu des emails avec des réductions ou des fidélisations à certains journaux, comme si il s'agissait de publicité pour le magazine Fashion. C'est intolérable et ce système doit évoluer. La recherche doit se faire en collaboration et doit être accessible à tous.

Bref, pour en revenir à ma thèse, après ces quatre années de recherche en Epigénétique, je crois pouvoir dire que nous n'en sommes qu'aux balbutiements de sa compréhension et cette biologie est d'autant plus complexe que ça la rend magnifique. Je suis fier d'avoir fait cette thèse et j'en remercie encore tous ceux qui y ont concouru.

Ghislain Auclair

ABREVIATIONS

5CaC	5-carboxylcytosine
5fC	5-formylcytosine
5mC	5-methylcytosine
5hmC	5-hydroxymethylcytosine
5hmU	5-hydroxymethyluracil
6mA	6-methyladenosine
ADD	Domaine protéique : ATRX-DNMT3-DNMT3L
ADN	Acide Désoxyribonucléique
ARN	Acide Ribonucléique
ARNnc	Acide Ribonucléique non codant
BAH	Domaine protéique : Bromo-Adjacent Homology
BER	Base Excision Repair
BSseq	Bisulfite-Sequencing
BWS	Beckwith-Wiedemann Syndrome
Cellules ES	Embryonic stem (cells)
CFP1	CXXC Finger Protein 1
CGi	Ilot CpG

CHG, CHH	Cytosine H Guanosine, Cytosine HH (H représente une adenosine, cytosine ou thymidine)
ChIP-seq	Chromatin Immunoprecipitation-sequencing
CMT3	Chromomethylase 3
COBRA	Combined Bisulfite Restriction Analysis
CpG	Cytosine-phosphate-Guanosine
CTCF	CCCTC-binding Factor
CXXC	Domaine protéique riche en cystéines
DBD	Domaine protéique : DNA Binding Domain
DIM	Domaine protéique : Dimerization Domain
DMAP	Domaine protéique : DNA Methyltransferase Associated Protein
DMR	Differentially Methylated Regions : sDMR = secondaire, gDMR = germinale
DNMT	DNA Methyl Transferase (1,2, 3L, 3a et 3b)
E2F1	Protéine E2F numéro 1
E2F6	Protéine E2F numéro 6
ERV	Endogenous Retroviruse
EZ	Enhancer of Zeste
FMR1	Maladie : fragile X mental retardation 1
G9a	Aussi appelée Ehmt2 ou KMT1C
GLP	Aussi appelée Ehmt1 ou KMT1D

H1	Histone H1
H2A	Histone H2A
H2B	Histone H2B
H3	Histone H3
H3K4m3	Histone H3, 3 fois méthylée sur la lysine (K) 4
H3K9m1,2,3	Histone H3, 1,2 ou 3 fois méthylée sur la lysine (K) 9
H3K23m3	Histone H3, 3 fois méthylée sur la lysine (K) 23
H3K27m3	Histone H3, 3 fois méthylée sur la lysine (K) 27
H3K36m3	Histone H3, 3 fois méthylée sur la lysine (K) 36
H4	Histone H4
HCP	High CpG Promoter
HDAC	Histone deacetylase
HiSeq	High-throughput sequencing
HP1	Heterochromatin protein 1
HSAN	Maladie : Hereditary Sensory and Autonomic Neuropathy
HSC	Hematopoietic Stem Cell
IAP	Intracisternal A-particle
ICF	Syndrome ICF: Immunodeficiency, Centromeric instability, Facial anomalies
ICM	Inner cell Mass
ICP	Intermediate CpG promoter

ICR	Imprinting Control region
Jpc/dpc	Jour <i>post-coïtum</i> /day <i>post-coïtum</i>
KDM1 B	Lysine (K) specific demethylase
KO	Knock Out
LCP	Low CpG Promoter
LINE	Long Interspersed Element
LSC	Leukemia stem cell
LTR	Long Terminal Repeat
MBD	Methyl Binding Domain (1,2,3 et 4)
MBD-cap	Methyl Binding Domain-capture
MBP	Methyl Binding Protein
MeCP2	Methyl CpG binding protein 2
MedIP	Methylated DNA immunoprecipitation
NER	Nucleotide Excision Repair
NuRD	Nucleosome Remodeling Deacetylase
OCT4	Octamer-binding transcription factor 4 aussi appelée POU5F1 (POU domain, class 5, transcription factor 1)
ORC	Origin Recognition Complex
oxBS-seq	Oxidative Bisulfite Sequencing
PBAT	Post Bisulfite Adapter Tagging
PCNA	Proliferating Cell Nuclear Antigen

PCR	Polymerase Chain Reaction
PGC	Primordial Germ Cells
PGC7	Aussi appelée Stella ou Dppa3
PHD	Domaine protéique: Pleckstrin homology domain
piARN	Piwi-Interacting ARN
pmCGis	Partially methylated CGis
PRDM1	PR domain zinc finger protein 1
PRDM14	PR domain zinc finger protein 14
PRMT5	Protein Arginine methyltransferase 5
PWS/AS	Prader-Willi Syndrome/Angelman Syndrome
PWWP	Domaine protéique riche en prolines (P) et thryptophanes (W)
RFTS	Domaine protéique: Replication Foci Targeting Sequence
RNA-seq	RNA-sequencing
RRBS	Reduced Recombinated Bisulfite Sequencing
RT-qPCR	Reverse Transcription quantitative Polymerase Chain Reaction
SAM	S-adenosyl-L-methionine
SINE	Short Interspersed Element
SMUG1	Single-Strand-Selective Monofunctional Uracil-DNA Glycosylase 1
SNP	Single-Nucleotide Polymorphisms
SRA	Domaine protéique: SET and Ring finger-associated

TAB-seq	Tet-assisted bisulfite sequencing
TBP	TATA-box protein
TDG	Thymine DNA Glycosylase
TE	Transposable Element
TET	Protéine : Ten-Eleven Translocation
TIF1β	Aussi appelée KAP1 ou Trim28
TSS	Transcription Start Site
UBE3A	Ubiquitin-protein ligase E3A
UHRF1	Ubiquitin-like, containing PHD and RING finger domains, 1, aussi appelé Np95 ou ICBP90
UHRF2	Ubiquitin-like, containing PHD and RING finger domains, 2, aussi appelé NIRF
UpG	Uracile-phosphate-Guanosine
WGBS	Whole Genome Bisulfite Sequencing
WT	Wild Type
ZFP	Zinc Finger Protein

SOMMAIRE

INTRODUCTION.....

I) Le génome Eucaryote.....	4
a) <u>Les éléments du génome</u>	4
b) <u>La chromatine</u>	5
c) <u>L'épigénétique</u>	6
II) La méthylation des cytosines de l'ADN.....	9
a) <u>Découverte</u>	9
b) <u>Mise en place moléculaire et contexte nucléotidique</u>	11
c) <u>La méthylation des ilots CpG chez les Mammifères</u>	13
1. Définition des ilots CpG	13
2. La méthylation d'ilots CpG des séquences promotrices	14
3. Méthylation des autres ilots CpG	15
d) <u>Les ADN méthyltransférases (DNMTs)</u>	16
1. DNMT1	18
2. DNMT2	19
3. DNMT3L	20
4. DNMT3A et DNMT3B	20
e) <u>Les voies de déméthylation de l'ADN</u>	23
1. La déméthylation de manière passive	25
2. La déméthylation de manière active	25
2.1 Les enzymes TET	25
2.2 Les enzymes AID/APOBEC	26
f) <u>Les protéines reconnaissant l'ADN méthylé</u>	27
1. Les protéines à domaine MBD	27

2. Les protéines à domaine ZFP.....	28
3. Les protéines à domaine SRA.....	28

III) Dynamiques et fonctions de la méthylation de l'ADN

.....	31
a) <u>Cartographie de la méthylation de l'ADN</u>	31
1. Techniques d'analyse de la méthylation de l'ADN de régions cibles	31
1.1 Par enzyme de restriction : COBRA.....	31
1.2 Par séquençage d'une région cible : BSseq	32
2. Techniques d'analyse de la méthylation à l'échelle du génome.....	33
2.1 MedIP (Methylated DNA Immunoprecipitation).....	33
2.2 RRBS (Reduced Representation Bisulfite Sequencing).....	34
2.3 WGBS (Whole Genome Bisulfite Sequencing).....	34
b) <u>Dynamique de la méthylation de l'ADN à l'échelle du génome</u>	38
1. Répartition globale de la méthylation de l'ADN.....	38
2. Dynamique embryonnaire de la méthylation de l'ADN chez la souris	
.....	39
2.1 De la fertilisation à l'implantation	40
2.2 De l'implantation à la fin de la gastrulation	43
2.3 Des PGCs aux cellules germinales	43
c) <u>Régulation de l'empreinte génomique parentale</u>	46
1. Une expression monoallélique en cluster	46
2. La régulation des ICRs	49
d) <u>Inactivation du chromosome X</u>	49
e) <u>Pathologies liées à des perturbations de la méthylation de l'ADN</u>	50
1. Syndromes développementaux.....	50
2. Les cancers.....	52

IV) Modes de ciblage de la méthylation de l'ADN.....

a) <u>Rôles des ARN</u>	54
b) <u>Rôles des modifications d'histones</u>	56

1. Les modifications d'histones de chromatine active.....	56
2. Les modifications d'histones de chromatine inactive.....	57
c) <u>Rôles des facteurs de transcription</u>	60
1. E2F6, un facteur recrutant la méthylation de l'ADN	60
2. Les facteurs protégeant l'ADN de la méthylation	61
3. Les facteurs protégeant l'ADN de la déméthylation.....	62
d) <u>Autres possibilités de ciblage de la méthylation</u>	64
V) Rôle de G9a dans la régulation de la méthylation de l'ADN	65
a) <u>G9a et son homologue GLP</u>	65
b) <u>Le rôle de G9a dans le recrutement de la méthylation de l'ADN</u>	68
1. Les cibles potentielles de méthylation de l'ADN par G9a.....	68
2. Les mécanismes potentiels de régulation de la méthylation de l'ADN par G9a.....	69
3. Les facteurs influençant l'activité de G9a.....	70
3.1 Les facteurs de transcription WIZ et ZNF644.....	70
3.2 Le facteur de transcription E2F6.....	71
3.3 Le facteur de transcription LSH.....	71
RESULTATS	74
➤ RESULTATS DU PROJET 1 : Etablissement de la méthylation de l'ADN au cours du développement embryonnaire murin	76
• INTRODUCTION	76
• TRAVAUX REALISES.....	77
• ARTICLE 1	81
• CONCLUSIONS.....	108
➤ RESULTATS COMPLEMENTAIRES DU PROJET 1 : Analyse du double mutant <i>Dnmt3a</i> ^{-/-} <i>Dnmt3b</i> ^{-/-}	109
I) <u>Obtention des souris double Knock Out pour DNMT3A et DNMT3B</u>	109

II)	<u>Les souris double Knock Out perdent toute trace de méthylation globale</u>	111
III)	<u>La méthylation globale n'affecte presque pas le transcriptome</u>	114
•	CONCLUSIONS	115
➤	RESULTATS DU PROJET 2 : Identification du rôle de G9a dans le recrutement de la méthylation de l'ADN durant le développement embryonnaire murin.....	117
•	INTRODUCTION	117
•	TRAVAUX REALISES	119
•	ARTICLE 2	121
•	CONCLUSIONS	161
➤	RESULTATS DU PROJET 3 : Identification du rôle d'E2F6 dans le recrutement de la méthylation de l'ADN durant le développement embryonnaire murin.....	163
•	INTRODUCTION	163
•	TRAVAUX REALISES	164
•	RESULTATS	165
I)	<u>Expression d'E2F6 dans l'embryogénèse et rôle dans la méthylation de gènes candidats</u>	165
II)	<u>Le méthylome des embryons E2F6^{-/-} révèle une conservation globale de la méthylation de l'ADN et quelques régions hypométhylées</u>	168
III)	<u>L'influence d'E2F6 sur la méthylation de l'ADN semble directe</u>	172
IV)	<u>E2F6 réprime des gènes méiotiques</u>	173
V)	<u>E2F6 et G9a ne régulent pas la méthylation des mêmes régions</u>	174
•	CONCLUSIONS	176

DISCUSSION	180
BIBLIOGRAPHIE	194
ANNEXE	218
• ARTICLE 3	220

INTRODUCTION

La méthylation de l'ADN est une modification dite épigénétique. Elle prend place durant le développement embryonnaire sur tout le génome des Mammifères. Les objectifs de ma thèse sont de comprendre où et comment la méthylation se met en place, et quels sont les facteurs qui permettent de réguler cette modification chimique de l'ADN. Dans cette introduction, je présenterai dans une première partie le génome eucaryote et l'épigénétique. En deuxième et troisième parties je développerai nos connaissances actuelles sur la méthylation de l'ADN. En quatrième partie, je traiterai des acteurs de la méthylation et de la déméthylation de l'ADN. Et enfin en cinquième partie, je présenterai G9a, la modification d'histone associée à cette protéine, ainsi que nos connaissances sur leurs rôles dans le contrôle de la méthylation de l'ADN.

I) Le génome Eucaryote

a) Les éléments du génome

Chez les Eucaryotes, le génome est l'ensemble du matériel génétique de l'organisme. Il est codé par son ADN (Acide Désoxyribonucléique) et contient à la fois des séquences codantes et non-codantes. Les séquences codantes sont appelées gènes et leur nombre varie d'un organisme à l'autre. D'après la base de données Ensembl (juillet 2015), il y a chez la Souris 22.547 gènes codants (et 12.583 gènes non codants) pour une taille de génome d'environ 3482 Mpb (Million de paires de base). Chez l'Homme, on trouve 20296 gènes codants (et 25.173 gènes non codants) pour une taille de génome d'environ 3545 Mpb. Les séquences non-codantes regroupent les introns situés dans les gènes, les régions inter-géniques, les pseudogènes, et les régions répétées. Les régions répétées se composent de différents éléments : les répétitions en tandem (satellites et micro-satellites) et les

éléments transposables (TEs). Les éléments transposables se subdivisent en deux classes. La classe 1 comprend les rétrotransposons, composée de LTRs (Long Terminal Repeat), par exemple les ERVs (Endogenous Retroviruses) et les IAPs (Intracisternal A-particle) et de non-LTR, par exemple les LINEs (Long Interspersed Elements) et les SINEs (Short Interspersed Elements). La classe 2 comprend les transposons ADN, qui contrairement aux rétrotransposons, ne nécessitent pas d'intermédiaire ARN pour se dupliquer dans le génome.

b) La chromatine

L'ADN des cellules eucaryotes est enroulé dans une structure nucléoprotéique appelée chromatine. Il 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 deux histones H3, deux histones H4, deux histones H2A et deux histones H2B, tous organisés par paires. Un neuvième histone de liaison, l'histone H1, se lie à l'ADN entre les nucléosomes et permet d'aboutir à une structure complexe composée d'histones et d'ADN appelée la fibre chromatinienne (Luger et al., 1997). Les fibres de chromatine peuvent se compacter donnant naissance au dernier niveau de compaction de la chromatine : le chromosome métaphasique.

Il existe deux types de chromatine, l'euchromatine et l'hétérochromatine. Il s'agit d'états plus ou moins condensés de la chromatine, pouvant affecter la transcription des gènes. 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 dites épigénétiques (Woodcock and Ghosh, 2010).

c) L'épigénétique

L'Épigénétique est un terme désignant des changements persistants mais réversibles dans la chromatine. Ces états de chromatine sont réversibles car ils peuvent subir un effacement en réponse à un stimulus cellulaire ou durant le développement embryonnaire, permettant ainsi de réinitialiser les états chromatiniens acquis durant la vie des cellules parentes. Ces modifications peuvent affecter l'expression des gènes sans modification de la séquence de l'ADN et peuvent être transmises à la descendance cellulaire par mitose ou méiose (Berger et al., 2009). Il peut s'agir de modifications post-traductionnelles des histones (ex : l'acétylation, la méthylation, la sumoylation,...), de modification chimique de l'ADN (ex : la méthylation de l'ADN), de remplacement d'histones par des variants d'histones (ex : H1.1, H3.1, H2Ax,...), de remodelage de la chromatine par des complexes ATP-dépendant, ou de l'utilisation de petits ou long ARNs non codants capables de dégrader les transcrits ou de réguler l'expression des gènes (**Figure 1**).

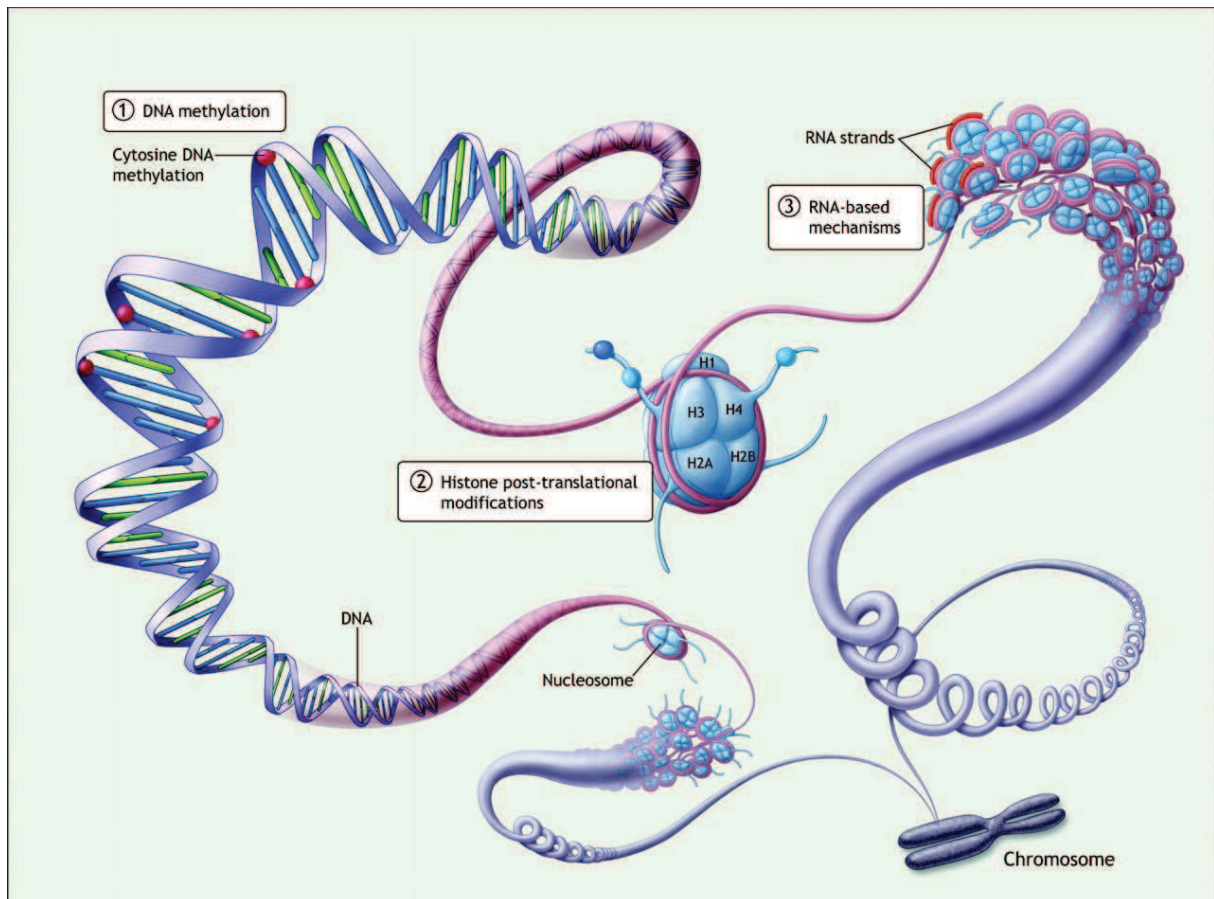


Figure 1 : La chromatine et ses différentes modifications épigénétiques.

La molécule d'ADN peut être modifiée chimiquement par méthylation (en rouge). Elle s'enroule autour d'octamères comprenant les dimères d'histones H3, H4, H2A et H2B, et l'histone H1. Ce complexe nucléoprotéique est alors appelé nucléosome. Dans ce nucléosome, les histones peuvent subir des modifications post-traductionnelles qui permettent de compacter ou de décompacter la chromatine. Les différents nucléosomes se regroupent pour former la fibre chromatinienne. Des micro-ARNs peuvent interagir à la fibre chromatinienne. Les fibres de chromatine peuvent se compacter à un degré supérieur pour former le chromosome métaphasique. Adaptée de (Fraineau et al., 2015).

Ainsi, au cours du développement embryonnaire, la chromatine subit de nombreuses modifications, visant à exprimer ou à réprimer des gènes pour permettre aux cellules de se différencier en un type cellulaire précis, et à l'embryon de se développer correctement. Le concept de l'épigénétique vient du biologiste Conrad Hal Waddington (Waddington, 2012) qui, en 1942, suggère selon ces mots un « epigenetic landscape » qui entourerait les gènes et leur permettrait de choisir un destin de différenciation adapté à leur environnement. Continuant sa métaphore, il décrit un système irréversible où la cellule descend une montagne à travers certaines vallées qui la dirigeront vers un destin cellulaire plutôt qu'un autre, sans possibilité de retour (**Figure 2**).

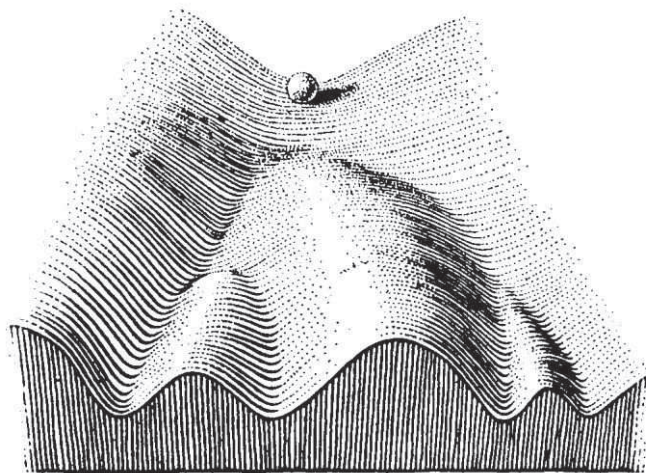


Figure 2 : Représentation classique de la vue de C.H. Waddington du paysage épigénétique.

La sphère représente une cellule embryonnaire qui se divise et se différencie au cours du développement embryonnaire (représenté conceptuellement par la descente). Les différentes cellules filles prendront des voies/vallées différentes selon leurs acquisitions épigénétiques. Adaptée du livre *The Strategy of the genes*, 1957.

II) La méthylation de l'ADN

a) Découverte

C'est en 1925 que fut découverte la possibilité de l'existence de la 5-méthyl cytosine (Johnson and Coghill, 1925). Ces chercheurs découvrent la présence de cette cytosine modifiée dans l'acide tuberculinique, un acide nucléique isolé de la bactérie *Mycobacterium tuberculosis*. Une vingtaine d'années plus tard, le biochimiste Rollin Hotchkiss découvre l'existence naturelle de cytosines méthylées chez les Eucaryotes à partir d'expériences de chromatographie réalisées avec de l'ADN de thymus de veau (Hotchkiss, 1948). Ce n'est qu'une trentaine d'années plus tard que la méthylation de l'ADN fut révélée comme étant impliquée dans des processus biologiques tels que la régulation des gènes et la différenciation cellulaire (Holliday and Pugh, 1975; Compere and Palmiter, 1981). Aujourd'hui, on sait que la méthylation de l'ADN est impliquée dans de nombreux processus physiologiques via la régulation de l'expression des gènes et la stabilité du génome. De plus, il fut observé que la perturbation des profils de méthylation est souvent associée à des pathologies, en particulier des cancers (voir chapitre III, partie e). Certains de ces cancers sont d'ailleurs causés par des mutations dans les enzymes responsables de la méthylation des cytosines, les ADN méthyl transférases DNMT (DNA Methyltransferase), ou de la déméthylation des cytosines, les protéines TET (Ten-eleven Translocation methylcytosine dioxygenase).

On trouve ainsi de la méthylation des cytosines de l'ADN dans la plupart des organismes chez les vertébrés, invertébrés, plantes et champignons. Pourtant, certains organismes Eucaryotes en sont dépourvus. On trouve par exemple de la méthylation de l'ADN chez le champignon filamenteux *Neurospora Crassa*, mais elle est absente chez la levure *Saccharomyces Cervisiae*. De même, chez les insectes, la méthylation des cytosines existe en faible quantité chez *Apis mellifera* (Feng et al., 2010; Lyko et al., 2010), mais pas chez *Drosophila Melanogaster* (Raddatz et al.,

2013) (**Figure 3**). La méthylation des cytosines est également indétectable chez le ver nématode *Caenorhabditis elegans*. Cela indique que la méthylation des cytosines n'est pas indispensable à la vie d'organismes multicellulaires complexes et que son absence pourrait être compensée par d'autres marques épigénétiques chez certains organismes. A ce propos, il est intéressant de mentionner que la présence de méthylation sur l'azote 6 des Adénines (adenine N(6)-methylation, 6mA) vient d'être décrite très récemment chez *Drosophila Melanogaster* et *Caenorhabditis elegans* (Greer et al., 2015; Zhang et al., 2015), suggérant que cette forme de méthylation pourrait compenser l'absence de méthylation des cytosines dans ces organismes (**Figure 3**). Pendant ma thèse, je me suis plus particulièrement intéressé à la méthylation de l'ADN chez les Mammifères.

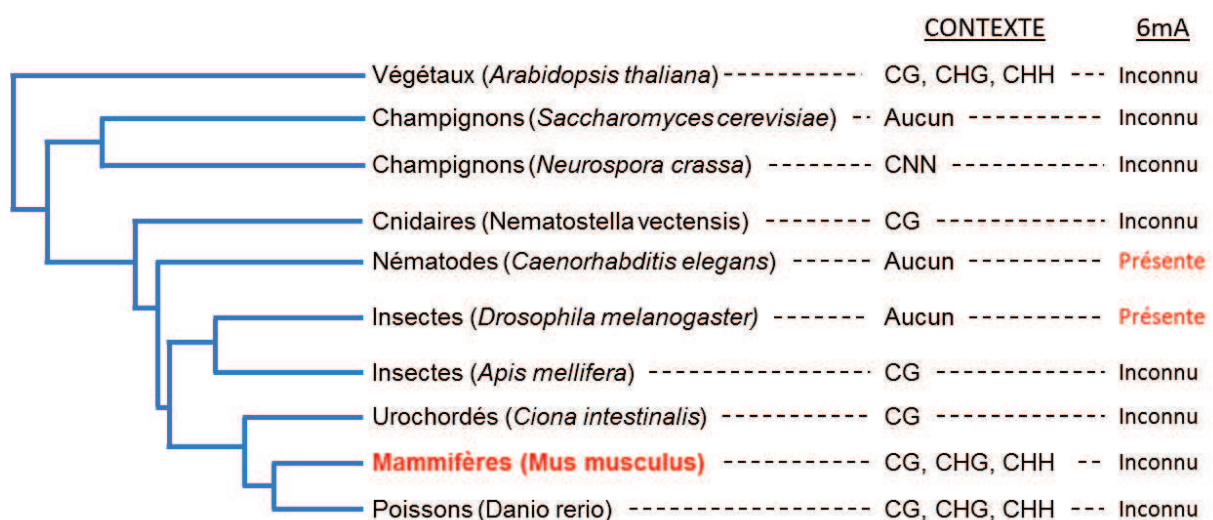


Figure 3 : Distribution phylogénétique de la méthylation de l'ADN mises en évidence chez les Eucaryotes.

La méthylation de l'ADN sur une cytosine peut se trouver dans plusieurs contextes différents comprenant CG, CHG (H= A,C ou T), CHH ou CNN (N=A,C,T ou G). La méthylation sur l'azote 6 des adénines (6mA) est mise en évidence dans certains organismes qui n'ont pas de méthylation des cytosines. Adaptée de (Feng et al., 2010; Ratel et al., 2006 ; Zemach et al., 2010).

b) Mise en place moléculaire et contexte nucléotidique

La méthylation de l'ADN est une modification chimique appliquée par les enzymes de la famille des ADN (cytosine-5-)-methyltransferases. A l'aide du cofacteur SAM (S-adenosyl-L-méthionine), ces enzymes transfèrent un groupement méthyl sur le carbone cinq de la cytosine, créant ainsi la 5-méthyl cytosine (5mC) (**Figure 4**).

Chez les plantes et les champignons, la méthylation des cytosines existe dans le contexte de dinucléotides CpG (Cytosine-phosphate-Guanosine) ou de trinucléotides CHG ou CHH (La lettre H représente un A, C ou T) (Feng et al., 2010; Lister et al., 2008) (**Figure 3**).

Chez les animaux comme l'abeille et les vertébrés, la méthylation de cytosines de l'ADN prend place dans la majorité des cas dans un contexte CpG (Feng et al., 2010; Lister et al., 2009; Stadler et al., 2011). En effet, à l'échelle du génome, la méthylation des cytosines ne représente que 3 à 5% de l'ensemble des cytosines, alors qu'elle est présente sur environ 60 à 80% des dinucléotides CpG. Ces CpG méthylés sont la plupart du temps localisés dans des régions pauvres en CpG et couvrent très souvent les éléments répétés qui représentent près de 40% du génome (Gama-Sosa et al., 1983).

Cependant elle existe aussi, chez certains animaux dont les Mammifères, dans d'autres contextes comme les séquences CHG et CHH (H=A, C ou T). La méthylation de l'ADN sur CpG peut être maintenue pendant la réplication de l'ADN car elle forme un palindrome sur les deux brins de l'ADN. En revanche, les autres modifications sur les contextes CHG et CHH ne peuvent être maintenues pendant la réplication de l'ADN et doivent donc être rétablies *de novo* à chaque division cellulaire. Chez les animaux, la méthylation non-CG a été trouvée dans les ovocytes et les neurones, cellules qui ne se divisent pas, et dans les cellules ES humaines et murines (Lister et al., 2009; Shirane et al., 2013; Xie et al., 2012; Arand et al., 2012; Ziller et al., 2011). Dans les cellules ES, la présence de méthylation non-CpG a été liée à la forte activité des enzymes DNMT3A et DNMT3B (Arand et al., 2012; Ziller et

al., 2011). Le rôle régulateur de cette méthylation non-canonique en dehors des CpG commence à être exploré. Par exemple, deux équipes ont récemment obtenu des résultats suggérant que la méthylation non-CG pouvait être reconnue par MeCP2 et régulaient l'expression de gènes dans les neurones postmitotiques (Chen et al., 2015; Guo et al., 2014b).

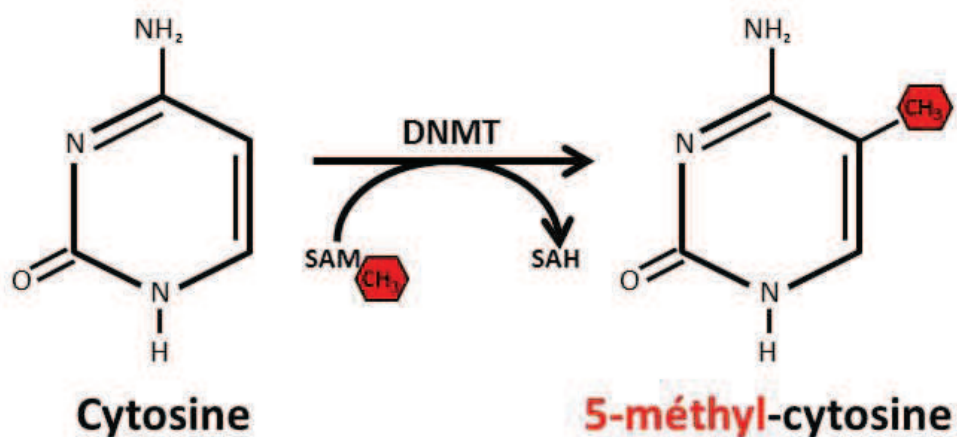


Figure 4 : La méthylation d'une cytosine de l'ADN.

Une DNMT (DNA Methyltransferase) transfère un groupement méthyle (CH₃) via l'utilisation du cofacteur SAM (S-adenosyl-L-méthionine) sur le carbone 5 de la base azotée Cytosine. La nouvelle base méthylée est appelée 5-méthyl-cytosine.

c) La méthylation des îlots CpG chez les Mammifères

1. Définition des îlots CpG

Les CpG sont parfois regroupés en régions à forte densité en CpG appelées îlots CpG (CGIs, CpG islands). Ces îlots CpG de taille variable entre 200 et 2000pb, peuvent contenir 5 à 10 CpG par 100 paires de base, ce qui représente une densité environ dix fois supérieure comparée aux autres régions de l'ADN. Ces îlots CpG sont présents dans environ 60-70% des promoteurs de gènes dans les génomes des Mammifères (Saxonov et al., 2006), mais seuls 2 à 4% d'entre eux sont méthylés dans les cellules somatiques (Meissner et al., 2008; Auclair et Weber, 2012).

Les îlots CpG peuvent être divisés en deux grandes classes : ceux situés sur les promoteurs de gènes annotés et ceux, appelés îlots « orphelins », qui se situent dans des régions intra et intergéniques et représentent environ la moitié des îlots CpG (Illingworth et al., 2010) (**Figure 5**). Une grande proportion de ces îlots orphelins portent la marque d'histone H3K4m3, une marque de chromatine transcriptionnellement active, et sont des sites d'initiation de transcription caractérisant potentiellement des promoteurs de gènes non-décrits, d'ARN non-codants ou des promoteurs secondaires de transcrits alternatifs (Illingworth et al., 2008, 2010). Parmi les transcrits des îlots « orphelins », plusieurs ARN non codants ont été mis en évidence. *Air* est un ARNnc dont la transcription s'initie dans un îlot CpG situé dans un intron du gène *Igf2r* et est nécessaire pour réprimer l'allèle paternel de ce gène (Sleutels et al., 2002). *HOTAIR* est un ARNnc qui est transcrit à partir d'un îlot CpG et qui permet de réguler les gènes *Hox* (Rinn et al., 2007). Enfin, un îlot CpG dans l'intron 1 du gène *Kcnq1* est à l'origine d'un ARNnc (*Kcnq1ot1*) requis pour la régulation du domaine *Kcnq1* (Mancini-DiNardo et al., 2006). Ainsi, environ 40% des îlots CpG « orphelins » ont une activité promotrice chez l'Homme et la souris. Certains sont cependant espèce-spécifiques (Illingworth et al., 2010). Seulement la moitié des îlots CpG non méthylés co-localisent entre le génome de l'Homme et celui de la souris (Illingworth et al., 2010). Cette perte évolutive chez la souris pose la question de leur importance régulatrice.

Enfin, Il est intéressant de remarquer que 83% des ilots CpG co-localisent avec les sites d'origine de réplication (Delgado et al., 1998; Sequeira-Mendes et al., 2009). Une étude a d'ailleurs montré que le complexe ORC reconnaît spécifiquement certaines régions à ilots CpG, quand celles-ci sont méthylées (Bartke et al., 2010).

Les ilots CpG des promoteurs et les ilots CpG « orphelins » se distinguent fortement par leur susceptibilité à la méthylation de l'ADN. En effet, il a été montré que 20 à 34% des ilots de régions intragéniques sont méthylés dans un ou plusieurs lignages cellulaires. A l'inverse, seuls 3% des ilots CpG situés aux promoteurs de gènes annotés sont méthylés (Illingworth et al., 2010; Maunakea et al., 2010).

2. La méthylation d'ilots CpG des séquences promotrices

Il existe trois grandes catégories d'ilots CpG méthylés liés à des promoteurs : les ilots du chromosome X, les ilots liés à des gènes soumis à l'empreinte parentale et les ilots liés à des gènes de la lignée germinale.

Au niveau du chromosome X, des centaines d'ilots CpG sont méthylés sur la copie inactive du chromosome X dans les cellules femelles. Dans ces cas, la méthylation est recrutée par un ARN non codant *Xist* sur un seul des deux chromosomes X des femelles mammifères. Cette méthylation se produit après l'inactivation des gènes associés et durant l'implantation embryonnaire (voir *chapitre III, partie d*).

Des gènes soumis à l'empreinte parentale peuvent également présenter une méthylation d'ilots CpG de promoteur sur un seul des deux allèles parentaux. Cette méthylation mono-allélique provient soit du gamète paternel soit du gamète maternel (voir *chapitre III, partie c*).

Enfin, il a été montré que les gènes de la lignée germinale et certains gènes du développement comme les gènes *Hox* portent fréquemment une méthylation au niveau des promoteurs liée à une répression transcriptionnelle dans les cellules somatiques (Illingworth et al., 2008, Weber et al., 2007). Cependant, il est important de remarquer que la majorité des promoteurs à ilots CpG sont constitutivement non

méthylés, ce qui n'empêche pas tout de même une répression du gène associé. Cela démontre que la méthylation de l'ADN n'est pas un mécanisme universel de répression des promoteurs de gènes. Des îlots CpG non méthylés peuvent ainsi être réprimés par des modifications spécifiques d'histones. Il a d'ailleurs été montré que certains îlots possèdent à la fois les marques H3K4me3 et H3K27me3 (Bernstein et al., 2006; Mikkelsen et al., 2007). Cette bivalence de marques permet à ces îlots CpG d'être dans deux états alternatifs. Pendant la différenciation, ils peuvent perdre la marque H3K27me3 et devenir actifs, ou bien perdre la marque H3K4me3 et devenir inactifs. Ces îlots bivalents représentent environ un cinquième des îlots dans les cellules ES (Ku et al., 2008) et dans d'autres types cellulaires (Mikkelsen et al., 2007).

De plus, des données montrent que la méthylation des îlots CpG des promoteurs n'a pas un rôle dans l'initiation de la répression du gène. En effet, des îlots ont été montrés comme acquérant de la méthylation après l'arrivée de marques d'histones répressives, et après l'arrêt de la transcription des ARN associés (Csankovszki et al., 2001). La méthylation des îlots aurait donc plutôt un rôle de maintien de l'état réprimé pour stabiliser la répression à long terme.

3. Méthylation des autres îlots CpG

Les îlots CpG « orphelins » sont plus fréquemment méthylés dans les cellules différenciées que ceux présents dans les promoteurs de gènes annotés (Davies et al., 2012; Illingworth et al., 2008, 2010). C'est aussi le cas dans les cellules cancéreuses ou ces îlots CpGs intragéniques et intergéniques peuvent être hyperméthylés, ce qui est associé à l'inactivation d'ARN non-codants (Lujambio et al., 2010).

Il a été proposé que la méthylation de séquences intragéniques puisse influencer l'épissage d'exons alternatifs. Selon un modèle proposé, la méthylation de l'ADN présente dans les exons serait le recruteur de la modification d'histone H3K9me3 et des protéines HP1 (Heterochromatin Protein 1), qui recruteraient des protéines impliquées dans l'épissage alternatif d'exons (Lev Maor et al., 2015). Selon

un autre modèle, la méthylation de l'ADN permettrait le recrutement du répresseur transcriptionnel CTCF qui pourrait alors induire une pause locale de l'ARN polymérase et favoriser ainsi l'assemblage du spliceosome au niveau du site d'épissage alternatif (Shukla et al., 2011).

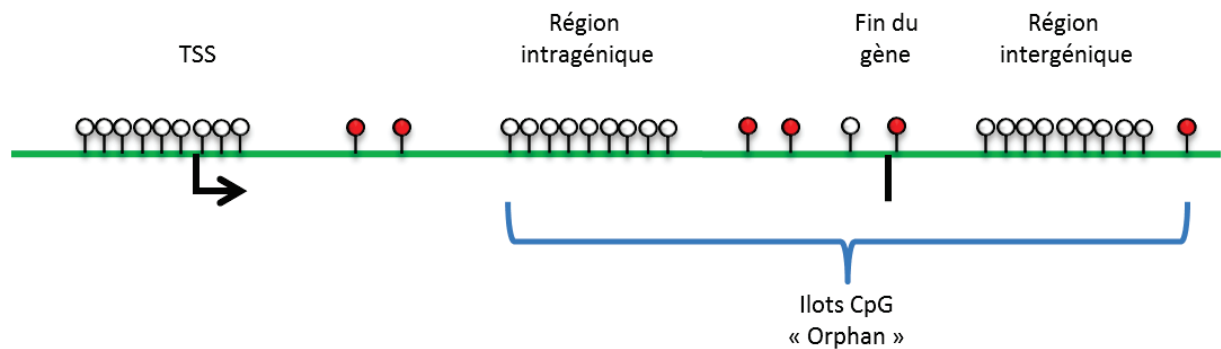


Figure 5 : Représentation schématique des îlots CpG le long du génome.

Les îlots CpG se trouvent au niveau du TSS dans 60-70% des promoteurs de gènes annotés. On peut aussi trouver des îlots CpG dans les régions intra et intergéniques, ce sont les îlots CpG « orphelins ». Ici les cercles blancs représentent des cytosines non méthylées, et les cercles rouges représentent des cytosines méthylées.

d) Les ADN méthyltransférases : les DNMTs

Chez les Mammifères, il existe cinq membres paralogues de la famille des ADN méthyltransférases (DNMT) : DNMT1, DNMT2 (ou TRDMT1), DNMT3A, DNMT3B et DNMT3L (**Figure 6**). Parmi ces cinq molécules, seules DNMT1, DNMT3A et DNMT3B sont capables de méthyler l'ADN sur le carbone 5 de la cytosine (Hermann et al., 2004). Il est à souligner qu'hormis DNMT1 et DNMT2, les protéines DNMT possèdent un domaine ADD (ATRX-DNMT3-DNMT3L) qui leur permet d'interagir avec les histones H3 non modifiées.

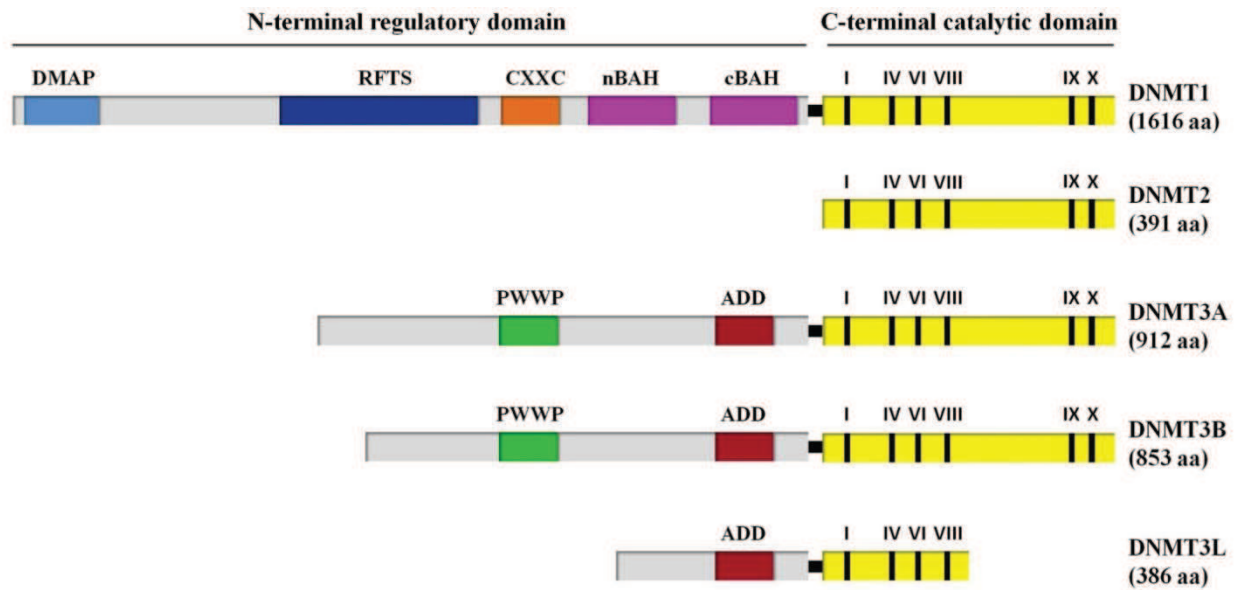


Figure 6 : Représentation schématique de la structure des cinq enzymes de la famille des ADN méthyl transférases (DNMT) humaines.

Les enzymes DNMTs disposent toutes un domaine catalytique en C-terminal. Elles possèdent également différents domaines régulateurs en commun ou non en N-terminal qui leur permettent d'interagir à l'ADN ou à des protéines. Seule DNMT3L détient un domaine catalytique tronqué et ne possède donc plus la capacité de 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 aussi un domaine CXXC, riche en cystéines et permettant la liaison de DNMT1 à des séquences d'ADN comportant des CpGs 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 et leur donnant la possibilité d'interagir avec certaines histones modifiées. Enfin, les protéines DNMT3A, DNMT3B et DNMT3L 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. Adaptée de (Maresca et al., 2015).

1. DNMT1

DNMT1 est la plus grande ADN méthyltransférase, avec un poids moléculaire de 184 kDa (**Figure 6**). L'absence de cette enzyme dans des souris conduit à une létalité embryonnaire autour de 9.5 jpc (jours *post coïtum*), soulignant son caractère vital (Lei et al., 1996). Ces embryons déficients pour DNMT1 montrent une perte globale de la méthylation de l'ADN sur le génome (Lei et al., 1996).

DNMT1 est responsable du maintien de la méthylation de l'ADN durant les divisions cellulaires. Lors de la division cellulaire, l'ADN est répliqué et les deux nouveaux duplex d'acides nucléiques ne portent de la méthylation que sur l'un des deux brins ; on dit que l'ADN est hémiméthylé. Selon le modèle communément admis, DNMT1 se localise aux sites de réplication de l'ADN en interagissant avec PCNA (Proliferating Cell Nuclear Antigen) et UHRF1 (Ubiquitin-like, containing PHD and RING finger domains, 1, aussi appelé Np95 ou ICBP90). UHRF1 interagit avec l'ADN hémiméthylé et favorise le recrutement de DNMT1 au niveau de l'ADN hémiméthylé (Bostick et al., 2007; Sharif et al., 2007). Une fois recrutée, DNMT1 catalyse le dépôt de méthylation sur les sites CpG hémiméthylés et copie donc fidèlement les profils de méthylation de l'ancien brin sur le nouveau brin d'ADN (**Figure 7**). Ce modèle est également renforcé par des études montrant que DNMT1 a une activité catalytique préférentielle sur les sites CpG hémiméthylés par rapport aux sites CpG non méthylés (Vilkaitis et al., 2005). Des études structurales ont permis d'éclaircir le mécanisme de la préférence de DNMT1 pour les sites hémiméthylés. En effet, la structure de DNMT1 en liaison avec l'ADN a montré que la liaison de DNMT1 à des CpG hémiméthylés rend l'ADN accessible pour le domaine catalytique de DNMT1 alors qu'à l'inverse, la liaison de DNMT1 à des CpG non méthylés protège l'accès au site catalytique par le brin d'ADN par un mécanisme d'inhibition conformationnelle du domaine CXXC de DNMT1 (Song et al., 2011).

Le site CpG est un site palindromique, et c'est grâce à cette configuration symétrique que la méthylation des CpGs peut être copiée fidèlement sur les brins d'ADN néosynthétisés et donc constituer une modification épigénétique héritable au cours des divisions cellulaires. A l'inverse, la méthylation en contexte non-symétrique CHG ou CHH n'est pas héritable selon ce modèle, ce qui explique qu'on la détecte

en conditions physiologiques principalement dans des cellules qui ne se divisent pas (ovocytes ou neurones).

Des travaux montrent que DNMT1 pourrait tout de même aussi être impliquée dans la méthylation *de novo* sur des régions non méthylées. En effet, les enzymes DNMT3A et DNMT3B (que je présente ci-après) sont capables de mettre en place de la méthylation sur de l'ADN non méthylé, mais uniquement sur l'un des deux brins. Ce serait donc en coopération avec DNMT1 que la méthylation pourrait se mettre en place sur les deux brins d'ADN (Kim et al., 2002 ; Fatemi et al., 2002). Enfin, il semblerait que DNMT1 puisse aussi participer à la méthylation *de novo* des régions répétées de l'ADN, en l'absence des autres DNMTs, et y serait recrutée par UHRF1 (Arand et al., 2012). De même, l'inactivation de DNMT1 dans les ovocytes en croissance au moment de l'acquisition de la méthylation de l'ADN engendre une baisse légère de la méthylation globale de l'ovocyte (Shirane et al., 2013). Puisque les ovocytes en croissance ne se divisent pas au moment d'acquérir la méthylation, cela suggère également une contribution légère de DNMT1 à la méthylation *de novo* de l'ADN dans ce contexte (Shirane et al., 2013).

2. DNMT2

DNMT2 ne mesure que 45kDa. Contrairement aux autres enzymes de la famille des DNMTs, elle a perdu son domaine N-terminal de régulation, mais a conservé le domaine catalytique d'activité méthylase (**Figure 6**). Cependant, son absence dans les cellules de la souris n'induit aucun effet sur la méthylation de l'ADN (Okano et al., 1998) et aucun effet phénotypique embryonnaire n'a été observé chez la souris *Dnmt2*^{-/-} (Goll et al., 2006). Cela a rapidement suggéré que DNMT2 n'est pas impliquée dans l'établissement ou le maintien de la méthylation de l'ADN génomique chez la souris. Par la suite il a été découvert que DNMT2 est capable de méthyler des ARN de transfert (Goll et al., 2006). Cette méthylation permet de protéger ces ARN des clivages par des ribonucléases (Schaefer et al., 2010). De récentes découvertes révèlent que cette méthylation permettrait aussi une régulation fine de la synthèse protéique (Tuorto et al., 2012 ; Tuorto et al., 2015).

3. DNMT3L

DNMT3L (DNMT3-like protein) possède une séquence homologue à celle de DNMT3A et de DNMT3B, mais ne possède pas de site catalytique ni de domaine PWWP (**Figure 6**). Elle ne peut donc pas méthyler l'ADN. DNMT3L est abondante dans les embryons précoces et les cellules germinales (Bourc'his et al., 2001a; Guenatri et al., 2013). Malgré l'absence d'activité catalytique, l'inactivation de DNMT3L chez la souris inhibe complètement l'établissement *de novo* de la méthylation de l'ADN dans les cellules germinales, y compris au niveau des gènes soumis à l'empreinte (voir *chapitre III, partie c*) (Bourc'his et al., 2001b; Kobayashi et al., 2012; Smallwood et al., 2011). De plus, DNMT3L est requise pour la méthylation et l'extinction des rétrotransposons dans les cellules germinales mâles (Bourc'his and Bestor, 2004). Il a été démontré par la suite que DNMT3L est un cofacteur qui interagit en complexe avec les méthyltransférases *de novo* DNMT3A et DNMT3B et stimule leur activité (Chédin et al., 2002; Holz-Schietinger and Reich, 2010; Jia et al., 2007; Suetake et al., 2004).

4. DNMT3A et DNMT3B

DNMT3A et DNMT3B sont, quant à elles, indispensables pour le développement embryonnaire et la mise en place de la méthylation de l'ADN. Ces enzymes ont une taille intermédiaire comparées à DNMT2 et DNMT1. En effet, elles avoisinent 100 et 130 kDa pour DNMT3A et DNMT3B respectivement, car elles possèdent un domaine N-terminal plus petit (**Figure 6**). DNMT3A est plutôt exprimée dans les stades embryonnaires tardifs et dans les cellules différenciées. A l'inverse, DNMT3B est plus présente dans les stades précoces du développement embryonnaire, et est la principale enzyme responsable de l'acquisition de la méthylation de l'ADN durant l'implantation embryonnaire (Borgel et al., 2010).

Les souris mutantes déficientes pour DNMT3A se développent normalement jusqu'à la naissance. Cependant elles apparaissent chétives quelques jours après la naissance et meurent quatre semaines plus tard (Okano et al., 1999). Les souris déficientes pour DNMT3B, quant à elles, se développent normalement jusqu'à 9.5

jpc, puis montrent de multiples défauts de développement embryonnaire et une croissance défectueuse après mi-gestation, ainsi que des anomalies au niveau du tube neural rostral. Ces embryons meurent à 15.5 jpc (Okano et al., 1999). Les embryons doubles mutants pour ces deux enzymes montrent une plus petite taille et une morphologie anormale entre 8.5 et 9.5 jpc (Okano et al., 1999). Leur croissance est stoppée après la gastrulation et ils meurent avant 11.5 jpc. De plus, le niveau de méthylation global des embryons double mutants à 9.5 jpc est équivalent à celui des blastocystes (3.5 jpc), ce qui démontre bien qu'en absence de ces deux enzymes, il n'y a pas de mise en place de la méthylation de l'ADN. Ces données phénotypiques montrent que DNMT3A et DNMT3B ont des fonctions vitales, et certainement partiellement redondantes, dans l'embryogénèse. En effet, Okano et ses collaborateurs ont suivi la méthylation de novo d'un ADN rétroviral inséré dans des cellules ES simples mutants pour chacune des deux enzymes ou doubles mutants *Dnmt3a*^{-/-} *Dnmt3b*^{-/-} (Okano et al., 1999). Alors que la méthylation de l'ADN rétroviral se met en place normalement dans les cellules simples mutantes, elle est complètement abolie dans les cellules doubles mutantes, ce qui démontre que ces deux enzymes ont des fonctions chevauchantes. Cela n'exclue pas qu'elles aient également des fonctions spécifiques. Par exemple, il a été montré que la méthylation des promoteurs à îlots CpG sur le chromosome X inactif implique uniquement l'action de DNMT3B dans les embryons de souris (Gendrel et al., 2012).

Enfin, il n'est pas à exclure que les enzymes DNMT3A et DNMT3B aient également une participation dans le maintien de la méthylation après la division cellulaire. Dans les cellules ES, il a été montré qu'une coopération entre les DNMTs est nécessaire pour la maintenance de la méthylation de l'ADN. En effet, les équipes de Peter Jones et En Li ont montré que l'inactivation ectopique de DNMT3A et DNMT3B entraîne une perte de méthylation au niveau de séquences de gènes uniques, des régions soumises à l'empreinte, ainsi que des séquences répétées de transposons, en dépit de la présence fonctionnelle de DNMT1 (Chen et al., 2003; Liang et al., 2002). D'autre part, l'équipe de J. Walter a également montré que dans les cellules ES de souris, alors que les gènes soumis à l'empreinte comme *Snrpn* et *H19* dépendent exclusivement de DNMT1 pour maintenir la méthylation de l'ADN, d'autres séquences comme le gène *Igf2* ou des éléments répétés requièrent une

coopération entre DNMT3A, DNMT3B et DNMT1 pour maintenir ces régions méthylées (Arand et al., 2012). Ces données suggèrent un mécanisme biphasique de propagation de la méthylation selon lequel l'action inefficace de DNMT1 à la fourche de réplication serait compensée par DNMT3A et DNMT3B qui viendraient par la suite compléter la méthylation des CpG hémiméthylés qui ont été manqués par DNMT1 (Jones and Liang, 2009) (**Figure 7**).

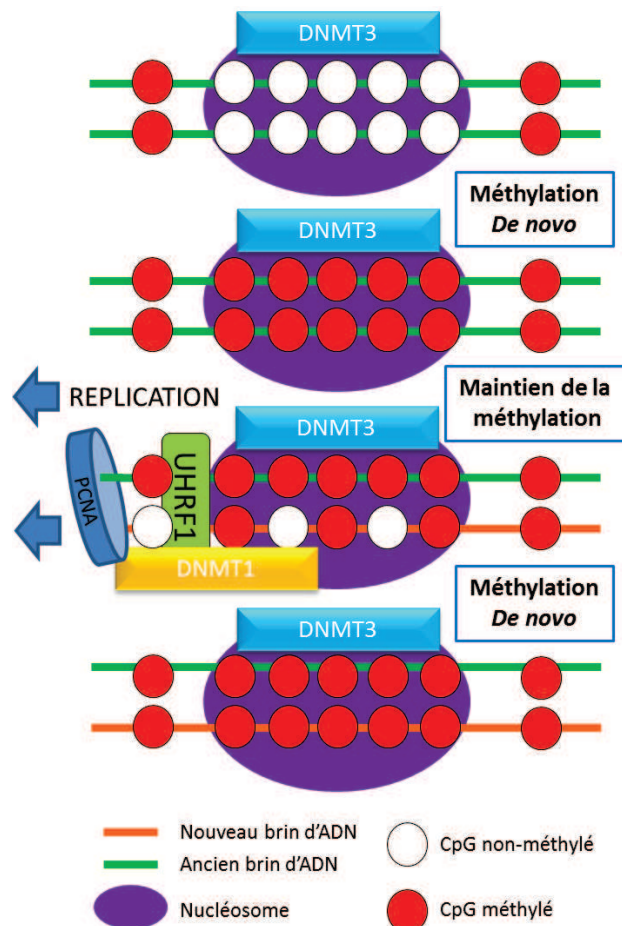


Figure 7 : Représentation schématique de la mise en place et du maintien de la méthylation de l'ADN.

Durant la méthylation *de novo* de l'ADN, la méthylation est mise en place par les enzymes DNMT3 à des régions spécifiques. Lors de la division cellulaire, DNMT1, recrutée par UHRF1 et PCNA, maintient la méthylation de l'ADN déjà en place en reconnaissant spécifiquement les CpG hémiméthylés. DNMT1 peut cependant manquer certains CpG lors de la réplication et ces oublis pourraient être compensés par les enzymes DNMT3 pour rétablir la méthylation de l'ADN après le passage de DNMT1. Adaptée de (Jones and Liang, 2009).

e) Les voies de déméthylation de l'ADN

La méthylation de l'ADN est une modification réversible. Durant la vie d'un individu, la méthylation de l'ADN subit deux grandes vagues d'effacement : une première après la fertilisation du zygote, et une seconde lors de la différenciation des cellules germinales. De plus, certains gènes peuvent être déméthylés lors de différenciations cellulaires. Les mécanismes de déméthylation de l'ADN sont très importants car ils reflètent la capacité de la méthylation de l'ADN à être une modification épigénétique réversible. Pour l'heure les mécanismes de déméthylation ne sont pas encore formellement élucidés et peuvent mettre en jeu des voies de déméthylation passive ou de déméthylation active (**Figure 8**).

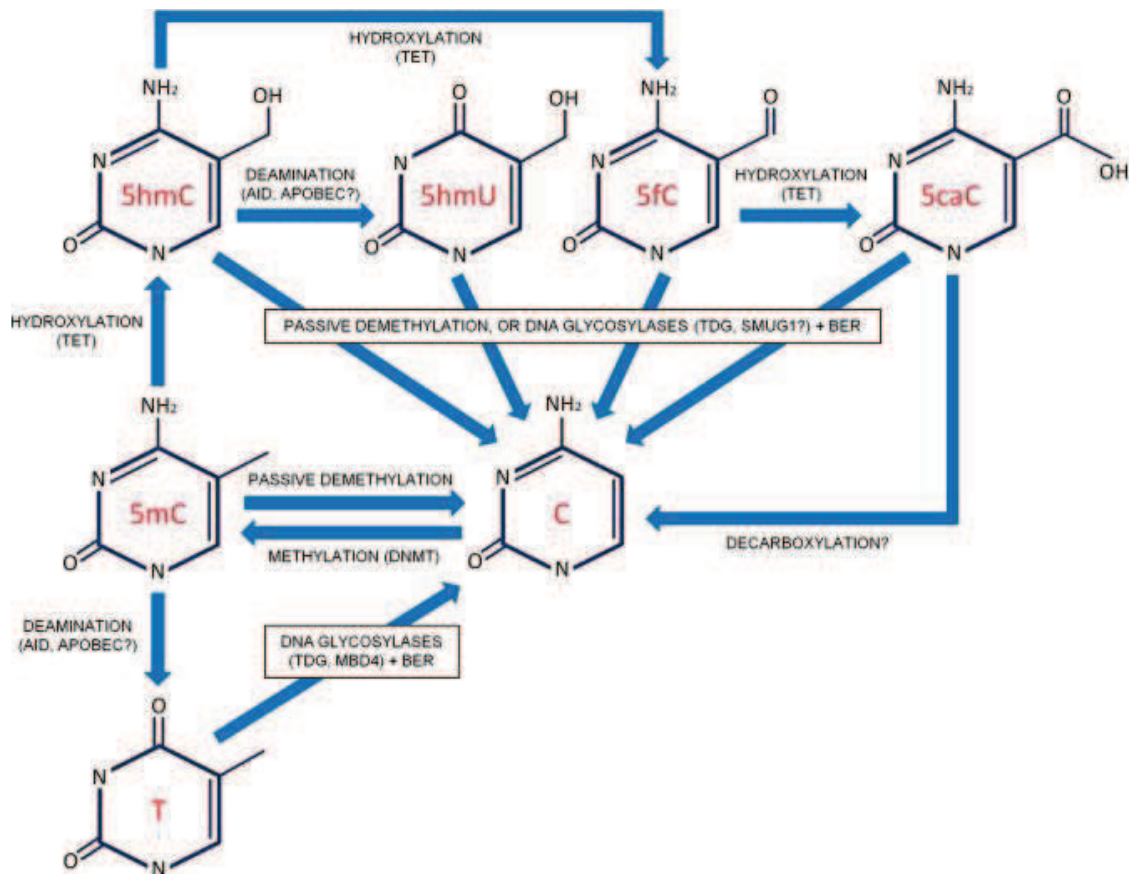


Figure 8: Représentation schématique des différentes voies chimiques de déméthylation des cytosines.

La 5-méthylcytosine (5mC) est maintenue au cours des divisions cellulaires grâce à DNMT1. Si cette maintenance n'est pas effectuée, la méthylation est progressivement perdue au cours des divisions cellulaires ; c'est la déméthylation passive. La méthylation peut être supprimée plus rapidement, de manière dite active, par désamination en thymine par AID ou APOBEC puis excision par des ADN glycosylases. La 5mC peut aussi être hydroxylée par une enzyme TET en 5-Hydroxyméthylcytosine (5hmC), 5-formylcytosine (5fC) et 5-carboxylcytosine (5CaC). La 5hmC peut être désaminée par AID/APOBEC pour former la 5-hydroxyméthyluracil (5hmU). Les cytosines modifiées 5hmC, 5hmU, 5fC et 5CaC pourraient être déméthylées de manière passive, ou excisées par des ADN glycosylases et le système de réparation BER pour être remplacées par des cytosines. Adaptée de (Auclair et Weber, 2012).

1. La déméthylation de manière passive

On parle de déméthylation passive lorsque la méthylation des cytosines n'est tout simplement pas maintenue au cours de la réplication de l'ADN. Ainsi, la présence de 5mC se dilue progressivement au cours des divisions cellulaires pour devenir absente dans les cellules restantes. Ce mécanisme semble contribuer grandement à la perte de méthylation globale du génome après la fertilisation ainsi que dans les PGCs (voir chapitre III, partie b.2).

2. La déméthylation de manière active

On parle de déméthylation active lorsque la déméthylation met en jeu des enzymes indépendamment de la réplication cellulaire. La suppression directe du groupement méthyl de la cytosine est une réaction thermodynamiquement non favorable, cependant d'autres voies enzymatiques ont été mises en évidence.

2.1 Les enzymes TET

Les enzymes TET (Ten-eleven translocation) catalysent l'oxydation de la 5mC en 5hmC (5-hydroxyméthylcytosine), 5fC (5-formylcytosine) et 5caC (5-carboxylcytosine) (**Figure 8**). La 5hmC a été découverte en 1972 (Penn et al., 1972) et est une marque épigénétique présente en quantité détectable sur l'ADN des Mammifères (Tahiliani et al., 2009). En effet, on trouve environ 0.1% de 5hmC dans les différents tissus de Mammifères (Globisch et al., 2010) mais jusqu'à 1% dans le cerveau de souris (Kriaucionis et Heintz, 2009).

Il existe actuellement trois protéines TET connues chez les vertébrés, TET1, TET2 et TET3. Les enzymes TET1 et TET2 sont très exprimées dans les cellules ES, mais ne semblent pas jouer de rôle important dans la pluripotence des cellules, ni dans le développement embryonnaire (Dawlaty et al., 2011). De même, les souris déficientes pour TET3 ne meurent qu'à la naissance (Gu et al., 2011). Par contre, des résultats ont montré que l'absence simultanée des trois enzymes TET bloque la

reprogrammation de fibroblastes en cellules pluripotentes induites (iPS) (Hu et al., 2014). Il se pourrait donc que les protéines TET aient des rôles redondants dans le contrôle du développement et de la pluripotence.

Ces dérivés d'oxydation 5hmC, 5fC et 5caC semblent être des intermédiaires de la déméthylation. Une des voies possibles est que les dérivés oxydatifs induisent la déméthylation passive. En effet les hemi-hydroxymethyls ne sont pas reconnus efficacement par DNMT1 (Valinluck et Sowers, 2007). Alternativement, les bases oxydées peuvent induire d'autres voies de déméthylation active impliquant par exemple les mécanismes de réparation de l'ADN. Les bases 5fC et 5caC peuvent être sélectivement reconnues et excisées de l'ADN par l'enzyme TDG (Thymine DNA Glycosylase), puis le système de réparation BER (Base Excision Repair) introduit une cytosine non modifiée pour compléter le cycle de déméthylation (He et al., 2011; Shen et al., 2013). De plus, des souris déficientes pour TDG montrent une hyperméthylation des gènes développementaux, conduisant à une létalité embryonnaire (Cortázar et al., 2011, Cortellino et al., 2011). En dehors du contexte des PGCs, cette déméthylation induite par les protéines TET a lieu essentiellement au niveau d'éléments de régulation distants tels que les enhanceurs (Lu et al., 2014).

2.2 Les enzymes AID/APOBEC

En dehors des enzymes TET, il a été proposé que la déméthylation active implique également les désaminases de la famille AID/APOBEC. Celles-ci peuvent directement désaminer les 5mC pour former des thymines. Les thymines forment alors un mésappariement G:T qui pourra ensuite être repéré puis excisé par le système BER (Cortellino et al., 2011). Aussi, elles peuvent désaminer les dérivés oxydés 5hmC en créant cette fois une uracile oxydée 5hmU. Celle-ci pourrait à son tour, être excisée par TDG ou par l'uracil-DNA glycosylase 1 (SMUG1) avant d'être remplacée par le système BER (Guo et al., 2011).

Il est possible que les cellules utilisent un ou plusieurs de ces systèmes afin de reprogrammer leur information épigénétique. Des études au niveau génomique de ces modifications à différents stades embryonnaires précoces, ou à différents stades

de différenciation des PGCs, nous permettraient d'éclaircir quels sont les différents mécanismes qui induisent la déméthylation à un endroit ou à un autre du génome.

f) Les protéines reconnaissant l'ADN méthylé

Il existe des protéines capables de se lier spécifiquement aux cytosines méthylées ; ce sont les MBPs (Methyl Binding Proteins). Ces protéines recrutent des corépresseurs qui peuvent modifier les histones et induire une répression chromatinienne. Il en existe trois catégories qui se distinguent par leur domaine de liaison à l'ADN méthylé : les protéines à domaine MBD (Methyl Binding Domain), les protéines qui reconnaissent l'ADN méthylé par un domaine en doigt de Zinc, et les protéines à domaine SRA (SET and Ring finger-associated). Ces protéines sont impliquées dans la diffusion de l'état transcriptionnellement inactif (Clouaire and Stancheva, 2008) (**Figure 9**).

1. Les protéines à domaine MBD

La première catégorie comprend les protéines utilisant un domaine de liaison aux CpG méthylés appelé MBD (Methyl Binding Domain), telles que MBD1, MBD2, MBD4 et MeCP2. MBD3 est, quant à lui, connu pour être un composant du complexe NuRD (Nucleosome Remodelling et histone Deacetylation), mais ne possède pas de domaine de liaison MBD fonctionnel (Zhang et al., 1999; Hendrich et Bird, 1998). MBD2 s'associe aussi au complexe NuRD (Clouaire and Stancheva, 2008). Ce complexe est impliqué directement dans la répression de la transcription dans un contexte d'ADN méthylé (Feng et Zhang, 2001). MBD4 est la seule protéine qui n'a pas un rôle de répression de la transcription, mais qui est un facteur de réparation du génome. En effet, MBD4 se lie et participe à la réparation des mésappariements T:G qui résultent de désaminations spontanées qui convertissent les 5mC en thymines (Wong et al., 2002). MeCP2 interagit à l'ADN méthylé et y recrute des histones deacetylases et methyltransferases (Fuks et al., 2003a; Jones et al., 1998). MBD1 interagit avec SUV39h1 et HP1 (Fujita et al., 2003). D'autres protéines possèdent des domaines MBD telle que l'histone méthyltransférase SETDB1, indiquant un rôle

potentiel de cette protéine en aval de la méthylation de l'ADN (Schultz et al., 2002). D'après leur structure, les protéines à domaine MBD peuvent interagir à tout type de séquences méthylées (Ohki et al., 2001). Une étude récente de leur distribution génomique dans les cellules ES a confirmé le peu de spécificité des MBD en montrant qu'elles se lient par défauts aux régions méthylées avec une relation linéaire entre la probabilité de liaison et la densité en 5mC (Baubec et al., 2013). Cette même étude a montré que MBD3 n'est pas enrichi au niveau des séquences méthylées, ce qui est compatible avec les données *in vitro* montrant que son domaine MBD muté n'est pas capable de se lier à l'ADN méthylé

2. Les protéines à domaine ZFP

La deuxième catégorie comprend des protéines ZFP (Zinc Finger Protein), protéines utilisant un domaine à doigt de Zinc, telles que Kaiso, ZBTB4 ou ZBTB38 (Filion et al., 2006; Klose et Bird, 2006 ; Daniel and Reynolds, 1999). Kaiso est un composant du complexe répresseur NCoR et son recrutement engendre l'hypoacétylation et la méthylation de l'histone H3K9, faisant de Kaiso un répresseur transcriptionnel dépendant de la méthylation d'ADN (Yoon et al., 2003). ZBTB4 et ZBTB38 répriment la transcription d'autres promoteurs de même dans un contexte où ce promoteur est méthylé (Filion et al., 2006; Klose and Bird, 2006). ZBTB38 interagit avec des complexes contenant des histones deacetylases et methyltransferases. La manière dont ZBTB4 réprime les promoteurs auxquels il peut s'associer est encore inconnue à ce jour. On sait cependant qu'il peut s'associer avec le complexe Sin3/HDAC afin de réprimer le promoteur du gène *p21* mais cette répression est toutefois indépendante de la méthylation de l'ADN (Weber et al., 2008).

3. Les protéines à domaine SRA

La troisième catégorie comprend les protéines à domaine SRA. Parmi elles UHRF1 (Np95 ou ICBP90) et UHRF2 (NIRF). UHRF1 co-localise aux origines de

réplication en fin de phase S et est requis pour la formation de l'hétérochromatine chez les Mammifères (Papait et al., 2007). UHRF1 peut s'associer avec PCNA et DNMT1 à ces sites de réplication afin de recruter spécifiquement DNMT1 à l'ADN hémiméthylé. La fonction d'UHRF2 reste encore à éclaircir. Dans les cellules ES, UHRF2 n'est pas capable de restaurer la méthylation perdue par l'absence de UHRF1 (Zhang et al., 2011). Une équipe suggère que UHRF2 pourrait participer à l'induction de l'apoptose sous le contrôle du facteur E2F1 (Lu and Hallstrom, 2013).

Malgré leur potentiel de liaison à l'ADN méthylé et de recrutement de co-répresseurs transcriptionnels comme des histones déacétylases et des histones méthyltransférases (Fuks et al., 2003a), le rôle de ces protéines dans la régulation transcriptionnelle n'est pas encore clairement défini. En effet, l'inactivation des protéines MBD seules ou en combinaison chez la souris n'induit pas de phénotype sévère (Divanlian et al., 1989). De plus bien d'autres protéines interagissant avec l'ADN méthylé ont été découvertes (Bartke et al., 2010) (Spruijt et al., 2013) mais leurs rôles dans la régulation épigénétique des gènes restent encore à déterminer.

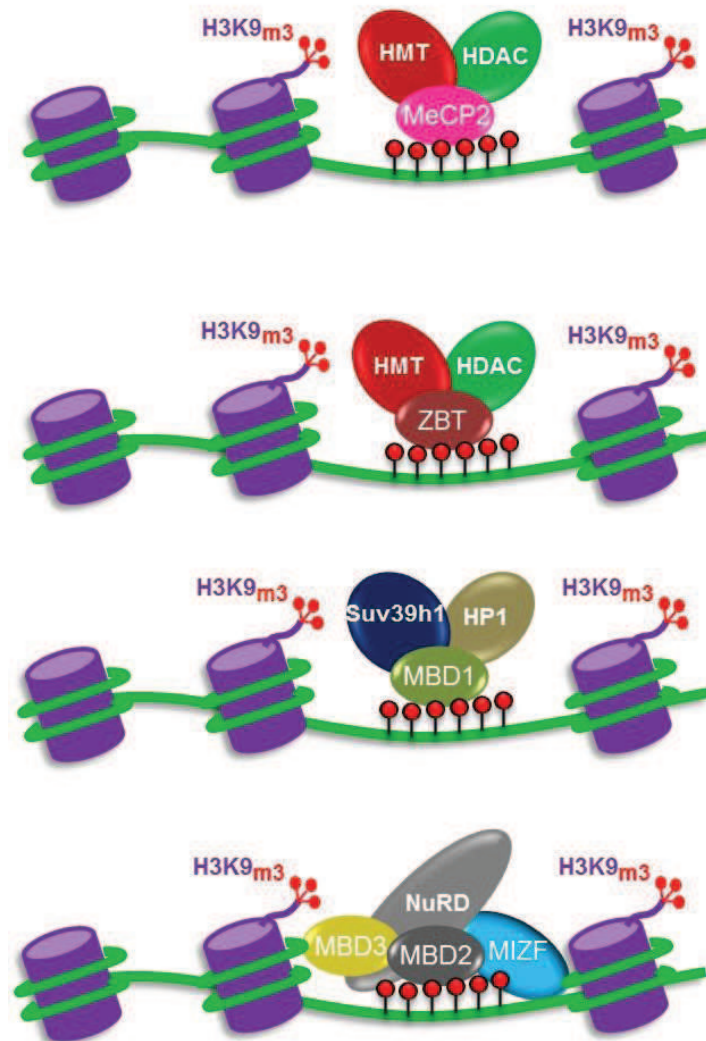


Figure 9: Modèle représentant schématiquement différentes protéines pouvant interagir avec l'ADN méthylé (cercles rouges).

La protéine MBD1 interagit à l'ADN méthylé et y recrute SUV39h1 qui va méthyler les histones au niveau de H3K9. Ces méthylations peuvent recruter les protéines HP1 pour obtenir un état inactif de la chromatine. MeCP2 et les protéines ZBT (ZBTB4 et Kaizo) peuvent interagir à l'ADN méthylé afin d'y recruter des histones méthyltransférases et histones désacétylases. MBD2 peut recruter le complexe NuRD qui contient des histones méthyl et désacétylases, ainsi que MBD3. MBD2 peut aussi interagir à MIZF qui reconnaît des séquences spécifiques de l'ADN. Adaptée de (Sekimata and Homma, 2004).

III) Dynamiques et fonctions de la méthylation de l'ADN

a) Cartographie de la méthylation de l'ADN

1. Techniques d'analyse de la méthylation de l'ADN de régions cibles

1.1 Par enzyme de restriction : COBRA

Le COBRA (Combined Bisulfite Restriction Analysis) est une technique permettant de visualiser si une région spécifique de l'ADN est méthylée. Cette technique combine la conversion au bisulfite de sodium avec une amplification PCR et une restriction enzymatique (Xiong et Laird, 1997) (**Figure 10**). Tout d'abord l'ADN génomique est traité avec du bisulfite de sodium qui converti les cytosines non méthylées en uraciles, alors que les cytosines méthylées sont résistantes à la conversion. Ensuite, une amplification par PCR est réalisée en ciblant la région d'intérêt. La PCR va diluer les uraciles et les remplacer par des thymines dans les séquences. Le produit de PCR est ensuite purifié puis digéré par une enzyme de restriction qui possède un site CpG dans son site de restriction, par exemple *TaqI* (T^ACGA). Cette enzyme ne va donc couper que si le CpG présent dans son site de restriction était initialement méthylé et donc résistant à la conversion en UpG lors du traitement au bisulfite de sodium. Après migration sur un gel d'agarose, on peut alors visualiser si le ou les sites de restriction présents sur la région amplifiée étaient méthylés ou non.

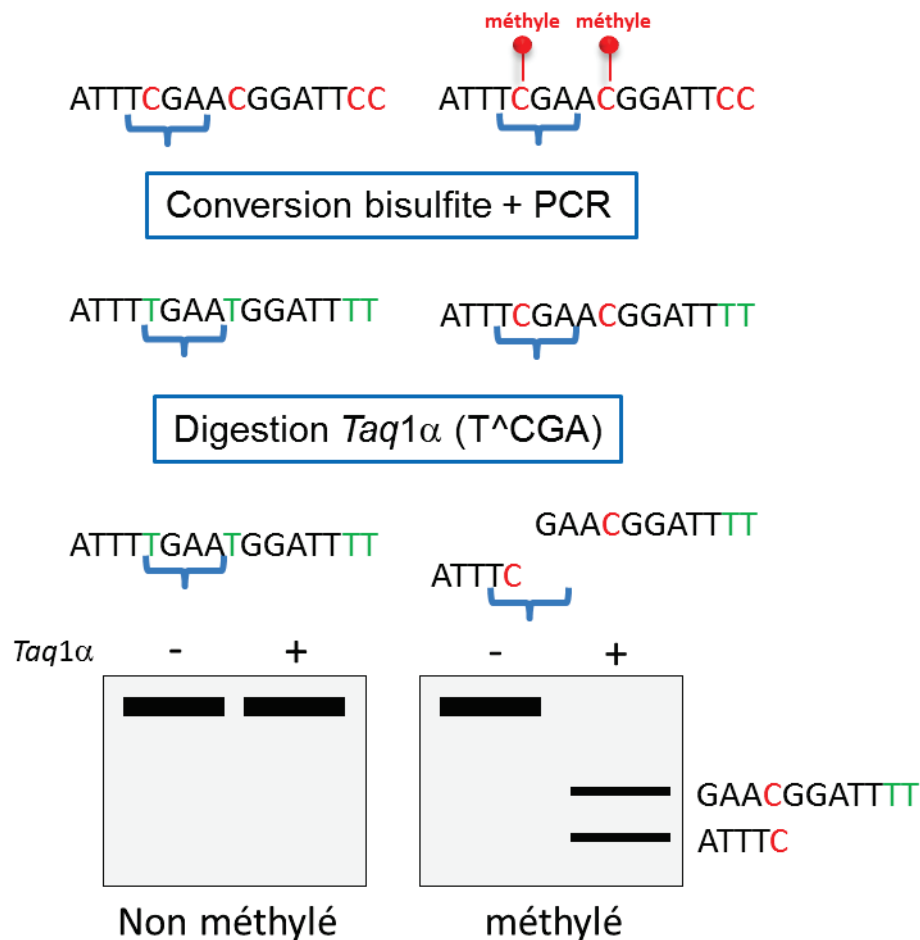


Figure 10 : Méthode d'analyse de la méthylation de l'ADN par COBRA.

Le COBRA consiste à convertir les cytosines non méthylées en T par traitement au bisulfite de sodium puis à digérer les sites TCGA par l'enzyme *Taq1*α. Si les cytosines des sites TCGA sont méthylées, les sites sont coupés. Si les cytosines ne sont pas méthylées, les sites ne sont plus présents et l'enzyme ne peut donc plus couper.

1.2 Par séquençage d'une région cible : BSseq

Le BSseq (Bisulfite-Séquençage) est une technique alternative au COBRA permettant de quantifier le taux de méthylation de la région amplifiée par PCR à l'échelle de cytosines individuelles (Frommer et al., 1992). Pour cela, la séquence amplifiée par PCR à partir de l'ADN converti au bisulfite de sodium est clonée dans un plasmide bactérien. Plusieurs colonies de bactéries différentes sont ensuite

récupérées et séquencées, permettant d'accéder à l'état de méthylation des CpGs du fragment amplifié à l'échelle de molécules individuelles en comparant les séquences obtenues avec la séquence de référence non convertie.

2. Techniques d'analyse de la méthylation de l'ADN à l'échelle du génome

Au cours des dix dernières années, de nouvelles techniques ont été développées permettant de cartographier la méthylation de l'ADN à l'échelle génomique. Ces cartographies permettent d'accéder à une information beaucoup plus complexe et permettent de découvrir de nouvelles régions différentiellement méthylées, mais aussi de suivre la reprogrammation de la méthylation sur le génome en fonction des stades de développement ou de changement environnementaux. Je présente ici en particulier deux techniques utilisées pour mes travaux de thèse, le MedIP (Methylated DNA immunoprecipitation) et le RRBS (Reduced Recombinated Bisulfite Sequencing).

2.1 MedIP (Methylated DNA Immunoprecipitation)

Cette technique consiste à soniquer l'ADN génomique puis à immunoprécipiter les fragments comportant de l'ADN méthylé avec un anticorps monoclonal anti-5 méthylcytosine (**Figure 11**). L'ADN ainsi immunoprécipité et l'ADN non traité (Input) peuvent être couplés à des fluorochromes différents et co-hybridés sur des puces à ADN, ou bien séquencés par séquençage à haut-débit. Cette technique offre une résolution à environ 250 paires de bases de la méthylation de l'ADN (Weber et al., 2005). Son inconvénient majeur est qu'elle est peu quantitative et 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 au nucléotide près (Rauch et al., 2008).

2.2 RRBS (Reduced Representation Bisulfite Sequencing)

Une autre technique, que j'ai beaucoup utilisée pendant ma thèse, est le RRBS (**Figure 11**). Cette technique consiste à réaliser une conversion au bisulfite de sodium du génome suivie d'une digestion par l'enzyme *MspI* (C^A CGG) et une sélection sur gel d'agarose des fragments de petites tailles qui permet d'enrichir en fragments génomiques riches en CpG. 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 ensuite alignés sur le génome de référence et la méthylation est estimée pour chaque CpG en rapportant le nombre de fragments séquencés contenant un CpG 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 (Meissner et al., 2005). 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. Une équipe a amélioré cette technique pour cartographier l'ADN méthylé à partir de cellules uniques (Guo et al., 2013).

2.3 WGBS (Whole Genome Bisulfite Sequencing)

Enfin, la technique de référence est le WGBS (Whole Genome Bisulfite Sequencing) qui permet de générer des cartes de méthylome du génome entier à la résolution du nucléotide (**Figure 11**). 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 une coupure du génome au hasard par sonication et un effort de séquençage bien plus important avec plus de 500 millions de lectures par échantillon contre 20 millions dans le RRBS. Cette différence fait du WGBS une

méthode plus intéressante car elle apporte bien plus d'éléments génomiques que le RRBS, mais son coût est aussi bien supérieur. Elle permet notamment de mesurer la méthylation sur les régions intra et intergéniques, qui sont innaccessibles avec la technique RRBS.

Récemment, cette technologie a été améliorée pour éviter la dégradation d'ADN « template » durant le traitement au bisulfite en réalisant ce traitement avant la ligation des adaptateurs (Miura et al., 2012). L'avantage est que cela permet d'appliquer le protocole à des quantités beaucoup plus réduites d'ADN de départ. Ce protocole a même été adapté pour pouvoir analyser la méthylation de l'ADN sur une cellule unique (Smallwood et al., 2014). Cette méthode permet de caractériser l'hétérogénéité cellulaire et d'explorer la méthylation sur des embryons une cellule par exemple.

Le désavantage majeur des technologies utilisant la conversion au bisulfite de sodium est qu'elles ne permettent pas de distinguer 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. C'est pourquoi des modifications ont été apportées au protocole de conversion au bisulfite pour pouvoir quantifier spécifiquement la présence de 5mC et 5hmC.

L'**oxBS-seq** (Oxidative Bisulfite Sequencing) permet d'analyser la présence des cytosines hydroxylées (5hmC) (**Figure 12**). 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 oxydation *in vitro* par les protéines TET (**Figure 12**). 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).

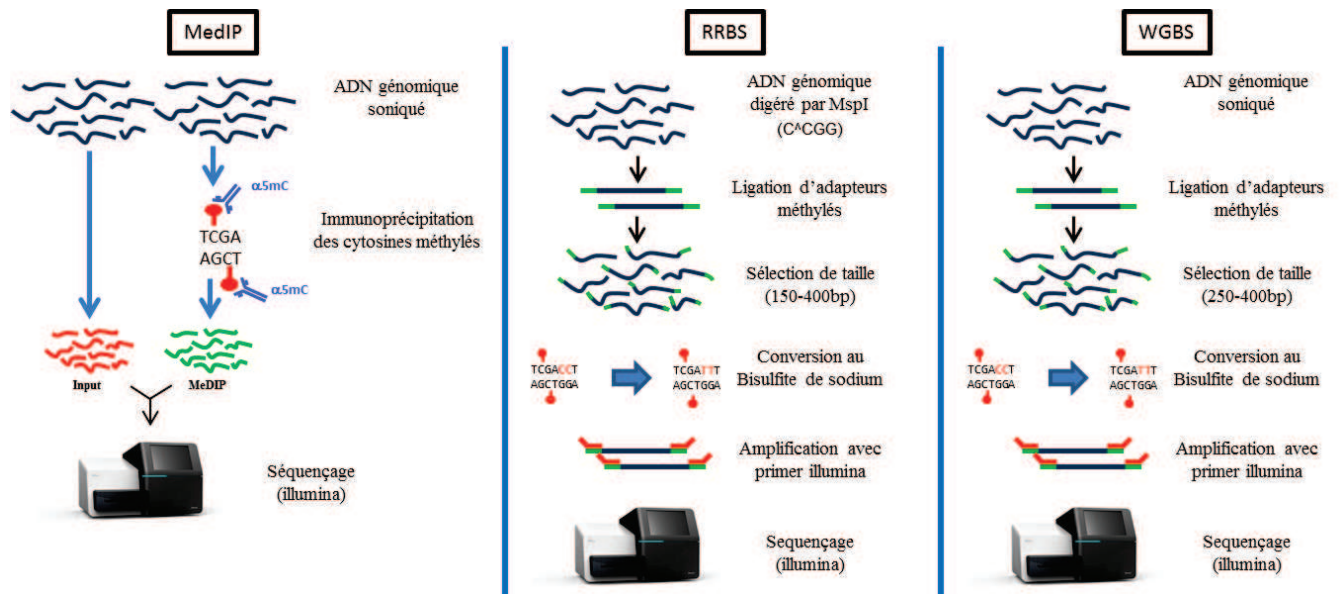


Figure 11 : Méthodes d'analyse du méthylome.

Le **MeDIP** consiste à soniquer puis à isoler l'ADN méthylé en utilisant des anticorps anti-5mC. Cet ADN et l'ADN total (Input) sont ensuite séquencés à haut débit pour obtenir une couverture totale du génome. Le **RRBS** consiste à couper l'ADN à l'aide d'une enzyme de restriction, dans des régions riches en CpG. Puis l'ADN génomique est converti au bisulfite de sodium, transformant toutes les cytosines non méthylées en Thymidines après amplification PCR. L'ADN ainsi amplifié est séquencé à haut débit pour obtenir une mesure de méthylation à la résolution du nucléotide sur la portion riche en CpG du génome. Le **WGBS** consiste à fragmenter l'ADN puis séquencer l'ADN converti au bisulfite de sodium sur l'ensemble du génome.

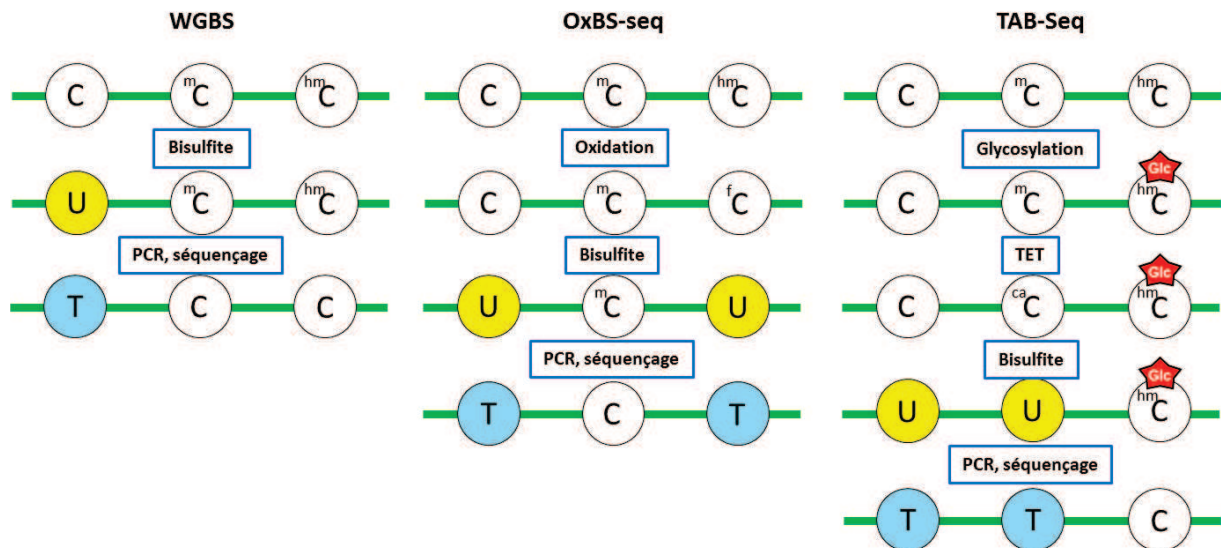


Figure 12 : Méthodes d'analyses de la 5mC.

Dans le protocole classique de WGBS, le **Bisulfite de sodium** transforme les cytosines non méthylées en U après traitement au bisulfite puis en T après PCR. Les 5mC et les 5hmC sont résistantes à la conversion et donc indiscernables l'une de l'autre au séquençage. Dans le protocole de bisulfite oxydatif (oxBS-Seq), les 5hmC sont transformées en 5fC par oxidation chimique puis en T après traitement au bisulfite et PCR. La présence de 5hmC peut ainsi être quantifiée par soustraction des résultats de WGBS et oxBS-Seq. Dans le protocole de TAB-Seq, les 5mC sont oxydées *in vitro* en 5CaC en présence de protéines TET. Les 5hmC sont préalablement sélectivement glycosylées afin de résister à l'action de la protéine TET. Les 5hmC seront donc les seules résistantes à la conversion au bisulfite de sodium et resteront des cytosines après PCR lors du séquençage.

b) Dynamique de la méthylation de l'ADN à l'échelle du génome

1. Répartition globale de la méthylation de l'ADN

Les techniques d'analyse du méthylome permettent désormais d'étudier la répartition de la méthylation des cytosines tout le long du génome. Elles permettent ainsi d'étudier la dynamique de la méthylation de l'ADN au cours des différenciations cellulaires ou en réaction à des stress environnementaux à l'échelle du génome. Ces techniques permettent donc une compréhension beaucoup plus importante du rôle et de la dynamique de la méthylation de l'ADN dans les génomes complexes des Mammifères.

Grace à ces cartographies, il fut révélé que la méthylation de l'ADN se trouve sur tout le génome, au niveau des régions intergéniques, codantes, répétées et de certains promoteurs. Ces études ont notamment permis de révéler que le niveau de méthylation des CpG est inversement proportionnel à la densité de CpG. En effet, les régions pauvres en CpG sont pour la plupart très méthylées alors que les îlots CpG restent non méthylés (Meissner et al., 2008). Cette relation inverse entre méthylation et densité en CpG a également été observée au niveau des promoteurs de gènes qui peuvent être catégorisés en trois classes : les LCP (Low CpG Promoter), les ICP (Intermediate CpG Promoter) et les HCP (High CpG Promoter). Les LCP sont fréquemment hyperméthylés, même dans le cas de gènes transcriptionnellement actifs (Weber et al., 2007). Il a été montré pour plusieurs LCP que la présence de faible densité de cytosines méthylées n'empêche pas la transcription et que les LCP subissent une déméthylation qui ne fait que succéder à l'activation de la transcription (Waterland et al., 2009 ; Métivier et al., 2008). En revanche, les HCP sont très rarement méthylés alors que les ICP ont une probabilité de méthylation intermédiaires entre les LCP et les HCP (Weber et al., 2007; Meissner et al., 2008).

Parallèlement, il a été montré que les variations de méthylation entre types cellulaires sont les plus fréquentes dans les régions intergéniques. Ces sites intergéniques correspondent le plus souvent à des régions activatrices distantes de

promoteurs, dites « enhancers » (Stadler et al., 2011; Sérandour et al., 2011 ; Hon et al., 2013) . Ces enhancers sont dans un état hypométhylé lorsqu'ils sont activés, ce qui a conduit à l'hypothèse que la méthylation de l'ADN participe à la régulation de l'activité des enhancers. Cependant, des études plus détaillées ont finalement montré que la déméthylation est un évènement secondaire à la fixation des facteurs de transcription (Stadler et al., 2011). Ces sites de méthylation intergéniques pourraient également réguler la transcription d'ARN non-codants qui peuvent réguler d'autres gènes (Illingworth et al., 2010).

Enfin, les expériences de cartographie à grand échelle du méthylome ont mis à jour une relation inattendue entre le niveau de méthylation intragénique et le niveau d'expression des gènes. En effet, un haut niveau de méthylation intragénique corrèle positivement avec un niveau élevé d'expression génique. Cette corrélation est très prononcée dans les ovocytes (Kobayashi et al., 2012) mais existe également dans les cellules somatiques (Lou et al., 2014; Varley et al., 2013). Un autre exemple parlant de cette corrélation est que la copie inactivée du chromosome X montre à la fois une hyperméthylation des promoteurs à îlots CpG mais également une hypométhylation des corps de gènes (Hellman et Chess, 2007).

2. Dynamique embryonnaire de la méthylation de l'ADN chez la souris

Au cours du développement embryonnaire chez les Mammifères, le génome subit plusieurs vagues de méthylation et de déméthylation (**Figure 13**). Une première vague de déméthylation a lieu durant le développement embryonnaire précoce, après la fertilisation et jusqu'au stade blastocyste. Une seconde vague de déméthylation a lieu dans les précurseurs de cellules germinales (PGCs, Primordial Germ Cells) entre 9.5 jpc et 13.5 jpc (Hemberger et al., 2009; Sasaki et Matsui, 2008). Ces déméthylations s'accompagnent d'importants remodelages de la chromatine, ce qui modifie les histones et déplace des variants d'histone, permettant à certains gènes d'être réprimés et à d'autres de s'exprimer.

2.1 De la fertilisation à l'implantation

Après fertilisation dans le zygote, le génome est globalement méthylé. Puis très vite, la méthylation baisse globalement jusqu'au stade blastocyste où le génome est globalement déméthylé (Smallwood et al., 2011; Smith et al., 2012) (**Figure 13**).

Le génome paternel est fortement méthylé dans le sperme, entre 80 et 90 % (Oswald et al., 2000). Il subit une déméthylation globale de manière active quelques heures après la fertilisation (Santos et al., 2002). Cette déméthylation rapide impliquerait l'oxydation par les protéines TET et le système de réparation BER (Oswald et al., 2000 ; Iqbal et al., 2011). En effet, dans le génome paternel, les 5mC disparaissent au profit de 5hmC (Gu et al., 2011), impliquant l'enzyme TET3 dans la déméthylation. De récentes découvertes ont cependant montré que la déméthylation du génome paternel ne serait pas uniquement due à TET3 (Guo et al., 2014a; Shen et al., 2014). En effet l'inhibition de la réplication de l'ADN bloque en partie la déméthylation du génome paternel malgré la présence de TET3. La réplication (système passif) est donc aussi impliquée dans la déméthylation du génome paternel (Shen et al., 2014). Dans le même sens que ces résultats, TET3 ne serait pas l'acteur principal de la déméthylation active. Dans des embryons à 3.5 jpc déficients en TET3, le niveau de méthylation atteint tout de même un niveau de méthylation proche d'un blastocyste WT (Inoue et al., 2015; Peat et al., 2014), ce qui suggère la présence d'autres systèmes de déméthylation.

Le génome maternel est méthylé à environ 45% (Howlett and Reik, 1991; Smallwood et al., 2011). Il est globalement déméthylé de manière passive par un non-maintien de la méthylation durant la réplication. Cette perte de méthylation dans le génome maternel est certainement due à l'exclusion de DNMT1 du noyau (Guo et al., 2014a). Cependant, de récentes découvertes montrent que le génome maternel semble aussi se déméthylé en partie de manière active, via la contribution de l'enzyme TET3 (Guo et al., 2014a; Shen et al., 2014).

Malgré la vague globale de déméthylation, certaines régions résistent tout de même à la déméthylation comme les régions de contrôle de l'empreinte (ICRs voir *chapitre III, partie c*) et des éléments transposables comme les LINE-1, LTR et IAPs

(Lane et al., 2003; Smith et al., 2012). Cela atteste d'une activité de maintien de la méthylation par DNMT1 ciblée à des endroits spécifiques du génome. Les mécanismes permettant de maintenir certaines régions méthylées sont encore assez obscures. Des évidences montrent un rôle de la protéine STELLA (PGC7) dans la protection contre la déméthylation. STELLA est capable d'interagir à la modification d'histone H3K9m2 et, en s'y fixant, elle protège l'ADN méthylé contre l'oxydation en 5hmC (Nakamura et al., 2012). Une autre protéine, ZFP57, est également requise pour maintenir la méthylation des ICRs durant ces stades préimplantatoires (voir chapitre II, partie c).

Quelle est le rôle de cette vague de déméthylation ? Une hypothèse est que la perte de méthylation permet aux gènes de facteurs de pluripotence comme *Oct4* (*Pou5f1*) et *Dppa3* (*Stella* ou *PGC7*) de s'exprimer, ce qui permet aux cellules de l'ICM (Inner cell Mass) de devenir pluripotentes.

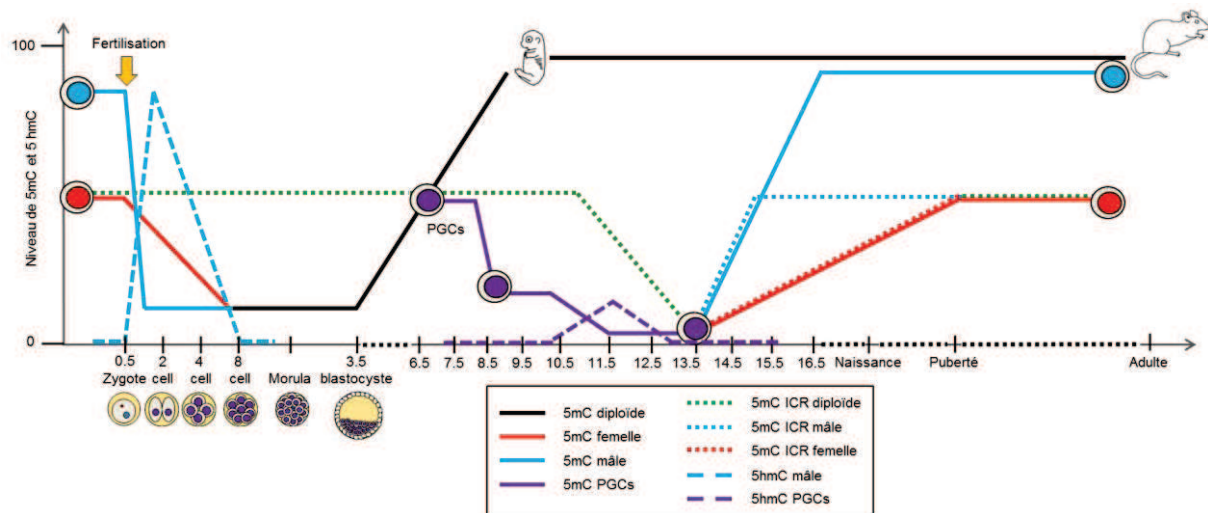


Figure 13: Représentation schématique des vagues de reprogrammation de la méthylation de l'ADN au cours du développement embryonnaire murin.

Après fertilisation, les génomes paternel et maternel se déméthylent 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, la méthylation de l'ADN reprend place sur tout le génome grâce aux enzymes DNMT3A et DNMT3B. Le génome restera ensuite méthylé dans les souris adultes. Seuls les PGCs (Primordial Germ Cells) contenues dans l'épiblaste, subissent une nouvelle vague de déméthylation entre 7.5 jpc et 13.5 jpc. Puis, ces cellules réacquièrent la méthylation en se différenciant en gamètes durant la méiose, et à des rythmes différents pour les gamètes mâles (en bleu) et les gamètes femelles (en rouge).

2.2 De l'implantation à la fin de la gastrulation

Après implantation du blastocyste à 4.5 jpc, le pool de cellules souches contenu dans l'ICM (Inner Cell Mass) va se différencier en un épiblaste qui contiendra la majorité des cellules souches embryonnaires qui seront la source de toutes les différentes lignées cellulaires de l'individu. Pendant cette phase, la méthylation de l'ADN se remet en place progressivement sur le génome jusqu'à la fin de la gastrulation à 8.5 jpc (**Figure 13**). En particulier, plusieurs gènes de pluripotence réacquièrent de la méthylation de l'ADN et s'éteignent dans l'embryon, ce qui stabilise la sortie de la pluripotence des cellules de l'embryon (Borgel et al., 2010; Feldman et al., 2006; Mohn et al., 2008). Ce processus requiert les ADN méthyltransferases *de novo* DNMT3A et DNMT3B, tandis que la maintenance qui en découle est toujours assurée par DNMT1 au cours des divisions cellulaires. Comme mentionné précédemment (voir chapitre IV, partie b), des embryons déficients pour DNMT3A et DNMT3B présentent une létalité après la gastrulation, ce qui suggère que la mise en place de la méthylation de l'ADN du génome au cours du développement est essentielle pour la morphogenèse. Plus tard dans les souris adultes, le niveau de méthylation est relativement stable d'un tissu à l'autre (Hon et al., 2013b).

2.3 Des PGCs aux cellules germinales

Une seconde vague de déméthylation se produit plus tard dans les cellules germinales (Saitou et al., 2012). Certaines cellules de l'épiblaste à 6.5 jpc vont former les précurseurs des cellules de la lignée germinale, les PGCs (Primordial Germ Cells). Ces PGCs vont former un groupe de cellules au stade 7.5 jpc dans l'épiblaste, puis vont rapidement proliférer et migrer le long de l'interface embryon/extra-embryon jusqu'aux gonades en développement pour coloniser les crêtes génitales à 10.5 jpc. Ensuite, elles vont continuer de proliférer jusqu'au stade 13.5 jpc, où elles atteignent leur plus bas niveau de méthylation (**Figure 13**).

Ces cellules sont caractérisées par l'activation de l'expression de marqueurs spécifiques des PGCs primordiaux que sont PRDM1 (PR domain zinc finger protein

1, appelé aussi BLIMP1), PRDM14, DPPA3 (appelé aussi PGC7 ou STELLA) et Ap2 γ à 6.5 jpc (Lane et al., 2003; Magnúsdóttir et al., 2013; Ohinata et al., 2005; Sato et al., 2002; Yamaji et al., 2008). Ces marqueurs spécifiques aux PGCs vont initier une cascade d'évènements qui incluent la répression de gènes somatiques, la ré-acquisition de la pluripotence et l'initiation de la reprogrammation épigénétique. PRDM1 par exemple va interagir avec l'arginine méthyltransferase PRMT5 pour éteindre les gènes somatiques dans ces cellules (Saitou, 2009). Cette reprogrammation va complètement transformer les cellules au niveau épigénétique entre 6.5 jpc et 13.5 jpc. Ces transformations comportent des changements de variants d'histones, des modifications d'histones et la perte de la méthylation de l'ADN, permettant un retour proche de la pluripotence de ces cellules (Hajkova et al., 2008; Seki et al., 2005).

Les cartographies de méthylome ont révélé que le taux global de méthylation baisse drastiquement dans les PGCs et que les cinétiques et mécanismes en jeu sont complexes. Il a été suggéré que les PGCs subissent en réalité deux phases de déméthylation : une première vague de manière globale lors de leur migration, et une seconde lors de leur entrée dans les gonades (Seisenberger et al., 2012). Les îlots CpG associés à des promoteurs de gènes de la lignée germinale, au chromosome X ou aux gènes soumis à empreinte semblent se déméthylent plus tardivement que le reste du génome entre 11.5 jpc et 13.5 jpc (Seisenberger et al., 2012). Les PGCs subissent un effacement global au niveau de tous les gènes. Le niveau de méthylation passe ainsi de 70% à 6.5 jpc à 30% dans les PGC à 9.5 jpc puis à moins de 10% dans les PGCs à 13.5 jpc (Seisenberger et al., 2012). Parmi les régions déméthylées se trouvent les promoteurs, les introns, les exons, les régions intergéniques, les îlots CpG et les éléments transposables (Guibert et al., 2012; Seisenberger et al., 2012; Lane et al., 2003; Lees-Murdock et al., 2003; Lees-Murdock et Walsh, 2008).

Plusieurs études suggèrent que la déméthylation dans les PGCs a une composante passive (Seisenberger et al., 2012, Ohno et al., 2013 ; Kagiwada et al., 2013). En effet durant cette phase, l'expression des gènes *Dnmt3a*, *Dnmt3b*, *Uhrf1*, *G9a* et *Glp* est réprimée, attestant d'une déméthylation globale qui est probablement passive (Ohno et al., 2013 ; Seki et al., 2007). La protéine UHRF1 détectée dans les

PGCs est localisée préférentiellement dans le cytoplasme (Seisenberger et al., 2012). De plus les PGCs sont caractérisées par une haute fréquence de sites CpG hémiméthylés aux stades 10.5 et 11.5 jpc (Ohno et al., 2013), ce qui est compatible avec une activité incomplète de DNMT1 aux fourches de réplifications. L'absence de GLP et G9a associée à la baisse de la modification H3K9m2 entre 6.5 jpc et 9.5 jpc (Ohno et al., 2013) pourrait aussi directement contribuer à la déméthylation passive de l'ADN car il a été suggéré que H3K9me2 stimule le recrutement du complexe UHRF1-DNMT1 à la chromatine (Liu et al., 2013; Rothbart et al., 2012). En revanche, la disparition de cette modification est suivie par une augmentation du niveau global de méthylation de H3K27 (Seki et al., 2005). L'équipe de Y. Matsui suggère que cette modification pourrait compenser l'absence de H3K9m2 pour continuer de réprimer certains gènes.

D'autres travaux suggèrent que la déméthylation dans les PGCs implique également des voies enzymatiques. En effet la désaminase APOBEC et les enzymes TET1 et TET2 sont exprimées dans les PGCs (Hajkova et al., 2010). Récemment, le rôle des enzymes TET dans la déméthylation des PGCs a été démontré. Les PGCs sont marquées par une accumulation transitoire de 5hmC (Hackett et al., 2013) et l'inactivation du gène *Tet1* induit une rétention de méthylation des cytosines sur les promoteurs riches en CpG de certains gènes méiotiques (Yamaguchi et al., 2013) et certains ICRs de loci soumis à empreinte (Yamaguchi et al., 2012). Les systèmes de réparation de la cellule tel que le BER ou le NER (Nucleotide Excision Repair) pourraient également être impliqués afin d'exciser les cytosines hydroxyméthylées produites. En effet, on sait que PARP1, APE1 et XRCC1, membres de ces systèmes, sont exprimés dans les PGC à 11.5 jpc (Hajkova et al., 2010).

Seuls certains éléments transposables comme les IAPs et LTR résistent à la déméthylation (Seisenberger et al., 2012), mais ils ne représentent que 7% des CpG. Il est probable que la méthylation de l'ADN soit maintenue au niveau de ces éléments afin d'éviter les effets délétères que pourrait avoir une réactivation de ces éléments mobiles dans les cellules germinales.

Après la fin de la reprogrammation épigénétique à 13.5 jpc, les PGCs entrent en prophase méiotique dans les gonades femelles, et au contraire, arrêteront leur

mitose dans les gonades mâles. Dans la lignée germinale mâle, l’empreinte paternelle va progressivement s’établir à partir de 14.5 jpc dans les spermatogonies en développement, et les éléments transposables vont réacquérir la méthylation (Lees-Murdock and Walsh, 2008; Li et al., 2004; Kato et al., 2007). A 16.5 jpc, la méthylation des PGCs mâles atteint 50% de méthylation alors qu’elle est encore à environ 10% dans les PGCs femelles (Seisenberger et al., 2012). Dans les cellules germinales femelles, la méthylation des ICRs ne s’établira qu’entre 10 et 25 jours après la naissance au moment où l’ovocyte entre en phase de croissance (Chotalia et al., 2009; Lucifero et al., 2004).

Toutes ces études montrent que la méthylation de l’ADN n’est que peu stable d’un point de vue héritabilité transgénérationnelle. L’héritabilité épigénétique se réfère à la transmission de variations épigénétiques stables au-delà d’une génération sans changements de la séquence de l’ADN (Heard et Martienssen, 2014). Ces effets passent d’une génération à l’autre en l’absence de l’effet les ayant mis en place dans la première génération sur le fœtus ou les cellules germinales de la seconde génération (Skinner, 2008). Certains éléments transposables semblent être les seules séquences dont la méthylation peut être transmise entre générations chez les Mammifères (Lane et al., 2003). Des variations stables de méthylation au niveau de l’un ou plusieurs de ces éléments pourraient influencer le phénotype en modifiant par exemple l’expression de gènes avoisinants.

c) Régulation de l’empreinte génomique parentale

1. Une expression monoallélique en cluster

Certains gènes, découverts dans les années 1980 chez les Mammifères, ont une expression monoallélique et dépendante de l’origine parentale de l’allèle (McGrath and Solter, 1984; Surani et al., 1984; DeChiara et al., 1991; Bartolomei et al., 1991). Ce phénomène a été nommé empreinte génomique parentale et est régulé en grande partie par la méthylation de l’ADN. Cette expression mono-allélique

dépend de la méthylation de l'ADN situé sur des régions dites de contrôle de l'empreinte génomique (ICR = Imprinting Control regions) (Neumann and Barlow, 1996). Elles sont différenciellement méthylées au niveau des deux allèles (DMRs = Differentially Methylated Regions), selon que l'allèle provienne du père ou de la mère. Cette méthylation différentielle est établie lors de la gamétogénèse dans les lignées germinales parentales, puis maintenue pendant la fusion des gamètes et durant le développement embryonnaire, avant d'être réinitialisée dans les cellules germinales (Arnaud, 2010). Il est intéressant de remarquer que certaines DMRs sont méthylées sur l'allèle exprimé et d'autres sur l'allèle non exprimé. On sait aussi que les ICRs permettent d'établir au cours du développement des méthylations dites « secondaires » au niveau de régions promotrices appartenant à un même cluster, souvent distantes, riches en CpG et méthylées sur un allèle (Reik and Walter, 2001).

Environ 80% des gènes sous contrôle d'une empreinte parentale sont regroupés en clusters de plusieurs centaines de kilobases, suggérant une régulation commune de l'ensemble de ces gènes (**Figure 14**). Ils sont impliqués dans de nombreux processus du développement embryonnaire tels que la formation du placenta ou la croissance fœtale. Chez l'Homme, des défauts de méthylation dans ces régions peuvent causer des syndromes de malformation tels que le syndrome de Prader-Wili/Angelman ou encore le syndrome Beckwith-Wiedemann (voir *chapitre III, partie e*).

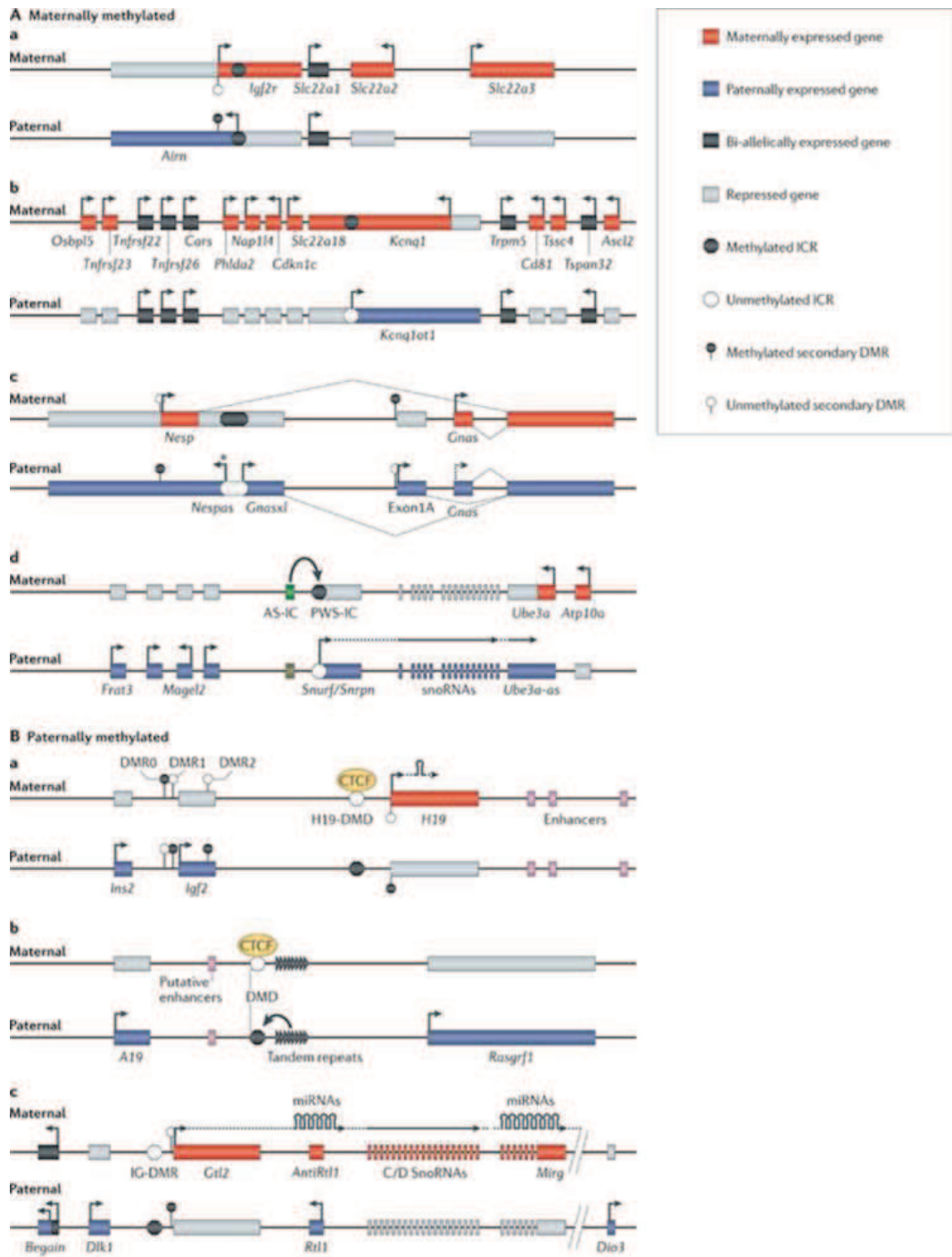


Figure 14 : Représentations schématiques de quatre clusters de gènes soumis à l’empreinte régulés par des ICR méthylés sur le chromosome maternel et trois clusters régulés par des ICR méthylés sur le chromosome paternel.

A.a. Cluster *IGF2R*. **A.b.** Cluster *Kcnq1*. **A.c.** Cluster *Gnas*. **A.d.** Cluster *Snrpn*. **B.a.** Cluster *IGF2-H19*. **B.b.** Cluster *Rasgrf1*. **B.c.** Cluster *Dlk1-Dio3*. Adaptée de (Ferguson-Smith, 2011).

2. La régulation des ICRs

Les ICRs peuvent réguler l'expression monoallélique des gènes soumis à empreinte de différentes manières. 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 14**).

Dans des cellules ES déficientes en DNMT1, il a été observé que les ICRs perdent leur méthylation et il s'en suit une perte d'expression mono-allélique de la majorité des gènes soumis à l'empreinte (Casparly et al., 1998; Tucker et al., 1996). Ces résultats ont pu être confirmés dans des embryons déficients en DNMT1 (Hirasawa et al., 2008). De plus, des résultats obtenus dans les cellules ES et dans des souris déficientes pour les différentes DNMT montrent aussi que DNMT1 semble être l'enzyme la plus importante pour propager la méthylation des ICRs dans les cellules embryonnaires (Arand et al., 2012; Okano 1999 ; Oda et al., 2006; Hirasawa et al., 2008). Enfin, il a été montré que la mise en place de la méthylation des ICR maternelles et paternelles dans les gamètes implique DNMT3A (Arima et al., 2006; Kaneda et al., 2004), et DNMT3L y agit comme cofacteur (Bourc'his et al., 2001b; Hata et al., 2002).

d) Inactivation du chromosome X

Chez les femelles mammifères, les cellules possèdent deux chromosomes X. Afin d'équilibrer l'expression du Chromosome X entre les individus mâles et femelles, une copie du chromosome X est inactivée dans les cellules femelles. Cette inactivation est établie très tôt durant l'embryogénèse, entre 3.5 et 8.5 jpc, et est vitale pour la survie de l'organisme (Takagi and Abe, 1990). Elle est aussi associée à l'abandon de la pluripotence et est mise en place par l'action d'un ARN non-codant appelé Xist (Augui et al., 2011). L'ARN *Xist* s'active sur l'un des deux chromosomes X et induit une cascade d'évènements épigénétiques en *cis* qui incluent le recrutement de protéines de la famille Polycomb, de modifications d'histones, du

variant d'histone macroH2A 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 DNMTs sur la mise en place et/ou le maintien de la méthylation de l'ADN du chromosome X inactif n'est pas encore éclairci. Une étude récente sur quelques gènes candidats a suggéré que DNMT3B seule est responsable de la mise en place de la méthylation des îlots CpG sur le chromosome X inactif (Gendrel et al., 2012). D'autre part, Il a été montré que l'acquisition de la méthylation de l'ADN sur les îlots CpG est un évènement tardif qui arrive après l'inactivation des gènes (Prissette et al., 2001). De plus, l'inactivation d'un chromosome X a lieu normalement dans des embryons déficients pour les *de novo* méthyltransférases DNMT3A et DNMT3B (Sado et al., 2004). Ceci suggère 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.

e) Pathologies liées à des perturbations de la méthylation de l'ADN

La méthylation de l'ADN, lorsqu'elle est anormalement mise en place sur le génome, peut déréguler l'expression de certains gènes et les fonctions cellulaires, et donc favoriser l'apparition de pathologies. Il est également important de souligner que la méthylation de l'ADN permet de réprimer les éléments répétés et de maintenir la stabilité du génome. Les maladies liées à la méthylation de l'ADN peuvent être de causes génétiques ou environnementales. Il en existe plusieurs connues à ce jour.

1. Syndromes développementaux

Le syndrome ICF (Immunodeficiency, Centromeric instability, Facial anomalies) est une maladie autosomale récessive causée par des mutations hypomorphes du gène *Dnmt3b* (Hansen et al., 1999; Xu et al., 1999). Il est associé à une instabilité génomique (Smeets et al., 1994) et une hypométhylation de plusieurs régions satellites péri-centromériques dont DNMT3B est responsable (Portela and

Esteller, 2010; Okano et al., 1999). Cette maladie est associée, comme son nom l'indique, à une immunodéficiência, et des anomalies faciales. Cependant les mutations dans le gène *Dnmt3b* ne sont retrouvées que dans 50% des cas de syndromes ICF. D'autres formes du syndrome ICF sont liées à des mutations dans d'autres gènes tels que *Zbtb24* (30% des cas) (de Greef et al., 2011; Nitta et al., 2013), *Cdca7* ou *Hells* (Thijssen et al., 2015). Il reste à déterminer si ces gènes ont un lien fonctionnel avec DNMT3B et la méthylation de l'ADN.

Un autre exemple, le syndrome Rett, est dû à des mutations dans le gène *MeCP2* (methyl CpG binding protein 2) sur le chromosome X (Wan et al., 1999; Amir et al., 1999) (voir chapitre II, partie f.1). Ces mutations conduisent à des problèmes de développement du cerveau (Willard and Hendrich, 1999). Il a été suggéré que le mécanisme impliquerait des dérégulations géniques causées par un défaut de la protéine MECP2 à lier l'ADN méthylé, en premier lieu dans les contextes non-CpG (Chen et al., 2015; Gabel et al., 2015).

Le syndrome du X fragile est associé à une absence de la protéine FMR1 (fragile X mental retardation 1). Cette protéine possède, au niveau de son promoteur, une région répétée de 29 CGG non transcrite. Cette région est méthylée et permet de réguler l'expression de *FMR1*. Chez les mutants, cette séquence est répétée 200 à 600 fois, ce qui a pour conséquence une hyperméthylation anormale de la région promotrice, et une inactivation constante du gène (Kremer et al., 1991; Gheldof et al., 2006). Cette maladie est associée à un déficit intellectuel léger à sévère qui peut être associé à des troubles du comportement et à des signes physiques caractéristiques.

Il existe aussi des syndromes causés par des méthylations anormales au niveau de centres de contrôle de l'empreinte parentale, comme les syndromes de Prader-Willi et d'Angelman. Ces deux syndromes impliquent une méthylation aberrante de la région soumise à empreinte q11-q13 du chromosome 15. Cette modification épigénétique engendre une perte d'expression paternelle des gènes de cette région dans le syndrome de Prader-Willi, alors qu'elle engendre une perte de l'expression maternelle du gène *Ube3a* (Ubiquitin-protein ligase E3A) dans le syndrome d'Angelman (Portela et Esteller, 2010). Le syndrome de Prader-Willi est caractérisé par une hypotonie sévère à la naissance et des difficultés alimentaires.

Le syndrome d'Angelman est quant à lui un trouble sévère du développement neurologique.

Enfin, la neuropathie HSAN (Hereditary Sensory and Autonomic Neuropathy) de type IE, implique des mutations hétérozygotes du gène *Dnmt1* (Klein et al., 2011). Les mutations sont concentrées dans les exons 20 et 21 qui codent pour le domaine RFTS (Replication Foci Targeting Sequence) de la protéine qui lui permet d'interagir aux fourches de réplication. Elles engendrent une hypométhylation globale du génome et des hyperméthylations locales (Sun et al., 2014). Cette maladie est associée à une perte de sensibilité à la douleur et aux températures dans les membres inférieurs, une perte d'audition, des mutilations ulcéreuses et de la démence (Auer-Grumbach et al., 2012; Houlden et al., 2006).

2. Les cancers

La dérégulation de la méthylation de l'ADN peut aussi être impliquée dans le développement de cancers. Tous les types de cancers présentent des perturbations du méthylome caractérisées à la fois par des pertes et des gains de méthylation. L'hypométhylation peut engendrer l'activation anormale de gènes, conduisant à un potentiel oncogénique. Aussi, l'hypométhylation de séquences répétées peut conduire à l'expression de ces éléments pouvant engendrer une instabilité génomique (Timp et Feinberg, 2013). A l'inverse, l'hyperméthylation d'ilots CpG conduit fréquemment à l'inactivation de gènes suppresseurs de tumeurs ou de gènes impliqués dans l'homéostasie cellulaire.

Dans tous les cancers, on trouve des anomalies de profils de méthylation de l'ADN. Cependant, il est encore difficile aujourd'hui de déterminer si certains de ces changements de méthylation sont une cause ou une conséquence de processus de transformation cellulaire. Un argument en faveur du rôle causal de la méthylation de l'ADN dans certains cancers est l'observation d'une forte fréquence de mutations dans les gènes contrôlant la méthylation de l'ADN. C'est le cas notamment dans plusieurs types de leucémies où des mutations de *Dnmt3a* et de *Tet2* y ont été retrouvées à haute fréquence (Walter et al., 2011; Yan et al., 2011 ; Chan and Majeti, 2013; Couronné et al., 2012; Delhommeau et al., 2009; Langemeijer et al., 2011)

Au-delà du débat sur le rôle causal de la méthylation de l'ADN, les chercheurs ont mis en place de nombreux tests cliniques qui utilisent désormais la méthylation de l'ADN comme un marqueur prédictif de cancer.

Par exemple, l'hypométhylation du gène suppresseur de tumeur *Brca1* (Breast Cancer 1) est un marqueur prédictif de cancer du sein (Wong et al., 2011). Beaucoup d'autres marqueurs ont été mis en évidence comme étant hypo ou hyperméthylés dans des tumeurs du sein (Radpour et al., 2009). Trouver des biomarqueurs indiquant la présence de premières lésions est très important pour prédire la possibilité d'un développement de cancer et en particulier pour ceux du pancréas qui sont foudroyants. L'hyperméthylation du promoteur de *Nptx2* (Neuronal Pentraxin II) est par exemple un marqueur de cancer pancréatique (Park et al., 2012).

Les cancers colorectaux montrent aussi d'importantes aberrations de méthylation de l'ADN. Ces méthylations aberrantes seraient impliquées dans les stades précoces de la progression du phénotype malin (Lao et Grady, 2011; Mitchell et al., 2014). On peut y trouver des hypométhylations de certaines régions du génome comme les éléments répétés (Antelo et al., 2012), et le gène soumis à empreinte *Igf2* (Insulin-like Growth Factor 2) (Naito et al., 2014). On y trouve aussi des hyperméthylations locales d'îlots CpG situées au niveau de régions promotrices (Irizarry et al., 2009). Alors qu'il est estimé que chaque tumeur du colon contient en moyenne une centaine de mutations de séquences codantes de gènes, l'hyperméthylation d'îlots CpG affecte également plusieurs centaines de gènes dans certaines tumeurs du colon, illustrant que les mutations épigénétiques (« épimutations ») sont aussi fréquentes que les mutations génétiques. Les cancers gastriques, quant à eux, ont clairement été montrés comme étant des maladies dont les facteurs sont multiples. Au niveau environnemental, la bactérie *Helicobacter pylori* a été reconnue comme pathogène pouvant affecter la méthylation de plusieurs gènes dont *Cdh1* (Cadherin 1), et de nombreux éléments répétés et transposables (Alvarez et al., 2013; Hudler, 2012). Certains adénocarcinomes gastriques présentent des hyperméthylations d'une centaine de gènes en particulier (Balassiano et al., 2011; Loh et al., 2014).

L'identification de ces marqueurs épigénétiques est très importante mais aussi très difficile, et les méta-analyses semblent être une solution pour les détecter (Sapari et al., 2012). D'importants efforts doivent être consentis pour identifier les

marqueurs de méthylation les plus pertinents et développer leur utilisation en tests diagnostiques et prédictifs pour détecter précocement les risques de cancers ou mieux orienter les thérapies.

IV) Modes de ciblage de la méthylation de l'ADN

Chez les Plantes, le système a été bien analysé. On sait que l'ADN méthyltransférase CMT3 (CHROMOMETHYLASE 3, qui méthyle l'ADN dans un contexte uniquement non-CpG) peut directement interagir avec la queue N-terminale de l'histone H3, mais seulement quand elle est simultanément méthylée sur H3K9 et H3K27 (Lindroth et al., 2004). Les résultats suggèrent que les méthylations d'histones H3K9 et H3K27 permettent de créer un code histone pour le recrutement de CMT3 à certains loci de l'ADN à méthyler. Les résultats suggèrent aussi une boucle de renforcement propre entre la méthylation de H3 et la méthylation de l'ADN. Si l'une des deux modifications est absente, CMT3 ne peut atteindre la région cible. Par ailleurs les plantes possèdent également un système de recrutement de la méthylation de l'ADN par des petits ARN interférents (Matzke and Mosher, 2014).

Actuellement chez les Mammifères, le ou les mécanismes permettant de recruter la méthylation de l'ADN au niveau du génome ne sont pas clairs. La réponse à cette question est très importante car elle nous permettrait de comprendre comment les remodelages s'effectuent lors de la différenciation cellulaire, et quelles sont les causes de défauts de méthylation pouvant mener à diverses maladies et cancers. Il existe actuellement plusieurs théories impliquant des ARN, les histones modifiées, ou les facteurs de transcription.

a) Rôles des ARN

La classe des Piwi-Interacting ARN, dits piRNAs, est une classe d'ARN de 24 à 32 nucléotides très abondants dans les cellules germinales. Dans ces cellules, le

génomique est globalement hypométhylé et les transposons peuvent y être transcrits en piARN primaires (Aravin et al., 2008). Ces ARN interagissent alors avec des protéines MILI qui les guident au niveau des transcrits anti-sens complémentaires. S'en suit un clivage du duplex ARN, qui produit des piARN secondaires. Ces piARN secondaires interagissent ensuite avec les protéines MIWI2 (Kuramochi-Miyagawa et al., 2008), lesquelles vont ensuite utiliser les piARN secondaires pour cibler et cliver les transcrits sens. Ce processus engendre une amplification des piARN primaires et une dégradation des ARN provenant des transposons. Par ailleurs, il a été montré que les piRNA sont nécessaires pour la méthylation *de novo* des éléments transposables dans les cellules germinales. On sait que la suppression des protéines MILI ou MIWI2 a un impact négatif sur la méthylation *de novo* des promoteurs des éléments répétés LINE et LTR (Carmell et al., 2007; Kuramochi-Miyagawa et al., 2008; Molaro et al., 2014). Très récemment, l'équipe d'A. Aravin a examiné en détails la contribution des deux protéines dans l'établissement de la méthylation de l'ADN dans les spermatogonies et a montré que MILI contribue à la méthylation d'un plus grand nombre de transposons de type LINE-1 ou LTR que MIWI2 (Manakov et al., 2015). D'autres études doivent être menées pour comprendre si des petits ARNs non-codant pourraient participer au ciblage de la méthylation de l'ADN à d'autres stades du développement embryonnaire.

De grands ARN non codants sont aussi impliqués dans la régulation épigénétique. En effet, des ARN anti sens de transcrits comme *Air* et *Kcnq1ot1* sont requis pour la méthylation de l'ADN de régions différentiellement méthylées (DMRs) de gènes soumis à empreinte parentale (Reik et Walter, 2001). Même si leur implication n'est qu'indirecte, elle n'en est pas moins indispensable. D'autre part, il a été montré que la transcription d'ARN messagers peut influencer sur la méthylation de leur propre gène. Dans l'ovocyte, les DMRs des locus *Gnas* et *Snrpn* ont ainsi besoin de la transcription des ARN messagers du gène qui les englobe pour acquérir de la méthylation (Chotalia et al., 2009; Smith et al., 2011). A l'échelle génomique, il a été confirmé que la transcription des gènes stimule fortement la méthylation dans les séquences intragéniques des ovocytes en croissance (Kobayashi et al., 2012; Smallwood et al., 2011). A ce stade, il reste à déterminer si la méthylation de l'ADN est stimulée par les molécules d'ARN ou bien par l'action de transcription et les

modifications d'histones associées comme la méthylation de H3K36 (voir ci-dessous). Ainsi en conclusion, à l'instar des Plantes, les ARN chez les Mammifères semblent aussi jouer d'importants rôles dans la régulation de la méthylation de l'ADN.

b) Rôles des modifications d'histones

Aujourd'hui, il est clair que la présence de méthylation de l'ADN dans le génome mammifère corrèle avec la présence de certaines histones modifiées, alors que l'absence de méthylation de l'ADN corrèle avec la présence d'autres histones modifiées. Le lien entre la méthylation de l'ADN et ces modifications n'est pas encore clairement établi.

1. **Les modifications d'histones de chromatine active**

Chez les Eucaryotes, la méthylation de l'histone H3 sur la lysine 36 (H3K36) est une modification mise en place par SETD2 (Edmunds et al., 2008). Elle est localisée de manière intragénique dans les gènes transcriptionnellement actifs (Blackledge et al., 2010; Kolasinska-Zwierz et al., 2009). Cette marque corrèle avec la présence de méthylation d'ADN. Le domaine PWWP de DNMT3A peut interagir avec H3K36me, ce qui souligne la possibilité d'un lien fonctionnel entre la méthylation de H3K36 et la méthylation des cytosines (Dhayalan et al., 2010) (**Figure 15.A**). Ce lien a été renforcé par le travail récent de l'équipe du Dr Dirk Schübeler qui a montré dans les cellules ES murines que l'enzyme DNMT3B est recrutée dans les corps de gènes transcrits et que ce recrutement est inhibé en l'absence d'un domaine PWWP fonctionnel ou de l'enzyme SETD2 (Baubec et al., 2015). Cependant en l'absence de SETD2, la perte de H3K36m3 dans des cellules cancéreuses, n'induit pas de diminution du niveau de méthylation sur les exons (Hahn et al., 2011). Le domaine PWWP pourrait donc reconnaître une autre modification d'histone.

La méthylation de l'histone H3K4 est une marque associée aux promoteurs de gènes et aux îlots CpG non méthylés. Plusieurs études ont montré que cette marque anti-corrèle très fortement avec la méthylation de l'ADN au niveau des îlots CpGs (Meissner et al., 2008), ce qui suggère qu'elle pourrait jouer un rôle protecteur contre la méthylation de l'ADN (**Figure 15.A**). En effet, des études biochimiques ont montré que la méthylation de H3K4 inhibe la liaison de DNMT3A à la chromatine via le domaine ADD (Zhang et al., 2010). De même, une autre équipe a montré que DNMT3L interagit avec l'histone H3 mais que cette interaction n'est possible que si l'histone H3 n'est pas méthylée au niveau de la lysine 4 (Otani et al., 2009). Enfin, le retrait de la modification d'histone H3K4me par l'enzyme déméthylase KDM1B (Lysine (K) specific demethylase) est indispensable pour la mise en place de la méthylation des DMRs des gènes soumis à empreinte dans les oocytes (Ciccone et al., 2009).

2. Les modifications d'histones de chromatine inactive

Les méthylations sur les lysines 27 (H3K27) et 9 (H3K9) sont des marques d'une chromatine condensée et transcriptionnellement inactive (Greer et Shi, 2012). On retrouve ainsi les marques H3K9me2 et H3K9me3 sur les rétrotransposons ERVs par exemple (Mikkelsen et al., 2007). Chez le champignon *Neurospora crassa* et la plante *Arabidopsis thaliana*, la méthylation de l'ADN est dépendante de la présence de la modification d'histone H3K9me (Jackson et al., 2002; Lindroth et al., 2004; Tamaru and Selker, 2001; Tamaru et al., 2003).

Chez les Mammifères, la méthylation de H3K9 a également été proposée comme étant un signal de recrutement de la méthylation de l'ADN. En effet, il a été montré que H3K9me2 est reconnue par UHRF1, le cofacteur de DNMT1, ce qui suggère un modèle selon lequel la présence de H3K9me2 stimule la méthylation de l'ADN en favorisant le recrutement du complexe UHRF1-DNMT1 (Liu et al., 2013; Rothbart et al., 2012) (**Figure 15.B**). De plus, des études ont montré que G9a, une histone méthyltransférase de H3K9, est impliquée dans le recrutement de la méthylation de l'ADN dans les cellules ES, même si il semblerait que cela soit indépendant de son activité catalytique (Dong et al., 2008; Epsztejn-Litman et al.,

2008; Tachibana et al., 2008) (voir chapitre V). Inversement, des études suggèrent que le complexe UHRF1/DNMT1 contribue à la maintenance des histones H3K9 méthylées (Estève et al., 2006; Kim et al., 2009). De plus, UHRF1 peut ubiquitinyler H3K23 et cette modification facilite le recrutement de DNMT1 via son domaine RFTS, lui permettant ainsi de méthyler l'ADN (Nishiyama et al., 2013) (**Figure 15.B**).

La marque d'histone H3K9m3 quant à elle, catalysée par les enzymes SUV39h1 et SUV39h2 au niveau de l'hétérochromatine péri-centrique (Peters et al., 2001), montre aussi un lien avec la méthylation de l'ADN. Dans des cellules ES déficientes pour ces deux enzymes, la méthylation de l'ADN est altérée au niveau de certaines séquences répétées de régions péri-centriques et il a été montré que DNMT3A et DNMT3B interagissent directement avec HP1 α qui lui-même interagit avec H3K9m3 grâce à son chromodomaine (Lehnertz et al., 2003) (**Figure 15.C**). On sait aussi que grâce à leur domaine ADD, les DNMT3 peuvent interagir avec SUV39h1 et SETDB1 (Fuks et al., 2003b; Li et al., 2006). Enfin, des expériences *in vitro* ont aussi montrée que DNMT1 peut interagir avec les protéines HP1 qui se lient avec H3K9 triméthylé, suggérant que les protéines HP1 pourraient jouer un rôle de pont moléculaire entre la méthylation des histones sur H3K9 et la méthylation de l'ADN (Smallwood et al., 2007).

La modification H3K27m3 est quant à elle associée à des régions réprimées par le complexe protéique polycomb. Cette marque est catalysée par Ezh1/2 de PRC2. Elle se localise majoritairement au niveau d'ilots CpG non méthylés dans les cellules ES. Il semblerait donc que ces deux modifications soient exclusives (Statham et al., 2012). Cependant, on en retrouve tout de même au niveau de certains ilots CpG qui acquièrent de la méthylation lors de la différenciation des cellules ES (Mohn et al., 2008). La méthylation de l'ADN pourrait donc être initiée par le complexe PRC qui recrute les DNMT (Viré, 2006). Un autre résultat va à l'encontre de ces premières évidences, en effet il a été montré que PRC2 peut s'associer avec TET1 dans les cellules ES (Neri et al., 2013). Le lien entre H3K27m3 et la méthylation de l'ADN reste donc à éclaircir.

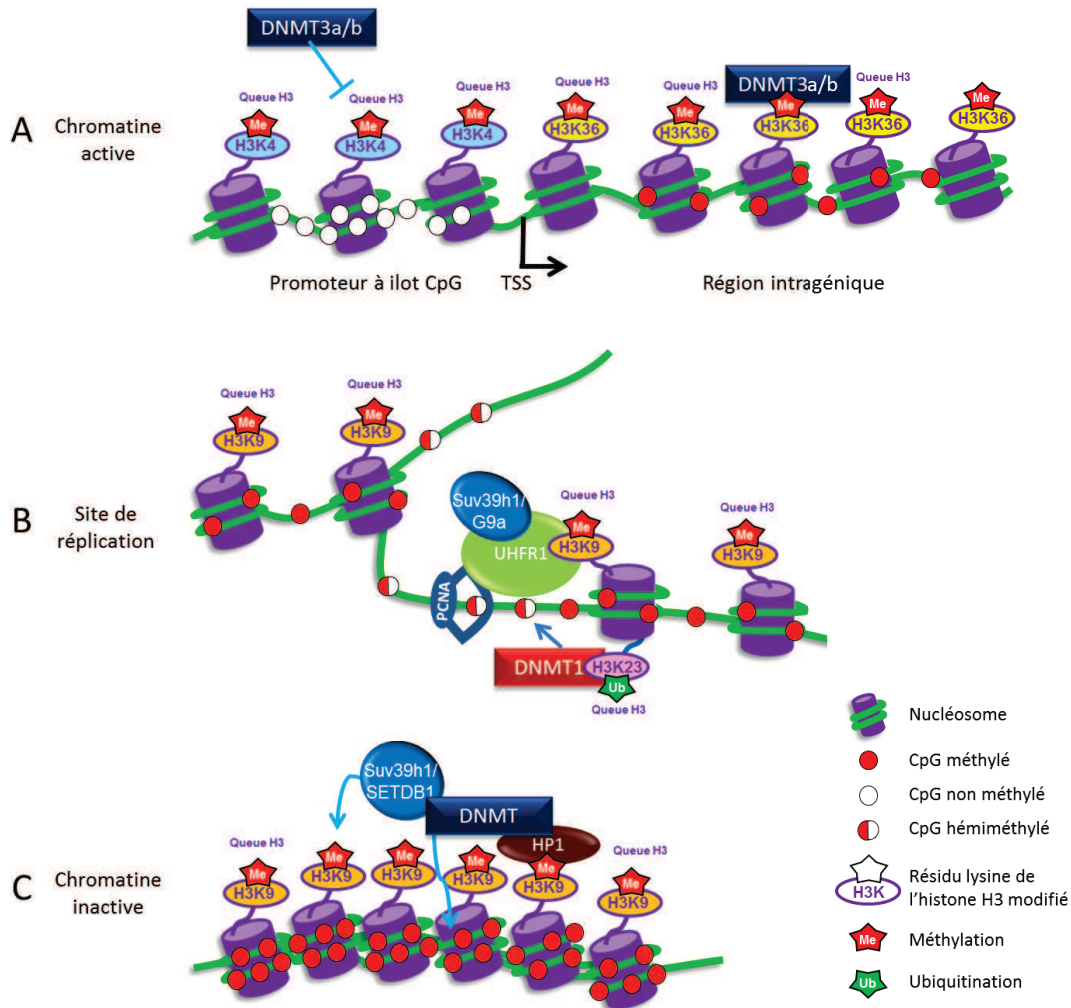


Figure 15: Modèles schématiques du lien entre la méthylation de l'ADN et les modifications d'histones.

A : Les gènes transcrits sont caractérisés par la méthylation des histones H3K4 aux promoteurs et H3K36 aux régions intragéniques. Le recrutement des enzymes DNMTs à la chromatine est inhibé par la présence de H3K4 méthylées. En revanche les DNMTs interagissent avec les H3K36 méthylées grâce à leur domaine PWWP.

B : Aux sites de réplication, UHRF1 reconnaît PCNA, l'ADN hémiméthylé et les modifications d'histone H3K9m2 et H3K9m3. UHRF1 ubiquitinye H3K23, créant ainsi une plateforme de recrutement pour DNMT1 qui ajoute un groupement méthyle au brin néosynthétisé. Les protéines SUV39h1 et G9a interagissent avec UHRF1 afin de propager les modifications d'histones H3K9m2 durant la réplication.

C : Les régions d'hétérochromatine péri-centriques sont caractérisées par la triméthylation de H3K9. Les protéines HP1 reconnaissent H3K9m3 et servent de « pont » afin de recruter les enzymes DNMT3 qui peuvent méthyler l'ADN. Les protéines SUV39H1 et SETDB1 reconnaissent ce complexe afin de méthyler les H3K9 alentours.

c) Rôles des facteurs de transcription

Même dans le cas où des modifications d'histone recrutent la méthylation de l'ADN, comment les enzymes de modification des histones sont-elles recrutées au niveau de certains gènes plutôt qu'à d'autres? Une des hypothèses possibles est que des facteurs de transcription reconnaissant des motifs spécifiques de l'ADN soient impliqués pour initier la cascade de recrutement de la méthylation de l'ADN (**Figure 16**). Des équipes ont montré en effet qu'il est possible d'observer un lien direct entre la séquence de l'ADN et la méthylation de l'ADN. Par exemple de petites mutations SNP (Single-Nucleotide Polymorphisms) peuvent suffire pour influencer le recrutement de la méthylation de l'ADN (Kerkel et al., 2008). Les facteurs de transcription pourraient ainsi être les premiers acteurs de la mise en place de la méthylation de l'ADN, en recrutant des enzymes modificatrices d'histones, et/ou en recrutant directement les DNMT aux régions cibles.

1. E2F6, un facteur recrutant la méthylation de l'ADN

E2F6 est l'un de ces facteurs de transcription potentiellement impliqué dans le recrutement de la méthylation de l'ADN (**Figure 16.A**). Membre de la famille des facteurs E2F, il est capable de se fixer à une séquence consensus de l'ADN : 5'-TTTCCCGC-3' (Cartwright et al., 1998). Les souris mutantes déficientes pour E2F6 sont viables et fertiles mais montrent des transformations homéotiques qui, de manière remarquable, sont identiques à celles observées dans des souris porteuses d'une mutation hypomorphe de *Dnmt3b* (Courel et al., 2008 ; Velasco et al., 2010). Au début de la différenciation cellulaire des cellules souches pluripotentes, les gènes de la lignée germinale sont rapidement éteints. Une équipe a montré que le facteur de transcription E2F6 est justement recruté au niveau des promoteurs de ces gènes chez la souris (Kehoe et al., 2008). De plus des modèles cellulaires tels que les cellules MEFs (Mouse Embryonic Fibroblasts) et les cellules ES montrent que l'inactivation d'E2F6 entraîne une réactivation de quatre gènes de la lignée germinale normalement hyperméthylés (Kehoe et al., 2008; Storre et al., 2005; Velasco et al.,

2010; Laseva et al., 2013 ; Storre et al., 2005). Enfin, il a aussi été montré q'E2F6 favorise le recrutement de l'enzyme DNMT3B au niveau de ces gènes (Velasco et al., 2010).

2. Les facteurs protégeant l'ADN de la méthylation

Un autre modèle possible est que des facteurs de transcription soient recrutés afin de protéger l'ADN de la méthylation. Par exemple le facteur Sp1, se fixant à la séquence 5'-(G/T)GGGCGG(G/A)(G/A)(C/T)-3' a été proposé comme l'un de ces facteurs (**Figure 16.B**). En effet, des expériences de transgénèse avec la séquence du promoteur du gène *Aprt* mutée au niveau des sites SP1 montrent que SP1 est nécessaire pour empêcher la méthylation du transgène (Brandeis et al., 1994). Plus récemment, des expériences d'insertions de séquences dans les cellules ES de souris ont confirmé l'importance des motifs de reconnaissance de facteurs de transcription dans la protection contre la méthylation de l'ADN (Krebs et al., 2014). Cependant, l'inactivation du gène *Sp1* chez la souris n'a pas d'effet sur la méthylation du gène *Aprt* endogène au stade 9.5 jpc (Marin et al., 1997), ce qui remet en question ce modèle dans le contexte de la chromatine endogène.

Aussi, au niveau des ilots CpG non méthylés, la protéine CFP1 (CXXC finger protein 1) se lie à 5'-(A/C)CG(A/C)-3'. Elle y recrute des enzymes modificatrices d'histones H3K4 qui, en méthylant ces histones, permettent de maintenir l'état non méthylé de l'ADN (Deaton et Bird, 2011). Il est cependant important de souligner qu'une fois encore, l'inactivation de *Cfp1* dans des cellules ES n'induit pas de changement majeur de méthylation de l'ADN (Clouaire et al., 2012). Ces résultats suggèrent que les mécanismes de protection des ilots CpG de la méthylation de l'ADN sont complexes et potentiellement redondants.

3. Les facteurs protégeant l'ADN de la déméthylation

A l'inverse, un autre exemple est le facteur de transcription ZFP57 (Zinc Finger Protein 57) (**Figure 16.C**). Il a été montré que ce facteur interagit directement à l'ADN méthylé au niveau du motif 5'-TGCCGC-3' et qu'il est nécessaire pour protéger les ICRs des gènes soumis à empreinte de la vague de déméthylation dans les embryons précoces (Li et al., 2008). Une équipe a montré que dans les cellules ES, ZFP57 interagit avec l'ADN méthylé et y recrute TIF1 β (KAP1 ou TRIM28) qui peut alors recruter les DNMTs au niveau des ICRs. Ils observent aussi qu'en absence de TIF1 β ou de ZFP57, la modification d'histone H3K9m3 diminue fortement à ces régions. Ces résultats suggèrent ainsi que ce facteur permet non seulement de recruter les DNMTs, mais aussi les enzymes capables de modifier les histones sur H3K9 (Quenneville et al., 2011).

Ces différents facteurs de transcription capables de recruter les enzymes responsables de la méthylation, de protéger l'ADN de la méthylation ou encore de la maintenir, peuvent co-exister dans une même cellule et agir en systèmes indépendants. Cependant, ces systèmes complexes sont encore très mal compris à ce jour.

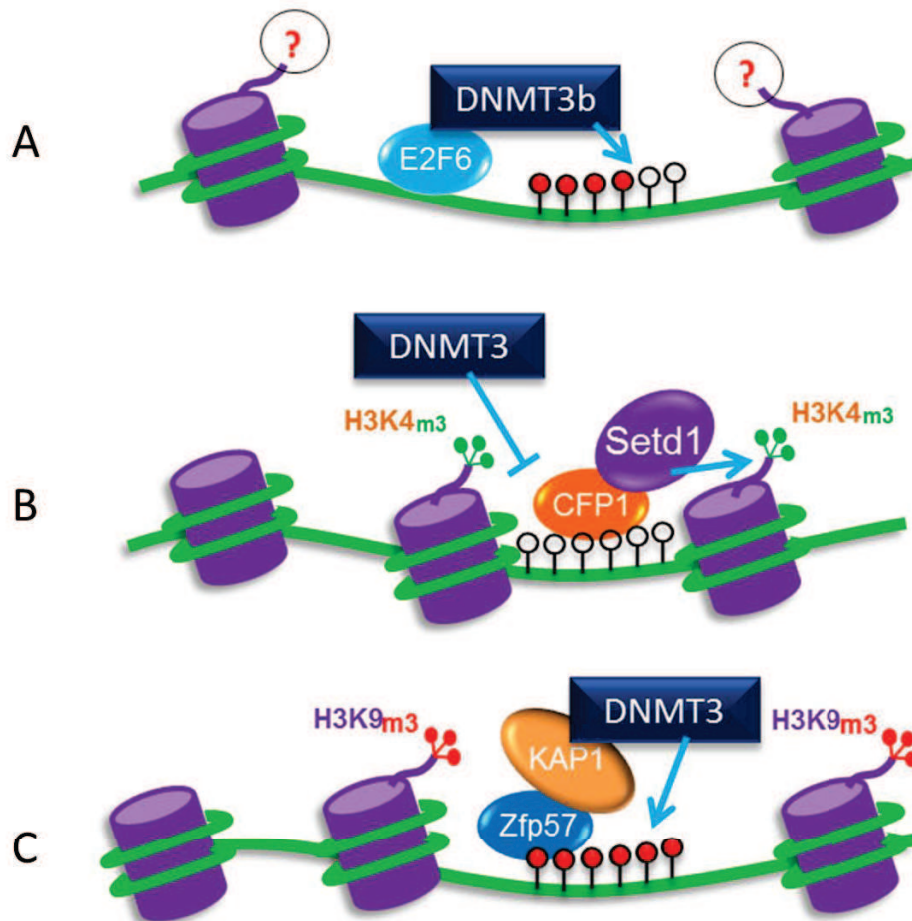


Figure 16: Modèle représentant schématiquement différents exemples d'interactions entre facteurs de transcription et méthylation de l'ADN.

A : Le facteur E2F6 interagit au niveau d'une séquence spécifique de l'ADN et pourrait y recruter DNMT3b. **B :** CFP1 interagit à l'ADN non méthylé et peut y recruter Setd1 qui, en méthylant les histones H3K4, pourrait contribuer à empêcher le recrutement des DNMT3. **C :** ZFP57 interagit à l'ADN méthylé au niveau des ICRs de certaines régions soumises à l'empreinte. ZFP57 recrute KAP1 et les DNMT, contribuant ainsi à maintenir l'état méthylé.

4. Autres possibilités de ciblage de la méthylation

D'autres voies moléculaires pourraient être concomitantes à celles décrites ci-dessus et participeraient au recrutement de la méthylation de l'ADN. L'une des idées est que seules les régions accessibles de l'ADN ou contenant des histones pourraient être méthylées. Cela impliquerait que certaines hélicases/ATPases pourraient influencer les profils de méthylation de l'ADN en libérant les cibles des DNMTs à des endroits spécifiques, comme cela a été observé pour les hélicases de la famille SNF2 telles que ATR-X et LSH (Dennis et al., 2001, Gibbons et al., 2000).

V) Rôle de G9a dans la régulation de la méthylation de l'ADN

Chez les Mammifères, on trouve cinq enzymes capables de méthyler la lysine 9 de l'histone H3 : G9a, GLP, SETDB1, SUV39h1 et SUV39h2. Elles font toutes parties de la grande famille moléculaire des protéines à domaine SET qui permet de catalyser la méthylation des histones. Dans le génome de la souris, on trouve plus de cinquante protéines possédant un domaine SET (Kouzarides, 2002). SETDB1 est essentielle pour le développement précoce de l'embryon murin de par son activité de catalyse de la triméthylation de H3K9 (Dodge et al., 2004). Cette enzyme est impliquée dans différents processus biologiques. En particulier il a été montré qu'elle est nécessaire pour réprimer les rétrovirus de type LTR dans les cellules ES via H3K9m3 et le recrutement de TIF1b (Matsui et al., 2010). SUV39h1 et son homologue SUV39h2 permettent quant à elles la triméthylation de H3K9 au niveau de l'hétérochromatine péri-centrique (Peters et al., 2001). Ces enzymes sont également des régulateurs potentiels de la méthylation de l'ADN. En effet, dans des cellules ES déficientes pour ces deux enzymes, la méthylation de l'ADN est altérée au niveau de certaines séquences répétées de régions péri-centriques (Lehnertz et al., 2003). Dans cette partie, je me concentre sur G9a et son partenaire GLP ainsi que leur interaction potentielle avec la méthylation de l'ADN.

a) G9a et son homologue GLP

G9a est essentielle pendant la gestation murine (Tachibana et al., 2002) car en son absence, les souris meurent entre 9.5 et 10.5 jpc (**figure 17**).

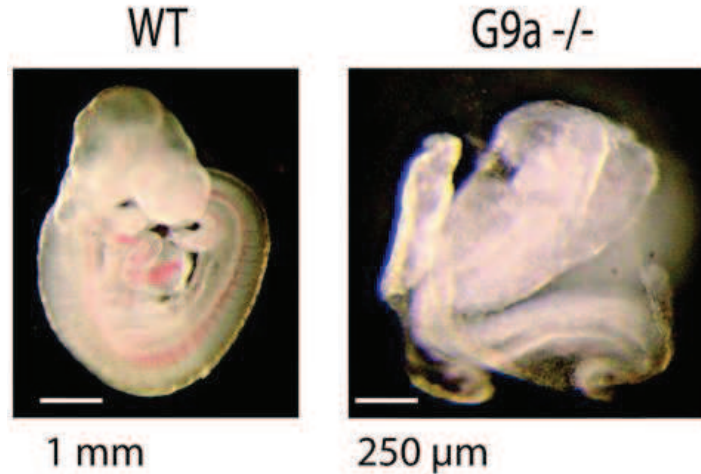


Figure 17: Comparaison d'embryons murins sauvage et déficient pour G9a 9.5 jours *post-coïtum*.

Les embryons déficients en G9a montrent des défauts de gastrulation et de croissance, et meurent 10 jours après la fertilisation. G9a semble donc avoir un rôle crucial lors du développement embryonnaire murin. Adaptée de (Wagschal et al., 2008).

G9a possède un domaine ANK (**Figure 18**) qui peut lui permettre de former un complexe hétérodimérique avec GLP *in vivo*. Toutes deux co-localisent dans le noyau. Il a été montré qu'une délétion de l'une ou l'autre de ces deux enzymes engendre une diminution dramatique des modifications d'histone H3K9m1 et H3K9m2 dans les cellules ES ou les embryons (Tachibana et al., 2002, 2005). Par contre, leur absence ne semble pas affecter la modification H3K9m3 dans ces cellules. Comme les souris *G9a*^{-/-}, les souris déficientes pour GLP meurent aussi vers 10.5 jpc (Tachibana et al., 2005). De plus, certains gènes comme *Magea2* qui se ré-expriment dans les mutants G9a (Tachibana et al., 2002), sont également ré-exprimés aussi dans les mutants GLP (Tachibana et al., 2005). Enfin dans ces deux mutants, on observe une baisse du niveau de H3K9m2 et une hausse de H3K4m3 au niveau des mêmes régions (Tachibana et al., 2005). En fait, tous les résultats obtenus indiquent donc que l'inactivation de G9a ou GLP conduit à des effets sévères très similaires, ce qui suggère que l'hétérodimère G9a-GLP est la forme active de la méthylase de H3K9 *in vivo*.

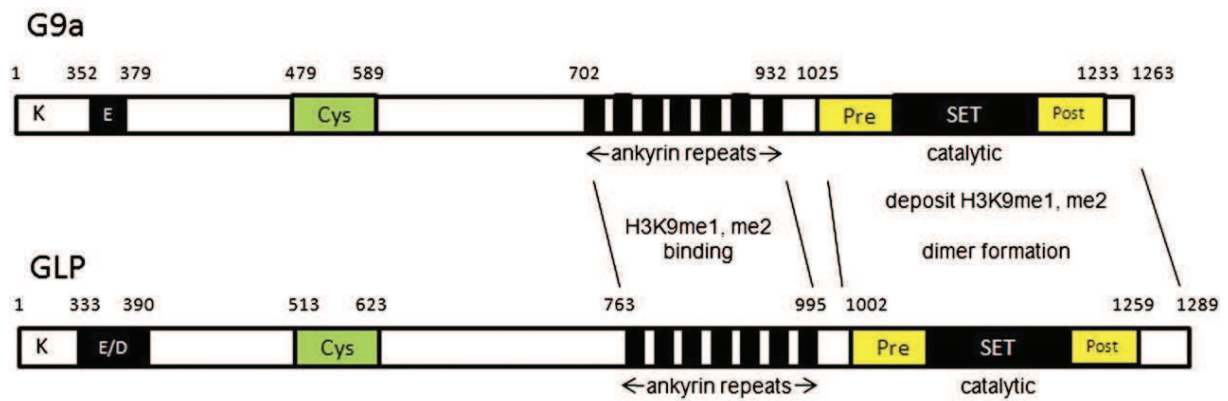


Figure 18 : Représentation schématique des deux enzymes G9a et GLP murines.

Les enzymes G9a et GLP possèdent une lysine (K) en N-terminal, qui est un site d'autométhylation, un domaine E qui est une région riche en glutamines ou E/D (région riche en glutamines et asparagines), un domaine Cys (région riche en cystéines), un domaine ANK (répétition d'anhydrines) qui permet l'interaction avec les histones modifiées H3K9m1 et H3K9m2, ainsi qu'un domaine SET responsable de l'activité catalytique de méthylation de H3K9 et H3K9me1. Le domaine SET est également nécessaire pour l'hétérodimérisation de G9a avec GLP. Adaptée de (Shinkai et Tachibana, 2011).

En plus de son activité de méthylase de H3K9, G9a peut catalyser la mono et la diméthylation de la lysine 27 de l'histone H3 *in vitro* (Patnaik et al., 2004; Tachibana et al., 2001). La triméthylation de H3K27 est mise en place par le complexe Polycomb. Ce complexe est composé de quatre protéines, SUZ12, EED, RbAp46/48, et EZH2. EZH2 est la sous unité catalytique qui permet l'établissement de la di et de la triméthylation de H3K27. Pour renforcer le lien entre G9a et la méthylation de H3K27, il a été montré que G9a interagit avec PRC2 et que G9a et PRC2 sont recrutées et co-régulent environ 175 gènes impliqués dans la différenciation cellulaire et le développement embryonnaire dans les cellules ES (Mozzetta et al., 2014). Les mêmes auteurs ont aussi montré qu'en l'absence de G9a et de GLP, EZH2 ne se lie plus aux cibles communes, ce qui est associé à une baisse de la modification d'histone H3K27m3. Cela suggère que le complexe G9a/GLP contrôle le recrutement de PRC2 et la mise en place de H3K27m3 au

niveau de ces cibles. Ils ont aussi utilisé un mutant catalytique de G9a (C1668A, (Tachibana et al., 2008)) qui est toujours capable d'interagir avec GLP mais ne peut plus catalyser H3K9me2. L'expression de cette protéine mutante dans des cellules ES déficientes pour G9a n'est pas capable de restaurer la liaison de EZH2 et H3K27m3 au niveau des cibles.

b) Le rôle de G9a dans le recrutement de la méthylation de l'ADN

1. Les cibles potentielles de méthylation de l'ADN par G9a

Grâce à son domaine ANK, G9a peut interagir avec les enzymes DNMT3a et DNMT3b (Epsztejn-Litman et al., 2008). Il a été montré dans les cellules ES que l'inactivation de G9a induit une perte de méthylation de l'ADN à des promoteurs riches en CpGs, au niveau de gènes de la lignée germinale (Dong et al., 2008) ou de gènes du développement comme les gènes *Magea2* (Tachibana et al., 2008) et de la famille *Rhox* (Myant et al., 2011). De même, une analyse par restriction landmark genomic scanning (RLGS) a conclu que dans les cellules ES, G9a contribue à la méthylation de l'ADN d'environ 220 gènes (Ikegami et al., 2007).

Toujours dans les cellules ES, une équipe a démontré que G9a est aussi requise pour l'acquisition de la méthylation de l'ADN au niveau d'un grand nombre de rétrotransposons : LTR et non-LTR (Dong et al., 2008).

D'autres chercheurs ont montré que G9a est aussi impliquée dans la méthylation de l'ADN de régions ICRs. De plus, dans des cellules ES déficientes pour G9a, cette perte de méthylation est concomitante avec la perte de H3K9m2 au niveau de ces régions (Xin et al., 2003).

Ces résultats obtenus dans les cellules ES suggèrent un rôle très important de G9a dans la régulation de la méthylation de l'ADN. Cependant, aucune recherche n'a été effectuée pour confirmer ces résultats *in vivo*, durant l'embryogénèse des Mammifères.

2. Les mécanismes potentiels de régulation de la méthylation de l'ADN par G9a

Il existe de nombreuses évidences d'un lien fonctionnel entre la modification d'histone H3K9me2 et la méthylation de l'ADN. Des résultats chez *Neurospora Crassa* et *Arabidopsis Thaliana* montrent que des mutations des enzymes modificatrices de H3K9 affectent fortement la méthylation de l'ADN (Jackson et al., 2002; Tamaru and Selker, 2001). Chez les Mammifères, il a été montré que dans l'ovocyte fertilisé ou dans les PGCs, la vague de déméthylation suit la perte de H3K9me2 (Liu et al., 2004). Ceci signifie que la perte de méthylation de H3K9 pourrait avoir un rôle dans la perte de méthylation de l'ADN.

D'autres données biochimiques ont montré un lien entre G9a, H3K9me2 et le complexe UHRF1-DNMT1. La modification H3K9 diméthylée est reconnue par UHRF1, favorisant ainsi la méthylation par DNMT1. De plus, G9a co-localise avec DNMT1 et UHRF1 aux sites de réplication de l'ADN, où il serait utile pour la propagation de la modification H3K9me2 après la réplication (voir chapitre IV, partie b.2) (Estève et al., 2006; Kim et al., 2009; Rothbart et al., 2012).

Afin de comprendre le rôle des modifications d'histones mises en place par G9a pour le recrutement de la méthylation de l'ADN, des chercheurs ont regardé l'effet de l'expression du mutant catalytique C1668A dans des cellules ES déficientes en G9a. Ils observent que la méthylation de l'ADN est en grande partie restaurée malgré le fait que le mutant C1668A ne soit pas capable de restaurer la modification H3K9me2 (Dong et al., 2008; Epsztejn-Litman et al., 2008; Tachibana et al., 2008). Ces résultats laissent à penser que l'influence de G9a sur la méthylation de l'ADN pourrait finalement être indépendante de son activité catalytique.

3. Les facteurs influençant l'activité de G9a

3.1 Les facteurs de transcription WIZ et ZNF644

Une équipe a montré que le complexe G9a/GLP pouvait interagir directement en complexe avec les protéines WIZ et ZNF644 dans les cellules humaines 293T (Bian et al., 2015). WIZ est un partenaire connu de G9a (Ueda et al., 2006) alors que la fonction de ZNF644 est inconnue. En effet, G9a et GLP possèdent un domaine TAD (trans-activating domain) dans leur partie N-terminale qui leur permet d'interagir avec WIZ pour GLP et ZNF644 pour G9a. Ces deux molécules possèdent plusieurs domaines à doigts de zinc qui leur permettent de reconnaître des motifs spécifiques de séquence d'ADN. Grâce à ces domaines, ces deux protéines recrutent le complexe G9a/GLP au niveau de certaines régions promotrices spécifiques de l'ADN (Figure 19).

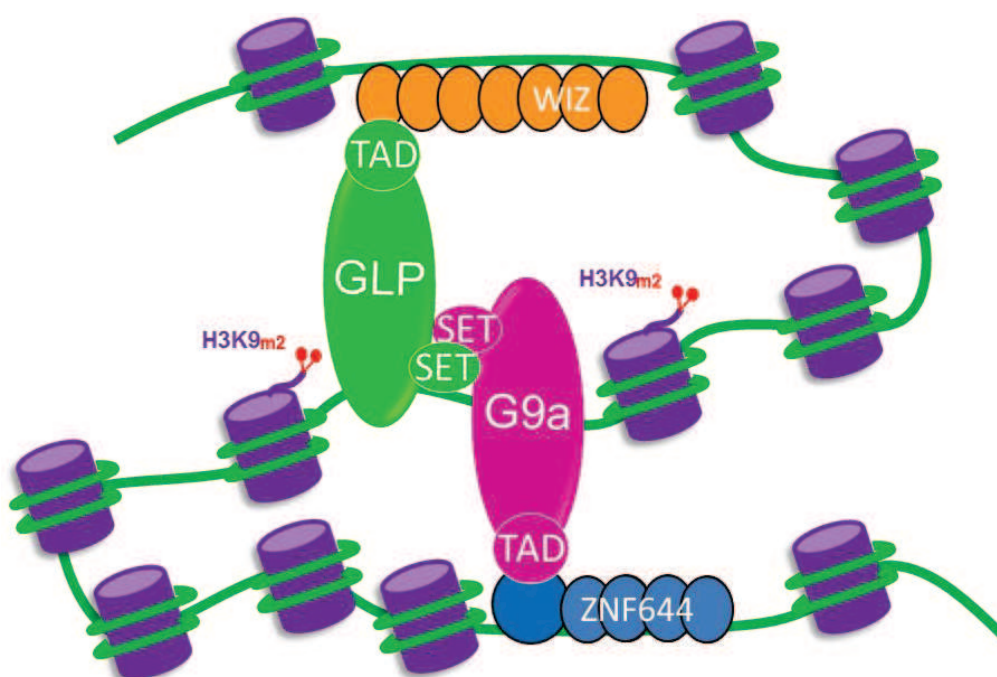


Figure 19: Représentation schématique de l'interaction de l'hétérodimère G9a/GLP l'ADN.

G9a et GLP peuvent interagir à l'ADN grâce à l'intervention de facteurs de transcription tels que WIZ et ZNF644. Une fois le complexe créé, G9a et/ou GLP peuvent alors méthyler les lysines 9 des histones H3 alentours. Adaptée de (Bian et al., 2015).

3.2 Le facteur de transcription E2F6

D'autres facteurs de transcription pourraient aussi entrer en jeu et participer au recrutement de G9a/GLP au niveau de régions cibles du génome. C'est le cas par exemple d'E2F6 (voir *chapitre IV, partie c.1*) qui a été purifié dans un même complexe que G9a (Ogawa et al., 2002). Cette interaction comprendrait aussi PRC2 et HP1 γ . Ensemble, ces molécules pourraient former un complexe de répression en recrutant la méthylation de l'ADN (Ogawa et al., 2002; Mozzetta et al., 2014)

3.3 Le facteur de transcription LSH

Dans les cellules ES, il a été montré que G9a et LSH (Lymphoid Specific Helicase) sont toutes deux requises pour inactiver des gènes de pluripotence et que LSH recrute G9a au niveau de leurs promoteurs. De plus, LSH recruterait des histones HDAC qui permettraient ensuite à G9a de méthyler H3K9 et ainsi d'inactiver ces gènes (Myant et al., 2011). Une équipe a montré que lors de la différenciation de cellules ES, LSH va recruter G9a afin de méthyler les histones. Ces modifications vont permettre le recrutement de protéines HP1 qui à leur tour recrutent les enzymes DNMT3 qui vont méthyler les promoteurs des gènes de pluripotence (Feldman et al., 2006).

RESULTATS

RESULTATS DU PROJET 1 :

Etablissement de la méthylation de l'ADN au cours du développement embryonnaire murin.

INTRODUCTION

Chez la souris, la méthylation du génome est mise en place à partir du stade blastocyste au moment de l'implantation de l'embryon. Cependant, aucune étude cinétique n'a encore été effectuée pour connaître à quel moment et à quelle vitesse la méthylation de l'ADN se met en place au cours du développement embryonnaire.

On sait que le blastocyste est globalement déméthylé. C'est à partir de ce stade, à 3.5 jpc, que la méthylation de l'ADN va reprendre place grâce aux enzymes *de novo*, DNMT3A et DNMT3B (Borgel et al., 2010; Feldman et al., 2006; Mohn et al., 2008). Cette méthylation est indispensable pour le bon développement de l'embryon comme le montre les phénotypes des embryons invalidés pour DNMT3A et DNMT3B (Okano et al., 1999). Nous avons donc cherché à déterminer les cinétiques d'acquisition de la méthylation au cours du développement embryonnaire. De plus, notre analyse nous a permis de déterminer si l'acquisition de la méthylation se fait de manière uniforme sur le génome ou bien si certaines régions acquièrent de la méthylation plus ou moins rapidement que d'autres. Pour cela, nous avons mis en application les nouvelles technologies de séquençage à haut débit (telles que le RRBS) qui permettent d'étudier la méthylation de cytosines de manière quantitative, à l'échelle génomique, et à la résolution du nucléotide. Ainsi nous avons pu étudier l'évolution de nombreuses régions telles que les îlots CpG des gènes autosomaux ou du chromosome X, les DMRs de domaines de gènes soumis à empreinte, ou encore de différents éléments transposables.

Les deux enzymes DNMT3A et DNMT3B sont indispensables pour la mise en place de la méthylation de l'ADN *de novo*. L'absence de ces enzymes conduit à une

létalité à 15.5 jpc dans les souris déficientes en DNMT3B et quatre semaines après la naissance pour les souris déficientes en DNMT3A (Okano et al., 1999). Cependant, leur contribution respective à la mise en place des patrons de méthylation *in vivo* n'a pas été étudiée. Grâce à ces mêmes méthodes de séquençage, nous avons aussi pu déterminer quelle est la contribution de chacune des enzymes DNMT3A et DNMT3B dans des embryons à 8.5 jpc et si elles possèdent des cibles spécifiques ou redondantes dans le génome.

L'objectif de ces recherches était donc de cartographier la cinétique d'acquisition de la méthylation de l'ADN au cours du développement embryonnaire, puis de déterminer les cibles des enzymes DNMT3A et DNMT3B sur le génome ainsi que leur importance sur la régulation de l'expression de ces gènes.

Ces résultats ont été publiés dans l'**article 1**.

TRAVAUX REALISES

Pour atteindre ces objectifs, j'ai réalisé les travaux suivants :

1-Disséquer des embryons et préparer des échantillons d'ADN et d'ARN d'embryons murins à des stades consécutifs de développement (**Figure 1**) :

- Blastocystes à 3.5 jpc et 4.5 jpc.
- Epiblastes d'embryons à 5.5, 6.5, et 7.5 jpc. Seul l'ADN des épiblastes a été extrait pour suivre la mise en place de la méthylation dans les cellules de l'embryon et pas des tissus extra-embryonnaires.
- Embryons totaux à 8.5 jpc.

2-Générer des lignées de souris mutantes déficientes en DNMT3A ou DNMT3B (**Figures 2 et 3**) afin d'extraire des échantillons d'ADN et ARN d'embryons sauvages et mutants aux stades 6.5, 8.5 et 11.5 jpc.

3-Préparer des bibliothèques de séquençage RRBS sur l'ADN des embryons récoltés.

4-Etudier le transcriptome par des expériences de RNA-seq sur les embryons *Dnmt3b*^{-/-} et contrôles.

5-Réaliser des RT-qPCR pour déterminer l'expression de DNMT3A et DNMT3B au cours du développement.

6-Participer aux analyses bio-informatiques des résultats. Ce travail a majoritairement été réalisé par Sylvain Guibert, Michael Dumas et Michaël Weber.

7-Valider les résultats de RRBS par traitement au bisulfite de sodium suivi de séquençages ou suivi de digestions enzymatiques (COBRA).

8-Valider les résultats de RNA-seq par RT-qPCR

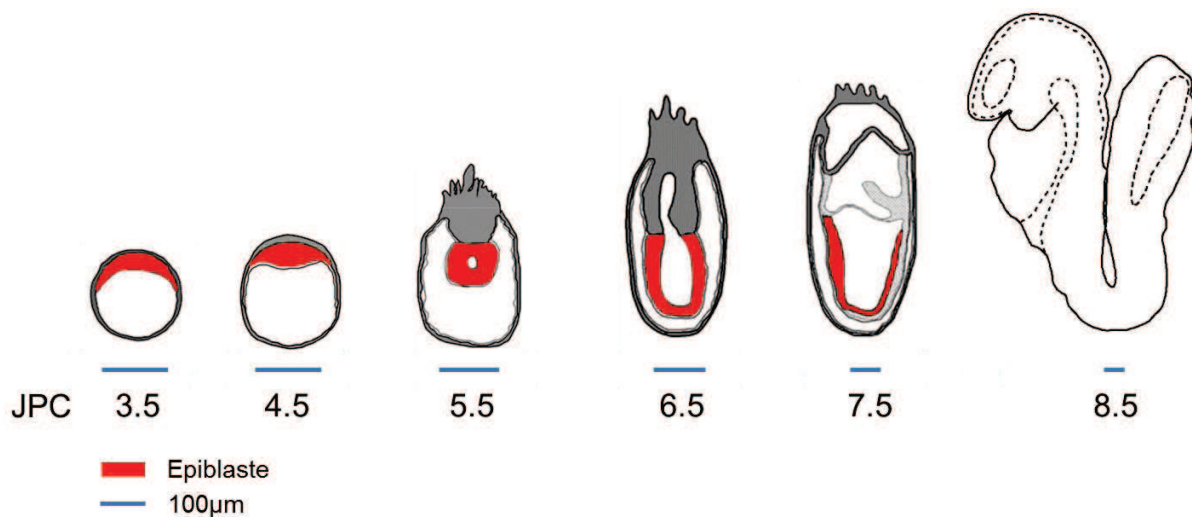
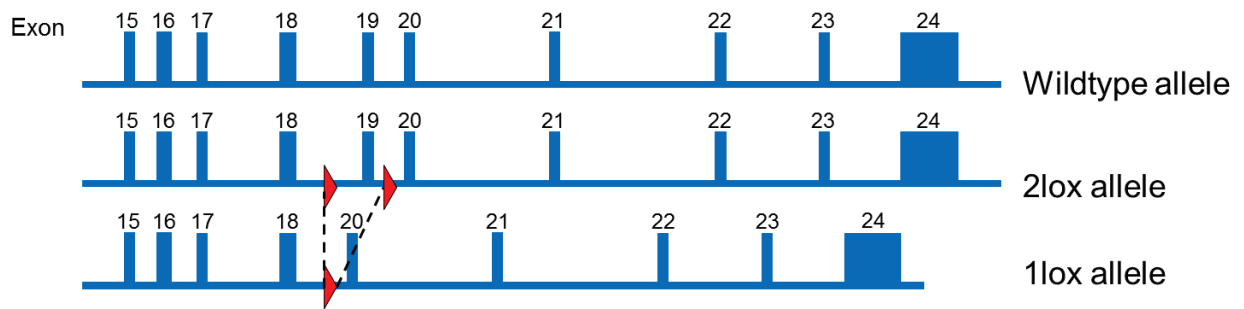


Figure 1 : Représentation schématique des embryons disséqués.

Les cellules pluripotentes de l'embryon sont représentées en rouge. A 3.5 jpc, elles forment l'ICM (Inner Cell Mass) entouré des cellules du trophoctoderme. A partir de 4.5 jpc, l'ICM forme l'épiblaste et à 5.5 jpc, se forme le tissu extra-embryonnaire à partir du trophoctoderme. A 7.5 jpc, la gastrulation commence et se forme alors le mésoderme ainsi que la cavité amniotique. A 8.5 jpc l'épiblaste disparaît, il y a apparition des somites et de nombreux tissus tels que le muscle cardiaque, le tube neural et l'allantoïde.

DNMT3a Excision of the catalytic domain



DNMT3b Excision of the catalytic domain

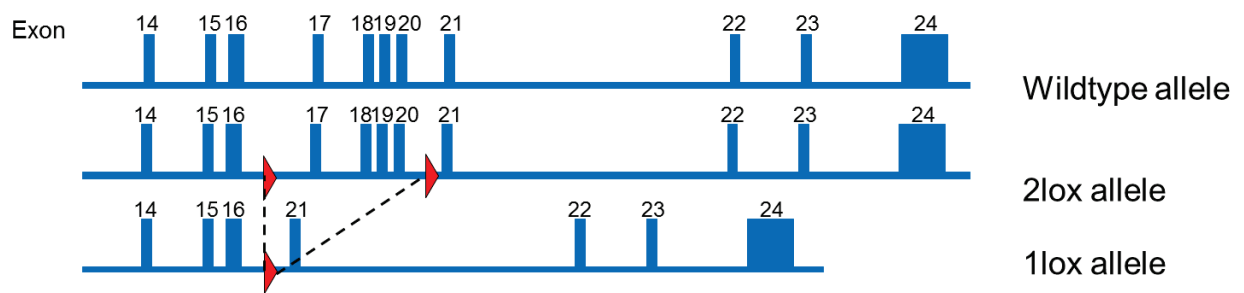


Figure 2 : Représentation schématique des allèles mutants générés pour les gènes *Dnmt3a* et *Dnmt3b*.

Nous avons généré des allèles mutants par délétion d'exons essentiels dans les domaines catalytiques. Pour *Dnmt3a*, deux sites lox permettent la délétion de l'exon 19 après excision par la recombinaise Cre. Pour *Dnmt3b*, deux sites lox permettent la délétion des exons 17 à 20 après excision par la recombinaise Cre. Les allèles mutés sont dits 1lox.

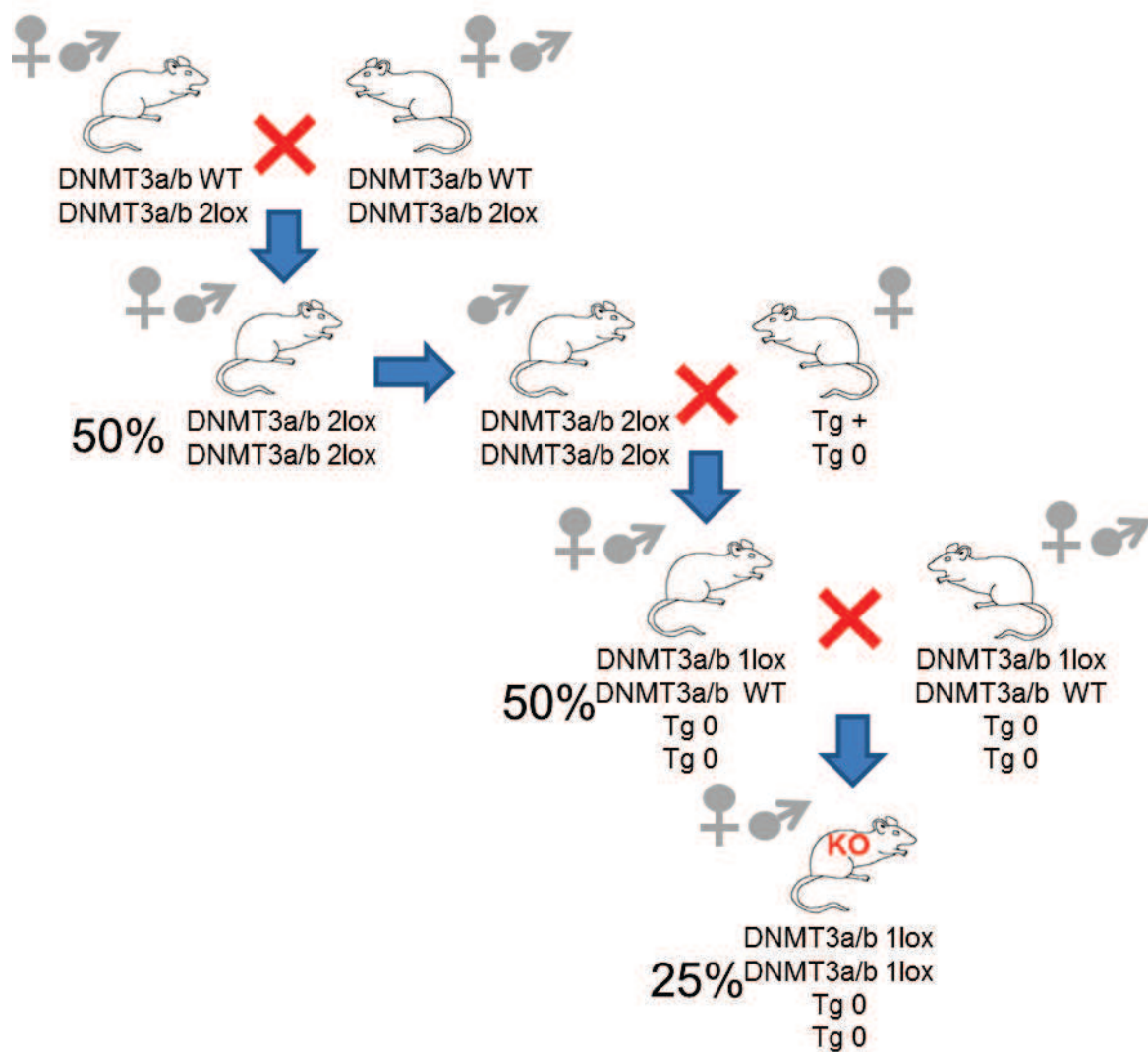


Figure 3 : Représentation schématique des croisements effectués pour obtenir des souris déficientes en *Dnmt3a* ou *Dnmt3b*.

Les premières souris croisées sont hétérozygotes 2lox/wt pour *Dnmt3a* ou *Dnmt3b*. Elles sont croisées pour obtenir des souris homozygotes 2lox/2lox avec 50% de chance. Les souris mâles homozygotes 2lox/2lox sont ensuite croisées avec des souris femelles portant un allèle *Actb-Cre* (Tg+/Tg0) qui code pour l'enzyme Cre sous le contrôle du promoteur du gène *Actin beta*. Cet allèle induit une expression forte de l'enzyme Cre et une recombinaison des séquences 2lox dès le stade du zygote dans 100% des embryons en raison de la transmission maternelle de la Cre par l'ovocyte. Les souris 1lox/wt sont ensuite croisées entre elles afin d'obtenir des souris homozygotes 1lox/1lox avec 25% de chance.

ARTICLE 1

RESEARCH

Open Access

Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse

Ghislain Auclair, Sylvain Guibert, Ambre Bender and Michael Weber*

Abstract

Background: In the mouse, the patterns of DNA methylation are established during early embryonic development in the epiblast. We quantified the targets and kinetics of DNA methylation acquisition in epiblast cells, and determined the contribution of the *de novo* methyltransferases DNMT3A and DNMT3B to this process.

Results: We generated single-base maps of DNA methylation from the blastocyst to post-implantation stages and in embryos lacking DNMT3A or DNMT3B activity. DNA methylation is established within two days of implantation between embryonic days 4.5 and 6.5. The kinetics of *de novo* methylation are uniform throughout the genome, suggesting a random mechanism of deposition. In contrast, many CpG islands acquire methylation slowly in late epiblast cells. Five percent of CpG islands gain methylation and are found in the promoters of germline genes and in exons of important developmental genes. The onset of global methylation correlates with the upregulation of *Dnmt3a/b* genes in the early epiblast. DNMT3A and DNMT3B act redundantly to methylate the bulk genome and repetitive elements, whereas DNMT3B has a prominent role in the methylation of CpG islands on autosomes and the X chromosome. Reduced CpG island methylation in *Dnmt3b*-deficient embryos correlates with gene reactivation in promoters but reduced transcript abundance in gene bodies. Finally, DNMT3B establishes secondary methylation marks at imprinted loci, which distinguishes *bona fide* germline from somatic methylation imprints.

Conclusions: We reveal that the DNMT3 *de novo* methyltransferases play both redundant and specific functions in the establishment of DNA methylation in the mouse embryo.

Background

Methylation of cytosines is an epigenetic mark of DNA with crucial functions in mammalian development and diseases. In mammals, methylation occurs almost exclusively in the context of cytosine-guanine (CpG) dinucleotides, which are found at higher frequency in short regions termed CpG islands (CGIs). Paradoxically, the majority of CGIs remain unmethylated in all cell lineages [1-3]. The contrast between the low methylation at CGIs and the high methylation in CpG-poor sequences arises from the accelerated mutational loss of methylated CpGs over evolutionary time [2].

DNA methylation is reprogrammed during embryonic development. The spermatozoa and oocyte acquire differential methylation at many sequences, which establishes a

strong epigenetic asymmetry between the gametes that is not limited to imprinted regions [4-6]. After fertilization, these gametic profiles are globally erased to reach a low point of methylation at the blastocyst stage [4-8]. Yet numerous sequences maintain partial methylation in the blastocysts, primarily on the oocyte-derived allele [4-8]. In the case of imprinted regions, this differential allelic methylation is stably maintained throughout embryogenesis and adulthood at a small number of germline differentially methylated regions (gDMRs). After implantation of the embryo, DNA methylation is restored to high levels in epiblast cells throughout the genome as well as at a small number of CGIs [6-8].

Cytosine methylation is catalyzed by DNA methyltransferases (DNMTs). DNMT1 copies methylation on the new DNA strand at hemimethylated CpG sites after DNA replication, which mediates epigenetic inheritance in dividing cells. In contrast, DNMT3A and DNMT3B mediate *de*

* Correspondence: michael.weber@unistra.fr

CNRS, University of Strasbourg, UMR 7242 Biotechnology and Cell Signaling, 300 Bd Sébastien Brant, BP 10413, 67412 Illkirch, France

de novo methylation by targeting previously unmethylated CpGs. Another member of the family, DNMT3L, lacks enzymatic activity but acts as a cofactor that stimulates the activity of DNMT3A and DNMT3B in germ cells [4,5]. The knockout of DNMT1 and DNMT3B in mice leads to mid-gestation lethality, indicating that DNA methylation is essential for development [9,10]. DNMT3A knockout animals survive until birth but die at around 4 weeks of age [10].

Advances with high-throughput methods led to a better characterization of the distribution of DNA methylation in mouse embryos, yet several points remain to be clarified, such as (i) the timing of acquisition of DNA methylation in embryos, (ii) the identity and role of CGIs that gain methylation, and (iii) the contribution of DNMT3A and DNMT3B to *de novo* methylation in the embryo. So far the single-gene studies have indicated that DNMT3B methylates the promoters of a few germline genes [7,11] and protocadherin genes [12], whereas DNMT3A and DNMT3B cooperate to methylate other sequences in the mouse embryo [10,13,14]. This suggests that DNMT3A and DNMT3B have both specific and overlapping functions in embryonic methylation, which has not been investigated in a systematic way.

To answer these questions, we generated a single base atlas of cytosine methylation by reduced representation bisulfite sequencing (RRBS) throughout mouse embryonic development starting from the blastocyst stage, as well as in *Dnmt3a* and *Dnmt3b* mutant embryos. We show that methylation is established rapidly at the time of implantation by the combined action of DNMT3A and DNMT3B. In contrast, CGIs behave as a functionally distinct class of sequences that acquire methylation slowly mediated primarily by DNMT3B. We provide a comprehensive analysis of the targets of CGI methylation in development and studied their impact on gene expression in embryos. Our study provides insights into the role and target specificities of the DNMT3 enzymes in mouse development.

Results

Temporal mapping of DNA methylation acquisition during murine embryogenesis

To characterize the wave of *de novo* DNA methylation in the mouse embryo, we generated single-base profiles of cytosine methylation by RRBS at consecutive stages of development between embryonic day (E)3.5 and E8.5 (Figure S1A,B in Additional file 1). We quantified methylation for approximately 1,300,000 CpGs per sample at an average sequencing depth of 68× (Figure S1A in Additional file 1). As shown previously [4-7], the genome of E3.5 blastocysts is globally hypomethylated but contains sequences with partial methylation caused by incomplete erasure of gametic methylation (Figure 1A). After E3.5, cytosine methylation progressively accumulates

after implantation exclusively in a CG sequence context (Figure S1C in Additional file 1). During this period, gene bodies, transposable elements and CpG-poor promoters are *de novo* methylated, whereas CpG-rich sequences are protected from methylation (Figure 1B,C). 'Canyons', a class of extended regions of low methylation that span loci of developmental transcription factors [15], also form after implantation through protection from *de novo* methylation (Figure 1C). Interestingly, most sequences with partial methylation in blastocysts gain full methylation in post-implantation embryos (Figure 1D); thus, resistance to demethylation after fertilization predisposes to *de novo* methylation in post-implantation embryos. We then investigated the dynamics of methylation by averaging methylation in 400 bp tiles and searching for tiles that gain or lose methylation at each developmental transition. The most dramatic wave of *de novo* methylation occurs in early epiblast between E4.5 and E5.5 (Figure 1E). In contrast there are very few demethylation events at any of the developmental transitions (Figure 1E). This demonstrates that *de novo* methylation occurs rapidly at the time of implantation and is unidirectional. To precisely quantify the kinetics of *de novo* methylation, we selected all tiles that gain methylation in post-implantation embryos and plotted their methylation as a function of the developmental stage (Figure 1F). Methylation increases rapidly from 12% to 62% between E4.5 and E5.5, and reaches almost maximum levels at E6.5 (Figure 1F). As a control, we show that the imprinted gDMRs have stable methylation between 40 and 50% at all stages (Figure 1F). To ask if the rate of *de novo* methylation varies along the genome, we monitored methylation separately in exons, introns and transposable elements and found that they acquire methylation with similar kinetics (Figure 1G). We also monitored methylation in classes of transposable elements and found that, while they have varying degrees of methylation in blastocysts, they all gain methylation with similar kinetics in post-implantation embryos (Figure 1H). This shows that *de novo* methylation occurs rapidly within two days around implantation and is uniform throughout the genome.

Kinetics of CpG island methylation in developing embryos

Next we focused on CGIs. Out of the 16,023 UCSC CGIs, 89% are covered in each sample (Figure S1D in Additional file 1) and 14,085 have methylation data in all samples. Out of these 14,085 CGIs, 713 (5%) acquire more than 50% methylation in post-implantation embryos (Additional file 2). Remarkably, the proportion of methylated CGIs with more than 50% methylation is much lower at transcription start sites (TSSs; 0.6%, 69/10,694 in total, 63/10,422 on autosomes) than in intergenic and intragenic regions, especially for CGIs covering exons (34%, 505/

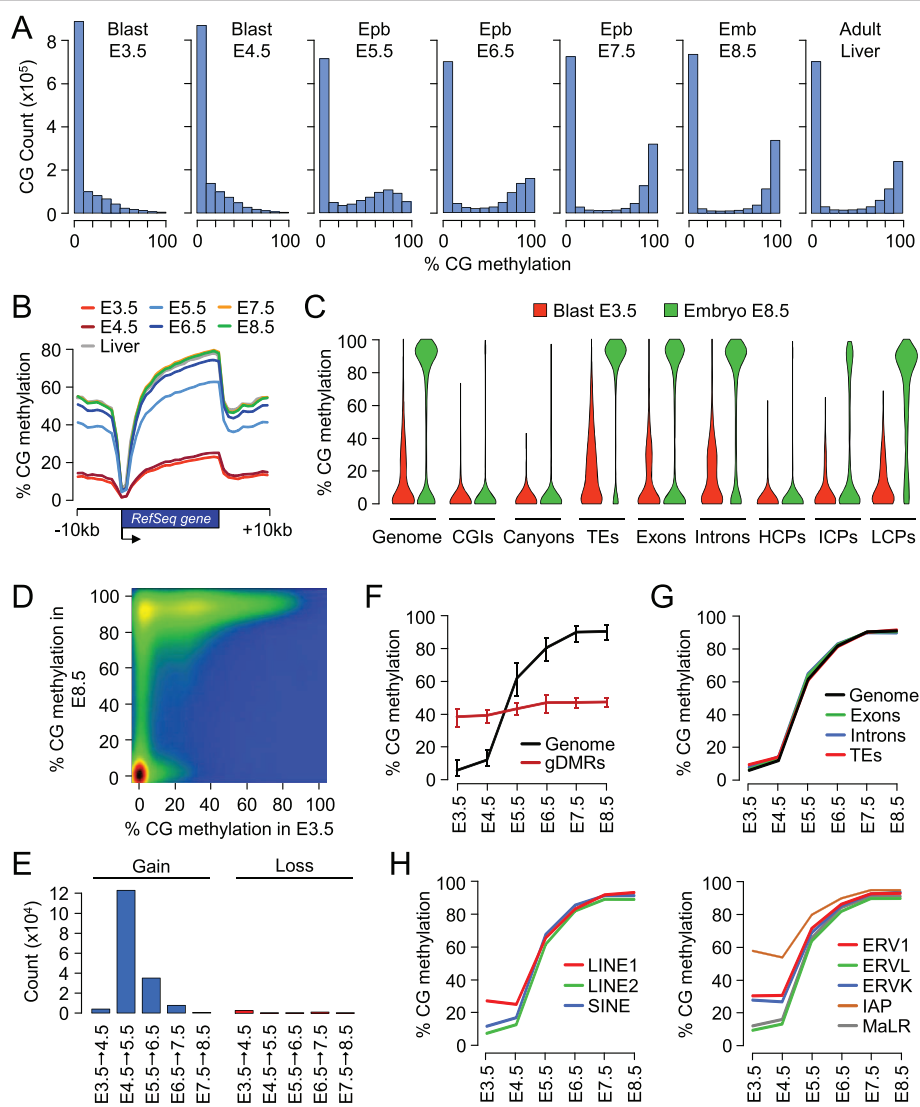
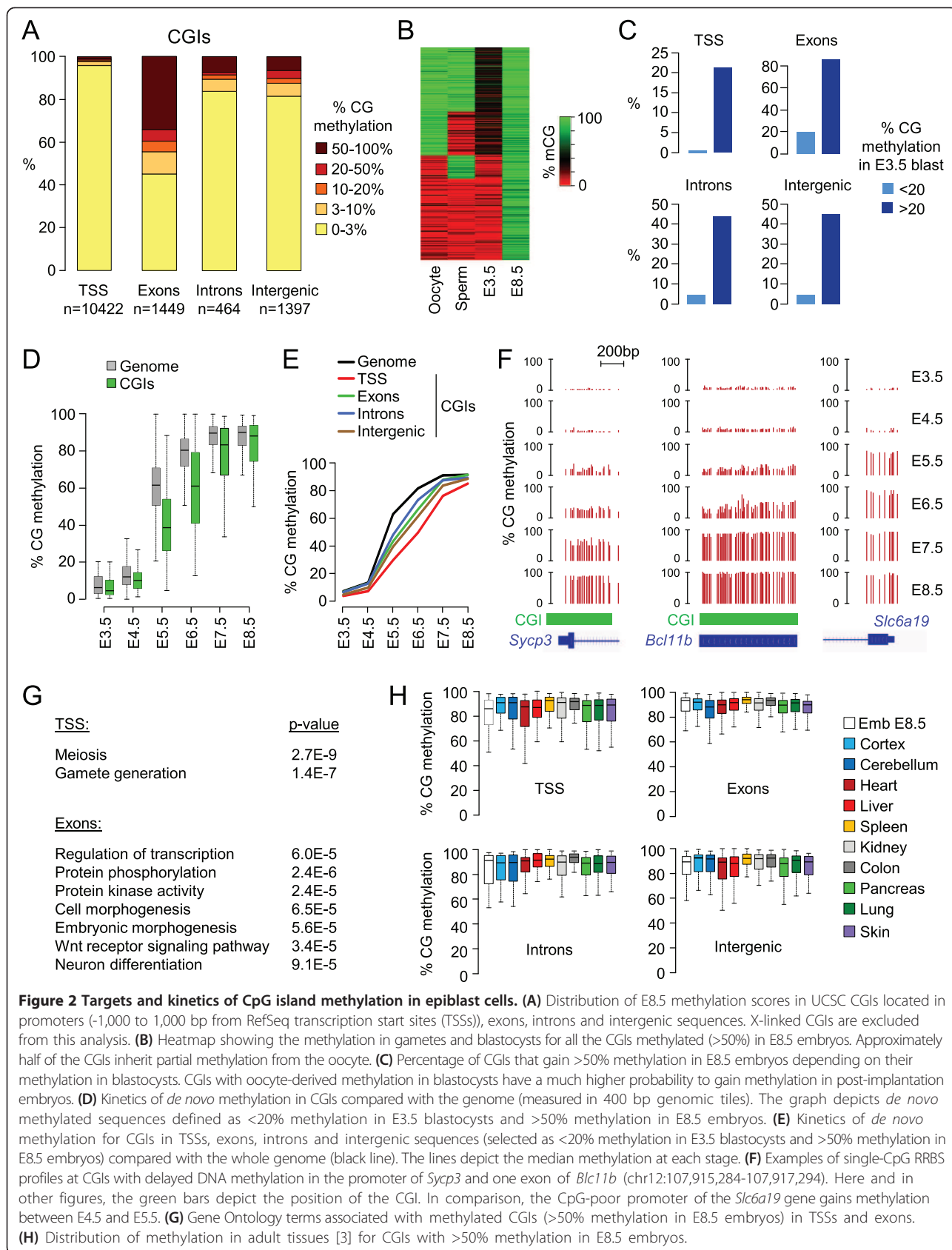


Figure 1 Acquisition of CpG methylation occurs at implantation in mouse embryos. (A) Density histograms showing the distribution of methylation levels measured at individual CpGs throughout embryonic development. Blast, blastocyst; Epb, epiblast; Emb, embryo. **(B)** Distribution on CpG methylation in RefSeq genes and 10 kb of flanking sequences throughout embryonic development. For each protein-coding RefSeq gene (excluding the X and Y chromosomes), we calculated methylation in 20 equal-sized windows within the gene and 10 1-kb windows of flanking sequences. **(C)** Violin plots showing the acquisition of CpG methylation in post-implantation embryos compared with blastocysts in various genome elements. TEs, transposable elements; HCPs, high CpG promoters; ICPs, intermediate CpG promoters; LCPs, low CpG promoters. **(D)** Pairwise comparison of CpG methylation (measured in 400 bp tiles) in E3.5 blastocysts and E8.5 post-implantation embryos. The density of points increases from blue to dark red. **(E)** Number of 400 bp tiles that gain or lose more than 20% CpG methylation at each developmental transition. **(F)** Kinetics of *de novo* DNA methylation in development. We selected all the genomic tiles (400 bp) that gain methylation in post-implantation embryos (defined as <20% methylation in E3.5 blastocysts and >50% methylation in E8.5 embryos) and then plotted their methylation as a function of the developmental stage (black line). The red line shows the methylation for 17 imprinted germline DMRs. The lines represent the median methylation and the error bars represent the 25th and 75th percentiles. **(G)** Kinetics of *de novo* DNA methylation in exons, introns and transposable elements (TEs). **(H)** Kinetics of *de novo* DNA methylation in classes of transposable elements. LINE, long interspersed nuclear element; SINE, short interspersed nuclear element. In **(G,H)**, the lines depict the median methylation measured at each developmental stage.

1,471 in total, 502/1,449 on autosomes) (Figure 2A). Because the UCSC annotations underestimate the number of CGIs [16], we repeated this analysis with a less stringent custom CGI annotation and found a similar repartition of CGI methylation (Figure S2A,B in Additional file 1). We explored the relationship between CGI methylation

in pre- and post-implantation stages and found that half of the CGIs methylated in post-implantation embryos already have persistent gametic methylation in blastocysts (Figure 2B). This reflects that CGIs that escape complete demethylation before implantation are more likely to reacquire methylation after implantation (Figure 2C).



Consequently transient imprinted CGI methylation in blastocysts rarely translates into lifelong imprinted methylation [7,17]. We then monitored the kinetics of *de novo* methylation at CGIs and found that they acquire methylation at a slower rate compared with the bulk genome (Figure 2D). A similar delay in methylation is observed when we use our extended set of custom annotated CGIs (Figure S2C in Additional file 1). This delay is most evident at TSS-proximal CGIs (Figure 2E), as illustrated by the *Sycp3* promoter (Figure 2F) and several other promoters (Figure S3A in Additional file 1). Methylation of intergenic and intragenic CGIs is, on average, less delayed than at promoters (Figure 2E); nevertheless, many of these CGIs also gain delayed methylation as exemplified by intragenic CGIs in the *Bcl11b*, *Dact1* and *Cux1* genes (Figure 2F; Figure S3B in Additional file 1).

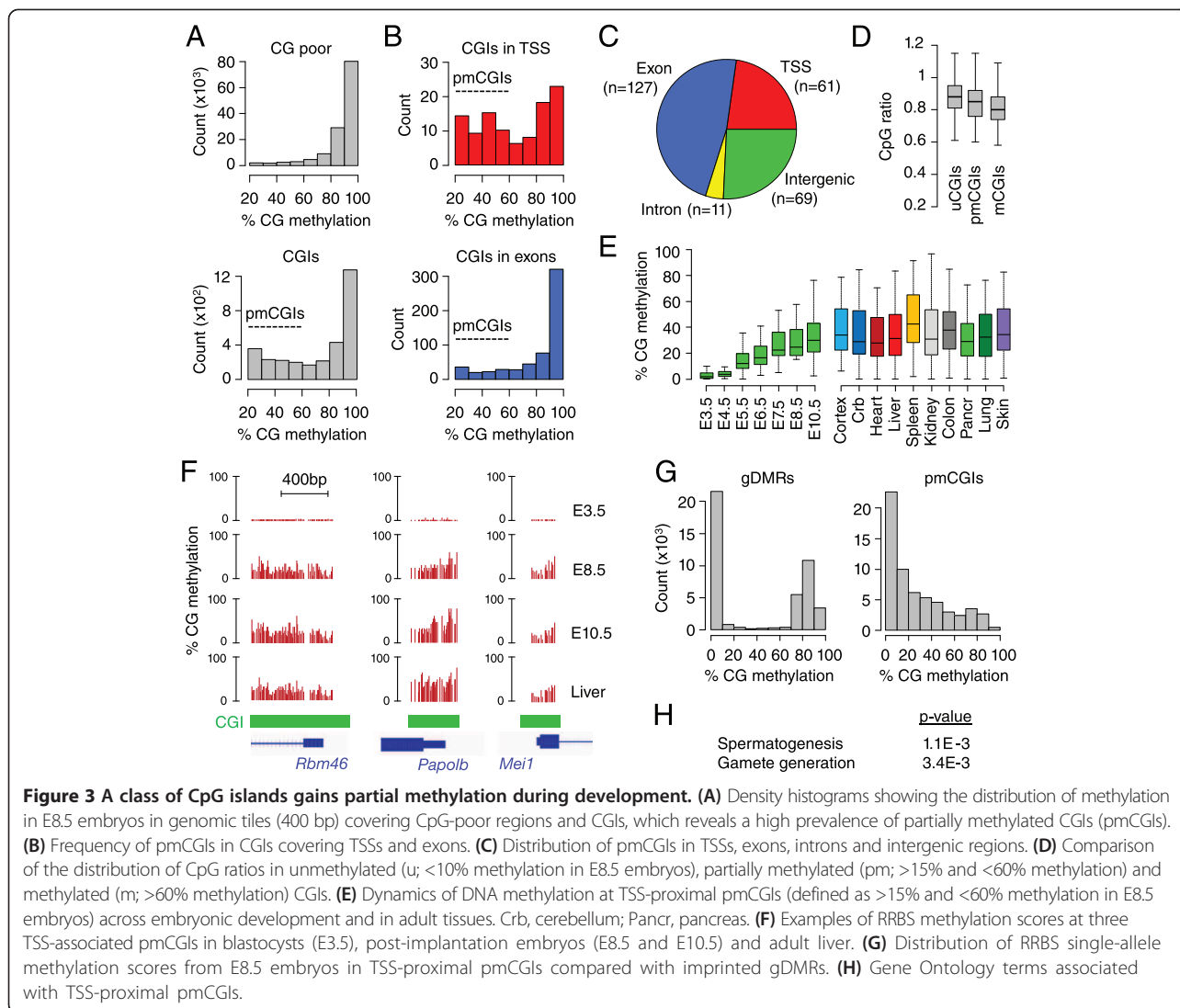
CpG island methylation is recruited to important developmental genes

To gain insights into the function of CGI methylation, we performed ontology analyses on genes that gain CGI methylation after implantation. In accordance with previous data [7,11], promoter CGI methylation is enriched at genes involved in gamete functions (Figure 2G; Figure S4A in Additional file 1). Using our custom CGI annotation, we identified 87 CGI promoters with more than 50% methylation in E8.5 embryos and found that 79% (69/87) are associated with germline genes (Additional file 3). The targets identified here and in previous studies [7] reveal that promoter CGI methylation is recruited in particular to genes involved in gamete chromatin (*Brdt*, *H1fnt*, *Hist1h2aa/ba*, *H2afy3*), meiosis (*Spo11*, *Sycp1/2/3*, *Syce1/3*, *Msh4*, *Hormad1/2*) and the Piwi-interacting RNA (piRNA) pathway (*Piwi1/2/4*, *Mov10l1*, *Fkbp6*, *Mael*, *Tdrd1/9/12*, *Rnf17*, *Ddx4*, *Asz1*). In contrast, exon CGI methylation is targeted to genes with important developmental functions in the regulation of transcription, morphogenesis, signaling pathways and neuronal development (Figure 2G). As examples, key transcription factor genes (*Cux1*, *Bcl11b*, *Klf3*, *Daxx*, *Foxo3*, *Zfp64*) gain exon CGI methylation in the epiblast, as well as several genes of the Wnt/ β -catenin signaling pathway that plays pivotal roles in embryogenesis and gastrulation. In contrast, we found no ontology category associated with intron CGI methylation (data not shown). To investigate if intragenic CGI methylation correlates with transcription of the surrounding gene as is the case in oocytes [4,18], we compared CGI methylation with RNA-Seq in E8.5 embryos and found a tendency for methylated CGIs in exons, but not introns, to be located within active transcription units (Figure S4B in Additional file 1). To ask if CGI methylation persists in the adult, we interrogated a published dataset from mouse adult tissues [3] and found that CGIs in promoters, exons, introns or intergenic regions remain

highly methylated in all tissues (Figure 2H). Thus, CGI methylation in the epiblast is targeted to important developmental genes and constitutes a stable epigenetic signature of all somatic lineages.

A class of CpG islands gains partial methylation in somatic lineages

We noted that CGIs have a peculiar distribution of CpG methylation in post-implantation embryos characterized by a high prevalence of intermediate methylation (Figure 3A). We defined partially methylated CGIs (pmCGIs) as having methylation of between 15 and 60% in E8.5 embryos (excluding the known imprinted differentially methylated regions (DMRs) and the X chromosome). The pmCGIs are found within and outside genes but their relative frequency is the highest in TSSs (Figure 3B,C). Interestingly their CpG ratio is intermediate between those of methylated and unmethylated CGIs (Figure 3D). To investigate the possibility that the pmCGIs gain methylation at later stages of development, we generated RRBS methylomes in E10.5 embryos and adult liver, and investigated methylomes from adult tissues [3]. Most promoter pmCGIs retain partial methylation throughout development and in adult tissues (Figure 3E,F). Similarly, a high proportion of intergenic and intragenic pmCGI harbors partial methylation in all tissues while some occasionally are fully methylated in some tissues (data not shown). We then asked if this partial methylation represents an allele-specific methylation or a low methylation per allele. To this end we extracted single-allele methylation data from the sequencing reads and found that, in contrast to the allele-specific methylation of imprinted gDMRs, the partial methylation at pmCGIs results from a low density of methylated cytosines per allele (Figure 3G). This is confirmed by bisulfite cloning and sequencing of larger amplicons (400 to 500 bp) in four pmCGI promoters in adult liver (Additional file 1). Interestingly, whole-genome bisulfite sequencing and the bisulfite cloning reveal that pmCGIs contain patches of low and high susceptibility to methylation (Additional file 1), which could reflect differential susceptibility caused by the positioning of nucleosomes. To further characterize the promoter pmCGIs, we performed an ontology analysis and found that, similar to the fully methylated CGI promoters, they are enriched for germline-specific genes (Figure 3H). Using our custom annotation pipeline, we identified 50 high confidence pmCGI promoters and counted that 60% (30/50) are linked to germline genes (Additional file 3). Representative examples include *Smc1b*, *Pap0lb*, *Boll*, *Mei1*, *Rbmxl2* and *Rbm46* (Figure 3F). This identifies a novel class of methylated promoter CGIs and extends the repertoire of germline genes targeted by DNA methylation in embryogenesis.



DNMT3A and DNMT3B cooperate to methylate the genome

Next we investigated the contribution of the *de novo* enzymes DNMT3A and DNMT3B to methylation in embryos. We first followed the expression of *Dnmt3a/b* mRNAs by quantitative RT-PCR (RT-qPCR) and found that both genes are upregulated in early epiblast cells, which coincides with the onset of genome-wide methylation (Figure 4A). Notably, *Dnmt3b* mRNAs reach higher levels of expression than *Dnmt3a* (Figure 4A). RNA-Seq indicates that embryos express predominantly the short *Dnmt3a2* isoform and the full length *Dnmt3b1* isoform (Figure S6A in Additional file 1). We generated embryos homozygous for catalytically inactive alleles of *Dnmt3a* and *Dnmt3b* (referred to as *Dnmt3a*^{-/-} and *Dnmt3b*^{-/-}) and performed RRBS at the E8.5 stage, which led to highly reproducible data in independent embryos (Figure S1E in Additional file 1). Importantly, we verified that

the inactivation of one *Dnmt3* gene does not modify the expression of the other *Dnmt* genes in embryos (Figure S6B,C in Additional file 1). We found that the inactivation of *Dnmt3a* or *Dnmt3b* leads to a partial reduction in global methylation, indicating that the inactivation of one enzyme is compensated for by the other and that both enzymes cooperate to methylate the bulk genome (Figure 4B,C). The decrease in methylation is unidirectional with no signs of gain of methylation, confirming that these enzymes solely act as methylases (Figure 4D). Overall, the inactivation of *Dnmt3b* leads to a higher number of hypomethylated sequences and an increased amplitude in the loss of methylation compared with *Dnmt3a* (Figure 4E). Detailed quantification is given in Figure 4F and shows that the median methylation of methylated sequences in E8.5 embryos drops from 91% in wild type (WT) to 86% in *Dnmt3a*^{-/-} and 72% in *Dnmt3b*^{-/-} embryos. These variations are equally distributed in exons,

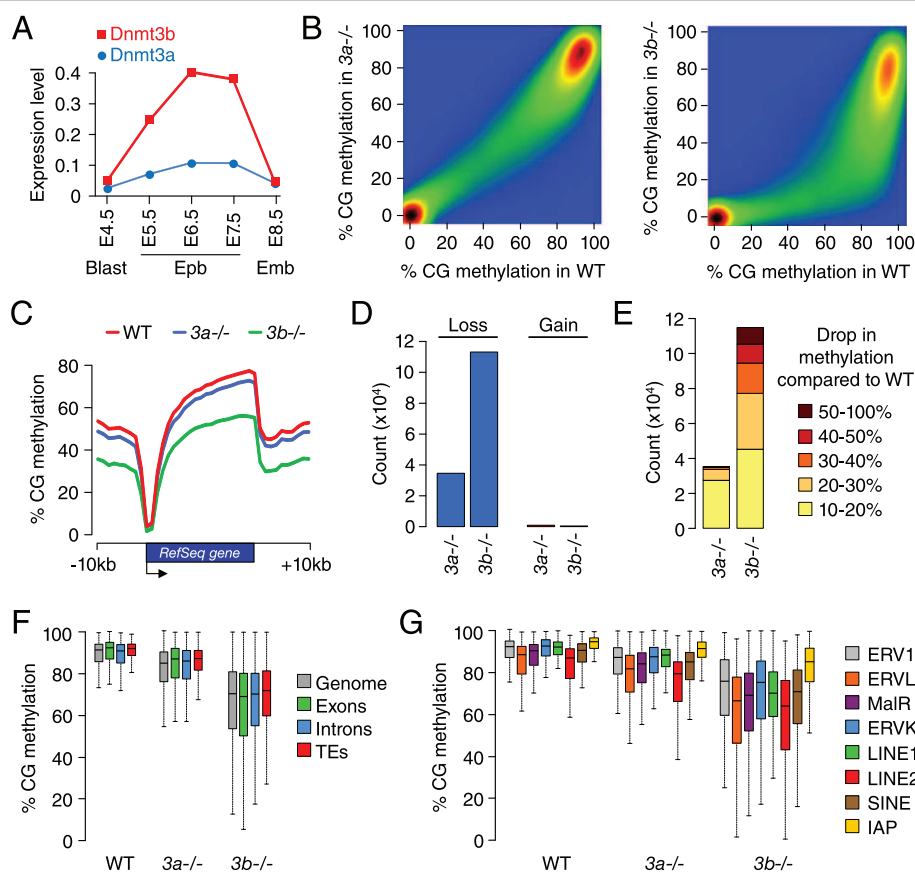


Figure 4 Methylome profiling in DNMT3A and DNMT3B-deficient embryos. (A) mRNA expression of *Dnmt3a/b* genes in embryos. Expression was measured by RT-qPCR on 5 to 10 pooled embryos and is depicted as a ratio relative to the expression of two housekeeping genes (*Actb* and *Rpl13a*). The primers used for RT-qPCR were designed in the last exons to amplify all isoforms. Blast, blastocyst; Epb, epiblast; Emb, embryo. (B) Pairwise comparison of CpG methylation in 400 bp tiles in wild-type (WT) compared with *Dnmt3a*^{-/-} and *Dnmt3b*^{-/-} E8.5 embryos. (C) Distribution of CpG methylation in RefSeq genes and 10 kb of flanking sequences in WT, *Dnmt3a*^{-/-} or *Dnmt3b*^{-/-} E8.5 embryos. (D) Number of 400 bp tiles that lose or gain more than 10% methylation in *Dnmt3a*^{-/-} and *Dnmt3b*^{-/-} compared with WT E8.5 embryos. (E) Detailed representation of the extent of methylation loss detected in 400 bp tiles in *Dnmt3a*^{-/-} and *Dnmt3b*^{-/-} compared with WT E8.5 embryos. (F) Boxplot representing the distribution of CpG methylation in the whole genome (400 bp tiles), exons, introns and transposable elements (TEs) in WT, *Dnmt3a*^{-/-} and *Dnmt3b*^{-/-} E8.5 embryos. Only sequences with >50% methylation in WT E8.5 embryos are considered. (G) Distribution of CpG methylation in several families of transposable elements in WT, *Dnmt3a*^{-/-} and *Dnmt3b*^{-/-} E8.5 embryos. IAP, intracisternal A-particle; LINE, long interspersed nuclear element; SINE, short interspersed nuclear element.

introns and transposable elements (Figure 4F). We also monitored methylation of various classes of transposable elements (long interspersed nuclear elements (LINEs), short interspersed nuclear element (SINEs), long terminal repeats) and found that they follow the same trend with a small decrease in methylation (approximately 5%) in *Dnmt3a*^{-/-} embryos and a more pronounced decrease (approximately 20%) in *Dnmt3b*^{-/-} embryos (Figure 4G). The exception is intracisternal A-particle elements, which are marginally affected in *Dnmt3a*^{-/-} and *Dnmt3b*^{-/-} embryos, which is consistent with previous data [10,14] and reflects that intracisternal A-particles maintain high methylation in pre-implantation stages. We conclude that DNMT3A and DNMT3B cooperate to establish DNA

methylation in embryos, with DNMT3B having a greater contribution than DNMT3A.

CpG islands are preferentially methylated by DNMT3B

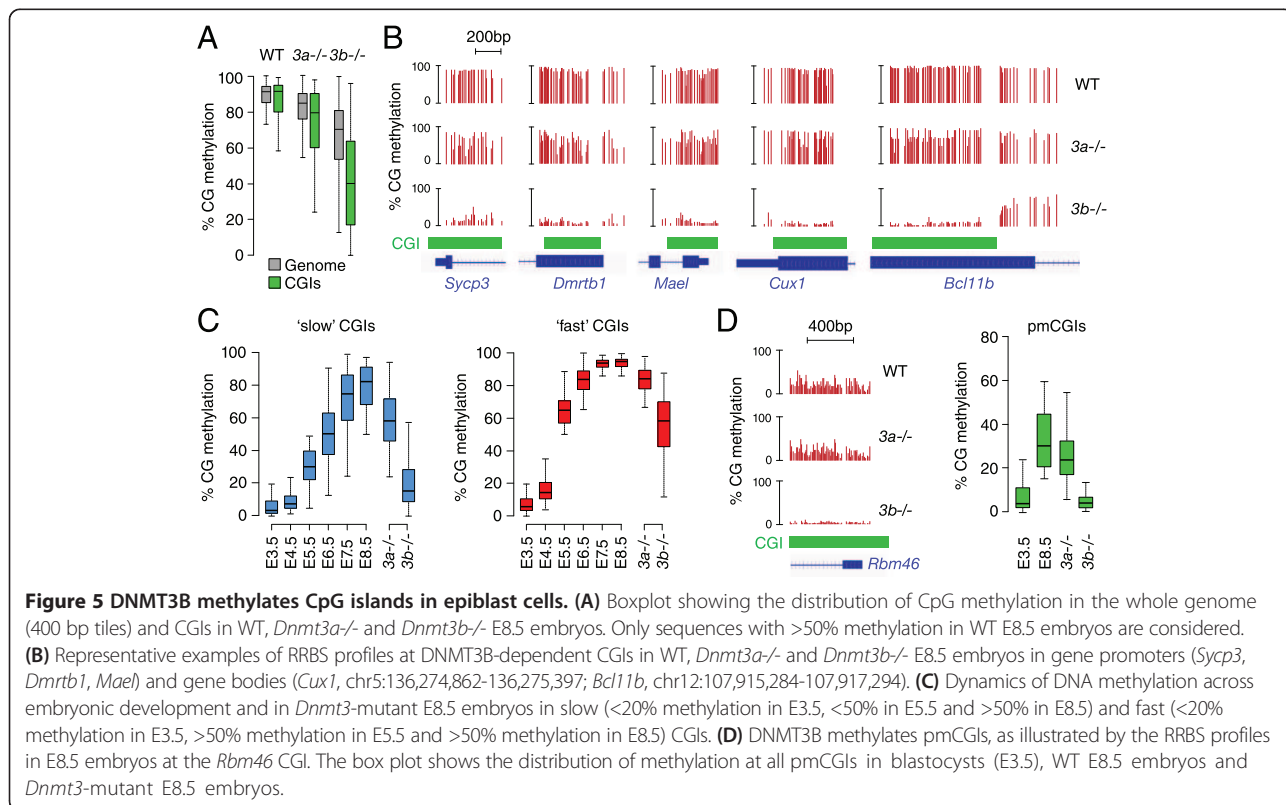
We next investigated if DNMT3A and DNMT3B have specific targets. As shown in Figure 4E, we identified severely hypomethylated sequences in *Dnmt3b*^{-/-} but not *Dnmt3a*^{-/-} embryos, indicating that only DNMT3B has specific targets for methylation. We identified 1,759 *Dnmt3b*-dependent targets defined as losing more than 60% methylation in *Dnmt3b*^{-/-} compared with WT embryos (Additional file 4). These targets are distributed in promoters, gene bodies and intergenic regions (Figure S7A in Additional file 1) and have an increased CpG density

(Figure S7B in Additional file 1), suggesting a preferential role for DNMT3B at CGIs. To verify this hypothesis, we monitored methylation at CGIs and found that they are more severely demethylated in *Dnmt3b*^{-/-} embryos compared with the bulk genome (Figure 5A). Many CGIs are markedly hypomethylated in *Dnmt3b*^{-/-} compared with WT embryos, such as in the promoters of *Sycp3*, *Dmrtb1*, *Mael*, and gene bodies of *Cux1* and *Bcl11b* (Figure 5B). Remarkably, the DNMT3B-dependent CGIs overlap with the ones that acquire delayed methylation in late epiblast cells (for example, *Sycp3* and *Bcl11b* in Figures 2F and 5B). We investigated this observation on a global scale and confirmed that the methylation of the ‘slow’ CGIs returns to levels close to those of blastocysts in *Dnmt3b*^{-/-} E8.5 embryos (Figure 5C, left panel). In contrast the methylation of the ‘fast’ CGIs is less affected by the inactivation of DNMT3B and can be partly compensated for by DNMT3A (Figure 5C, right panel). The methylation of the pmCGIs, which can be viewed as extreme cases of ‘slow’ CGIs, also strictly depends on DNMT3B activity (Figure 5D). Lastly, we show that the DNMT3B-dependent targets identified in E8.5 embryos are hypomethylated to a similar extent in limbs from E11.5 *Dnmt3b*^{-/-} embryos (Figure S7C in Additional file 1), demonstrating that DNMT3A is incapable of compensating DNMT3B at these targets even over a prolonged period of development. Altogether this reveals a specific

role for DNMT3B in the methylation of CpG-rich DNA in epiblast cells.

Influence of DNMT3B-dependent methylation on gene expression in embryos

To explore the influence of DNMT3B-dependent methylation on gene transcription, we conducted RNA-Seq in three WT and *Dnmt3b*^{-/-} E8.5 embryos (Figure S6D in Additional file 1). Overall we found relatively similar transcriptome profiles (Figure 6A). We identified 306 upregulated and 528 downregulated genes in *Dnmt3b*^{-/-} compared with WT embryos, but genes with the highest fold change are mostly upregulated (Figure 6A; Additional file 5). The reduction of promoter CGI methylation in *Dnmt3b*^{-/-} embryos strongly correlates with gene upregulation (Figure 6B), which leads to the ectopic activation of many full length germline transcripts that constitute 81% of the genes upregulated more than five-fold in *Dnmt3b*^{-/-} embryos (Figure 6C; Figure S8A,B in Additional file 1). The only other genes strongly upregulated in *Dnmt3b*^{-/-} embryos are genes of the *Rhox* cluster (*Gm9*, *Rhox4g*, *Rhox4e*, *Rhox9*) [14] and members of a family of X-linked imprinted genes (*Xlr3a/b/c*, *Xlr4a/b/c*) (Figure S8A in Additional file 1; Additional file 5). Thus, DNMT3B represses a small number of genes mainly associated with germline function. We validated the potent upregulation of germline genes by RT-qPCR in E8.5 *Dnmt3b*^{-/-}



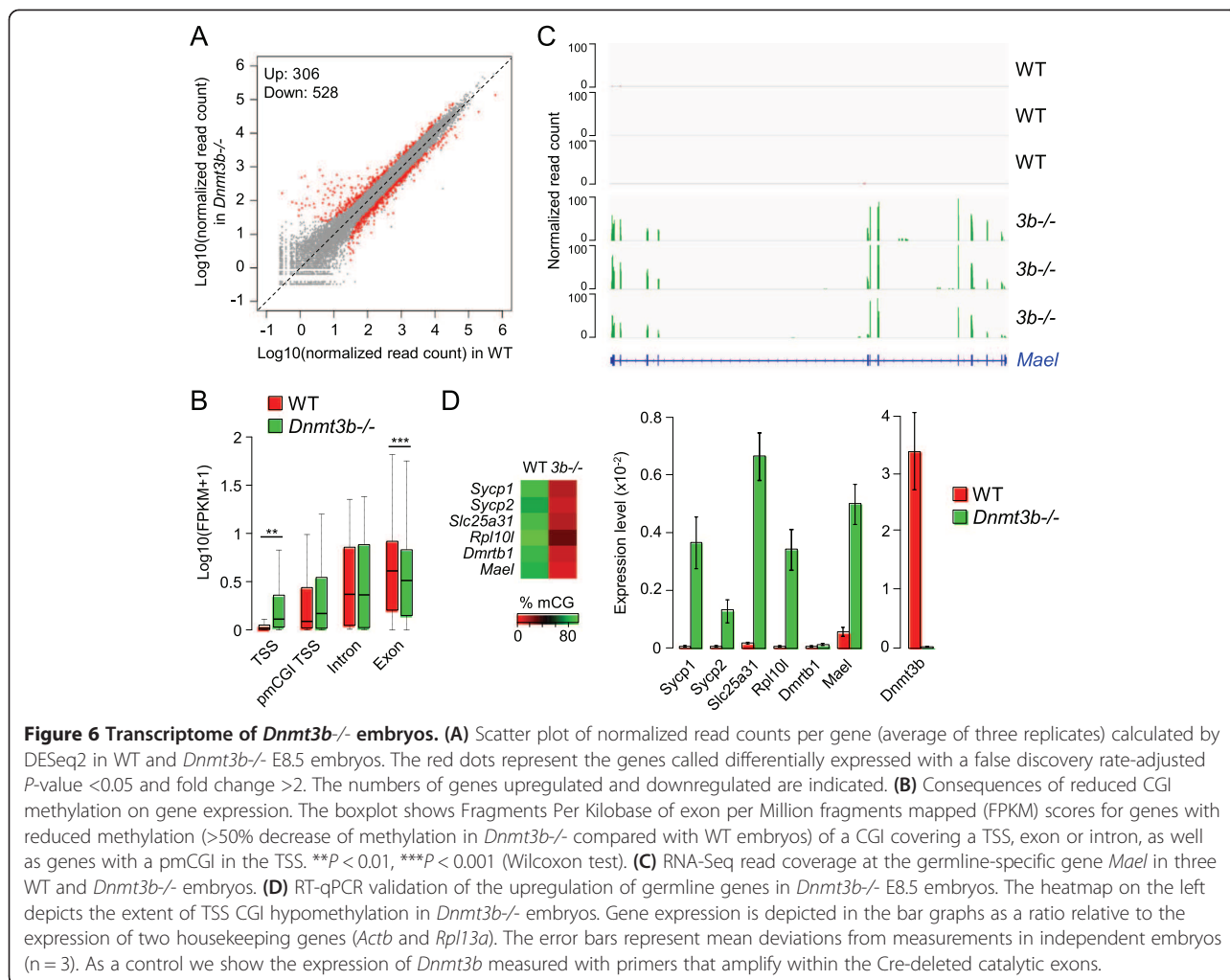


Figure 6 Transcriptome of *Dnmt3b*^{-/-} embryos. (A) Scatter plot of normalized read counts per gene (average of three replicates) calculated by DESeq2 in WT and *Dnmt3b*^{-/-} E8.5 embryos. The red dots represent the genes called differentially expressed with a false discovery rate-adjusted *P*-value <0.05 and fold change >2. The numbers of genes upregulated and downregulated are indicated. (B) Consequences of reduced CGI methylation on gene expression. The boxplot shows Fragments Per Kilobase of exon per Million fragments mapped (FPKM) scores for genes with reduced methylation (>50% decrease of methylation in *Dnmt3b*^{-/-} compared with WT embryos) of a CGI covering a TSS, exon or intron, as well as genes with a pmCGI in the TSS. ***p* < 0.01, ****p* < 0.001 (Wilcoxon test). (C) RNA-Seq read coverage at the germline-specific gene *Mael* in three WT and *Dnmt3b*^{-/-} embryos. (D) RT-qPCR validation of the upregulation of germline genes in *Dnmt3b*^{-/-} E8.5 embryos. The heatmap on the left depicts the extent of TSS CGI hypomethylation in *Dnmt3b*^{-/-} embryos. Gene expression is depicted in the bar graphs as a ratio relative to the expression of two housekeeping genes (*Actb* and *Rpl13a*). The error bars represent mean deviations from measurements in independent embryos (*n* = 3). As a control we show the expression of *Dnmt3b* measured with primers that amplify within the Cre-deleted catalytic exons.

embryos as well as limbs from E11.5 *Dnmt3b*^{-/-} embryos (Figure 6D; Figure S8C in Additional file 1), demonstrating that CGI methylation establishes long-term silencing of germline promoters throughout development. Interestingly, the absence of partial methylation at pmCGI promoters is also associated with a minor increase of transcript abundance in *Dnmt3b*^{-/-} embryos (Figure 6B), which was verified by RT-qPCR at three germline pmCGI genes (Figure S8C in Additional file 1). Thus, partial promoter methylation has a small contribution to promoter silencing at some genes. Consistent with a contribution of DNA methylation to the silencing of pluripotency genes in development [19], reduced promoter methylation is associated with a modest increase in transcript abundance of several pluripotency genes in *Dnmt3b*^{-/-} embryos, in particular *Dppa3/Stella* and *Dppa4* (Figure S8D in Additional file 1). Finally, we investigated the influence of intragenic CGI methylation on transcript abundance. Interestingly, RNA-Seq indicates that genes with reduced CGI methylation over an exon are significantly downregulated in *Dnmt3b*^{-/-} embryos (Figure 6B), consistent

with a role of intragenic CGI methylation in influencing the expression of the surrounding gene [20]. Overall, this shows that DNMT3B-dependent CGI methylation has a repressive function at promoters and a putative positive influence on transcript abundance in the body of genes.

DNMT3 knockouts distinguish germline from somatic imprinted differentially methylated regions

It has been suggested that the DNMT3 enzymes participate in the maintenance of DNA methylation in mammalian cells [21]. To ask if DNMT3A/B contribute to the maintenance of DNA methylation imprints, we monitored methylation at 17 imprinted gDMRs and found no signs of hypomethylation in *Dnmt3a*^{-/-} or *Dnmt3b*^{-/-} embryos (Figure 7A; Figure S9A in Additional file 1). This agrees with previous results obtained at the *Igf2r* gDMR [14] and two paternal gDMRs [22], and indicates that the individual DNMT3 enzymes are dispensable for the maintenance of methylation imprints *in vivo*. The possibility remains that DNMT3A/B are redundant for the maintenance of DNA methylation, which needs to be tested in double mutants.

In contrast to gDMRs, secondary somatic DMRs (sDMRs) acquire allele-specific DNA methylation after implantation. We profiled DNA methylation at several known sDMRs and observed that, in accordance with published data [23,24], they acquire methylation with variable kinetics in development (Figure 7B). Interestingly, these sDMRs all acquire methylation in a DNMT3B-dependent manner (Figure 7B). Accordingly, this is associated with a less than two-fold increase of some of these imprinted transcripts (*H19*, *Meg3*, *Cdkn1c*, *Mkrn3*) in *Dnmt3b*^{-/-} embryos (Figure S8E in Additional file 1). We then reasoned that our data can serve to accurately distinguish gDMRs from sDMRs: gDMRs are stable throughout embryogenesis and unaffected in *Dnmt3* mutants, whereas sDMRs gain DNMT3B-dependent methylation after implantation. Using these criteria, we reassessed the known gDMRs and confirmed that they inherit stable gametic methylation except for the *Exon1A* gDMR at the *Gnas* locus. The *Gnas* locus is unusual as previous studies identified two gDMRs in the *GnasXL* promoter and *Exon1A* (Figure 7C) [25,26]. We now reveal that the *Exon1A* DMR is a sDMR: it

inefficiently maintains maternal methylation in blastocysts and undergoes DNMT3B-dependent *de novo* methylation after implantation (Figure 7C; Figure S9B in Additional file 1). This suggests a revised model of imprinting at the *Gnas* locus under the control of only one gDMR.

DNMT3B methylates CpG islands on the inactive X chromosome in females

CGIs gain DNA methylation on the inactive copy of the X chromosome in female XX embryos to stabilize X-inactivation. To explore the role of DNMT3 enzymes in X-linked CGI methylation, we examined RRBS methylomes in female embryos. A visual inspection at promoters of X-inactivated genes shows that they acquire partial CGI methylation in female but not male E8.5 embryos (Figure 7D). We investigated this in a systematic way and found that all the CGIs on the X chromosome show a concordant gain of methylation in female embryos (Figure 7E). The methylation in female E8.5 embryos is only slightly lower than the one observed in a female adult liver (Figure 7E), indicating that most of the

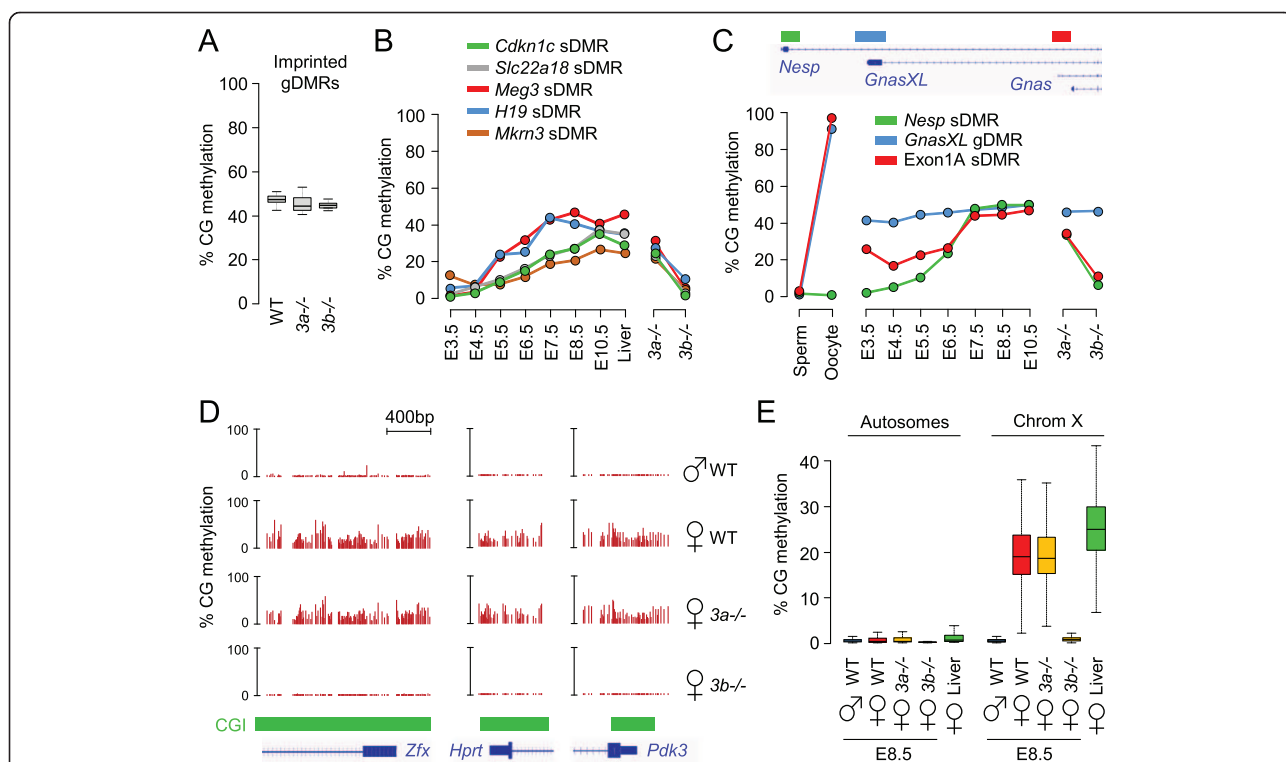


Figure 7 Role of DNMT3B at imprinted differentially methylated regions and X-linked CpG islands in embryos. **(A)** Box-plot representing the methylation of 17 imprinted germline DMRs in WT, *Dnmt3a*^{-/-} and *Dnmt3b*^{-/-} E8.5 embryos. **(B)** Dynamics of methylation throughout development and in *Dnmt3* knockout E8.5 embryos at somatic DMRs: *Cdkn1c* (chr7:143,459,734-143,460,383), *Slc22a18* (chr7:143,465,018-143,465,543), *Meg3* (chr12:109,540,809-109,541,073), *H19* (chr7:142,578,145-142,578,462) and *Mkrn3* (chr7:62,419,498-62,420,497). **(C)** Dynamics of methylation throughout development and in *Dnmt3* knockout E8.5 embryos at the three DMRs of the *Gnas* locus. The genomic organization of the locus and the position of the DMRs are shown on top of the graph. The detailed RRBS profiles are shown in Figure S9B in Additional file 1. **(D)** Examples of RRBS profiles at CGI promoters of three X-inactivated genes in female WT and *Dnmt3*-mutant embryos at the E8.5 stage. **(E)** Box plots of the global distribution of CGI methylation in autosomes (left) and on the X chromosome (right) in WT E8.5 embryos, female *Dnmt3*-mutant embryos and female adult liver.

X-linked CGI methylation is already established in early post-implantation embryos. In accordance with human data [27], we found that X-linked CGI methylation is not restricted to promoters but occurs at all CGIs in promoters, gene bodies and intergenic regions (Figure S10A in Additional file 1). In contrast, CGI methylation is not found at some of the genes known to escape X-inactivation in the mouse (Figure S10B in Additional file 1). Out of 13 described escapee genes in the mouse [28], 10 have their promoter covered in our dataset and 5 show no signs of promoter methylation (Figure S10C in Additional file 1). The other five genes might be mis-annotated escapees or could have alternative promoters. We then investigated the contribution of DNMT3 enzymes to X-linked methylation by examining RRBS methylomes generated in *Dnmt3a*^{-/-} and *Dnmt3b*^{-/-} female E8.5 embryos. Strikingly, X-linked CGI methylation is unaffected in *Dnmt3a*^{-/-} embryos but completely absent in *Dnmt3b*^{-/-} embryos (Figure 7D,E). DNMT3B methylates all the CGIs of the inactive X chromosome in promoters but also intragenic and intergenic regions (Figure S10A in Additional file 1). These data extend results obtained on candidate X-linked CGIs [29,30] and demonstrate that DNMT3B catalyzes methylation of all CGIs on the inactive X chromosome in female embryos.

Discussion

Using quantitative profiling of cytosine methylation at single-base resolution, we determined the kinetics of DNA methylation and the specificity of the *de novo* DNMT3 enzymes in mouse embryos *in vivo*. This extends previous findings on the dynamics of DNA methylation in mouse embryogenesis [4-8]. Our data provide a useful resource for investigating the inheritance and reprogramming of DNA methylation in embryos, notably by accurately distinguishing gDMRs from sDMRs at imprinted loci. We reassessed all the known gDMRs at imprinted loci and confirmed that they inherit gametic DNA methylation with no contribution of *de novo* methylation after fertilization. The exception is the *Exon1A* DMR at the *Gnas* locus, which was previously identified as a gDMR that maintains maternal methylation in blastocysts and post-implantation embryos [25]. Another gDMR was described in the *GnasXL* promoter, which made the *Gnas* locus an unusual case of imprinted loci with two gDMRs. We now reveal that the *Exon1A* DMR is only partially resistant to methylation reprogramming in blastocysts and gains DNMT3B-dependent methylation after implantation, which classifies it as a sDMR. This suggests a revised model of imprinting at the *Gnas* locus under the control of only one gDMR covering the *GnasXL* promoter, which then controls the establishment of the sDMRs *in cis*. In accordance with this model, the deletion of the *GnasXL* DMR influences methylation

of the *Exon1A* DMR in embryos, whereas the opposite is not true [26,31,32].

We show that the embryo acquires global DNA methylation within a short period of time between 4.5 and 6.5 days *post coitum* in epiblast cells. This is faster than when it is recapitulated *in vitro* in embryonic stem (ES) cells switched from 2i to serum conditions. In this case, bisulfite sequencing profiling revealed that global methylation is completed only 5 to 7 days after adapting ES cells from 2i to serum [33]. These differences could be due to different levels of expression of the DNMTs. One important question is whether *de novo* methylation occurs by default or is triggered by pre-existing factors to specific sites in the genome. Our kinetic study supports the default model of methylation because (1) methylation increases with similar kinetics throughout the genome, and (2) methylation increases rapidly as soon as DNMT3 genes are upregulated in early epiblast cells. As we studied whole populations of cells, however, we cannot exclude the existence of cell-to-cell heterogeneity in the kinetics of acquisition. This hints at a random mechanism of deposition of methylation rather than initiation and spreading from defined methylation centers, as has also been observed in ES cells switched from 2i to serum [33]. This model is consistent with studies showing that DNMT3 enzymes are recruited by default to chromatin via their PWWP domain [34-36].

Our results highlight the distinct regulation of DNA methylation at CGIs compared with the bulk genome: CGIs are mostly resistant to methylation at implantation, acquire delayed methylation in late epiblast cells, and frequently acquire partial methylation. Only 5% (713/14,085) of the annotated UCSC CGIs gain more than 50% methylation in the post-implantation embryo. The rarity of methylation at CGIs suggests that they are intrinsically protected from default methylation at the time of implantation. This could be mediated by proteins with CXXC domains that recognize stretches of CpG-rich DNA and could impose a chromatin structure that makes DNA refractory to methylation. The methylation-free state of CGIs could also be related to their activity as platforms for the binding of transcription factors. In line with this idea, mutations in transcription factor binding sites can relieve the protection from methylation at CGIs [37,38]. This model would predict that all CGIs are transcriptionally active or at least bound by the transcription machinery in embryos at the time of *de novo* methylation. An alternative possibility is that CGIs are 'methylatable' but that methylation is constantly removed via demethylation. The maintenance of the hypomethylated state of CGIs by demethylation could involve the action of TDG (thymine DNA glycosylase) or TET (ten-eleven translocation) proteins. Indeed, the inactivation of TDG in the mouse or the combined deficiency of TET1/2/3 in mouse ES cells leads

to the hypermethylation of some CGIs and impairs developmental potential [39–41]. In this context, the methylation of a subset of CGIs probably requires specific molecular pathways necessary to override the intrinsic resistance to methylation before allowing the deposition of DNA methylation. This would explain that CGI methylation is delayed compared with the bulk genome and can be partial due to inefficient targeting. To date very little is known about these molecular mechanisms that recruit DNA methylation to a subset of CGIs. Existing evidence indicates that it could involve non-coding RNAs [42] or DNA-binding repressors such as E2F6 [43].

Interestingly, intragenic and intergenic CGIs are much more prone to developmental DNA methylation than in the TSS. This was observed previously in differentiated lineages and suggests important roles for intragenic methylation in gene regulation [16,44,45]. In line with this idea, we show that intragenic CGI methylation is recruited to genes with important developmental functions in embryonic morphogenesis, signaling pathways and transcription regulation. In addition, our RNA-Seq data show that reduced intragenic CGI methylation in *Dnmt3b*^{-/-} embryos is associated with small changes in transcript abundance of the surrounding genes, suggesting that intragenic CGI methylation plays a role in the fine tuning of expression of developmental genes in the embryo. A positive correlation between intragenic DNA methylation and gene expression has been documented previously in normal and malignant mammalian cells (reviewed in [20]). Functionally, it can be speculated that intragenic methylation facilitates the transcription elongation or influences the posttranscriptional processing of the surrounding mRNA. Another possibility is that DNA methylation regulates the activity of alternative intragenic CGI promoters or promoters of regulatory non-coding RNAs. However, we found so far little evidence in the RNA-Seq data for activated transcription in the sense or antisense orientation at demethylated intragenic CGIs in *Dnmt3b*^{-/-} embryos (data not shown). Hence additional investigations are needed to explore the possible functional impact of intragenic CGI methylation in gene regulation [20]. In the TSS, CGI methylation is extremely rare (less than 1% have more than 20% methylation) and occurs almost exclusively at the promoters of germline-specific genes. This remarkable specificity, together with the fact that germline genes are durably reactivated in methylation-deficient embryos and cultured cells as shown by us and others [11,46–48], indicates that the silencing of the germline program in soma is a major evolutionary function of DNA methylation in mammals.

Mammalian genomes encode two *de novo* methyltransferases (DNMT3A and DNMT3B) but their respective contribution to embryonic methylation remained poorly characterized. We generated methylomes in embryos with catalytically inactive mutants of *Dnmt3a* or *Dnmt3b*

and found that, in agreement with data at candidate loci [7,10,14,29,43], DNMT3A and DNMT3B have both redundant and specific functions. First, we show that the inactivation of *Dnmt3a* or *Dnmt3b* induces a partial hypomethylation of the genome. The sum of the methylation in the knockout embryos is much higher than the methylation in WT embryos, indicating that the catalytic activities of both enzymes can compensate for each other and cooperate to methylate the genome. We note, however, that DNMT3B makes a greater contribution than DNMT3A, probably because of the higher expression of *Dnmt3b* compared with *Dnmt3a* in epiblasts as observed here by RT-qPCR (Figure 4A) and previously by immunostaining [10,49]. Their redundant function might have evolved to ensure robust and efficient methylation in embryos. Despite the global redundancy, DNMT3B also has a specific role in the methylation of many CGIs on autosomes and the inactive X chromosome that are dramatically hypomethylated in *Dnmt3b*^{-/-} embryos. This highlights again that DNA methylation at CGIs is controlled by different molecular pathways compared with the bulk genome, which involve the preferential recruitment of DNMT3B. Our data contribute to explain the more severe phenotype of *Dnmt3b*^{-/-} compared with *Dnmt3a*^{-/-} mice [10]. DNMT3B makes a greater contribution to genome methylation and specifically methylates a set of CpG-dense sequences associated with developmental and imprinted genes, which leads to the deregulation of germline, pluripotency and imprinted genes and hundreds of other direct or indirect targets in *Dnmt3b*-deficient embryos. Which of these genes contribute the most to the embryonic lethality of *Dnmt3b*^{-/-} animals remains to be investigated. The preferential role of DNMT3B at CGIs is compatible with data showing that mice or human ICF syndrome (immunodeficiency, centromeric instability, facial anomalies) patients with DNMT3B mutations have reduced methylation at CGIs on autosomes and the inactive X chromosome [43,50,51], suggesting a functional conservation of DNMT3B function between mice and humans. Our findings also shed new light on the possible molecular dysfunctions caused by the mutations of DNMT3B in the human ICF syndrome and might help to identify new epigenetically deregulated targets for diagnosis [52].

Conclusions

We revealed the target specificities of the *de novo* methyltransferases DNMT3A and DNMT3B in mouse development *in vivo*. We show that they have redundant catalytic functions in global genome methylation at implantation, and that DNMT3B specifically methylates a defined set of CpG islands on autosomes and on the X chromosome. This indicates that DNMT3 enzymes evolved to play both redundant and specific functions in mammalian embryos.

Further studies are required to elucidate the molecular mechanisms responsible for the recruitment of DNMT3B activity to CpG-dense regions, which might identify new pathways inducing abnormal CGI methylation in cancer.

Materials and methods

Biological samples

All embryos were obtained from naturally mated C57BL/6 mice. We designated the morning of the vaginal plug as E0.5 and performed all the dissections at the same hour of the day (1 pm). Blastocysts (E3.5 to E4.5) were collected by flushing the uteri with M2 medium. After implantation, we manually dissected individual embryos in M2 medium. At E5.5 to E7.5 we separated the epiblast from the extra-embryonic tissues by manual dissection. We dissected whole embryos at E8.5 to E10.5, and dissected forelimb buds from E11.5 embryos. Sperm was isolated from the caput epididymis of adult CD-1 mice. We prepared genomic DNA by proteinase K digestion, phenol/chloroform extraction and precipitation with ethanol. To generate catalytically inactive mutants of *Dnmt3* genes, we crossed *Dnmt3a-2lox* [53] and *Dnmt3b-2lox* [54] mice (provided by the Massachusetts General Hospital, Boston, MA, USA) with a C57BL/6 ACTB-Cre deleter line [55] (provided by the Institut Clinique de la Souris, Illkirch, France). *Dnmt3* knockout embryos were then obtained by mating heterozygous males and females.

Preparation of RRBS libraries

We prepared RRBS libraries from 100 pooled E3.5 blastocysts, 50 pooled E4.5 blastocysts, 25 pooled E5.5 epiblasts, 15 pooled E6.5 epiblasts, and 10 pooled E7.5 epiblasts. At E8.5, we prepared RRBS libraries from a pool of embryos as well as two individual WT, *Dnmt3a-/-* and *Dnmt3b-/-* embryos. At E10.5, we prepared RRBS libraries from pools of WT embryos. At E11.5 we prepared RRBS libraries from limbs of two WT and two *Dnmt3b-/-* embryos. RRBS libraries were prepared according to a published protocol [56] with modifications. Briefly, we digested 25 to 100 ng of genomic DNA for 5 h with *MspI* (Thermo Scientific, Waltham, MA, USA) followed by end-repair, A-tailing (with Klenow fragment, Thermo Scientific) and ligation to paired-end methylated adapters (with T4 DNA ligase, Thermo Scientific) in Tango 1X buffer. We purified fragments in the range 150 to 400 bp (insert plus adapter size) by electrophoresis on a 3% (w/v) agarose 0.5X TBE gel with the MinElute gel extraction kit (Qiagen). We then performed two rounds of bisulfite conversion with the EpiTect kit (Qiagen) according to the manufacturer's instructions. Final RRBS libraries were PCR amplified with PfUTurbo Cx hotstart DNA polymerase (Agilent, Santa Clara, CA, USA) and indexed PE Illumina primers using the following PCR conditions: 95°C for 2 minutes, 14 to 16 cycles (95°C for 30 s, 65°C for 30 s, 72°C for 45 s),

72°C for 7 minutes. The libraries were purified with AMPure magnetic beads (Beckman Coulter, Brea, CA, USA), quantified with a Qubit fluorometer (Life Technologies, Carlsbad, CA, USA) and verified by loading 10 ng of the library on a 4-20% Criterion precast polyacrylamide gel (Biorad, Hercules, CA, USA) stained with SYBR Green. The libraries were paired-end sequenced (2 × 75 bp) in multiplex on an Illumina HiSeq2000 by Integragen SA (Evry, France) to generate an average of 30 million pairs of reads per sample (Figure S1A in Additional file 1).

Processing of RRBS sequencing reads

We performed quality control checks on sequencing reads with FastQC [57]. Reads were trimmed with Trim Galore (v0.2.1) [58] to remove adapter sequences and low-quality ends with a Phred score below 20. Trim Galore was run in *-rrbs* mode to remove two additional bases artificially introduced at the *MspI* restriction sites. We aligned sequencing reads to the mouse mm10 genome with BSMAP (v2.74) [59] using the RRBS mode. For the mapping, we allowed a maximum of two mismatches and an insertion size for paired-end sequences of between 30 and 400 bp. We extracted methylation scores as the ratio of the number of Cs over the total number of Cs and Ts. We combined CpG methylation ratios from both strands and filtered for a minimum sequencing depth of 8×. We estimated the bisulfite conversion efficiency by calculating the C to T conversion at non-CpG sites, which was in most cases greater than 99.5% (Figure S1A in Additional file 1). Methylation scores were visualized with the IGV browser [60].

RRBS data analysis

RefSeq genes, transposons and CGI annotations were retrieved from the UCSC mm10 annotation. We filtered transposons to have a minimum size of 200 bp. HCPs (high CpG promoters), ICPs (intermediate CpG promoters) and LCPs (low CpG promoters) were annotated as previously described [19]. The genomic coordinates of the imprinted DMRs were retrieved from the Wamidex imprinting database [61]. The genomic coordinates of canyons were retrieved from a published dataset [15]. For oocyte methylation, we retrieved a published RRBS dataset (GSE34864) [6] and averaged the methylation scores from all the oocyte replicates in the dataset. To annotate CGIs, we measured the distance between the middle of the CGI and the closest RefSeq TSS and overlapped CGIs with RefSeq exons and introns. We defined the CGI categories as follows: TSS, the CGI is less than 1,000 bp from a TSS; exon, the CGI is more than 1,000 bp from a TSS and overlaps at least 1 bp of an exon; intron, the CGI is more than 1,000 bp from a TSS and is entirely included in an intron; intergenic, the CGI does not meet any of the previous criteria. To follow the global dynamics

at the genome level, we averaged methylation in 400 bp tiles containing at least three CpGs. To identify DNMT3B-dependent targets, we selected genomic tiles with a difference in methylation (WT minus *Dnmt3b*^{-/-}) greater than 60%, and then merged overlapping tiles. To analyze methylation from single alleles, we mapped sequencing reads with Bismark [62], which returns mapping information on single reads. We then processed the Bismark output to extract methylation scores of individual sequenced molecules. We performed gene ontology analysis using the DAVID functional annotation tool [63]. All data processing and representation were performed with the R software using custom developed scripts.

Custom CGI annotation

To generate our custom CGI annotation, we split the genome into 150 bp sliding windows with a 25 bp offset and selected windows with a GC percentage greater than 55% and a CpG ratio (observed/expected) greater than 0.65. We then merged the windows closer than 50 bp and selected the windows larger than 250 bp. To identify high-confidence custom promoter CGIs (Additional file 3), we used the same procedure with a minimal size of 225 bp, and then filtered the custom CGIs to be less than 800 bp from a RefSeq TSS.

Bisulfite sequencing and cloning

Bisulfite conversion of genomic DNA was performed with the EpiTect kit (Qiagen). We performed PCR amplification of converted DNA followed by cloning as described [7]. Cloned PCR products were sequenced, aligned with the BISMAR software [64] and filtered to remove clonal biases. The sequences of the primers are provided in the Additional file 6.

RNA-Seq

We prepared RNA-Seq libraries from three WT and *Dnmt3b*^{-/-} littermate embryos collected at E8.5. We extracted total RNAs from the embryos with the RNeasy Protect Mini Kit (Qiagen) and verified the integrity of RNAs with a Bioanalyzer (Agilent Technologies). RNA-Seq libraries were prepared from 500 ng of total RNA by Integragen SA using ribosomal RNA depletion with the Ribo-Zero Gold kit (Epicentre, Madison, WI, USA) and the TruSeq Stranded Total RNA Sample preparation kit (Illumina), followed by paired-end sequencing (2 × 75bp) on an Illumina HiSeq2000. The total numbers of paired-end reads for each library are as follows: WT1, 73,493,661 pairs; WT2, 52,488,499 pairs; WT3, 64,138,495 pairs; KO1, 61,632,013 pairs; KO2, 55,357,040 pairs; KO3, 60,179,105 pairs. We performed quality control checks on sequencing reads with FastQC [57] and aligned reads to the mouse mm10 genome with TopHat2 (v2.0.12) [65]. For data visualization, we generated BigWig files of

normalized read counts per base with bam2wig.py in the RSeQC package (v2.4) [66] using only reads that map uniquely in the genome. We calculated raw read counts in mouse RefSeq exons from the BAM files with HTseq-count (v0.6.0) [67] and used these counts to identify differentially expressed genes with DESeq2 (v1.4.5) [68]. Genes were called differentially expressed if the false discovery rate-adjusted *P*-value between WT and knockout was lower than 0.05 and the fold change greater than 2. Normalized counts and Fragments Per Kilobase of exon per Million fragments mapped (FPKM) scores were calculated with the 'counts' and 'fpkm' functions of DESeq2.

RT-qPCR

RNAs were reverse transcribed with the Maxima first strand cDNA synthesis kit (Thermo Scientific) using a combination of oligo(dT) and random hexamer primers. RT-qPCR was performed with the Fast SYBR Green Master Mix (Life Technologies) on a StepOnePlus real-time PCR system (Life Technologies). We used fast PCR cycling conditions as follows: 95°C for 20 s, 40 cycles (95°C for 20 s, 64°C for 30 s), followed by a dissociation curve. We performed qPCR measurements in triplicate reactions and normalized to the expression of two house-keeping genes (*Rpl13a*, *Actb*). In parallel we systematically amplified no-RT controls to rule out the presence of contaminating genomic DNA. Primer sequences for qPCR are provided in Additional file 6.

Data access

The RRBS and RNA-Seq data have been deposited at the NCBI Gene Expression Omnibus database [69] under accession number GSE60334.

Additional files

Additional file 1: Figure S1. Features of RRBS methylomes in mouse embryos. **Figure S2.** Distribution and kinetics of DNA methylation at custom CpG islands. **Figure S3.** Kinetics of *de novo* methylation at CpG islands in embryos. **Figure S4.** Comparison of CpG island methylation with gene expression in embryos and adult tissues. **Figure S5.** Bisulfite cloning and sequencing of promoter pmCGIs in adult liver. **Figure S6.** Expression of *Dnmt* genes in WT and knockout embryos. **Figure S7.** Analysis of DNMT3B-specific targets in mouse embryos. **Figure S8.** Expression of germline, pluripotency and imprinted genes in *Dnmt3b*^{-/-} embryos. **Figure S9.** RRBS methylation profiles at imprinted loci throughout development and in *Dnmt3* knockout embryos. **Figure S10.** CpG island methylation on the X chromosome and at genes that escape X-inactivation in female embryos.

Additional file 2: Table S1. List of CpG island methylation scores.

Additional file 3: Table S2. High confidence list of methylated promoter CpG islands.

Additional file 4: Table S3. List of DNMT3B-dependent regions.

Additional file 5: Table S4. Differentially expressed genes in *Dnmt3b* knockout E8.5 embryos.

Additional file 6: Table S5. List of primer sequences.

Abbreviations

CGI: CpG island; DMR: differentially methylated region; DNMT: DNA methyltransferase; E: embryonic day; ES: embryonic stem; gDMR: germline differentially methylated region; pmCGI: partially methylated CpG island; RRBS: reduced representation bisulfite sequencing; RT-qPCR: quantitative RT-PCR; sDMR: somatic differentially methylated region; TSS: transcription start site; WT: wild type.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

GA performed embryo dissections and all experiments. SG developed R scripts and performed analysis of sequencing data. AB optimized the RRBS protocol and participated in sample preparation. MW supervised the research and performed data analysis. GA, SG and MW wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Jesus del Mazo and Florence Cammas for contributing samples. This research has received funding from the ATIP-AVENIR program, the Association pour la Recherche contre le Cancer (ARC SFI20121205729), the Ligue Contre le Cancer, the MEDDTL (11-MRES-PNRPE-9-CVS-072), the EpiGeneSys European Network of Excellence, and the European Research Council under the European Union's Seventh Framework Programme (ERC consolidator grant number 615371).

Received: 17 July 2014 Accepted: 14 November 2014

Published online: 03 December 2014

References

- Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, Wirbelauer C, Oakeley EJ, Gaidatzis D, Tiwari VK, Schubeler D: **DNA-binding factors shape the mouse methylome at distal regulatory regions.** *Nature* 2011, **480**:490–495.
- Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D: **Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome.** *Nat Genet* 2007, **39**:457–466.
- Hon GC, Rajagopal N, Shen Y, McCleary DF, Yue F, Dang MD, Ren B: **Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues.** *Nat Genet* 2013, **45**:1198–1206.
- Kobayashi H, Sakurai T, Imai M, Takahashi N, Fukuda A, Yayoi O, Sato S, Nakabayashi K, Hata K, Sotomaru Y, Suzuki Y, Kono T: **Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks.** *PLoS Genet* 2012, **8**:e1002440.
- Smallwood SA, Tomizawa S, Krueger F, Ruf N, Carli N, Segonds-Pichon A, Sato S, Hata K, Andrews SR, Kelsey G: **Dynamic CpG island methylation landscape in oocytes and preimplantation embryos.** *Nat Genet* 2011, **43**:811–814.
- Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, Regev A, Meissner A: **A unique regulatory phase of DNA methylation in the early mammalian embryo.** *Nature* 2012, **484**:339–344.
- Borgel J, Guibert S, Li Y, Chiba H, Schubeler D, Sasaki H, Forne T, Weber M: **Targets and dynamics of promoter DNA methylation during early mouse development.** *Nat Genet* 2010, **42**:1093–1100.
- Wang L, Zhang J, Duan J, Gao X, Zhu W, Lu X, Yang L, Li G, Ci W, Li W, Zhou Q, Aluru N, Tang F, He C, Huang X, Liu J: **Programming and inheritance of parental DNA methylomes in mammals.** *Cell* 2014, **157**:979–991.
- Lei H, Oh SP, Okano M, Juttermann R, Goss KA, Jaenisch R, Li E: **De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells.** *Development* 1996, **122**:3195–3205.
- Okano M, Bell DW, Haber DA, Li E: **DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.** *Cell* 1999, **99**:247–257.
- Hackett JA, Reddington JP, Nestor CE, Dunican DS, Branco MR, Reichmann J, Reik W, Surani MA, Adams IR, Meehan RR: **Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline.** *Development* 2012, **139**:3623–3632.
- Toyoda S, Kawaguchi M, Kobayashi T, Tarusawa E, Toyama T, Okano M, Oda M, Nakauchi H, Yoshimura Y, Sanbo M, Hirabayashi M, Hirayama T, Hirabayashi T, Yagi T: **Developmental epigenetic modification regulates stochastic expression of clustered protocadherin genes, generating single neuron diversity.** *Neuron* 2014, **82**:94–108.
- Li JY, Pu MT, Hirasawa R, Li BZ, Huang YN, Zeng R, Jing NH, Chen T, Li E, Sasaki H, Xu GL: **Synergistic function of DNA methyltransferases Dnmt3a and Dnmt3b in the methylation of Oct4 and Nanog.** *Mol Cell Biol* 2007, **27**:8748–8759.
- Oda M, Yamagiwa A, Yamamoto S, Nakayama T, Tsumura A, Sasaki H, Nakao K, Li E, Okano M: **DNA methylation regulates long-range gene silencing of an X-linked homeobox gene cluster in a lineage-specific manner.** *Genes Dev* 2006, **20**:3382–3394.
- Jeong M, Sun D, Luo M, Huang Y, Challen GA, Rodriguez B, Zhang X, Chavez L, Wang H, Hannah R, Kim SB, Yang L, Ko M, Chen R, Gottgens B, Lee JS, Gunaratne P, Godley LA, Darlington GJ, Rao A, Li W, Goodell MA: **Large conserved domains of low DNA methylation maintained by Dnmt3a.** *Nat Genet* 2014, **46**:17–23.
- Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr AR, James KD, Turner DJ, Smith C, Harrison DJ, Andrews R, Bird AP: **Orphan CpG islands identify numerous conserved promoters in the mammalian genome.** *PLoS Genet* 2010, **6**:e1001134.
- Proudhon C, Duffie R, Ajjan S, Cowley M, Iranzo J, Carbajosa G, Saadeh H, Holland ML, Oakey RJ, Rakyen VK, Schulz R, Bourc'his D: **Protection against de novo methylation is instrumental in maintaining parent-of-origin methylation inherited from the gametes.** *Mol Cell* 2012, **47**:909–920.
- Chotalia M, Smallwood SA, Ruf N, Dawson C, Lucifero D, Frontera M, James K, Dean W, Kelsey G: **Transcription is required for establishment of germline methylation marks at imprinted genes.** *Genes Dev* 2009, **23**:105–117.
- Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, Stadler MB, Bibel M, Schubeler D: **Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors.** *Mol Cell* 2008, **30**:755–766.
- Kulis M, Queiros AC, Beekman R, Martin-Subero JI: **Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer.** *Biochim Biophys Acta* 2013, **1829**:1161–1174.
- Jones PA, Liang G: **Rethinking how DNA methylation patterns are maintained.** *Nat Rev Genet* 2009, **10**:805–811.
- Hirasawa R, Chiba H, Kaneda M, Tajima S, Li E, Jaenisch R, Sasaki H: **Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development.** *Genes Dev* 2008, **22**:1607–1616.
- Sato S, Yoshida W, Soejima H, Nakabayashi K, Hata K: **Methylation dynamics of IG-DMR and Gt2-DMR during murine embryonic and placental development.** *Genomics* 2011, **98**:120–127.
- Bhogal B, Arnaudo A, Dymkowski A, Best A, Davis TL: **Methylation at mouse Cdkn1c is acquired during postimplantation development and functions to maintain imprinted expression.** *Genomics* 2004, **84**:961–970.
- Liu J, Yu S, Litman D, Chen W, Weinstein LS: **Identification of a methylation imprint mark within the mouse Gnas locus.** *Mol Cell Biol* 2000, **20**:5808–5817.
- Williamson CM, Turner MD, Ball ST, Nottingham WT, Glenister P, Fray M, Tymowska-Lalanne Z, Plagge A, Powles-Glover N, Kelsey G, Maconochie M, Peters J: **Identification of an imprinting control region affecting the expression of all transcripts in the Gnas cluster.** *Nat Genet* 2006, **38**:350–355.
- Sharp AJ, Stathaki E, Migliavacca E, Brahmachary M, Montgomery SB, Dupre Y, Antonarakis SE: **DNA methylation profiles of human active and inactive X chromosomes.** *Genome Res* 2011, **21**:1592–1600.
- Yang F, Babak T, Shendure J, Distcheu CM: **Global survey of escape from X inactivation by RNA-sequencing in mouse.** *Genome Res* 2010, **20**:614–622.
- Gendrel AV, Apedaille A, Coker H, Termanis A, Zvetkova I, Godwin J, Tang YA, Huntley D, Montana G, Taylor S, Giannoulatou E, Heard E, Stancheva I, Brockdorff N: **Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome.** *Dev Cell* 2012, **23**:265–279.
- Gendrel AV, Tang YA, Suzuki M, Godwin J, Nesterova TB, Grealley JM, Heard E, Brockdorff N: **Epigenetic functions of smchd1 repress gene clusters on the inactive X chromosome and on autosomes.** *Mol Cell Biol* 2013, **33**:3150–3165.
- Liu J, Chen M, Deng C, Bourc'his D, Nealon JG, Erlichman B, Bestor TH, Weinstein LS: **Identification of the control region for tissue-specific**

- imprinting of the stimulatory G protein alpha-subunit. *Proc Natl Acad Sci U S A* 2005, **102**:5513–5518.
32. Williamson CM, Ball ST, Nottingham WT, Skinner JA, Plagge A, Turner MD, Powles N, Hough T, Papworth D, Fraser WD, Maconochie M, Peters J: **A cis-acting control region is required exclusively for the tissue-specific imprinting of Gnas.** *Nat Genet* 2004, **36**:894–899.
 33. Habibi E, Brinkman AB, Arand J, Kroeze LI, Kerstens HH, Matarese F, Lepikhov K, Gut M, Brun-Heath I, Hubner NC, Benedetti R, Altucci L, Jansen JH, Walter J, Gut IG, Marks H, Stunnenberg HG: **Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells.** *Cell Stem Cell* 2013, **13**:360–369.
 34. Jeong S, Liang G, Sharma S, Lin JC, Choi SH, Han H, Yoo CB, Egger G, Yang AS, Jones PA: **Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA.** *Mol Cell Biol* 2009, **29**:5366–5376.
 35. Kashiwagi K, Nimura K, Ura K, Kaneda Y: **DNA methyltransferase 3b preferentially associates with condensed chromatin.** *Nucleic Acids Res* 2011, **39**:874–888.
 36. Ge YZ, Pu MT, Gowher H, Wu HP, Ding JP, Jeltsch A, Xu GL: **Chromatin targeting of de novo DNA methyltransferases by the PWWP domain.** *J Biol Chem* 2004, **279**:25447–25454.
 37. Brandeis M, Frank D, Keshet I, Siegfried Z, Mendelsohn M, Nemes A, Temper V, Razin A, Cedar H: **Sp1 elements protect a CpG island from de novo methylation.** *Nature* 1994, **371**:435–438.
 38. Lienert F, Wirbelauer C, Som I, Dean A, Mohn F, Schubeler D: **Identification of genetic elements that autonomously determine DNA methylation states.** *Nat Genet* 2011, **43**:1091–1097.
 39. Dawlaty MM, Breiling A, Le T, Barrasa MI, Raddatz G, Gao Q, Powell BE, Cheng AW, Faulk KF, Lyko F, Jaenisch R: **Loss of tet enzymes compromises proper differentiation of embryonic stem cells.** *Dev Cell* 2014, **29**:102–111.
 40. Cortazar D, Kunz C, Selfridge J, Lettieri T, Saito Y, MacDougall E, Wirz A, Schuermann D, Jacobs AL, Siegrist F, Steinacher R, Jiricny J, Bird A, Schar P: **Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability.** *Nature* 2011, **470**:419–423.
 41. Cortellino S, Xu J, Sannai M, Moore R, Caretti E, Cigliano A, Le Coz M, Devarajan K, Wessels A, Soprano D, Abramowitz LK, Bartolomei MS, Rambow F, Bassi MR, Bruno T, Fanciulli M, Renner C, Klein-Szanto AJ, Matsumoto Y, Kobi D, Davidson I, Alberti C, Larue L, Bellacosa A: **Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair.** *Cell* 2011, **146**:67–79.
 42. Schmitz KM, Mayer C, Postepska A, Grummt I: **Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes.** *Genes Dev* 2010, **24**:2264–2269.
 43. Velasco G, Hube F, Rollin J, Neuillet D, Philippe C, Bouzinba-Segard H, Galvani A, Viegas-Pequignot E, Francastel C: **Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues.** *Proc Natl Acad Sci U S A* 2010, **107**:9281–9286.
 44. Maunakea AK, Nagarajan RP, Bilienky M, Ballinger TJ, D'Souza C, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y, Turecki G, Delaney A, Varhol R, Thiessen N, Shchors K, Heine VM, Rowitch DH, Xing X, Fiore C, Schillebeeckx M, Jones SJ, Haussler D, Marra MA, Hirst M, Wang T, Costello JF: **Conserved role of intragenic DNA methylation in regulating alternative promoters.** *Nature* 2010, **466**:253–257.
 45. Deaton AM, Webb S, Kerr AR, Illingworth RS, Guy J, Andrews R, Bird A: **Cell type-specific DNA methylation at intragenic CpG islands in the immune system.** *Genome Res* 2011, **21**:1074–1086.
 46. Karimi MM, Goyal P, Maksakova IA, Bilienky M, Leung D, Tang JX, Shinkai Y, Mager DL, Jones S, Hirst M, Lorincz MC: **DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs.** *Cell Stem Cell* 2011, **8**:676–687.
 47. Fouse SD, Shen Y, Pellegrini M, Cole S, Meissner A, Van Neste L, Jaenisch R, Fan G: **Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation.** *Cell Stem Cell* 2008, **2**:160–169.
 48. Rutledge CE, Thakur A, O'Neill KM, Irwin RE, Sato S, Hata K, Walsh CP: **Ontogeny, conservation and functional significance of maternally inherited DNA methylation at two classes of non-imprinted genes.** *Development* 2014, **141**:1313–1323.
 49. Watanabe D, Suetake I, Tada T, Tajima S: **Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis.** *Mech Dev* 2002, **118**:187–190.
 50. Bourchis D, Miniou P, Jeanpierre M, Molina Gomes D, Dupont J, De Saint-Basile G, Maraschio P, Tiepolo L, Viegas-Pequignot E: **Abnormal methylation does not prevent X inactivation in ICF patients.** *Cytogenet Cell Genet* 1999, **84**:245–252.
 51. Hansen RS, Stoger R, Wijmenga C, Stanek AM, Canfield TK, Luo P, Matarazzo MR, D'Esposito M, Feil R, Gimelli G, Weemaes CM, Laird CD, Gartler SM: **Escape from gene silencing in ICF syndrome: evidence for advanced replication time as a major determinant.** *Hum Mol Genet* 2000, **9**:2575–2587.
 52. Velasco G, Walton EL, Sterlin D, Hedouin S, Nitta H, Yuya I, Fouyssac F, Megarbane A, Sasaki H, Picard C, Francastel C: **Germline genes hypomethylation and expression define a molecular signature in peripheral blood of ICF patients: implications for diagnosis and etiology.** *Orphanet J Rare Dis* 2014, **9**:56.
 53. Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, Sasaki H: **Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting.** *Nature* 2004, **429**:900–903.
 54. Dodge JE, Okano M, Dick F, Tsujimoto N, Chen T, Wang S, Ueda Y, Dyson N, Li E: **Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization.** *J Biol Chem* 2005, **280**:17986–17991.
 55. Birling MC, Dierich A, Jacquot S, Herault Y, Pavlovic G: **Highly-efficient, fluorescent, locus directed cre and FlpO deleter mice on a pure C57BL/6 N genetic background.** *Genesis* 2012, **50**:482–489.
 56. Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A: **Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling.** *Nat Protoc* 2011, **6**:468–481.
 57. FastQC. [<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>]
 58. Trim Galore. [http://www.bioinformatics.babraham.ac.uk/projects/trim_galore]
 59. BSMAP. [<https://code.google.com/p/bsmap>]
 60. Integrative Genomics Viewer. [<http://www.broadinstitute.org/igv/>]
 61. Wamindex Imprinting Database. [<https://atlas.genetics.kcl.ac.uk>]
 62. Bismark. [<http://www.bioinformatics.babraham.ac.uk/projects/bismark>]
 63. DAVID Functional Annotation Tool. [<http://david.abcc.ncifcrf.gov/>]
 64. BISMA software. [<http://services.ibt.uni-stuttgart.de/BDPC/BISMA>]
 65. TopHat read mapper. [<http://ccb.jhu.edu/software/tophat>]
 66. RSeQC Package. [<http://rseqc.sourceforge.net/>]
 67. HTseq-count. [<http://www-huber.embl.de/users/anders/HTSeq/doc/count.html>]
 68. DESeq2 Package. [<http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html>]
 69. NCBI Gene Expression Omnibus. [<http://www.ncbi.nlm.nih.gov/geo/>]

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



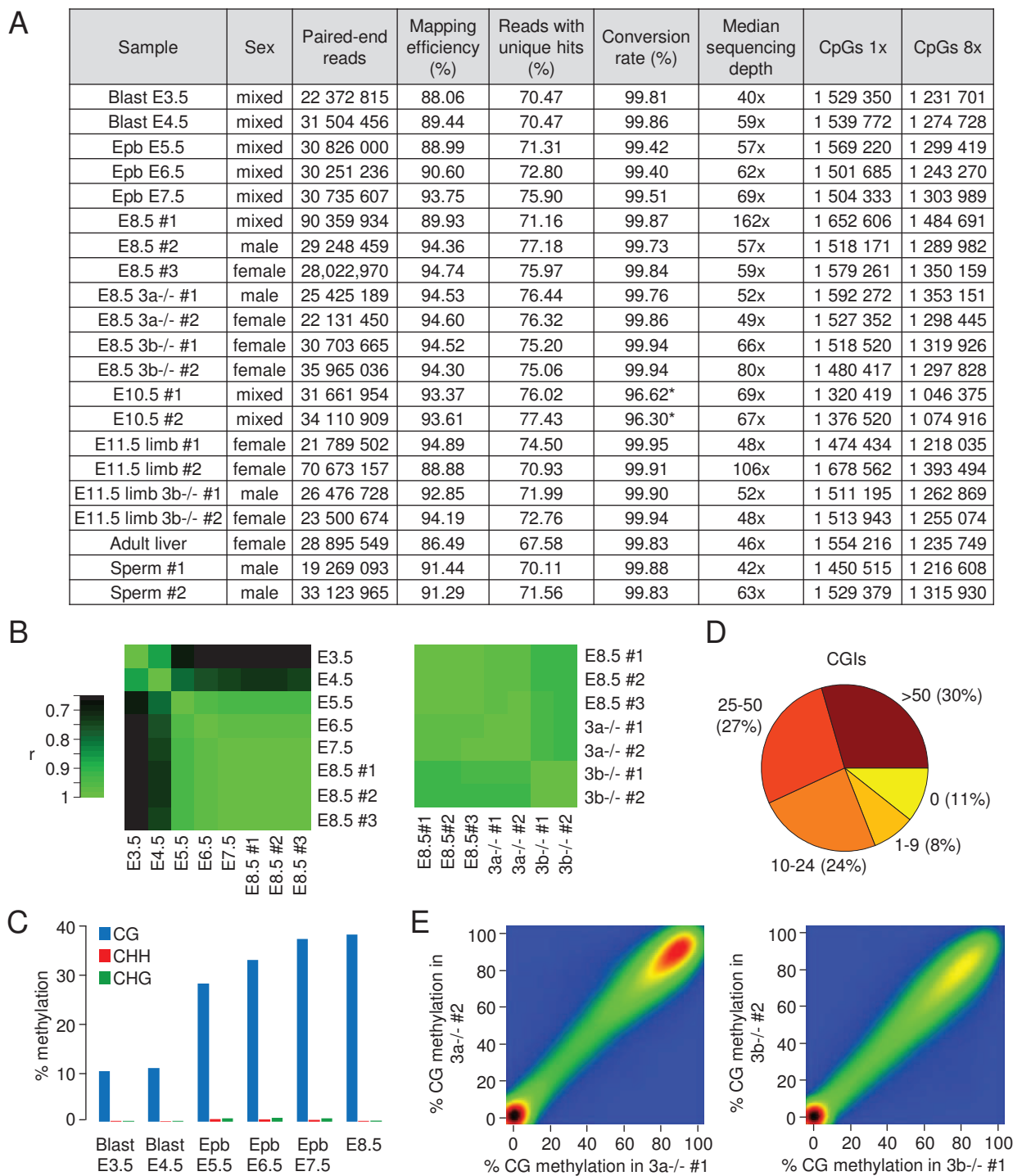


Figure S1. Features of RRBS methylomes in mouse embryos. **A)** Table summarizing the sequenced RRBS libraries. The last two columns indicate the number of CpGs sequenced at least 1x and 8x obtained from reads that mapped uniquely in the genome. The bisulfite conversion rate was estimated by calculating the C to T conversion at non-CpG sites. The samples marked with an asterisk (*) have a reduced conversion rate because they underwent only one round of bisulfite conversion. Blast=blastocyst; Epb=epiblast. **B)** Correlation matrix showing the Pearson correlation coefficient (r) of methylation at individual CGs between RRBS experiments. **C)** Evolution of the percentage of cytosine methylation at CG, CHH and CHG sites during development (H=A, C or T). **D)** Pie chart representing the coverage of UCSC CpG islands in the Epb E7.5 RRBS dataset. The numbers correspond to how many CpGs were sequenced at least 8x in the CGIs. **E)** Pairwise comparison of CpG methylation levels in 400bp tiles reveals an excellent correlation between biological replicates of RRBS in *Dnmt3a*^{-/-} and *Dnmt3b*^{-/-} embryos.

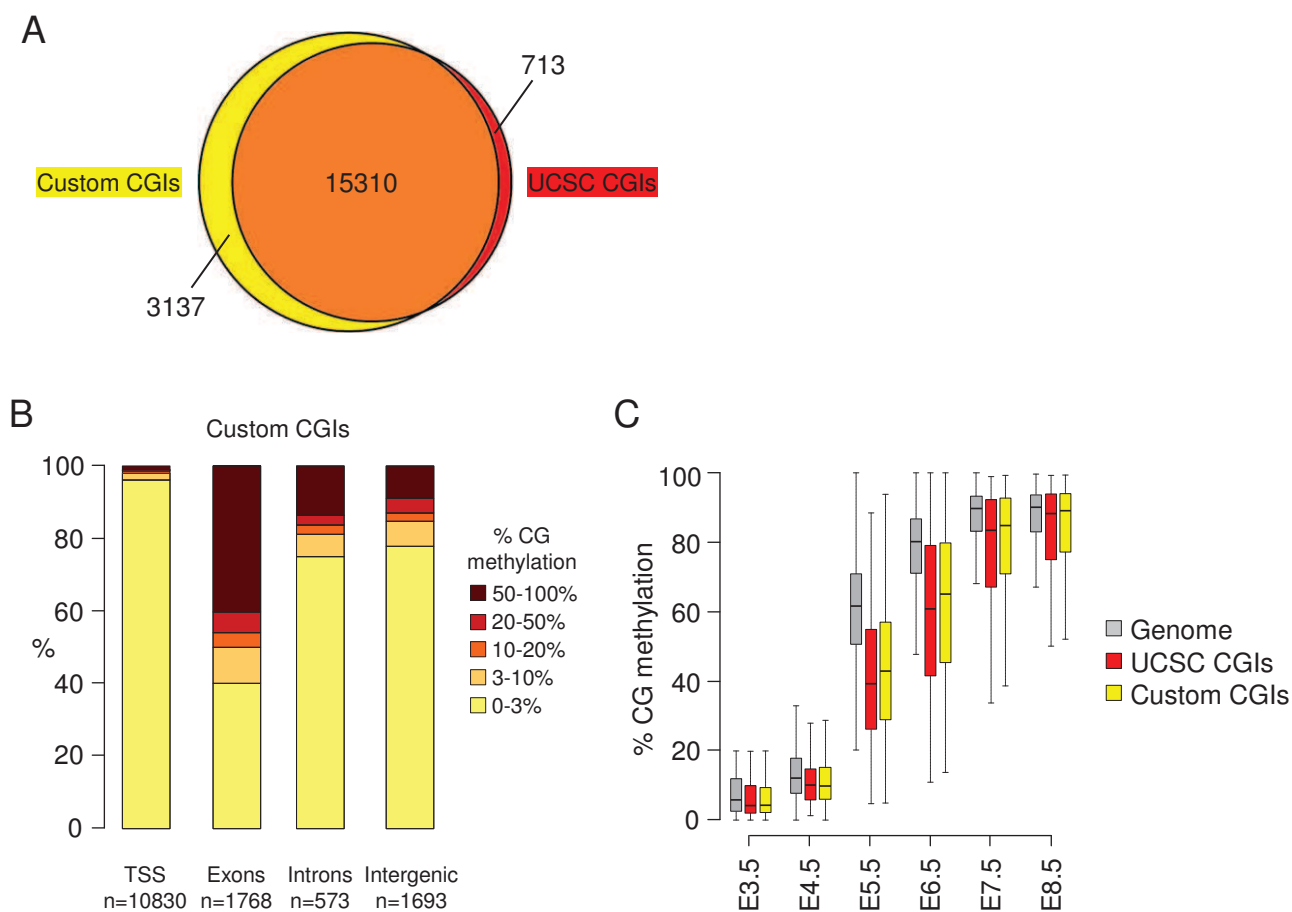


Figure S2. Distribution and kinetics of DNA methylation at custom CpG islands. **A)** Venn diagram showing the overlap between the UCSC CGIs and the CGIs identified by our custom annotation. **B)** Distribution of E8.5 methylation scores in custom CGIs located in promoters (-1000 to 1000bp from RefSeq TSS), exons, introns and intergenic sequences. X-linked CGIs are excluded from this analysis. **C)** Comparison of the kinetics of *de novo* methylation in genome tiles (400bp), UCSC CGIs and custom CGIs. The graph depicts only *de novo* methylated sequences defined as <20% methylation in E3.5 blastocysts and >50% methylation in E8.5 embryos.

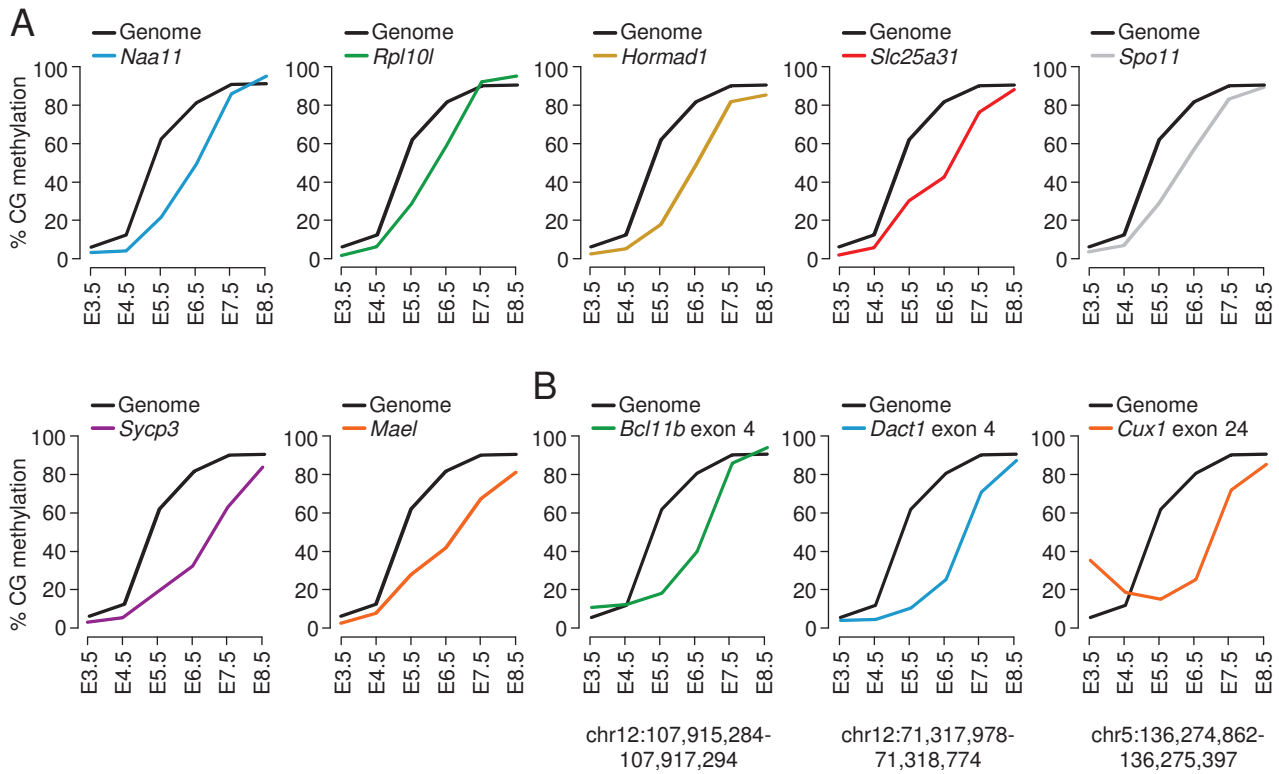


Figure S3. Kinetics of *de novo* methylation at CpG islands in embryos. The graphs show the CpG methylation levels measured at consecutive stages of development in selected CGIs in promoters (**A**) and exons (**B**). The genomic positions of the CGIs in exons are given below the graphs. In each panel, the black line represents the global rate of *de novo* methylation in the genome (calculated in all tiles with <20% methylation in E3.5 blastocysts and >50% methylation in E8.5 embryos).

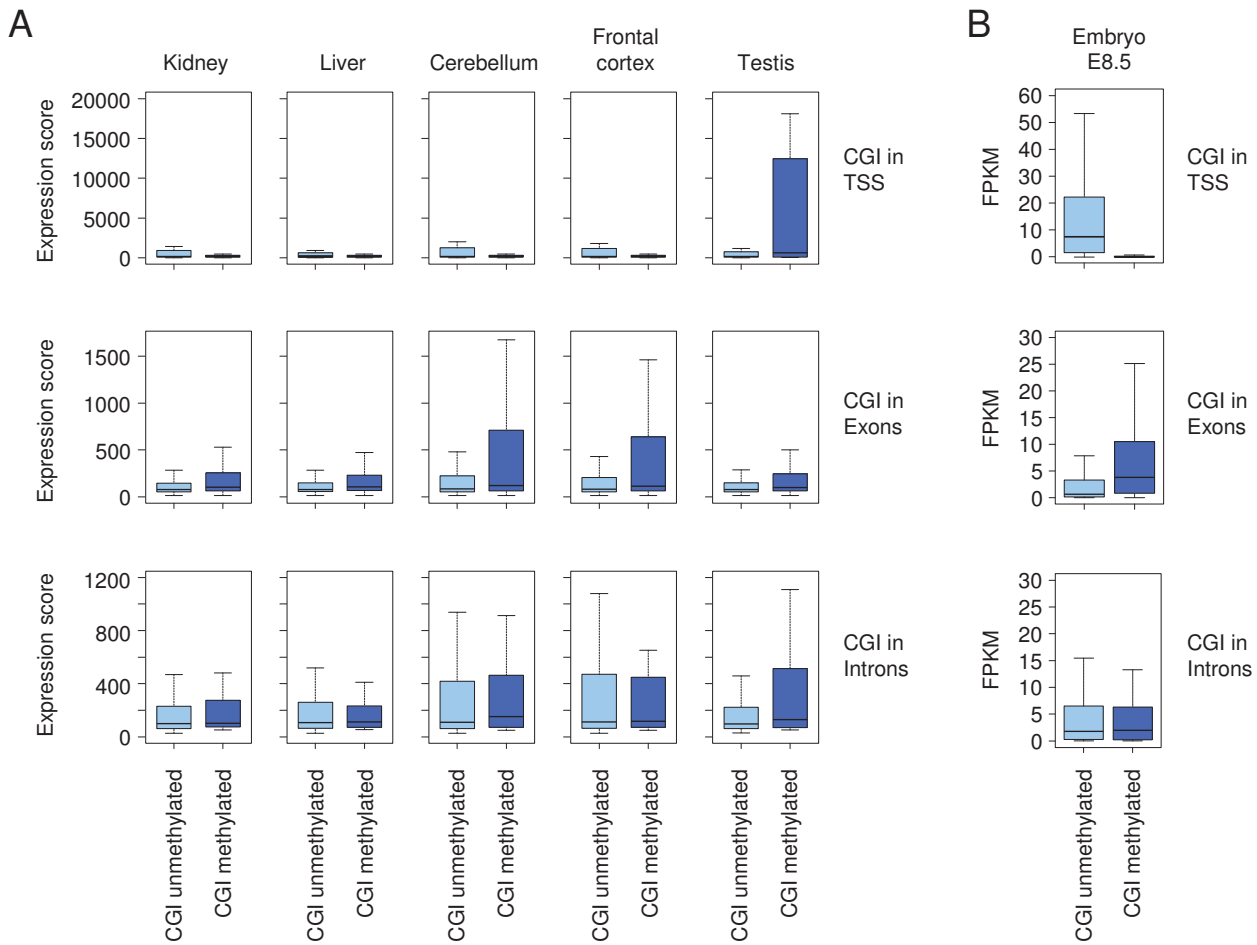


Figure S4. Comparison of CpG island methylation with gene expression in embryos and adult tissues. A) We extracted BioGPS expression scores in adult tissues for all genes with an annotated CGI in the TSS (top), in an exon (middle) or in an intron (bottom). The boxplots show the distribution of expression scores for genes with an unmethylated CGI (<15% methylation in E8.5 embryos, light blue) or a methylated CGI (>50% methylation in E8.5 embryos). This shows that genes that gain CGI methylation in the TSS are highly expressed in the testis, whereas genes that gain CGI methylation in an exon show the highest expression in brain tissues. **B)** Similar analysis as in (A) with RNA-Seq FPKM values obtained in wild-type E8.5 embryos.

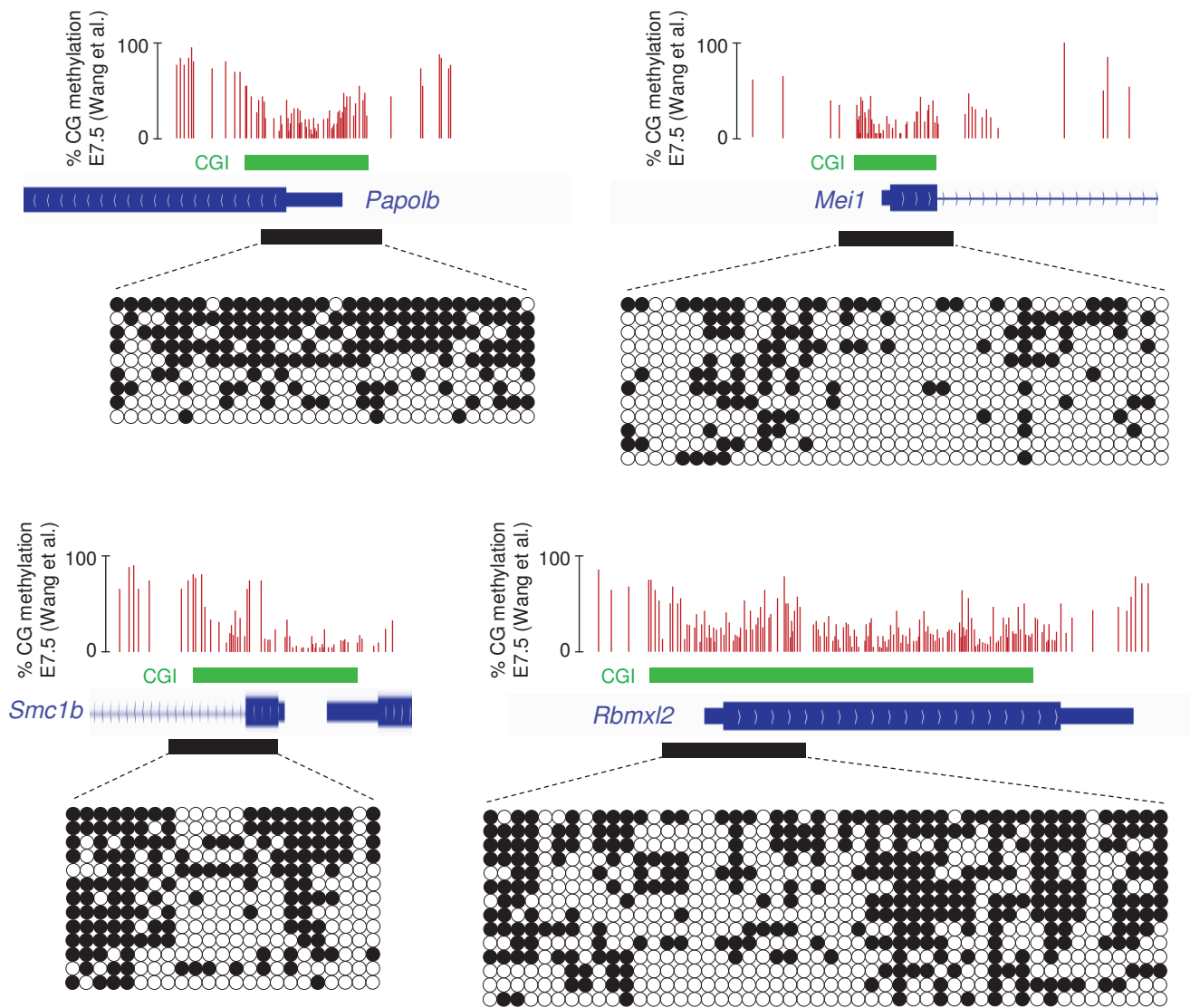


Figure S5. Bisulfite cloning and sequencing of promoter pmCGIs in adult liver. The green bars depict the position of the CGIs and the black bars depict the position of the bisulfite PCR fragments. The circles represent methylated (black) or unmethylated (white) CpG dinucleotides; each horizontal line is one sequenced clone. For comparison we show WGBS data generated in E7.5 embryos (Wang L *et al.* 2014, *Cell* **157**: 979-991).

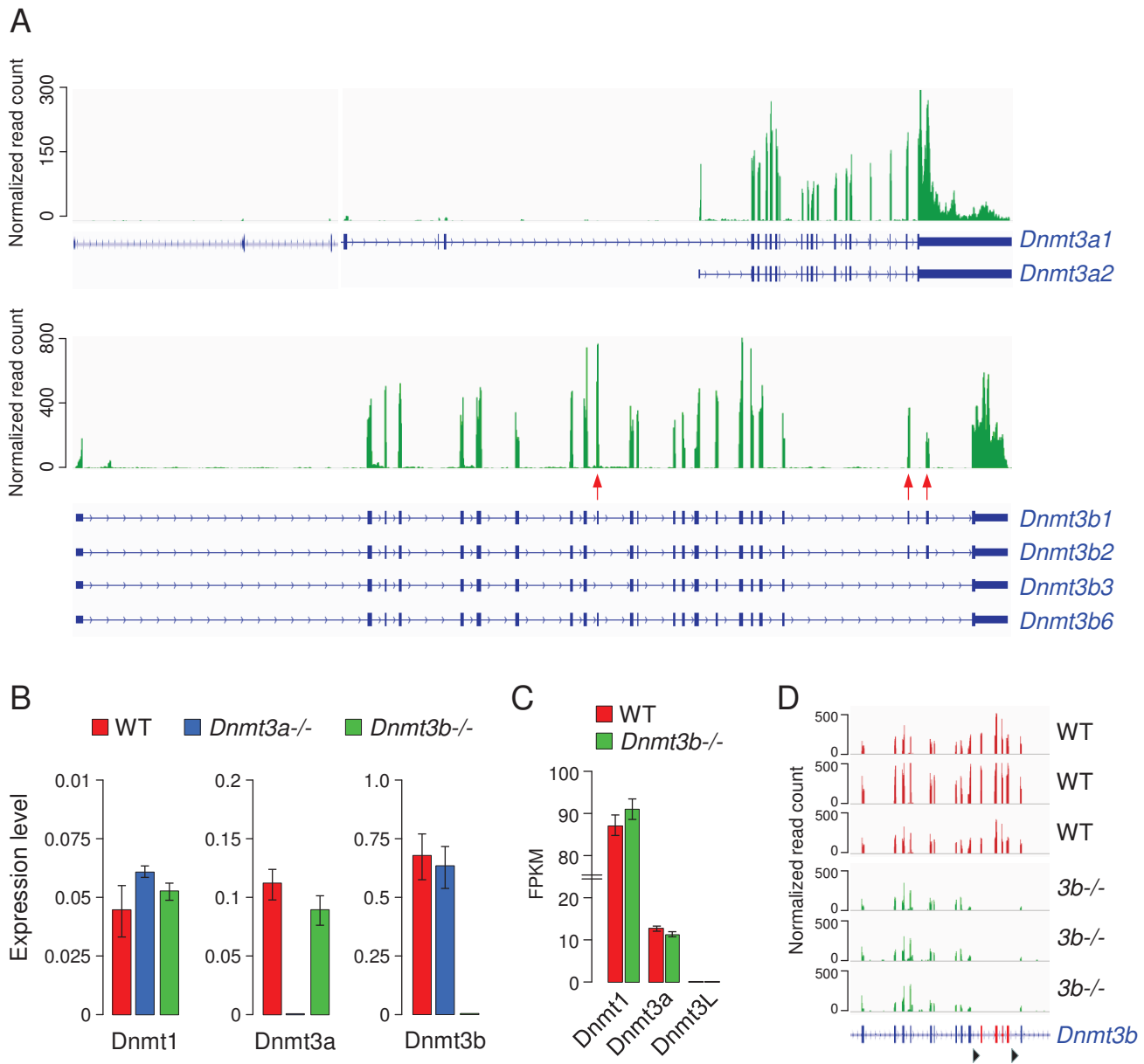


Figure S6. Expression of *Dnmt* genes in WT and knockout embryos. **A)** RNA-Seq profiles in a WT E8.5 embryo indicate that embryos predominantly express the short *Dnmt3a2* over the long *Dnmt3a1* isoform (top), and the full-length *Dnmt3b1* isoform over the alternative splicing isoforms lacking exon 10, 21 or 22 (marked by red arrows, bottom). **B)** The inactivation of *Dnmt3a* or *Dnmt3b* does not modify the expression of other *Dnmts* in epiblasts. The bar graphs show the expression of *Dnmts* measured by RT-qPCR in WT, *Dnmt3a*^{-/-} and *Dnmt3b*^{-/-} E6.5 epiblasts relative to the expression of the housekeeping gene *Actb*. The error bars represent mean deviations from independent embryos (n=3 for WT, n=2 for KO embryos). For *Dnmt3a* and *Dnmt3b* we used PCR primers that amplify within the Cre-deleted catalytic exons. **C)** Barplot of RNA-Seq FPKM values for *Dnmts* in WT and *Dnmt3b*^{-/-} E8.5 embryos. The error bars represent mean deviations from RNA-Seq in 3 embryos. **D)** Profiles of RNA-Seq read coverage at the *Dnmt3b* locus confirm that the four essential exons flanked by the loxP sites (in red between the black triangles) are absent in *Dnmt3b*^{-/-} embryos.

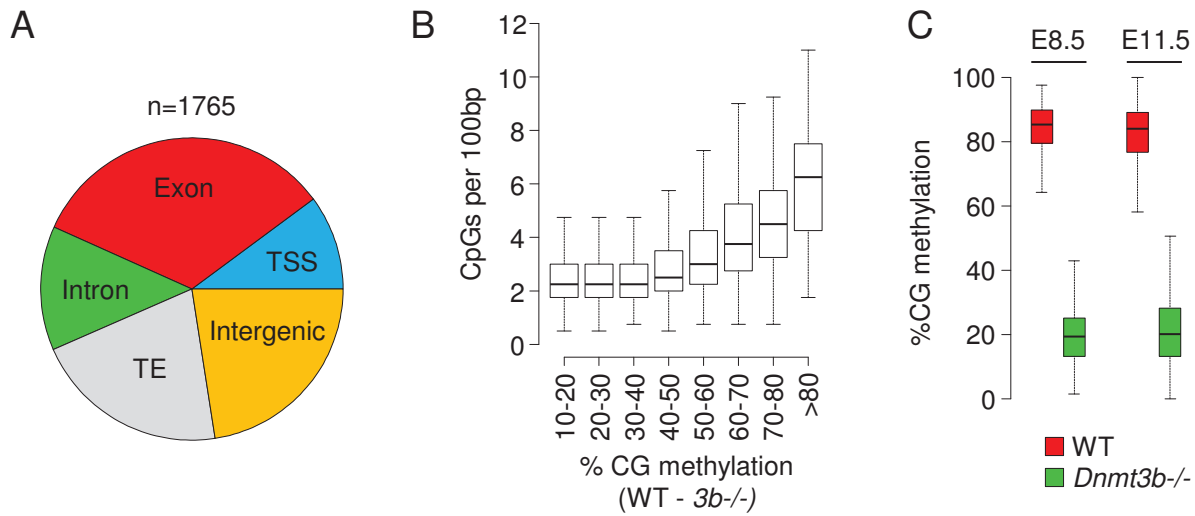


Figure S7. Analysis of DNMT3B-specific targets in mouse embryos. **A)** Pie chart showing the distribution of DNMT3B-dependent targets (defined as losing more than 60% methylation in *Dnmt3b*^{-/-} compared to WT E8.5 embryos) in TSS, exons, introns, transposable elements (TE) and intergenic regions. **B)** Box plot showing the CpG density (number of CpGs per 100bp, measured in 400bp tiles) as a function of the extent of methylation loss in *Dnmt3b*^{-/-} embryos. The sequences most affected by the inactivation of DNMT3B have an increased CpG density. **C)** Boxplot of CG methylation in E8.5 embryos and E11.5 limbs for all the DNMT3B-dependant targets identified at E8.5.

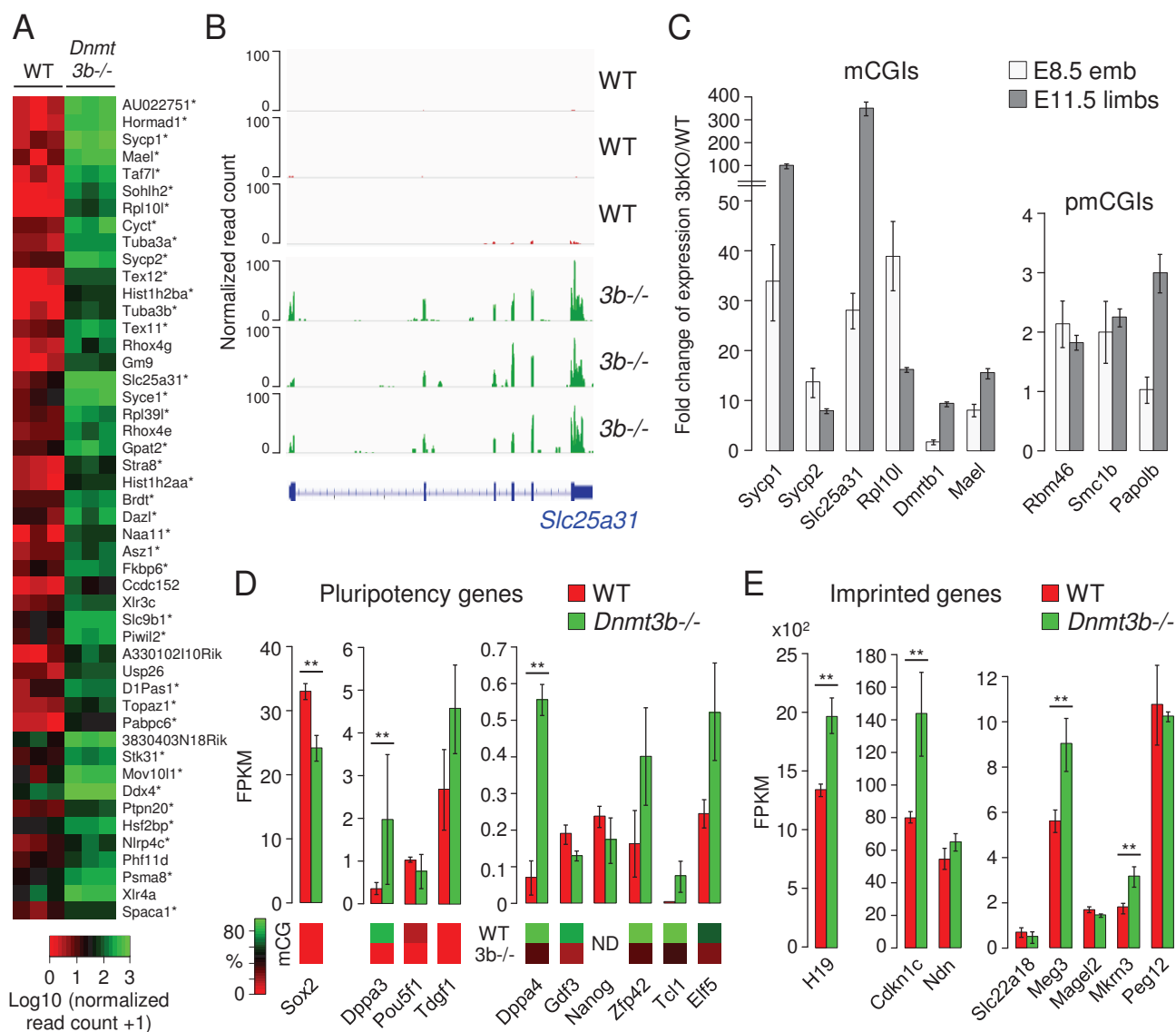


Figure S8. Expression of germline, pluripotency and imprinted genes in *Dnmt3b*^{-/-} embryos. **A)** Heatmap of DESeq2 normalized read counts for the 48 genes upregulated more than 5 fold in *Dnmt3b*^{-/-} versus WT E8.5 embryos. The asterisks (*) indicate germline-specific genes. **B)** RNA-Seq read coverage at the germline-specific gene *Slc25a31* in WT and *Dnmt3b*^{-/-} embryos. **C)** RT-qPCR analysis of the changes of expression of germline genes in *Dnmt3b*^{-/-} embryos. We measured the expression of germline genes containing a methylated (mCGI) or partially methylated (pmCGI) CGI promoter in E8.5 total embryos and E11.5 limbs. The bar graphs show the fold change of expression in three *Dnmt3b*^{-/-} compared to three WT samples after normalization to the expression of two housekeeping genes (*Actb* and *Rpl13a*). **D)** Barplot of RNA-Seq FPKM values for pluripotency genes in WT and *Dnmt3b*^{-/-} E8.5 embryos. The error bars represent mean deviations from RNA-Seq in 3 embryos. The heatmaps below the graph show the percentage of CG methylation measured by RRBS around the TSS (-1000 to +1000bp) in WT and *Dnmt3b*^{-/-} embryos. ND: not determined. (***) $p < 0.01$ (DESeq2 adjusted p -value). **E)** Barplot of RNA-Seq FPKM values in WT and *Dnmt3b*^{-/-} E8.5 embryos for imprinted genes with sDMRs. The error bars represent mean deviations from RNA-Seq in 3 embryos. (***) $p < 0.01$ (DESeq2 adjusted p -value).

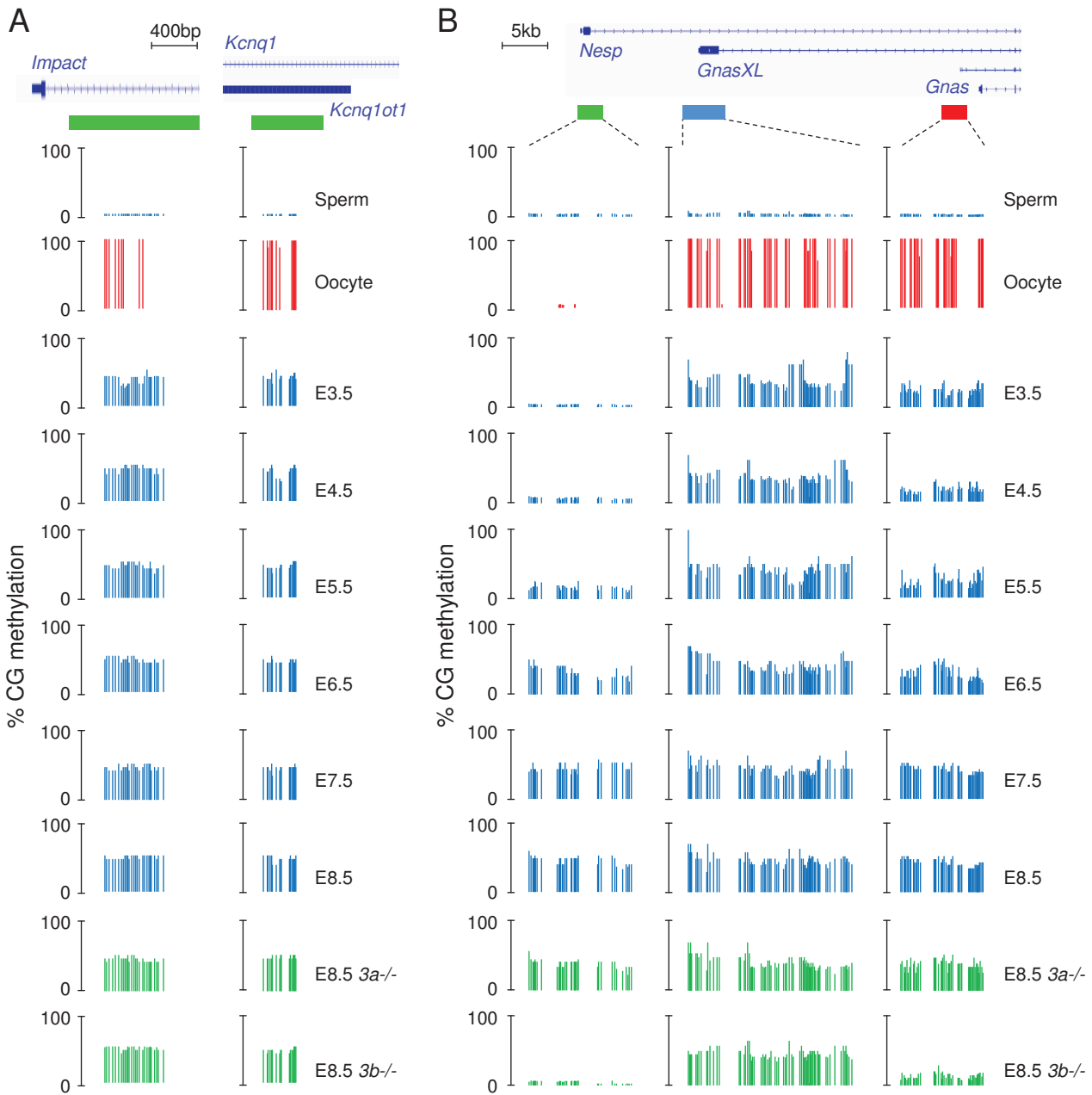


Figure S9. RRBS methylation profiles at imprinted loci throughout development and in *Dnmt3* knockout embryos. A) Single-CpG RRBS profiles at the *Impact* and *Kcnq1* gDMRs that maintain maternal allelic methylation throughout development. The genes are depicted on top of the graphs, and the green bars indicate the position of the gDMRs. B) RRBS profiles at the three DMRs of the *Gnas* locus. The structure of the locus is indicated on top of the graphs with the position of the *Nesp* sDMR (green bar), the *GnasXL* gDMR (blue bar) and the *Gnas* Exon1A DMR (red bar).

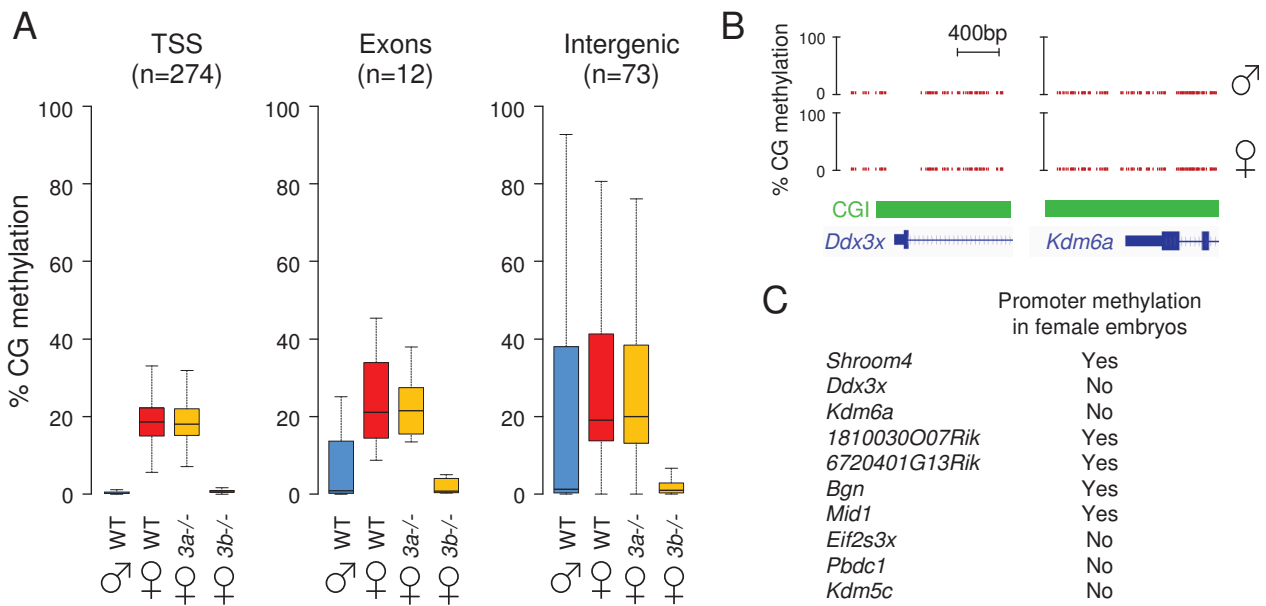


Figure S10. CpG island methylation on the X chromosome and at genes that escape X-inactivation in female embryos. **A)** Methylation of CGIs in TSS, exons and intergenic sequences on the X chromosome in male and female E8.5 embryos. **B)** RRBS methylation profiles in male and female E8.5 embryos at the CGI promoters of two genes known to escape X-inactivation in mice, which reveals an absence of X-linked methylation in female embryos. The green bars mark the position of CGIs. **C)** Summary of DNA methylation at 10 genes described to escape X-inactivation in mice (Yang F *et al.* 2010, *Genome Res* **20**: 614-622). The table indicates whether our RRBS data revealed a methylation of the gene promoter in female compared to male E8.5 embryos.

CONCLUSIONS

Les conclusions principales de ce travail sont :

- Nous avons pu déterminer les cinétiques de mise en place de la méthylation de l'ADN sur le génome après l'implantation et jusqu'à 8.5 jpc.
- Nos données montrent une régulation distincte de la méthylation *de novo* des îlots CpG comparés au reste du génome.
- Les îlots CpG qui acquièrent de la méthylation *in vivo* sont situés sur des promoteurs de gènes spécifiques de la lignée germinale et dans les exons de gènes importants du développement.
- Une nouvelle classe d'îlot CpG a été identifiée. Celle-ci acquiert une méthylation partielle dans les cellules somatiques. Nous les avons nommés pmCGIs (partially methylated CpG islands)
- DNMT3A et DNMT3B ont des fonctions redondantes dans la méthylation globale du génome *in vivo*.
- DNMT3B joue un rôle spécifique dans la méthylation *de novo* de certains îlots CpG des autosomes et du chromosome X inactif *in vivo*.
- Nos données procurent un outil efficace d'étude du devenir de la méthylation des DMRs de loci soumis à empreinte dans l'embryon. Nous montrons que DNMT3B semble indispensable pour la mise en place de la méthylation des DMRs secondaires et nos données ont permis de redéfinir les DMRs gamétiques au locus soumis à empreinte *Gnas*.

RESULTATS COMPLEMENTAIRES

DU PROJET 1 :

Analyse du double mutant *Dnmt3a*^{-/-} *Dnmt3b*^{-/-}

Les résultats de l'article 1 indiquent un rôle redondant des enzymes DNMT3A et DNMT3B. Cependant, nous n'avons pas pu déterminer si les deux enzymes DNMT3A et DNMT3B sont responsables de l'ensemble de la méthylation *de novo* dans l'embryon. Afin de mieux comprendre le rôle redondant des deux enzymes DNMT3A et DNMT3B, j'ai également généré et étudié des embryons de souris double Knock Out pour ces deux enzymes.

I) Obtention des souris double Knock Out pour *Dnmt3a* et *Dnmt3b*

Pour obtenir des embryons double KO *Dnmt3a*^{-/-} *Dnmt3b*^{-/-}, j'ai d'abord généré des souris double hétérozygotes *Dnmt3a*^{+/-} *Dnmt3b*^{+/-}. J'ai ensuite croisé ces souris entre elles afin d'obtenir des embryons double Knock Out (que je dénomme ensuite dKO) (**Figure 1**). D'après les travaux antérieurs de l'équipe de En Li, il avait été montré que ces embryons dKO ne sont viables que jusqu'à mi-gestation (Okano et al., 1999). Afin de comparer avec nos autres résultats dans les simples mutants DNMT3A et DNMT3B, j'ai extrait l'ADN et l'ARN d'embryons à 8.5 jpc pour mes analyses.

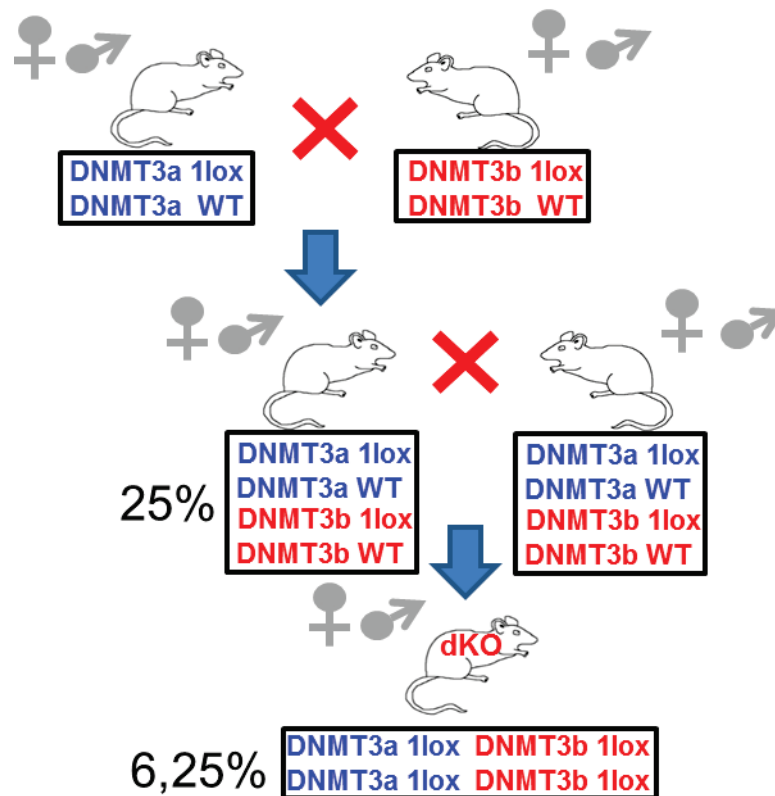


Figure 1 : Représentation schématique des croisements effectués pour obtenir des souris déficientes simultanément pour *Dnmt3a* et *Dnmt3b*.

Les souris hétérozygotes 1lox/wt pour *Dnmt3a* ou *Dnmt3b* sont croisées pour obtenir des souris double hétérozygotes 1lox/wt pour chacun des deux gènes avec une probabilité de 25%. Les souris double hétérozygotes sont ensuite croisées entre elles pour obtenir des souris double Knock Out (dKO) *Dnmt3a*^{-/-} *Dnmt3b*^{-/-} avec une probabilité de 6.25%.

II) Les souris double Knock Out perdent toute trace de méthylation globale

Pour étudier le méthylome en absence de DNMT3A et DNMT3B, j'ai réalisé des RRBS sur deux embryons dKO *Dnmt3a*^{-/-} *Dnmt3b*^{-/-} disséqués à 8.5 jpc. Les méthylomes de ces embryons dKO ont été comparés à des embryons contrôles double hétérozygotes *Dnmt3a*^{+/-} *Dnmt3b*^{+/-} et WT. Les résultats indiquent une perte globale de la méthylation du génome dans les embryons dKO (**Figure 2A**). La méthylation est absente sur tous les gènes et toutes les régions inter-géniques (**Figure 2B**), à l'exception de certains éléments transposables qui maintiennent un niveau détectable de méthylation (**Figure 2C**). Ces données démontrent que les enzymes DNMT3A et DNMT3B sont strictement nécessaires pour la mise en place de la méthylation dans l'embryon et que DNMT1 n'a que peu d'activité *de novo* dans ce contexte.

De manière surprenante, nous avons observé que même les ICRs de loci soumis à empreinte sont affectées et perdent les marques de méthylation dans les embryons dKO (**Figure 3**). DNMT3A et DNMT3B ont donc un rôle redondant dans le maintien de la méthylation des ICRs durant l'embryogénèse. Ce résultat est surprenant car les travaux passés des équipes de Okano et Sasaki avaient montré que les ICRs des gènes *H19*, *Dlk1/Gtl2* et *Igf2r* sont méthylées normalement dans les embryons dKO à 9.5 jpc (Oda et al., 2006 ; Hirasawa et al., 2008), ce qui suggérait que DNMT1 seule assure la propagation des empreintes de méthylation dans l'embryogénèse.

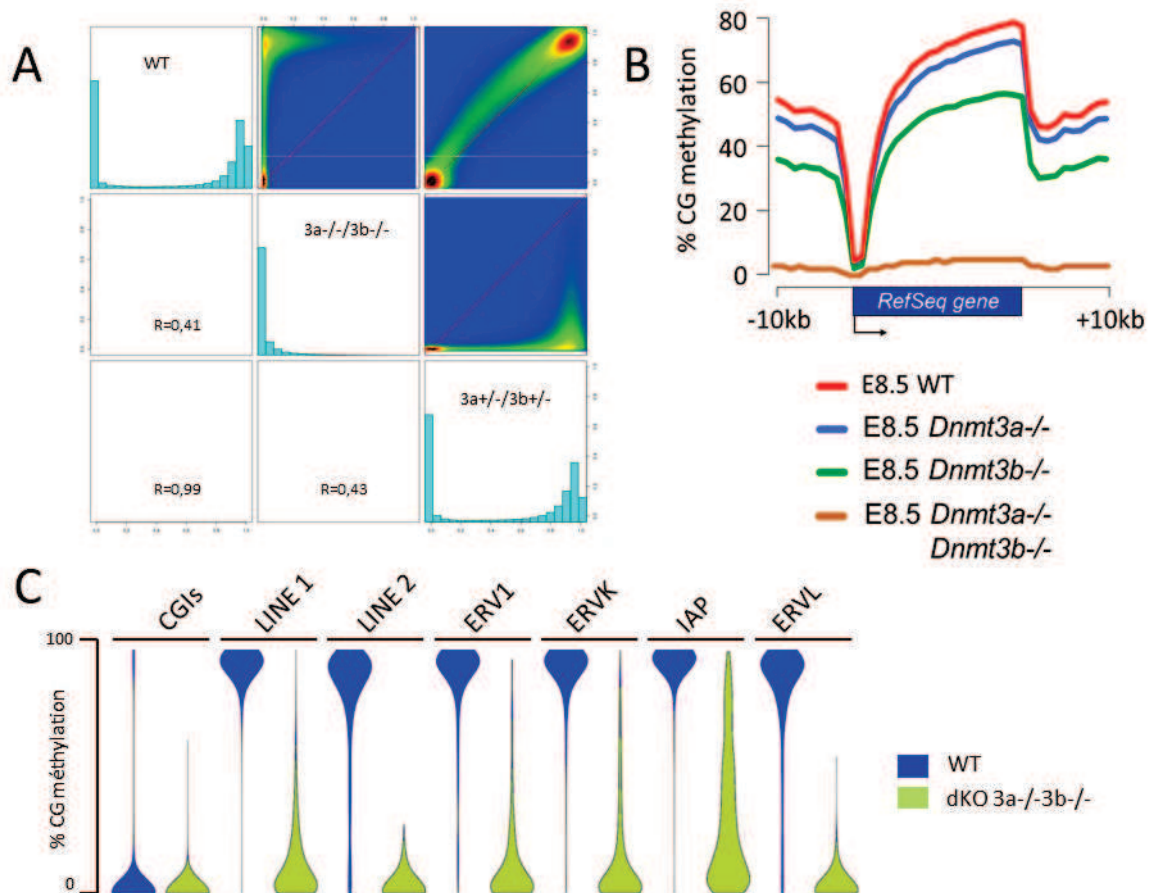


Figure 2 : Comparaison de la méthylation des CpG par RRBS dans les embryons WT, double hétérozygotes et dKO *Dnmt3a*^{-/-} *Dnmt3b*^{-/-} à 8.5 jpc.

A : Comparaison de la méthylation des CpG dans des fenêtres de 400 pb entre des embryons WT, double hétérozygotes et dKO *Dnmt3a*^{-/-} *Dnmt3b*^{-/-} à 8.5 jpc. Les graphiques montrent la corrélation de méthylation dans 50.000 fenêtres génomiques prises au hasard. Les histogrammes représentent la distribution de la méthylation de l'ADN dans toutes les fenêtres. Le coefficient de corrélation de Pearson (r) est indiqué. **B** : Distribution moyenne de la méthylation de l'ADN au niveau des gènes RefSeq flanqués de 10 kb dans les embryons WT, simples mutants et dKO *Dnmt3a*^{-/-} *Dnmt3b*^{-/-} à 8.5 jpc. **C** : Graphique en violon de la méthylation mesurée dans les îlots CpG (CGIs) et les rétrotransposons, dans des embryons WT et dKO *Dnmt3a*^{-/-} *Dnmt3b*^{-/-} à 8.5 jpc.

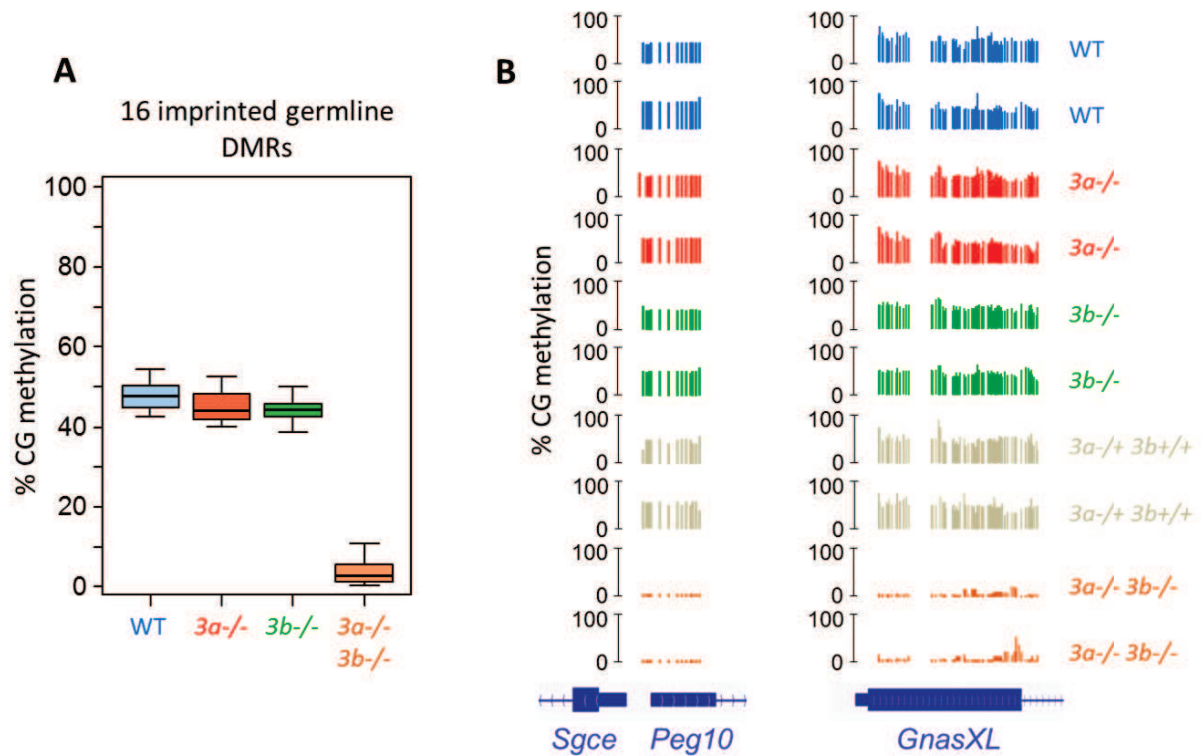


Figure 3 : Méthylation des DMRs de gènes soumis à empreinte dans les embryons dKO *Dnmt3a*^{-/-} *Dnmt3b*^{-/-} à 8.5 jpc.

A : Graphique en boîte à moustache représentant la méthylation de 16 DMRs soumises à empreinte génomique, dans des embryons mutants et WT à 8.5 jpc.

B : Exemples de profils RRBS au niveau des DMRs de deux loci soumis à empreinte génomique (*Peg10* et *Gnas*) dans des embryons mutants et WT à 8.5 jpc.

III) La méthylation globale n'affecte presque pas le transcriptome

Afin de connaître l'impact de l'absence de méthylation de l'ADN sur l'état du transcriptome des embryons déméthylés, j'ai réalisé des RNA-seq sur l'ARN de deux embryons dKO (**Figure 4**). Malgré la perte globale de méthylation de l'ADN, très peu de gènes sont dérégulés. Ces résultats suggèrent que la méthylation globale du génome ne semble pas avoir de rôle prépondérant dans la régulation du transcriptome *in vivo*. Les gènes surexprimés dans les embryons dKO correspondent majoritairement aux gènes spécifiques de la lignée germinale qui sont réprimés par la méthylation de leur promoteur et que nous avons également vu surexprimés dans les embryons simples mutants *Dnmt3b*^{-/-}. Cela indique que la fonction transcriptionnelle majeure de la méthylation de l'ADN dans l'embryon est la répression du programme germinale.

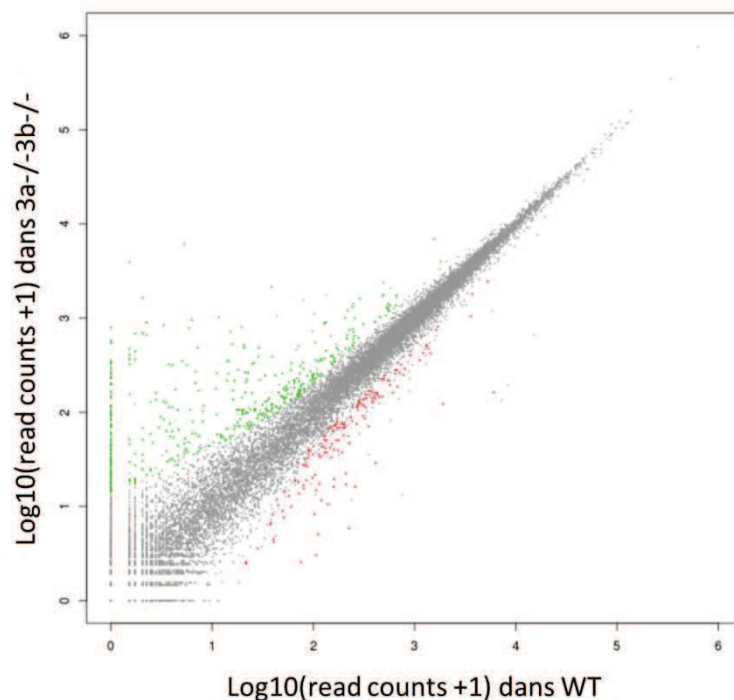


Figure 4 : Comparaison des profils d'expression de gènes par RNA-Seq dans les embryons WT et dKO à 8.5 jpc.

Le graphique montre le niveau d'expression des gènes RefSeq annotés (moyenne de deux embryons). Les gènes surexprimés (fold change > 2, p < 0.01) sont marqués en vert et les gènes réprimés (fold change > 2, p < 0.01) en rouge.

CONCLUSIONS

Les conclusions principales de ces travaux complémentaires sont :

- L'absence de DNMT3A et DNMT3B conduit à une absence globale de la méthylation du génome *in vivo*.
- DNMT3A et DNMT3B ont un rôle primordial et redondant sur le maintien de la méthylation des gDMRs *in vivo*.
- L'absence de méthylation globale a un impact limité sur le transcriptome et n'affecte pas l'expression de la majorité des gènes.

RESULTATS PROJET 2 :

Identification du rôle de G9a dans le recrutement de la méthylation de l'ADN durant le développement embryonnaire murin.

INTRODUCTION

Aujourd'hui, les connaissances sur le ciblage et la régulation de la méthylation de l'ADN sont insuffisantes chez les Mammifères. Le rôle des enzymes modificatrices d'histones dans ces processus a été bien étudié chez des organismes tels que le champignon *Neurospora crassa* ou la plante *Arabidopsis thaliana*. Chez ces organismes, on sait que la méthylation des histones H3 sur la lysine 9 (H3K9) permet le recrutement d'une grande partie de la méthylation des cytosines (Jackson et al., 2002; Lindroth et al., 2004; Tamaru and Selker, 2001; Tamaru et al., 2003). En revanche le rôle de la méthylation des histones dans le recrutement de la méthylation des cytosines chez les Mammifères n'est pas clair.

Nous nous sommes plus particulièrement intéressés à l'histone méthyltransférase G9a car un large panel de publications montre que cette enzyme semble liée à la méthylation de l'ADN dans les cellules ES murines (Dong et al., 2008 ; Tachibana et al., 2008 ; Myant et al., 2011 ; Ikegami et al., 2007 ; Xin et al., 2003). Chez les Mammifères, G9a catalyse les modifications chimiques H3K9me1 et H3K9me2, et les enzymes SUV39h1 et SETDB1 utilisent ces substrats pour créer H3K9me3 (Dodge et al., 2004 ; Peters et al., 2001). De plus, G9a peut interagir avec les protéines DNMT1, DNMT3A ou DNMT3B et pourrait donc réguler la méthylation *de novo* ou la propagation de la méthylation de l'ADN (Epsztejn-Litman et al., 2008 ; Estève et al., 2006). Il se pourrait également que G9a puisse recruter la méthylation de l'ADN par l'intermédiaire de la modification H3K9me2 car celle-ci est reconnue par UHRF1, un cofacteur essentiel pour l'action de DNMT1 (Epsztejn-Litman et al., 2008).

Des recherches dans les cellules ES ont mis en évidence que G9a semble y réguler la méthylation de l'ADN sur de nombreuses régions. En effet, des cellules ES déficientes pour G9a montrent des défauts de méthylation au niveau des locus soumis à empreinte, des transposons (LINE-1, LTR), des îlots CpG de gènes de la lignée germinale, de gènes du développement ou encore de gènes de pluripotence (Dong et al., 2008 ; Tachibana et al., 2008 ; Myant et al., 2011 ; Ikegami et al., 2007 ; Xin et al., 2003). Enfin, ces mêmes études ont montré que G9a a une influence sur la méthylation de l'ADN indépendante de son activité catalytique dans les cellules ES (Dong et al., 2008; Epsztejn-Litman et al., 2008; Tachibana et al., 2008).

Durant le développement embryonnaire, les embryons déficients en G9a meurent à 10.5 jpc, c'est-à-dire à peu près au même moment que les embryons déficients pour la méthylation de l'ADN (Tachibana et al., 2005). Il se pourrait donc que G9a ait un rôle primordial sur la méthylation de l'ADN et que cela participe au phénotype de létalité dans les embryons *G9a*^{-/-}.

Les résultats passés indiquent que G9a joue un rôle dans la régulation de la méthylation de l'ADN *in vitro*. Cependant, son implication dans la méthylation de l'ADN au cours de l'embryogénèse murine est encore inconnue. Nous avons utilisé ici un modèle de souris déficientes pour G9a afin d'étudier son rôle durant le développement embryonnaire. Ce projet a été initié en collaboration avec le laboratoire de Robert Feil qui a développé le modèle de souris invalidées pour G9a (Wagschal et al., 2008).

Le premier objectif de ces recherches était d'évaluer l'impact de l'absence de G9a sur la méthylation de l'ADN de l'embryon à l'échelle génomique. J'ai identifié et validé les gènes différentiellement méthylés dans les embryons WT et *G9a*^{-/-} au stade 8.5 jpc.

Le second objectif était de comprendre si G9a a un rôle direct sur la méthylation de l'ADN. Pour cela, nous avons mis au point les conditions d'immunoprécipitation de chromatine afin de tester si G9a se localise aux régions différentiellement méthylées dans les cellules embryonnaires.

Le troisième objectif était d'étudier le lien entre la modification d'histone H3K9me2, catalysée par G9a, et la méthylation de l'ADN, afin de comprendre si G9a influence la méthylation de l'ADN via son activité de méthylase des histones.

Enfin, le dernier objectif était de tester si la méthylation de l'ADN est importante pour la fonction biologique de G9a en comparant le transcriptome et la modification d'histone H3K9me2 dans les embryons déficients pour *G9a* ou *Dnmt3b*.

Ces résultats sont dans le manuscrit de l'**article 2**, actuellement en soumission.

TRAVAUX REALISES

Pour atteindre ces objectifs, les travaux suivants ont été réalisés:

1-Mesurer l'expression de G9a par RT-qPCR aux stades de développement embryonnaire 4.5 jpc jusqu'à 8.5 jpc.

2-Disséquer des embryons et préparer des échantillons d'ADN et d'ARN d'embryons WT et *G9a*^{-/-} aux stades 8.5 jpc et 9.5 jpc.

3-Réaliser des expériences de MeDIP-on-ChIP sur l'ADN des embryons 9.5 jpc WT et *G9a*^{-/-}. Ces expériences ont été réalisées par Julie Borgel qui a initié le projet dans l'équipe de Michael Weber.

4-Préparer des bibliothèques de séquençage RRBS dans les cellules ES mutantes pour *G9a*.

5-Préparer des bibliothèques de séquençage RRBS sur l'ADN des embryons *G9a*^{-/-} et WT à 8.5 jpc.

6-Réaliser des expériences de RNA-seq sur l'ARN des embryons *G9a*^{-/-} et WT à 8.5 jpc.

7-Valider les régions hypométhylées dans les embryons *G9a*^{-/-} par traitement au bisulfite de sodium suivi de séquençages ou de digestions enzymatiques (COBRA).

8-Valider les résultats de RNA-seq par RT-qPCR dans les embryons WT et *G9a*^{-/-}.

9-Réaliser des ChIP de G9a dans les MEFs et les embryons sauvages à 8.5 jpc. Ces expériences ont été réalisées par Judith Vallet qui a mis au point les conditions de ChIP de la protéine G9a.

10-Réaliser des ChIP de la modification H3K9me2 dans les MEFs sauvages ainsi que les embryons sauvages et *G9a*^{-/-} à 8.5 jpc

11-Réaliser des CHIP-seq de la modification H3K9m2 dans les embryons sauvages à 8.5 jpc pour déterminer la distribution globale de cette marque d'histone.

12-Participer aux analyses bio-informatiques des résultats. Ce travail a majoritairement été réalisé par Sylvain Guibert, Michael Dumas et Michaël Weber.

ARTICLE 2

1 **G9a/EHMT2 directs DNA methylation for efficient gene silencing in mouse embryos**

2

3 Ghislain Auclair¹, Julie Borge^{2,3,5}, Lionel A. Sanz^{2,4,5}, Judith Vallet¹, Sylvain Guibert¹, Michael
4 Dumas¹, Patricia Cavalier², Michael Girardot², Thierry Forné², Robert Feil^{2#}, Michael Weber^{1#}

5

6 ¹CNRS, University of Strasbourg, UMR7242 Biotechnology and Cell Signaling, 300 Bd Sébastien
7 Brant, BP 10413, 67412 Illkirch, France

8 ²Institute of Molecular Genetics, CNRS UMR5535, University of Montpellier, 1919 Route de Mende,
9 34293 Montpellier, France

10 ³Present address: MRC Clinical Sciences Centre, Imperial College London Faculty of Medicine,
11 Hammersmith Hospital Campus, London W12 0NN, UK.

12 ⁴Present address: Department of Molecular and Cellular Biology, University of California Davis,
13 Davis, USA.

14 ⁵These authors contributed equally to this work

15 #Corresponding authors: robert.feil@igmm.cnrs.fr; michael.weber@unistra.fr

16

17 **Running title:** G9a regulates DNA methylation in mouse embryos

18

19 **Keywords:** G9a, DNA methylation, CpG island, H3K9me2, embryogenesis

20

21 **ABSTRACT:**

22 The extent to which histone modifying enzymes contribute to DNA methylation in mammals remains
23 unclear. Previous studies suggested a link between the lysine methyltransferase G9a (also known as
24 EHMT2) and DNA methylation in the mouse. Here, we used a model of knock-out mice to explore the
25 role of G9a in DNA methylation during mouse embryogenesis. G9a is expressed in epiblast cells but is
26 dispensable for global DNA methylation in embryogenesis. In contrast, G9a regulates DNA
27 methylation at specific sequences that include CpG-rich promoters of germline-specific genes. These
28 loci are bound by G9a in embryonic cells, marked by H3K9 dimethylation and have strongly reduced
29 DNA methylation in *G9a*^{-/-} embryos. G9a also plays an essential role in the maintenance of germline-
30 derived DNA methylation at one imprinted locus, the *Slc38a4* gene. Finally, we show that DNA
31 methylation is instrumental for G9a-mediated gene silencing in embryogenesis. Our findings identify
32 G9a as a critical factor that facilitates repressive DNA methylation at specific genomic loci during
33 mammalian development.

34

35 **Reviewer link to GEO data:**

36 <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=yxmxgcuktdazbwn&acc=GSE71500>

37

1 INTRODUCTION

2
3 Cytosine methylation plays diverse roles in mammalian development. It contributes to genomic
4 imprinting, X chromosome-inactivation, as well as the stable repression of retroelements and
5 developmental genes (Smith and Meissner 2013). During development, most CpG island promoters
6 remain protected from DNA methylation, except for a small set associated with germline-specific
7 genes (Borgel *et al.* 2010; Auclair *et al.* 2014). The pluripotency genes *Pou5f1* (also known as *Oct4*)
8 and *Dppa3* also acquire CpG methylation in the post-implantation embryo, which stabilizes the exit
9 from pluripotency (Feldman *et al.* 2006; Borgel *et al.* 2010). This process requires the *de novo*
10 methyltransferases DNMT3A/B, whereas the subsequent maintenance of DNA methylation through
11 cell divisions is ensured by DNMT1.

12
13 While the targets of DNA methylation are well characterized, little is known about the molecular
14 determinants of DNA methylation in mammals. In plants and filamentous fungi, a large portion of
15 DNA methylation is directed by histone H3 methylated on lysine 9, and deletion of H3K9
16 methyltransferases has a major impact on DNA methylation (Saze *et al.* 2012). A link between H3K9
17 methylation and DNA methylation has been documented also in mammalian cells (reviewed in (Rose
18 and Klose 2014)). SUV39H1/2 and SETDB1, which mediate H3K9 trimethylation (H3K9me3) at
19 pericentric heterochromatin and ERV retrotransposons, interact with DNMTs (Fuks *et al.* 2003; Li *et*
20 *al.* 2006) and modulate DNA methylation at pericentric satellite repeats and ERV retrotransposons in
21 mouse embryonic stem cells (ESCs) (Lehnertz *et al.* 2003; Matsui *et al.* 2010). On the other hand, G9a
22 (also called EHMT2) and its closely-related partner ‘G9a-like protein’ (GLP, also called EHMT1)
23 catalyse H3K9 mono and di-methylation (H3K9me1 and me2) in euchromatin (Tachibana *et al.* 2002;
24 Tachibana *et al.* 2005). G9a and GLP play pivotal roles during early mouse development (Tachibana *et*
25 *al.* 2002; Tachibana *et al.* 2005). They exist mostly as a G9a/GLP heterodimeric complex, which is the
26 main functional H3K9 methyltransferase because inactivation of either G9a or GLP strongly affects
27 global H3K9me1/2 in embryonic cells (Tachibana *et al.* 2005). G9a interacts and co-localizes with
28 DNMT1 and UHRF1 at sites of DNA replication (Esteve *et al.* 2006; Kim *et al.* 2009). Inversely,
29 UHRF1 binds to chromatin containing H3K9me2/3, which may facilitate the maintenance of DNA
30 methylation at genomic sites containing methylated H3K9 (Karagianni *et al.* 2008; Rothbart *et al.*
31 2012; Liu *et al.* 2013). In mouse ESCs, G9a controls DNA methylation at the germline differentially
32 methylated regions (gDMRs) of imprinted loci (Xin *et al.* 2003; Dong *et al.* 2008), class I and II ERV
33 retrotransposons, LINE1 elements, satellite repeats and CpG-rich promoters of germline and
34 developmental genes (Ikegami *et al.* 2007; Dong *et al.* 2008; Tachibana *et al.* 2008; Myant *et al.* 2011).
35 G9a also interacts with the *de novo* methyltransferases DNMT3A and DNMT3B (Epsztejn-Litman *et*
36 *al.* 2008; Kotini *et al.* 2011) and participates in the *de novo* methylation of newly-integrated
37 retroviruses (Leung *et al.* 2011) and pluripotency genes in ESCs (Feldman *et al.* 2006; Epsztejn-

1 Litman et al. 2008; Athanasiadou et al. 2010). Several studies suggest that the influence of G9a on
2 DNA methylation in ESCs is independent of its catalytic activity (Dong et al. 2008; Epsztejn-Litman
3 et al. 2008; Tachibana et al. 2008).

4
5 These cell-based studies suggested that G9a is an important regulator of DNA methylation in
6 mammals, yet the contribution of G9a to DNA methylation in mammalian embryogenesis is unknown.
7 This prompted us to explore the role of G9a in DNA methylation during mouse embryogenesis using
8 mice deficient for G9a and a combination of locus-specific and genome-wide approaches. Our *in vivo*
9 studies show that G9a plays a role in the deposition of repressive DNA methylation at specific
10 genomic sites during embryonic development.

11 12 **RESULTS**

13 14 **G9a has a global influence on the DNA methylome of murine ES cells.**

15 Previous studies reported reduced DNA methylation in cultured murine *G9a*^{-/-} ESCs (Ikegami et al.
16 2007; Dong et al. 2008; Epsztejn-Litman et al. 2008; Tachibana et al. 2008; Myant et al. 2011). To
17 extend these findings and quantify the influence of G9a on the methylome of ESCs genome-wide, we
18 generated single base-resolution methylomes by ‘Reduced Representation Bisulfite Sequencing’
19 (RRBS) in WT TT2 ESCs, *G9a*^{-/-} TT2 ESCs and *G9a*^{-/-} cells rescued with a WT G9a transgene
20 (Tachibana et al. 2002). As expected, WT ESCs show a bimodal distribution of CpG methylation with
21 most of the hypomethylated CpGs residing in CpG islands located proximal to transcription start sites
22 (TSS) (**Figure 1A-B**). The inactivation of G9a in ESCs leads to a global and uniform decrease of CpG
23 methylation over all sequences of the genome (**Figure 1A-B**). In agreement with the earlier studies
24 (Dong et al. 2008; Tachibana et al. 2008; Myant et al. 2011), hypomethylation also affects the
25 promoters of germline and developmental genes such as *Dazl*, *Wfdc15a*, *Brdt*, *Tuba3a* and *Rhox* genes
26 (**Figure 1C**). In addition, all classes of transposable elements lose on average ~40% CpG methylation
27 in *G9a*^{-/-} compared to the parental TT2 ESCs (**Figure 1D**). The re-introduction of G9a in *G9a*^{-/-}
28 ESCs restores DNA methylation at all sequences including the transposable elements (**Figure 1A-D**).
29 The overall methylation is slightly higher in G9a rescued cells compared to the parental cell line,
30 which may be attributed to the higher expression of G9a in rescued cells (Mozzetta et al. 2014).

31 32 **Expression of G9a in embryos and role in DNA methylation at candidate genes.**

33 We next investigated the role of G9a in DNA methylation during embryogenesis *in vivo*. We first
34 followed the expression of *G9a* in mouse embryos by RT-qPCR and found that it is expressed at the
35 time of *de novo* methylation of DNA between embryonic day 4.5 (E4.5) and E8.5 (**Figure 2A**). To
36 study the role of G9a in DNA methylation, we used a G9a knock-out line in which a LacZ cassette
37 inserted after the exon 11 results in truncated transcripts lacking the Ankyrin (ANK) repeats and the

1 catalytic SET domain (**Figure S1A-D**) (Wagschal *et al.* 2008). In agreement with an earlier G9a
2 knock-out (Tachibana *et al.* 2002), *G9a*^{-/-} mice show a developmental delay and mid-gestation
3 lethality at around E10.5 (Wagschal *et al.* 2008). Using this model, we first analysed DNA
4 methylation at candidate genes by ‘Combined Bisulfite Restriction Analysis’ (COBRA) and bisulfite
5 sequencing. We assessed DNA methylation of the promoters of the germline genes *Dazl*, *Dpep3* and
6 *Tex12* and observed no hypomethylation in *G9a*^{-/-} embryos (**Figure 2B**). Since G9a controls the *de*
7 *novo* methylation of pluripotency genes in differentiating ESCs (Epsztejn-Litman *et al.* 2008), we
8 investigated DNA methylation at the *Pou5f1* and *Dppa3* promoters, but found no evidence for reduced
9 DNA methylation in E9.5 *G9a*^{-/-} embryos (**Figure 2C**) or in trophoblast cells of E9.5 *G9a*^{-/-} animals
10 (**Figure S2A**). Finally, because G9a cooperates with Dnmt3 to regulate eye development in zebrafish
11 (Rai *et al.* 2010), we measured methylation in the promoters of eye-specific genes (*Rho*, *Crygd*, *Cplx4*,
12 *Mfrp*) and found no sign of hypomethylation at these either (**Figure 2B**).

13

14 **Methylome profiling reveals a global conservation of DNA methylation in *G9a*^{-/-} embryos.**

15 Having found no changes in DNA methylation at candidate genes in *G9a*^{-/-} embryos, we generated
16 genome-scale methylomes using two approaches. We first profiled 5-methylcytosine with ‘Methylated
17 DNA Immunoprecipitation’ (MeDIP) coupled to micro-arrays covering 11kb of all gene promoters. A
18 direct comparison showed that the 5mC log₂ values correlate strongly between WT and *G9a*^{-/-}
19 embryos (**Figure 2D**). In parallel, we generated methylomes at single-base resolution by RRBS in WT
20 and *G9a*^{-/-} E8.5 embryos (**Figure S3A**). The RRBS data were highly reproducible between replicate
21 embryos (**Figure S3B**) and revealed no signs of hypomethylation at the genome level in *G9a*^{-/-}
22 embryos (**Figure 2E-F**). We extended this finding by assessing the methylation status of
23 retrotransposons and found that SINEs, LINEs and LTRs are globally methylated at normal levels in
24 *G9a*^{-/-} embryos (**Figure 2G**), which was confirmed by restriction analysis and conventional bisulfite
25 sequencing at IAP, MusD and MuLV elements (**Figure S2B-C**). The promoters of most germline and
26 developmental genes, including those hypomethylated in *G9a*^{-/-} ESCs, such as *Dazl*, *Wfdc15a*, *Brdt*,
27 *Tuba3a* and *Rhox*, are methylated to the same levels in *G9a*^{-/-} and WT embryos (**Figure 2H**; **Figure**
28 **S4A-B**). Finally, the absence of G9a does not impair CpG island methylation on the inactive X
29 chromosome in female embryos (**Figure S4C-D**) (Ohhata *et al.* 2004). Combined, these data indicate
30 that, in contrast to ESCs, G9a is dispensable for genome-wide DNA methylation in embryogenesis.

31

32 **Identification of hypomethylated sequences in *G9a*^{-/-} embryos.**

33 Despite global conservation, the methylomes of *G9a*^{-/-} embryos are clearly distinct and clustered
34 separately from WT embryos (**Figure S3B-C**). RRBS revealed that the inactivation of G9a leads to
35 focal changes with 956 regions losing more than 20% and 517 regions gaining more than 20%
36 methylation in *G9a*^{-/-} embryos (**Figure 3A**; **Table S1**). Hypomethylation occurs in promoters, gene
37 bodies, intergenic regions and occasionally in transposons of the LINE1 and LTR ERVK families

1 **(Figure 3B; Table S1)**. The *G9a* gene body itself is hypomethylated in *G9a*^{-/-} embryos after the site
 2 of LacZ insertion where transcription is aborted **(Figure S5)**, supporting the model that transcription
 3 promotes gene body methylation (Baubec et al. 2015). Strikingly, the absence of G9a also led to
 4 hypermethylation at a small set of CpG islands mostly within genes **(Figure 3B; Figure S6A)**.
 5 Because hypomethylation occurs more frequently and with higher amplitude, we focused on the
 6 regions losing methylation, which we called hypomethylated regions (HMRs). HMRs undergo *de novo*
 7 methylation at implantation **(Figure 3C)**, which excludes that their hypomethylation reflects the
 8 developmental delay of *G9a*^{-/-} embryos. The extent of DNA hypomethylation at HMRs in *G9a*^{-/-}
 9 embryos is similar to that caused by inactivation of the DNA methyltransferase DNMT3B **(Figure**
 10 **3C)**. Notably, promoter HMRs are enriched for promoters of germline-specific genes **(Figure 3D)**,
 11 including *Cyct*, *Naa11*, *Asz1*, *Hormad2*, *Morc2b*, *Pdha2*, *Ptpn20* and *Abca16* **(Figure 3E; Figure S7)**.
 12 Several other germline genes show a reduction of promoter methylation just below the threshold of
 13 20% (*Wdr20rt*, *Rpl10l*, *1700019A02Rik*) or, when not covered by RRBS, were identified as
 14 hypomethylated in the MeDIP data (*Slc9b1*, *4933427D06Rik*). HMRs also occur in promoter regions
 15 of somatic genes such as *Trim13*, *Tff3*, *Aplnr*, *Tenm2*, *Sh3tc2*, *Ano5* and genes of the *Xlr* imprinted
 16 cluster (*Xlr3a/b*, *Xlr4a/b/c*) **(Figure 3E; Figure S7)**. To validate HMRs, we performed COBRA on
 17 nine genes and confirmed the hypomethylation in *G9a*^{-/-} embryos compared to WT and *G9a*^{+/-}
 18 littermates **(Fig S8A-B)**. Bisulfite sequencing confirmed a decrease in methylation ranging from 32 to
 19 84% over multiple contiguous CpGs **(Figure 3F; Figure S8C)**. Taken together, our data show that
 20 G9a is required for the deposition of DNA methylation at specific genomic sites during
 21 embryogenesis, including the CpG-rich promoters of several germline genes.

22

23 **G9a participates in the maintenance of imprinted DNA methylation at the *Slc38a4* locus.**

24 We also examined the contribution of G9a to the maintenance of DNA methylation imprints *in vivo*.
 25 RRBS quantification of DNA methylation at 16 imprinted gDMRs revealed a methylation level close
 26 to 50% in WT embryos, and this is unchanged at all but one gDMR in *G9a*^{-/-} embryos **(Figure 4A;**
 27 **Figure S9A)**. This is confirmed by bisulfite sequencing and *McrBC* digestion at four gDMRs (*H19*,
 28 *Kcnq1ot1*, *Snrpn*, *Peg3*) in *G9a*^{-/-} embryos **(Figure 4B; Figure S9B)**. Remarkably, the absence of
 29 G9a leads to the specific hypomethylation of the *Slc38a4* gDMR, which carries allele-specific
 30 methylation inherited from the oocyte (Proudhon et al. 2012) **(Figure 4A)**. RRBS analysis shows that
 31 the extent of hypomethylation at the *Slc38a4* gDMR is variable between individual embryos **(Figure**
 32 **4C)**, indicating that there is a partially penetrant effect. Thus, G9a plays a locus-specific role in the
 33 maintenance of the DNA methylation imprint at the *Slc38a4* locus during embryogenesis.

34

35 **The influence of G9a on DNA methylation is direct.**

36 We next investigated the mechanisms by which G9a controls DNA methylation. RNA-Seq indicates
 37 that the expression of genes encoding DNMTs, UHRF1 and TETs is not perturbed in *G9a*^{-/-} embryos

1 (**Figure 5A**), which was confirmed by RT-qPCR (**Figure S10A**). We detected a mild upregulation of
2 *Dnmt3b* only, which likely is a consequence of the developmental delay of *G9a*^{-/-} embryos. This
3 finding agrees with data in ESCs (Dong et al. 2008; Tachibana et al. 2008) and suggests that *G9a* does
4 not influence DNA methylation indirectly by modulating the expression of the DNA methylation
5 machinery. To determine if *G9a* binds to HMRs, we performed chromatin immunoprecipitation (ChIP)
6 on primary mouse embryonic fibroblasts (MEFs) and found *G9a* enriched at several HMRs (**Figure**
7 **5B**). Importantly, we were also able to detect an enrichment of *G9a* at HMRs by ChIP *in vivo* in E8.5
8 embryos (**Figure 5C**). To further confirm the binding of *G9a* at HMRs in embryonic cells, we
9 explored a published *G9a* ChIP-Seq dataset in ESCs (Mozzetta et al. 2014) and found that peaks of
10 *G9a* frequently co-localize with HMRs, in particular in the promoters of germline genes (*Cyct*, *Naa11*,
11 *Asz1*, *Hormad2*, *Morc2b*, *Pdha2*, *Pgam2*, *Ptpn20*, *Abca16*, *Slc9b1*, *Wdr20rt*, *4933427D06Rik*,
12 *1700019A02Rik*) (**Figure 5D**; **Figure S11A**). Similarly, the ChIP-Seq data indicate a higher binding of
13 *G9a* in the vicinity of the *Slc38a4* gDMR compared to other imprinted gDMRs (**Figure S11B**). To
14 explore this in a systematic way, we computed the distribution of *G9a* ChIP-Seq signals around all
15 HMRs and observed that HMRs contain higher *G9a* signal than the surrounding regions (**Figure 5E**).
16 Conversely, hypermethylated DMRs rarely match peaks of *G9a*, suggesting that a proportion of them
17 might result from indirect effects (**Figure S6B**). Thus, *G9a* is frequently enriched in the proximity of
18 HMRs in embryonic cells, suggesting that it influences DNA methylation in *cis* at these targets.

19

20 **Interplay between H3K9me2 and DNA methylation in embryos.**

21 We then investigated if the influence of *G9a* on DNA methylation is related to its activity as a
22 methylase of H3K9. We performed ChIP-qPCR of H3K9me2 on WT E8.5 embryos and found that all
23 the tested HMRs are strongly enriched for H3K9me2 (**Figure 5F**). Conversely they are not marked by
24 the active histone mark H3K4me3 except for the *Slc38a4* imprinted gene (**Figure 5F**). A similar
25 pattern of histone marks was observed in primary embryonic fibroblasts (**Figure S12A**). For two
26 HMRs (*Naa11* and *Asz1*), we confirmed that H3K9me2 is reduced in *G9a*^{-/-} embryos, demonstrating
27 that *G9a* exerts its catalytic activity at these sites (**Figure 5G**). To extend these findings, we analyzed
28 ChIP-Seq profiles of H3K9me1/2 in ESCs (Liu et al. 2015) and also performed H3K9me2 ChIP-Seq
29 in E8.5 embryos. According to the ChIP-Seq data, HMRs are marked by H3K9 mono- and di-
30 methylation in ESCs and E8.5 embryos but the enrichment for these marks is minimal compared to the
31 surrounding sequences (**Figure 5E**). In fact, the ChIP-Seq signals indicate that H3K9me2 exists in
32 large blocks of chromatin covering most of the genome, suggesting that the presence of high levels of
33 H3K9me2 is not a hallmark of HMRs. ChIP-qPCR in E8.5 embryos confirmed that other gene
34 promoters carry H3K9me2 at levels comparable to HMRs and lose this mark in *G9a*^{-/-} embryos
35 without any detectable effect on their DNA methylation (**Figure 5G**). Thus, HMRs are marked by
36 H3K9me2, but there is no consistent correlation between the loss of H3K9me2 and DNA methylation
37 in *G9a*^{-/-} embryos.

1 DNA methylation is instrumental for G9a-mediated repression of germline genes in embryos.

2 Finally, we explored the impact of G9a on gene expression by conducting RNA-Seq in WT versus
 3 *G9a*^{-/-} E8.5 embryos. As expected (Tachibana et al. 2002), *Mage-a* genes are reactivated in *G9a*^{-/-}
 4 embryos (**Figure 6A**). We identified 252 genes as more than 3 fold up-regulated and 181 genes as
 5 more than 3 fold down-regulated in *G9a*^{-/-} embryos (**Figure 6A; Table S2**). The down-regulated
 6 genes are enriched for genes involved in neuronal and muscle morphogenesis (data not shown).
 7 Conversely, the up-regulated genes are enriched for testis-specific genes (**Figure 6B**) and the most up-
 8 regulated ones correspond to genes with reduced promoter DNA methylation: *Cyct*, *Naa11*, *Asz1*,
 9 *Pdha2*, *Ptpn20*, *Abca16*, *Hormad2* and *Slc9b1* (**Figure 6C-D; Table S2**). Notably, several genes with
 10 a promoter HMR (*Morc2b*, *Trim13*, *Tenm2*, *Sh3tc2*) are only modestly or not over-expressed in *G9a*^{-/-}
 11 embryos, indicating that promoter DNA hypomethylation is not a consequence of gene reactivation.
 12 This pattern of gene reactivation upon inactivation of G9a was validated by RT-qPCR on independent
 13 embryos (**Figure S10B**). To further test the role of DNA methylation in the derepression of these G9a
 14 targets, we analyzed RNA-Seq from embryos deficient for the DNA methyltransferase *Dnmt3b*
 15 (Auclair et al. 2014) and found that the germline genes are also overexpressed when hypomethylated
 16 in *Dnmt3b* mutants, albeit not to the extent seen in *G9a* mutants (**Figure 6D**). We monitored
 17 H3K9me2 by CHIP in *Dnmt3b*^{-/-} embryos and found that the reduced DNA methylation does not
 18 impair the deposition of H3K9me2 at HMRs (**Figure 6E; Figure S12A**). Furthermore, we found no
 19 incidence of the reduced DNA methylation on the binding of G9a at promoter HMRs in *Dnmt3b*^{-/-}
 20 fibroblasts (**Figure S12B**). We conclude that the dependence of CpG methylation on G9a is not
 21 accompanied by a reciprocal dependence of G9a on CpG methylation, and that H3K9me2 alone is not
 22 sufficient to induce an efficient silencing of the germline genes targeted by G9a. Collectively, these
 23 results demonstrate that G9a acts upstream of DNA methylation, and that G9a-guided DNA
 24 methylation is instrumental for the repression of these G9a targets *in vivo* (**Figure 6F**).

25

26 DISCUSSION

27

28 The molecular pathways guiding DNA methylation in mammalian genomes remain poorly understood.
 29 Here we addressed the role of the lysine methyltransferase G9a in the control of DNA methylation in
 30 the mouse *in vivo*. G9a controls DNA methylation at specific sites, including the CpG-rich promoters
 31 of germline-specific genes. The absence of G9a also affects the maintenance of the DNA methylation
 32 imprint at the *Slc38a4* gDMR, but not at gDMRs of other imprinted loci, which highlights the
 33 existence of locus-specific mechanisms of maintenance of imprinted DNA methylation. *Slc38a4*
 34 differs from other imprinted gDMRs in that it shows no ZFP57 binding (Saadeh and Schulz 2014;
 35 Strogantsev et al. 2015), displays no allelic ATRX binding and H3.3 incorporation (Voon et al. 2015)
 36 and also lacks H3K9me3 in ES cells. These marked differences could explain why this gDMR

1 uniquely relies on alternative, G9a-dependent, mechanisms for the maintenance of its germline-
2 derived allelic DNA methylation.

3

4 Does G9a promote DNA methylation *in cis*? Several arguments suggest that this is the case. First, G9a
5 binding is detected by ChIP at many of the regions that show DNA hypomethylation. Second, we did
6 not observe a deregulation of genes encoding DNMT or TET proteins in *G9a*^{-/-} embryos. Third,
7 several of the hypomethylated genes are not overexpressed in *G9a*^{-/-} embryos, indicating that their
8 hypomethylation is not a secondary consequence of gene activation. We therefore speculate that the
9 absence of G9a impairs the recruitment of DNA methylation *in cis* at these targets. This scenario,
10 however, does not exclude that DNA methylation may be affected indirectly at some sites.

11

12 What mechanisms link G9a to DNA methylation? H3K9me2 is recognized by UHRF1, a cofactor for
13 DNMT1, suggesting that G9a-mediated H3K9me2 could stimulate the maintenance of DNA
14 methylation by the UHRF1-DNMT1 complex (Karagianni et al. 2008; Rothbart et al. 2012; Liu et al.
15 2013). HMRs are marked by H3K9me2 and the loss of DNA methylation at HMRs correlates with a
16 decrease in H3K9me2 in *G9a*^{-/-} embryos. However this model is in contradiction with the observation
17 that H3K9me2 covers a large part of the genome and that DNA methylation at other sites of H3K9me2
18 loss is unaffected in *G9a*^{-/-} embryos. Thus, our results are in favour of a role of G9a in DNA
19 methylation that is independent of H3K9 methylation. G9a methylates other histone residues such as
20 H3K56 (Yu et al. 2012) and lysines in H1 variants (Trojer et al. 2009; Weiss et al. 2010), and has
21 several non-histone protein substrates (Rathert et al. 2008), including DNMT3A (Chang et al. 2011),
22 which could be necessary for effective methylation of DNA. Alternatively, as supported by studies in
23 ESCs (Dong et al. 2008; Epsztejn-Litman et al. 2008; Tachibana et al. 2008), G9a could regulate DNA
24 methylation independently of its catalytic activity. G9a physically interacts with the DNA
25 methyltransferases DNMT1, DNMT3A and DNMT3B (Esteve et al. 2006; Epsztejn-Litman et al.
26 2008; Kotini et al. 2011) and could stimulate their recruitment at specific genomic sites. It also
27 remains to be studied if the G9a heterodimeric partner GLP functionally overlaps with G9a in the
28 control of embryonic DNA methylation.

29

30 Our data shed new light on the role of G9a during embryogenesis. RNA-Seq demonstrated that the
31 inactivation of G9a leads to the reactivation of a few hundred genes. Besides known targets of G9a
32 such as the genes of the *Mage-a* family, many of the most upregulated genes correspond to germline
33 genes and genes of the imprinted *Xlr* cluster harboring reduced promoter DNA methylation. By
34 comparing embryos deficient for G9a or the DNA methyltransferase DNMT3b, we provide evidence
35 that the recruitment of DNA methylation is important for the biological functions of G9a. Indeed,
36 germline genes are reactivated in *Dnmt3b*^{-/-} embryos even though the binding of G9a and the levels of
37 H3K9me2 are unchanged, indicating that DNA methylation is important to achieve efficient silencing

1 of these G9a target genes. We note, however, that the extent of gene reactivation in *Dnmt3b*^{-/-}
2 embryos does not reach the levels seen in *G9a*^{-/-} embryos, suggesting that H3K9 and DNA
3 methylation cooperatively silence these genes. Moreover, other genes like *Magea2* and *Wfdc15a* are
4 reactivated in *G9a*^{-/-} embryos without a reduction in their promoter DNA methylation, illustrating that
5 G9a also regulates genes independently of DNA methylation.

6
7 The mechanisms that specify the genes repressed by G9a are still unclear. One hypothesis is that G9 is
8 recruited by long non-coding RNAs (lncRNA). This model has emerged from studies on imprinted
9 domains where specific recruitment of histone lysine methyltransferases controls the allelic repression
10 at placenta-specific genes of the *Kcnq1* and *Igf2r* loci in the mouse (Nagano et al. 2008; Wagschal et
11 al. 2008). In humans, similarly, expression of an antisense lncRNA mediates G9A recruitment and
12 local DNA methylation at the *DHRS4* gene locus (Li et al. 2012). Another recent study shows that
13 chromatin at the *Slc38a4* locus is controlled by a lncRNA (Monnier et al. 2013), again establishing a
14 possible link between non-coding RNA and G9a recruitment. Alternatively, it is known that G9a is
15 part of larger protein complexes with sequence-specific DNA binding factors that could guide G9a to
16 specific sites. In particular, G9a was identified as a member of the E2F6 complex (Ogawa et al. 2002),
17 which has been linked to the control of DNA methylation in mouse cells (Velasco et al. 2010).

18
19 We show that the influence of G9a on DNA methylation is much more restricted in embryos compared
20 to cultured ESCs. This is in line with single-gene studies showing that *Wfdc15a*, *Magea2* and *Snrpn*
21 are hypomethylated in *G9a*^{-/-} ESCs but not embryos (Xin et al. 2003; Tachibana et al. 2008). In
22 addition, the ectopic inactivation of G9a has little effect on DNA methylation in differentiated cells
23 (Link et al. 2009; Sharma et al. 2012). One hypothesis to explain this discrepancy is that G9a could
24 have a greater influence on the kinetics of DNA methylation at earlier timepoints of embryogenesis.
25 Unfortunately we failed to test this hypothesis because *G9a*^{-/-} embryos implant much later than WT
26 embryos, which precluded us from dissecting mutant and control embryos from the same mothers at
27 E6.5 and makes it difficult to compare the kinetics of methylation because of the confounding effect of
28 the developmental delay. Alternatively, the absence of G9a in ESCs could lead to indirect effects on
29 DNA methylation, for example by promoting a naïve pluripotent state associated with a global
30 reduction in DNA methylation (Ficz et al. 2013; Habibi et al. 2013). Interestingly, studies on other
31 epigenetic regulators also revealed that knock-out embryos have milder methylation defects compared
32 to ESCs. For example, loss of SUV39H1/H2 leads to a reduction in DNA methylation at satellite DNA
33 in ESCs (Lehnertz et al. 2003), but this is not apparent in embryo-derived fibroblasts (Pannetier et al.
34 2008). This indicates that the mechanisms controlling DNA methylation are less robust in cultured
35 stem cells and illustrates the remarkable epigenetic robustness of the mammalian embryo.

1 METHODS

2

3 **Mouse embryos.** The morning of the vaginal plug was designated E0.5. Pre-implantation blastocysts
4 (E3.5-E4.5) were collected by flushing the uteri with M2 medium (Sigma-Aldrich). Post-implantation
5 epiblasts (E5.5-E7.5) and embryos (E8.5-E9.5) were manually dissected in M2 medium. We prepared
6 genomic DNA samples by proteinase K digestion, phenol/chloroform extraction and precipitation with
7 ethanol. We maintained *G9a* knockout mice on a C57BL/6J background and obtained *G9a*^{-/-} embryos
8 by natural mating of *G9a*^{+/-} females and males. Genotyping was performed by two PCR reactions
9 using primers that amplify the LacZ reporter cassette and the exons 11-12 (**Figure S1B**). *Dnmt3b*^{-/-}
10 embryos were obtained as described (Auclair et al. 2014).

11

12 **Isolation and culture of mouse embryonic fibroblasts (MEFs).** We generated primary MEFs from
13 E13.5 embryos. The embryos were mechanically dissociated and incubated with Trypsin 0.25% at
14 37°C for 10 min. MEFs were isolated by differential attachment and cultured in Dulbecco's Modified
15 Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 50µg/mL
16 gentamycin in a humidified incubator at 37°C with 5% CO₂. We used MEFs at low passage (< p4).

17

18 **DNA Methylation analysis by restriction enzyme digestion and bisulfite treatment.** We digested 1
19 µg of genomic DNA with 20U of *Mcr*BC (New England Biolabs), an endonuclease which digests
20 methylated DNA in a sequence independent manner, followed by qPCR to measure the percentage of
21 digestion relative to the undigested control DNA. We normalized values to those of an intergenic
22 control sequence containing no CpGs. Bisulphite conversion of genomic DNA was performed with the
23 EpiTect Bisulfite kit (Qiagen). We performed PCR amplification of converted DNA followed by
24 COBRA or cloning as described (Borgel et al. 2010). Cloned PCR products were sequenced, aligned
25 with the BISMA software and filtered to remove clonal biases. Primer sequences for qPCR and PCR
26 amplification of bisulfite-treated DNA are provided in the **Table S3**.

27

28 **MeDIP-Chip.** We performed MeDIP on 200ng genomic DNA prepared from WT and *G9a*^{-/-}
29 littermate embryos collected at E9.5. Precipitates from three replicates of MeDIP and input samples
30 were hybridized to Roche Nimblegen mm9 HD2 2.1M Deluxe Promoter arrays covering -8200 to
31 +3000 bp from 23,517 TSS. Sample preparation and bio-informatic treatment of micro-array data were
32 performed as described (Borgel et al. 2010). For data representation we smoothed log₂ ratios over
33 200bp windows using the Ringo R package.

34

35 **RRBS.** We performed RRBS on single embryos, on four WT and three *G9a*^{-/-} littermate embryos
36 collected at E8.5. RRBS libraries were produced as described (Auclair et al. 2014). Briefly, we
37 digested 70ng of genomic DNA five hours with *Msp*I (Thermo Scientific), performed end-repair and

1 A-tailing (with 5U Klenow-fragment, Thermo Scientific), and ligated to methylated adapters (with
2 30U T4 DNA ligase, Thermo Scientific) in Tango 1X buffer. Fragments between 150 and 400 bp were
3 excised from a 3% agarose 0.5 X TBE gel with the MinElute gel extraction kit (Qiagen) and bisulfite-
4 converted with the EpiTect Bisulfite kit (Qiagen) with two consecutive rounds of conversion. Final
5 RRBS libraries were amplified by PCR with the PfUTurbo Cx hotstart DNA polymerase (Agilent)
6 using the following PCR conditions: 95°C for 2 min, 16 cycles (95°C for 30 s, 65°C for 30 s, 72°C for
7 45 s), 72°C for 7 min. We purified the libraries with AMPure magnetic beads (Beckman Coulter) and
8 performed paired-end sequencing (2x75bp) on an Illumina HiSeq2500 at Integragen SA (Evry,
9 France). ESCs lines were cultivated as described (Mozzetta et al. 2014) and RRBS was performed
10 with 100ng starting DNA and 14 cycles for the final PCR. We cleaned the sequencing reads with Trim
11 Galore (v0.2.1, parameters --rrbs --paired -r1 30 -r2 30 -q 20 --length 20 --retain_unpaired) and
12 aligned to the mm10 genome with BSMAP (v2.74, parameters -v 2 -w 100 -r 1 -x 400 -m 30 -D C-
13 CGG -n 1). Percent methylation values were calculated as the ratio of the number of Cs over the total
14 number of Cs and Ts with methratio.py (parameters -z -u -g). The bisulfite conversion efficiency was
15 estimated by calculating the C to T conversion in non-CpG sites. For all data analysis, we filtered
16 CpGs to have a minimum sequencing depth of 8x and visualized methylation values with the IGV
17 browser.

18

19 **RT-qPCR analysis.** Total RNAs were extracted from embryos with Trizol (Invitrogen), treated with
20 the RQ1 DNase, and reverse transcribed with the QuantiTect Reverse Transcription Kit (Qiagen). We
21 quantified the expression of target cDNAs by qPCR with the KAPA SYBR FASR qPCR kit on a
22 StepOnePlus realtime PCR system (Life technologies) using the standard curve method. qPCR
23 reactions were performed in triplicates and expression was normalized to the expression of two
24 housekeeping genes (*Rpl13a*, *Actb*). In parallel we amplified no-RT controls to rule out the presence
25 of contaminating DNA. Primer sequences for qPCR are provided in the **Table S3**.

26

27 **RNA-Seq.** We performed RNA-Seq on single embryos, on two WT and two *G9a*^{-/-} littermate
28 embryos collected at E8.5. Total RNAs were extracted with the RNeasy Protect Mini Kit (Qiagen).
29 RNA-Seq libraries were prepared with the Ovation RNA-Seq System V2 (NuGEN) and sequenced on
30 a HiSeq2500 (1x50bp). Raw sequences were aligned to the mouse mm10 genome by TopHat (v2.0.13,
31 default parameters plus --no-coverage-search --library-type fr-firststrand) using a transcriptome index
32 built from RefSeq mm10. Browser tracks in bigwig format were generated with bam2wig.py from the
33 RSeQC package (v2.4, parameters -u -t 5000000000) and visualized with the IGV browser. Read
34 counts in genes were computed with htseq-count (v0.6.0, parameters -t exon -s no) using a RefSeq
35 mm10 gtf file. We tested for differential gene expression using DESeq2 (v1.4.5) and defined
36 differentially expressed genes as having a fold change greater than 3 and an adjusted p-value below

1 0.01. Genes on the Y chromosome were excluded. Normalized counts and FPKM scores were
2 calculated with the 'counts' and 'fpkm' functions of DESeq2.

3
4 **ChIP of G9a.** We cross-linked 5.10⁶ MEFs with 2 mM EGS (Life Technologies) for 45 min and then
5 with 1% formaldehyde (Thermo Scientific) for 10 min in DPBS. The excess formaldehyde was
6 quenched with 125mM glycine for 5 min at RT. The cells were scraped and lysed in 200 μ l lysis
7 Buffer (1% SDS, 10 mM EDTA, 50 mM Tris pH 8.1) supplemented with protease inhibitors
8 (Diagenode) for 10 min on ice. We sonicated the chromatin using a Bioruptor waterbath sonicator
9 (Diagenode) to obtain fragments in the range of 0.2 to 1.5 kb. The sheared chromatin was cleared by
10 centrifugation at 14,000 rpm for 20 min at 4°C and diluted 1:10 in dilution Buffer (0.01% SDS, 1.1%
11 Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.1, 167 mM NaCl, protease inhibitors). 10% of
12 the chromatin was set aside as the input. The chromatin was immunoprecipitated overnight at 4°C with
13 DiaMag protein A-coated magnetic beads (Diagenode) pre-bound to a mouse monoclonal G9a
14 antibody (R&D Systems #PP-A8620A-00, Clone A8620A) or an IgG2a isotype control (R&D
15 Systems #MAB0031). For the pre-binding of beads, 5 μ g antibody were coupled to 17 μ L magnetic
16 beads in 90 μ L of dilution buffer for 3 h at 4°C. The beads-antibody complexes were washed once in
17 low salt buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.1, 150 mM NaCl),
18 once in high salt buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.1, 500
19 mM NaCl), once in LiCl buffer (0.25 M LiCl, 1% IGEPAL-CA630, 1% sodium deoxycholate, 1 mM
20 EDTA, 10 mM Tris-HCl pH 8.1), and twice in TE Buffer (10 mM Tris-HCl pH 8.1, 1 mM EDTA).
21 The beads were eluted twice with the elution buffer (1% SDS, 0.1 M NaHCO₃) at RT for 15 min and
22 the two eluates were combined. The cross-linking was reversed in 200 mM NaCl at 65°C for 4h. We
23 isolated DNA by proteinase K digestion at 45°C for 1h followed by phenol/chloroform extraction and
24 precipitation in EtOH. For ChIP on E8.5 embryos, we dissociated 8 embryos in Trypsin 0.25% EDTA
25 1mM for 5 min at RT, and cross-linked the cells in suspension with 1.5 mM EGS for 30 min followed
26 with 1 % formaldehyde for 10 min in 1 mL of DPBS. All the following steps are the same as described
27 above. The qPCR analysis was done with the KAPA SYBR FASR qPCR kit on a StepOnePlus
28 realtime PCR system (Life technologies), and we calculated for each gene the percentage of recovery
29 in the ChIP relative to the input. Primer sequences for qPCR are provided in the **Table S3**.

30
31 **ChIP of histone methylation.** 5-8 embryos collected at E8.5 were dissociated in Trypsin 0.25%
32 EDTA 1mM for 5 min at RT, washed in PBS and cross-linked with 1 % formaldehyde for 10 min at
33 RT. The excess formaldehyde was quenched with 125mM glycine for 5 min at RT. The embryos were
34 washed in PBS supplemented with protease inhibitors (Diagenode), and ChIP was performed with a
35 Bioruptor waterbath sonicator (Diagenode) and the LowCell ChIP kit (Diagenode). We used
36 antibodies against H3K9me2 (Abcam ab1220), H3K4me3 (Abcam ab8580) and an IgG2a control
37 (R&D Systems #MAB0031). ChIP on MEFs was performed on 200,000 cells with the LowCell ChIP

1 kit (Diagenode). For experiments on *G9a*^{-/-} and *Dnmt3b*^{-/-} embryos, we performed ChIP on single
2 embryos. The embryos were dissociated and cross-linked as described above and ChIP was performed
3 with the True MicroChIP kit (Diagenode). Primer sequences for qPCR are provided in the **Table S3**.

4
5 **H3K9me2 ChIP-Seq.** ChIP was performed as described above on three pools of 8 embryos collected
6 at E8.5. ChIP-Seq libraries were prepared with the NEXTflex ChIP-Seq Kit (Bioo Scientific) and
7 sequenced on a HiSeq2500 (1x50bp). Because of the broad distribution of H3K9me2, we merged the
8 Fastq files from the three experiments. Raw reads were aligned to the mm10 genome with bowtie2
9 (v2.2.4, default parameters). We flagged PCR duplicates with Picard MarkDuplicates (v1.136) and
10 generated density tracks in wig format with igvtools count (v2.3.3.2, parameters -w 25 -e 300 --
11 minMapQuality 30).

12
13 **Data analysis and characterization of HMRs.** Data processing and representation was performed
14 with the R software using custom developed scripts. To represent the distribution of methylation over
15 genes (**Figure 1F**), we calculated the average methylation in twenty equal-sized windows of each
16 protein-coding RefSeq gene and in ten 1 kb windows of flanking sequences. To calculate methylation
17 of repeat families, we averaged the methylation scores of all the CpGs contained in annotated repeats
18 with a size above 400bp. We used the eDMR algorithm from the methylKit R package (Li et al. 2013)
19 to identify RRBS DMRs between WT and *G9a*^{-/-} embryos with at least 3 CpGs, a difference in
20 methylation greater than 20% and an adjusted p-value below 0.01. We annotated DMRs relative to
21 exons, introns and repeats using the RefSeq gene annotation and annotated repeats with a size above
22 400bp. DMRs were considered promoter-proximal when the center of the DMR is distant less than
23 1000 bp from a RefSeq TSS. We performed gene ontology and tissue expression analyses with the
24 DAVID annotation tools. HMRs were compared with ChIP-Seq data for G9a in ESCs (GSE46545),
25 H3K9me1/2 in ESCs (GSE54412), H3K4me2 in ESCs (GSE30203) and H3K9me2 in E8.5 embryos
26 (this study). To generate the heatmaps, we counted the density of reads from the ChIP-Seq wig files in
27 200 bp windows and divided by the density of reads in the same windows in the corresponding input
28 sample.

29 30 **DATA ACCESS**

31 The MeDIP-Chip, RRBS, ChIP-Seq and RNA-Seq data have been deposited in NCBI's Gene
32 Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo>) under the accession number
33 GSE71500.

34 35 **ACKNOWLEDGEMENTS**

36 We are grateful to Dr Yoichi Shinkai for the authorization to use G9a KO cell lines. We thank
37 Lauriane Fritsch, Julien Pontis and Slimane Ait-Si-Ali for providing cell lines, sharing G9a ChIP-Seq

1 data and for scientific discussions. We also thank the staff of the IGBMC high-throughput sequencing
 2 facility. This work was supported by the Agence National de Recherche (ANR Blanc ‘EMPREINTE’),
 3 the Institut National du Cancer (INCa_5960, INCa_7889), the Association pour la Recherche contre le
 4 Cancer (ARC SFI20101201555, SFI20121205729), the Fondation pour la Recherche Médicale, the
 5 Agency for International Cancer Research (UK), the MEDDTL (11-MRES-PNRPE-9-CVS-072), the
 6 LABEX EpiGenMed, the ATIP-AVENIR program, the EpiGeneSys Network of Excellence and the
 7 European Research Council (ERC Consolidator grant n°615371). G.A. was supported by fellowships
 8 from the French Ministère de la Recherche and from the Ligue Contre le Cancer.

10 AUTHOR CONTRIBUTIONS

11 G.A., J.B., L.A.S, J.V., P.C. and M.G. conducted the experiments. S.G. and M.D. processed the
 12 sequencing data and performed data analysis. R.F. and M.W. designed the study and, together with
 13 T.F., participated in data analysis. R.F. and M.W. wrote the paper.

15 FIGURE LEGENDS

17 **Figure 1. Global impact of G9a on the DNA methylome of mouse ESCs.** **A.** Density histograms of
 18 RRBS methylation scores at CpGs in WT ESCs, *G9a*^{-/-} ESCs and *G9a*^{-/-}ESCs rescued with G9a. **B.**
 19 Average distribution of RRBS methylation over RefSeq genes and flanking sequences in WT, *G9a*^{-/-}
 20 and G9a-rescued ESCs. **C.** Heatmap representation of the RRBS methylation scores measured in the
 21 promoters (-1000 to +1000bp) of selected germline and developmental genes. **D.** Violin plots of RRBS
 22 methylation scores in retrotransposons.

24 **Figure 2. Impact of G9a on the methylome in mouse embryos.** **A.** Expression of G9a mRNA
 25 measured by RT-qPCR in early mouse embryos, depicted as a ratio relative to the expression of two
 26 housekeeping genes (*Actb* and *Rpl13a*) (mean ± SEM, n=2 technical replicates on 5-10 pooled
 27 embryos). **B.** Promoter DNA methylation of candidate genes measured by COBRA in WT and *G9a*^{-/-}
 28 E9.5 embryos. The restriction fragments marked with an asterisk are end products of the digestion
 29 (indicating DNA methylation). **C.** Promoter DNA methylation of the *Pou5f1* and *Dppa3* genes
 30 measured by bisulfite sequencing in WT and *G9a*^{-/-} E9.5 embryos. Circles represent methylated
 31 (black) or unmethylated (white) CpG dinucleotides; each horizontal line is one sequenced clone. **D.**
 32 Pairwise comparison of MeDIP log₂ ratios at individual oligos in WT and *G9a*^{-/-} E9.5 embryos. The
 33 density of data points increases from blue to dark red. The Pearson correlation coefficient (r) is
 34 indicated on the graph. **E.** Density histograms of RRBS CpG methylation scores in WT and *G9a*^{-/-}
 35 E8.5 embryos. **F.** Average distribution of RRBS methylation over RefSeq genes and flanking
 36 sequences in WT and *G9a*^{-/-} embryos. **G.** Violin plots of RRBS methylation measured in CpG islands
 37 (CGIs) and retrotransposons. **H.** Example of methylation profiles in the promoter of the *Dazl* gene in

1 WT and *G9a*^{-/-} embryos. The upper tracks depict smoothed MeDIP log₂ ratios of individual
 2 oligonucleotides and the bottom tracks depict RRBS methylation scores at individual CpGs. The CpG
 3 density is shown in black.

4

5 **Figure 3. Identification of hypomethylated regions (HMRs) in *G9a*-deficient embryos. A.**
 6 Histogram of the number of sequences that gain or lose more than 20% methylation in the RRBS from
 7 *G9a*^{-/-} compared to WT embryos. **B.** Pie charts representing the percentages of hypomethylated and
 8 hypermethylated sequences mapping to promoters, gene bodies, intergenic regions and transposable
 9 elements (TEs). **C.** Boxplot showing the dynamics of methylation at HMRs (filtered to lose more than
 10 30% methylation in *G9a*^{-/-} embryos) in early embryos, *G9a*^{-/-} E8.5 embryos (green) and *Dnmt3b*^{-/-}
 11 E8.5 embryos (red). **D.** Preferential site of expression of genes with a promoter-proximal HMR (-1000
 12 to +1000 bp relative to the TSS). The dashed line represents the position for a p-value of 0.05. **E.**
 13 Examples of promoter-proximal HMRs. The sites of hypomethylation identified by MeDIP (grey
 14 boxes) confirm the differences measured by RRBS. **F.** Bisulfite cloning and sequencing of selected
 15 promoter HMRs in WT and *G9a*^{-/-} E9.5 embryos. The values below the sequencing data indicate the
 16 percentage of methylated CpGs in the amplicon.

17

18 **Figure 4. Locus-specific control of DNA methylation imprints by *G9a* at the *Slc38a4* gDMR. A.**
 19 Methylation scores measured by RRBS in 16 known imprinted gDMRs in WT and *G9a*^{-/-} embryos. **B.**
 20 Bisulfite cloning and sequencing analysis of four gDMRs in WT and *G9a*^{-/-} E9.5 embryos. **C.** MeDIP
 21 profiles at the imprinted *Slc38a4* locus (top) and RRBS scores in four WT and three *G9a*^{-/-} embryos
 22 (bottom). The orange rectangle marks the position of the gDMR.

23

24 **Figure 5. HMRs are bound by *G9a* and marked by H3K9me2. A.** RNA-Seq quantification of the
 25 expression of genes encoding components of the DNA methylation machinery in *G9a*^{-/-} and WT E9.5
 26 embryos (mean FPKM ± SEM, n=2 embryos). **B.** ChIP-qPCR analysis of *G9a* binding at HMRs in
 27 primary MEFs, represented as the percentage of input (mean ± SEM, n=5). ChIP assays were
 28 performed with an antibody against *G9a* and a control rabbit IgG. *Actb* served as a negative control
 29 and *Rhox11* was chosen as a positive control (Myant *et al.* 2011). **C.** ChIP-qPCR analysis of *G9a*
 30 binding at HMRs in E8.5 embryos (mean ± SEM, n=4). **D.** Browser views of *G9a* ChIP-Seq profiles in
 31 ESCs (Mozzetta *et al.* 2014) reveal that peaks of *G9a* binding co-localize with promoter-proximal
 32 HMRs (red bars) at three germline genes. **E.** Heatmap representation of the distribution of *G9a* and
 33 H3K9 mono- and di-methylation at HMRs. The data represents the average density of ChIP-Seq reads
 34 for *G9a* in ESCs, H3K9me1/2 in ESCs, and H3K9me2 in E8.5 embryos normalized by the density of
 35 reads in the input control. **F.** ChIP-qPCR analysis of H3K9me2 and H3K4me3 at HMRs in E8.5
 36 embryos, represented as the percentage of input (mean ± SEM, n=4). The promoters of the

1 housekeeping genes *Actb* and *Ube2f* served as controls. **G.** ChIP-qPCR analysis of H3K9me2 at four
 2 gene promoters in WT and *G9a*^{-/-} E8.5 embryos (mean ± SEM, n=6 embryos for WT, n=4 embryos
 3 for *G9a*^{-/-}). The heatmap on the bottom indicates CpG methylation measured by RRBS in the same
 4 promoters. ** = p<0.01 (t-test).

5

6 **Figure 6. G9a represses germline genes via DNA methylation in mouse embryos.** **A.** Comparison
 7 of RNA-Seq expression levels for RefSeq genes in WT and *G9a*^{-/-} embryos. Genes of the *Mage-a*
 8 family and differentially expressed genes are highlighted in colors. **B.** Preferential tissue of expression
 9 of genes upregulated at least 3 fold in *G9a*^{-/-} embryos. **C.** Examples of RNA-Seq profiles at the *Cyct*
 10 and *Asz1* genes in two biological replicates of WT and *G9a*^{-/-} embryos. **D.** Activation of germline
 11 genes in *G9a*^{-/-} and *Dnmt3b*^{-/-} E8.5 embryos. The heatmap on the bottom indicates CpG methylation
 12 measured by RRBS in the corresponding promoters in *G9a*^{-/-} and *Dnmt3b*^{-/-} E8.5 embryos. **E.** ChIP-
 13 qPCR analysis of H3K9me2 in WT and *Dnmt3b*^{-/-} E8.5 embryos (mean ± SEM, n=3 embryos for WT,
 14 n=4 embryos for *Dnmt3b*^{-/-}), showing that the reduced DNA methylation does not impact the
 15 deposition of H3K9me2. **F.** Model. G9a deposits H3K9me2 and facilitates cytosine methylation at a
 16 subset of gene promoters in embryos. The inactivation of G9a inhibits H3K9me2 and leads to reduced
 17 cytosine methylation, leading to aberrant gene activation. In *Dnmt3b*^{-/-} embryos, G9a is able to bind to
 18 its target promoters but can no longer recruit cytosine methylation, which leads to incomplete gene
 19 silencing.

20

21 REFERENCES

22

23 Athanasiadou R, de Sousa D, Myant K, Merusi C, Stancheva I, Bird A. 2010. Targeting of de novo
 24 DNA methylation throughout the Oct-4 gene regulatory region in differentiating embryonic stem cells.
 25 *PloS one* **5**(4): e9937.

26 Auclair G, Guibert S, Bender A, Weber M. 2014. Ontogeny of CpG island methylation and specificity
 27 of DNMT3 methyltransferases during embryonic development in the mouse. *Genome biology* **15**(12):
 28 545.

29 Baubec T, Colombo DF, Wirbelauer C, Schmidt J, Burger L, Krebs AR, Akalin A, Schubeler D. 2015.
 30 Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation.
 31 *Nature* **520**(7546): 243-247.

32 Borgel J, Guibert S, Li Y, Chiba H, Schubeler D, Sasaki H, Forne T, Weber M. 2010. Targets and
 33 dynamics of promoter DNA methylation during early mouse development. *Nature genetics* **42**(12):
 34 1093-1100.

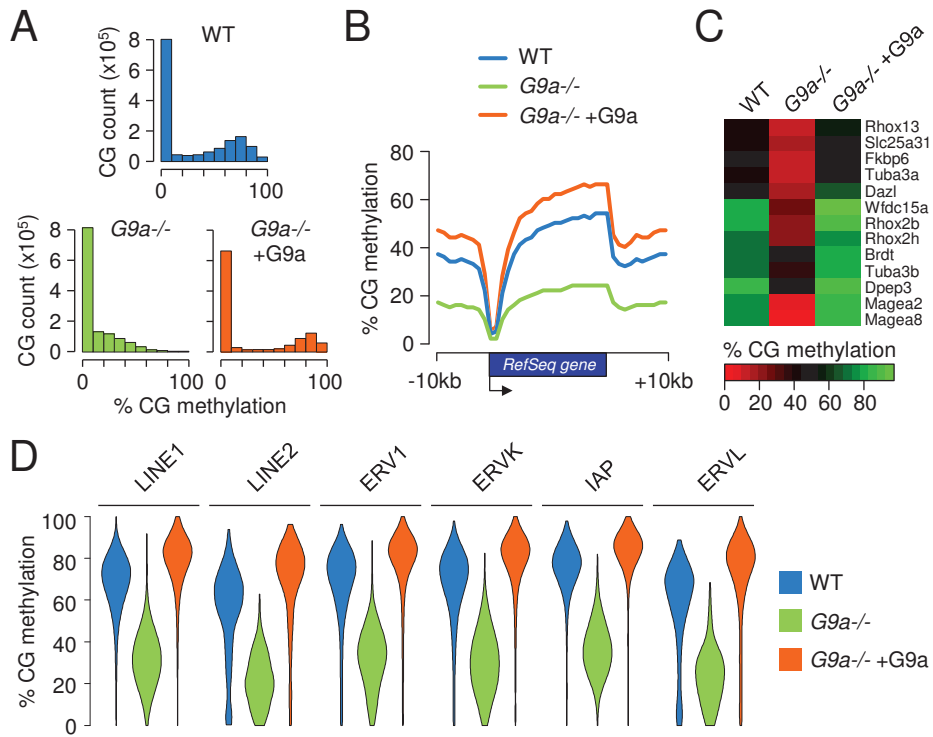
- 1 Chang Y, Sun L, Kokura K, Horton JR, Fukuda M, Espejo A, Izumi V, Koomen JM, Bedford MT,
2 Zhang X et al. 2011. MPP8 mediates the interactions between DNA methyltransferase Dnmt3a and
3 H3K9 methyltransferase GLP/G9a. *Nature communications* **2**: 533.
- 4 Dong KB, Maksakova IA, Mohn F, Leung D, Appanah R, Lee S, Yang HW, Lam LL, Mager DL,
5 Schubeler D et al. 2008. DNA methylation in ES cells requires the lysine methyltransferase G9a but
6 not its catalytic activity. *The EMBO journal* **27**(20): 2691-2701.
- 7 Epsztejn-Litman S, Feldman N, Abu-Remaileh M, Shufaro Y, Gerson A, Ueda J, Deplus R, Fuks F,
8 Shinkai Y, Cedar H et al. 2008. De novo DNA methylation promoted by G9a prevents reprogramming
9 of embryonically silenced genes. *Nature structural & molecular biology* **15**(11): 1176-1183.
- 10 Esteve PO, Chin HG, Smallwood A, Feehery GR, Gangisetty O, Karpf AR, Carey MF, Pradhan S.
11 2006. Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during
12 replication. *Genes & development* **20**(22): 3089-3103.
- 13 Feldman N, Gerson A, Fang J, Li E, Zhang Y, Shinkai Y, Cedar H, Bergman Y. 2006. G9a-mediated
14 irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nature cell biology* **8**(2):
15 188-194.
- 16 Ficiz G, Hore TA, Santos F, Lee HJ, Dean W, Arand J, Krueger F, Oxley D, Paul YL, Walter J et al.
17 2013. FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic
18 ground state of pluripotency. *Cell stem cell* **13**(3): 351-359.
- 19 Fuks F, Hurd PJ, Deplus R, Kouzarides T. 2003. The DNA methyltransferases associate with HP1 and
20 the SUV39H1 histone methyltransferase. *Nucleic acids research* **31**(9): 2305-2312.
- 21 Habibi E, Brinkman AB, Arand J, Kroeze LI, Kerstens HH, Matarese F, Lepikhov K, Gut M, Brun-
22 Heath I, Hubner NC et al. 2013. Whole-genome bisulfite sequencing of two distinct interconvertible
23 DNA methylomes of mouse embryonic stem cells. *Cell stem cell* **13**(3): 360-369.
- 24 Ikegami K, Iwatani M, Suzuki M, Tachibana M, Shinkai Y, Tanaka S, Grealley JM, Yagi S, Hattori N,
25 Shiota K. 2007. Genome-wide and locus-specific DNA hypomethylation in G9a deficient mouse
26 embryonic stem cells. *Genes to cells : devoted to molecular & cellular mechanisms* **12**(1): 1-11.
- 27 Karagianni P, Amazit L, Qin J, Wong J. 2008. ICBP90, a novel methyl K9 H3 binding protein linking
28 protein ubiquitination with heterochromatin formation. *Molecular and cellular biology* **28**(2): 705-
29 717.

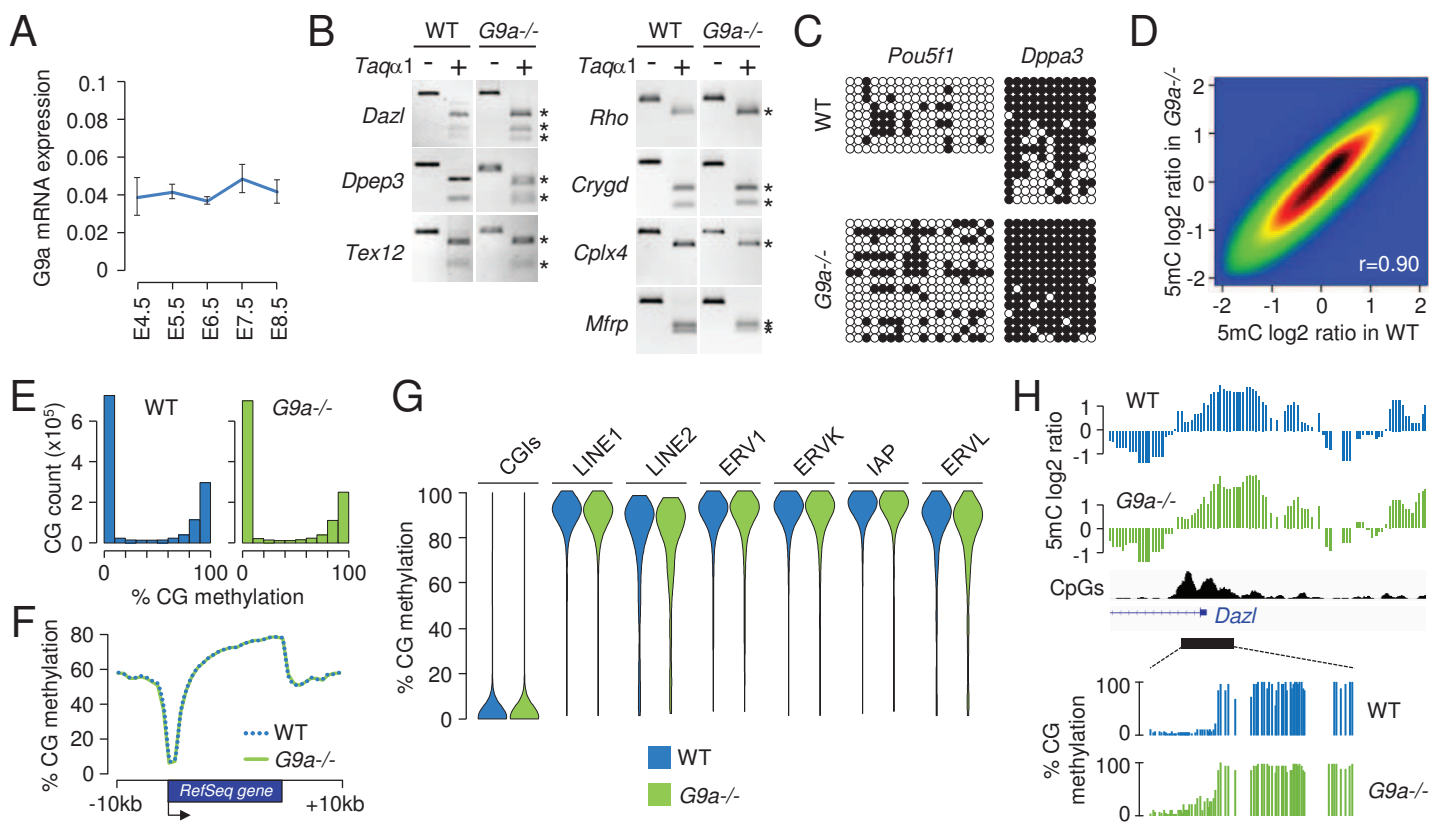
- 1 Kim JK, Esteve PO, Jacobsen SE, Pradhan S. 2009. UHRF1 binds G9a and participates in p21
2 transcriptional regulation in mammalian cells. *Nucleic acids research* **37**(2): 493-505.
- 3 Kotini AG, Mpakali A, Agalioti T. 2011. Dnmt3a1 upregulates transcription of distinct genes and
4 targets chromosomal gene clusters for epigenetic silencing in mouse embryonic stem cells. *Molecular
5 and cellular biology* **31**(7): 1577-1592.
- 6 Lehnertz B, Ueda Y, Derijck AA, Braunschweig U, Perez-Burgos L, Kubicek S, Chen T, Li E,
7 Jenuwein T, Peters AH. 2003. Suv39h-mediated histone H3 lysine 9 methylation directs DNA
8 methylation to major satellite repeats at pericentric heterochromatin. *Current biology : CB* **13**(14):
9 1192-1200.
- 10 Leung DC, Dong KB, Maksakova IA, Goyal P, Appanah R, Lee S, Tachibana M, Shinkai Y, Lehnertz
11 B, Mager DL et al. 2011. Lysine methyltransferase G9a is required for de novo DNA methylation and
12 the establishment, but not the maintenance, of proviral silencing. *Proceedings of the National
13 Academy of Sciences of the United States of America* **108**(14): 5718-5723.
- 14 Li H, Rauch T, Chen ZX, Szabo PE, Riggs AD, Pfeifer GP. 2006. The histone methyltransferase
15 SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters
16 silenced in cancer cells. *The Journal of biological chemistry* **281**(28): 19489-19500.
- 17 Li Q, Su Z, Xu X, Liu G, Song X, Wang R, Sui X, Liu T, Chang X, Huang D. 2012. AS1DHRS4, a
18 head-to-head natural antisense transcript, silences the DHRS4 gene cluster in cis and trans.
19 *Proceedings of the National Academy of Sciences of the United States of America* **109**(35): 14110-
20 14115.
- 21 Li S, Garrett-Bakelman FE, Akalin A, Zumbo P, Levine R, To BL, Lewis ID, Brown AL, D'Andrea
22 RJ, Melnick A et al. 2013. An optimized algorithm for detecting and annotating regional differential
23 methylation. *BMC bioinformatics* **14 Suppl 5**: S10.
- 24 Link PA, Gangisetty O, James SR, Woloszynska-Read A, Tachibana M, Shinkai Y, Karpf AR. 2009.
25 Distinct roles for histone methyltransferases G9a and GLP in cancer germ-line antigen gene regulation
26 in human cancer cells and murine embryonic stem cells. *Molecular cancer research : MCR* **7**(6): 851-
27 862.
- 28 Liu N, Zhang Z, Wu H, Jiang Y, Meng L, Xiong J, Zhao Z, Zhou X, Li J, Li H et al. 2015.
29 Recognition of H3K9 methylation by GLP is required for efficient establishment of H3K9
30 methylation, rapid target gene repression, and mouse viability. *Genes & development* **29**(4): 379-393.

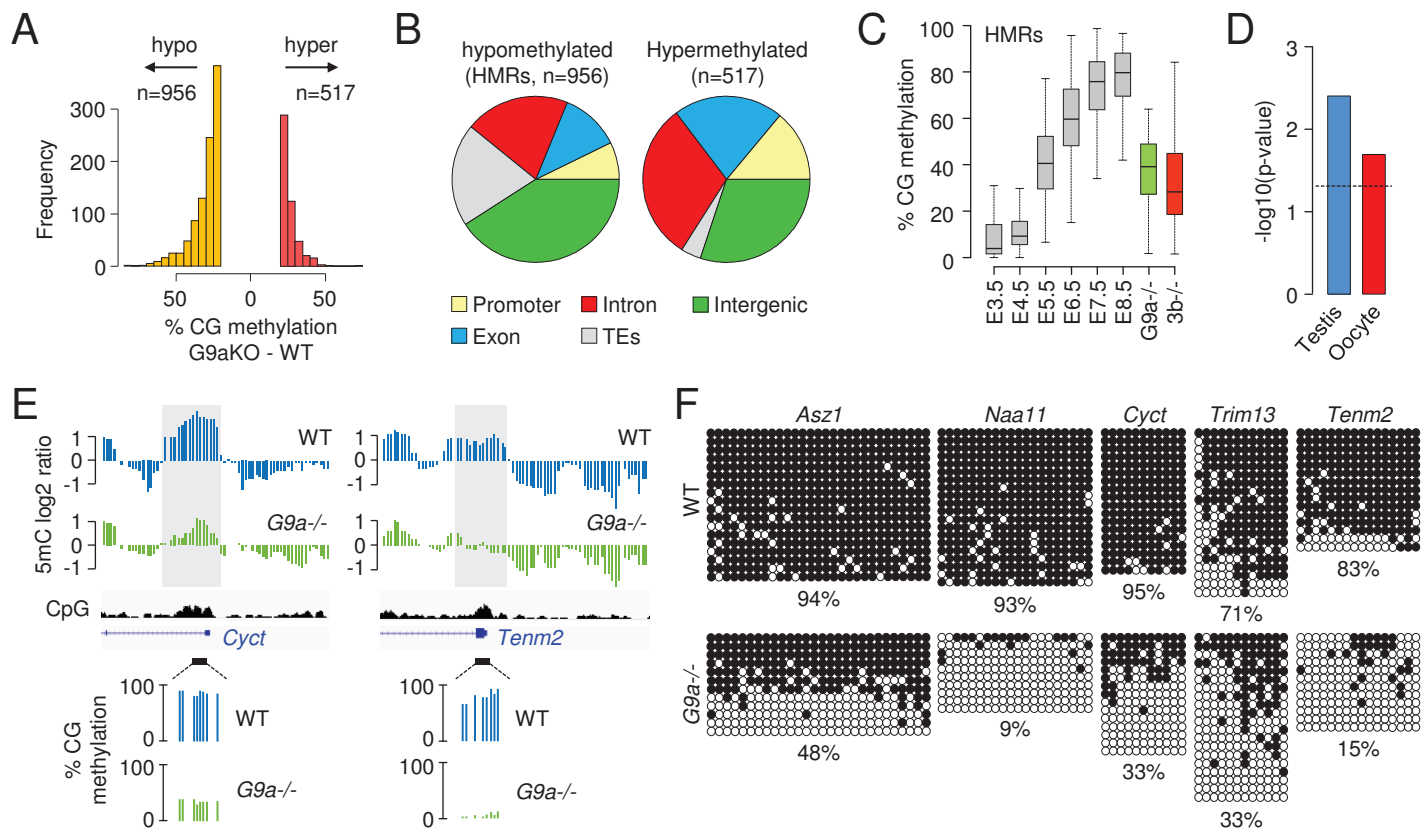
- 1 Liu X, Gao Q, Li P, Zhao Q, Zhang J, Li J, Koseki H, Wong J. 2013. UHRF1 targets DNMT1 for
2 DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9.
3 *Nature communications* **4**: 1563.
- 4 Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, Tachibana M, Lorincz MC,
5 Shinkai Y. 2010. Proviral silencing in embryonic stem cells requires the histone methyltransferase
6 ESET. *Nature* **464**(7290): 927-931.
- 7 Monnier P, Martinet C, Pontis J, Stancheva I, Ait-Si-Ali S, Dandolo L. 2013. H19 lncRNA controls
8 gene expression of the Imprinted Gene Network by recruiting MBD1. *Proceedings of the National*
9 *Academy of Sciences of the United States of America* **110**(51): 20693-20698.
- 10 Mozzetta C, Pontis J, Fritsch L, Robin P, Portoso M, Proux C, Margueron R, Ait-Si-Ali S. 2014. The
11 Histone H3 Lysine 9 Methyltransferases G9a and GLP Regulate Polycomb Repressive Complex 2-
12 Mediated Gene Silencing. *Molecular cell* **53**(2): 277-289.
- 13 Myant K, Termanis A, Sundaram AY, Boe T, Li C, Merusi C, Burrage J, de Las Heras JI, Stancheva I.
14 2011. LSH and G9a/GLP complex are required for developmentally programmed DNA methylation.
15 *Genome research* **21**(1): 83-94.
- 16 Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P. 2008. The Air
17 noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science*
18 **322**(5908): 1717-1720.
- 19 Ogawa H, Ishiguro K, Gaubatz S, Livingston DM, Nakatani Y. 2002. A complex with chromatin
20 modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science* **296**(5570): 1132-1136.
- 21 Ohhata T, Tachibana M, Tada M, Tada T, Sasaki H, Shinkai Y, Sado T. 2004. X-inactivation is stably
22 maintained in mouse embryos deficient for histone methyl transferase G9a. *Genesis* **40**(3): 151-156.
- 23 Pannetier M, Julien E, Schotta G, Tardat M, Sardet C, Jenuwein T, Feil R. 2008. PR-SET7 and SUV4-
24 20H regulate H4 lysine-20 methylation at imprinting control regions in the mouse. *EMBO reports*
25 **9**(10): 998-1005.
- 26 Proudhon C, Duffie R, Ajjan S, Cowley M, Iranzo J, Carbajosa G, Saadeh H, Holland ML, Oakey RJ,
27 Rakyen VK *et al.* 2012. Protection against de novo methylation is instrumental in maintaining parent-
28 of-origin methylation inherited from the gametes. *Molecular cell* **47**(6): 909-920.

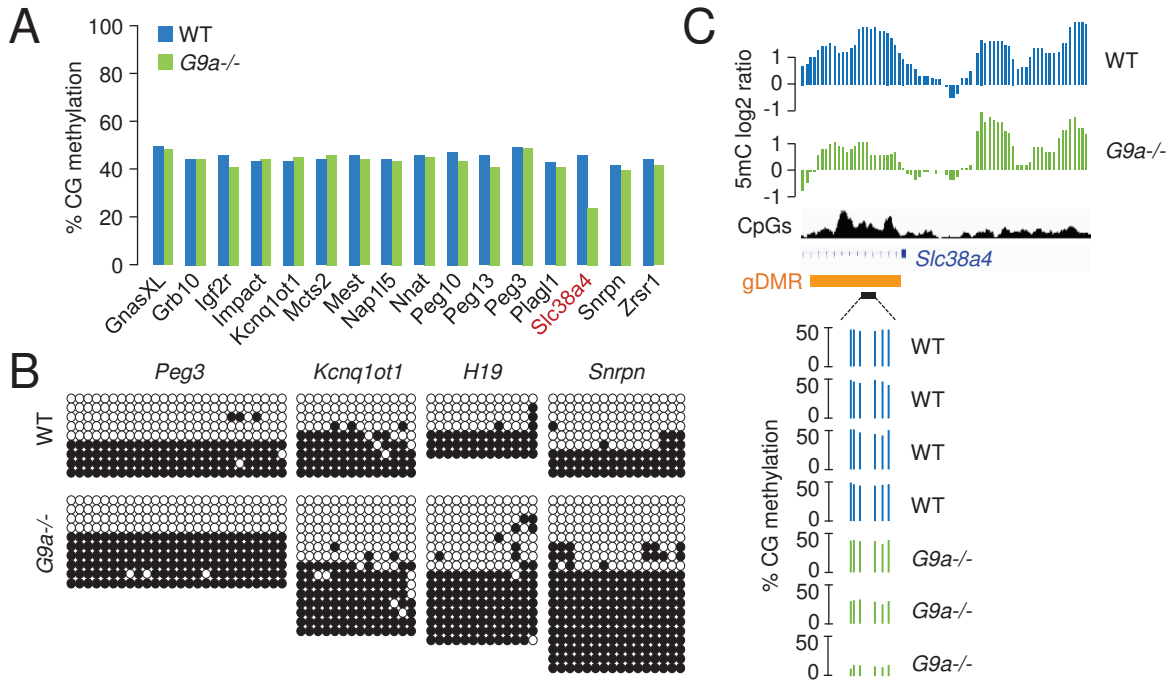
- 1 Rai K, Jafri IF, Chidester S, James SR, Karpf AR, Cairns BR, Jones DA. 2010. Dnmt3 and G9a
2 cooperate for tissue-specific development in zebrafish. *The Journal of biological chemistry* **285**(6):
3 4110-4121.
- 4 Rathert P, Dhayalan A, Murakami M, Zhang X, Tamas R, Jurkowska R, Komatsu Y, Shinkai Y,
5 Cheng X, Jeltsch A. 2008. Protein lysine methyltransferase G9a acts on non-histone targets. *Nature*
6 *chemical biology* **4**(6): 344-346.
- 7 Rose NR, Klose RJ. 2014. Understanding the relationship between DNA methylation and histone
8 lysine methylation. *Biochimica et biophysica acta* **1839**(12): 1362-1372.
- 9 Rothbart SB, Krajewski K, Nady N, Tempel W, Xue S, Badeaux AI, Barsyte-Lovejoy D, Martinez JY,
10 Bedford MT, Fuchs SM et al. 2012. Association of UHRF1 with methylated H3K9 directs the
11 maintenance of DNA methylation. *Nature structural & molecular biology* **19**(11): 1155-1160.
- 12 Saadeh H, Schulz R. 2014. Protection of CpG islands against de novo DNA methylation during
13 oogenesis is associated with the recognition site of E2f1 and E2f2. *Epigenetics & chromatin* **7**: 26.
- 14 Saze H, Tsugane K, Kanno T, Nishimura T. 2012. DNA methylation in plants: relationship to small
15 RNAs and histone modifications, and functions in transposon inactivation. *Plant & cell physiology*
16 **53**(5): 766-784.
- 17 Sharma S, Gerke DS, Han HF, Jeong S, Stallcup MR, Jones PA, Liang G. 2012. Lysine
18 methyltransferase G9a is not required for DNMT3A/3B anchoring to methylated nucleosomes and
19 maintenance of DNA methylation in somatic cells. *Epigenetics & chromatin* **5**(1): 3.
- 20 Smith ZD, Meissner A. 2013. DNA methylation: roles in mammalian development. *Nature reviews*
21 *Genetics* **14**(3): 204-220.
- 22 Strogantsev R, Krueger F, Yamazawa K, Shi H, Gould P, Goldman-Roberts M, McEwen K, Sun B,
23 Pedersen R, Ferguson-Smith AC. 2015. Allele-specific binding of ZFP57 in the epigenetic regulation
24 of imprinted and non-imprinted monoallelic expression. *Genome biology* **16**(1): 112.
- 25 Tachibana M, Matsumura Y, Fukuda M, Kimura H, Shinkai Y. 2008. G9a/GLP complexes
26 independently mediate H3K9 and DNA methylation to silence transcription. *The EMBO journal*
27 **27**(20): 2681-2690.
- 28 Tachibana M, Sugimoto K, Nozaki M, Ueda J, Ohta T, Ohki M, Fukuda M, Takeda N, Niida H, Kato
29 H et al. 2002. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9
30 methylation and is essential for early embryogenesis. *Genes & development* **16**(14): 1779-1791.

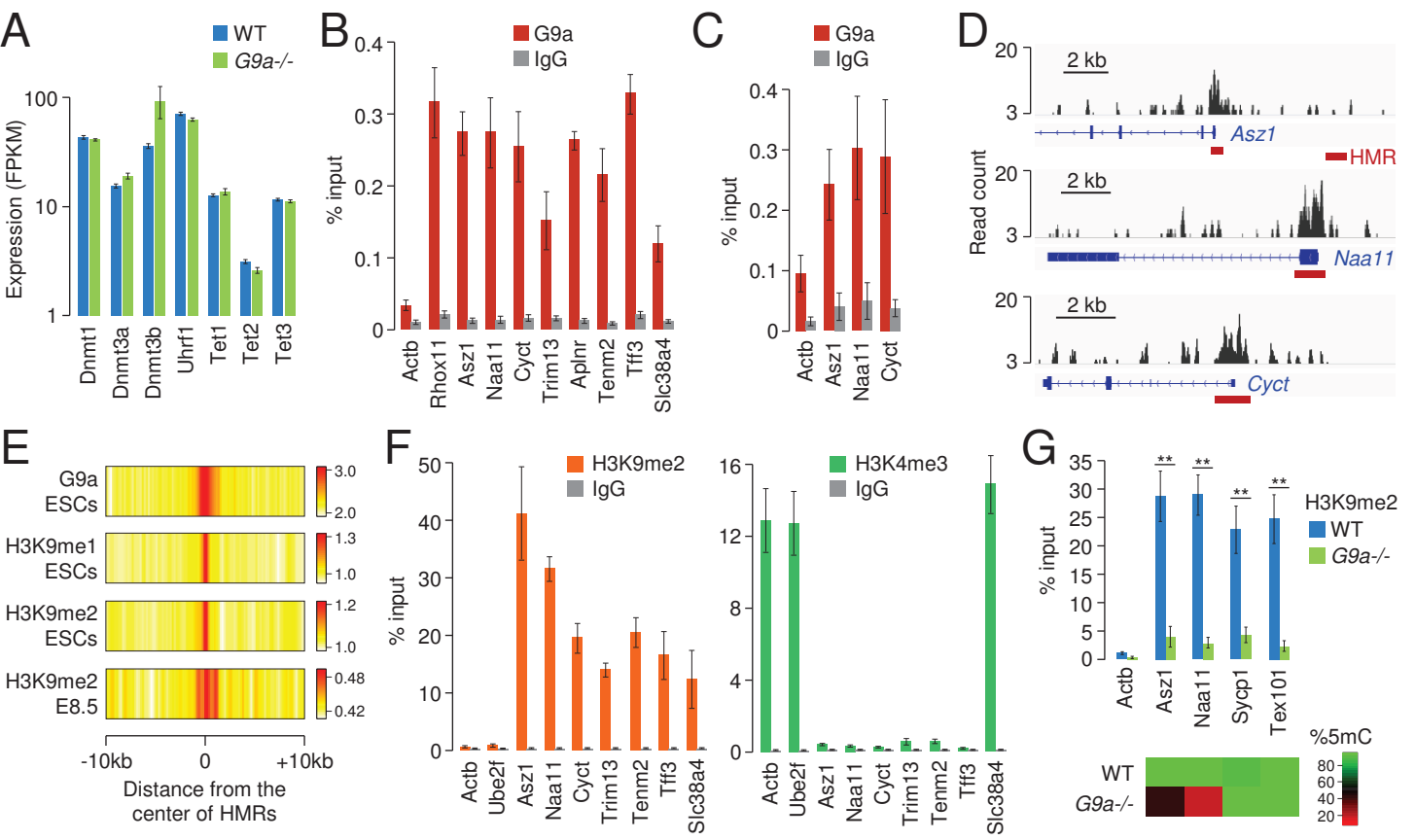
- 1 Tachibana M, Ueda J, Fukuda M, Takeda N, Ohta T, Iwanari H, Sakihama T, Kodama T, Hamakubo
2 T, Shinkai Y. 2005. Histone methyltransferases G9a and GLP form heteromeric complexes and are
3 both crucial for methylation of euchromatin at H3-K9. *Genes & development* **19**(7): 815-826.
- 4 Trojer P, Zhang J, Yonezawa M, Schmidt A, Zheng H, Jenuwein T, Reinberg D. 2009. Dynamic
5 Histone H1 Isotype 4 Methylation and Demethylation by Histone Lysine Methyltransferase
6 G9a/KMT1C and the Jumonji Domain-containing JMJD2/KDM4 Proteins. *The Journal of biological*
7 *chemistry* **284**(13): 8395-8405.
- 8 Velasco G, Hube F, Rollin J, Neuillet D, Philippe C, Bouzinba-Segard H, Galvani A, Viegas-
9 Pequignot E, Francastel C. 2010. Dnmt3b recruitment through E2F6 transcriptional repressor mediates
10 germ-line gene silencing in murine somatic tissues. *Proceedings of the National Academy of Sciences*
11 *of the United States of America* **107**(20): 9281-9286.
- 12 Voon HP, Hughes JR, Rode C, De La Rosa-Velazquez IA, Jenuwein T, Feil R, Higgs DR, Gibbons
13 RJ. 2015. ATRX Plays a Key Role in Maintaining Silencing at Interstitial Heterochromatic Loci and
14 Imprinted Genes. *Cell reports* **11**(3): 405-418.
- 15 Wagschal A, Sutherland HG, Woodfine K, Henckel A, Chebli K, Schulz R, Oakey RJ, Bickmore WA,
16 Feil R. 2008. G9a histone methyltransferase contributes to imprinting in the mouse placenta.
17 *Molecular and cellular biology* **28**(3): 1104-1113.
- 18 Weiss T, Hergeth S, Zeissler U, Izzo A, Tropberger P, Zee BM, Dundr M, Garcia BA, Daujat S,
19 Schneider R. 2010. Histone H1 variant-specific lysine methylation by G9a/KMT1C and
20 Glp1/KMT1D. *Epigenetics & chromatin* **3**(1): 7.
- 21 Xin Z, Tachibana M, Guggiari M, Heard E, Shinkai Y, Wagstaff J. 2003. Role of histone
22 methyltransferase G9a in CpG methylation of the Prader-Willi syndrome imprinting center. *The*
23 *Journal of biological chemistry* **278**(17): 14996-15000.
- 24 Yu Y, Song C, Zhang Q, DiMaggio PA, Garcia BA, York A, Carey MF, Grunstein M. 2012. Histone
25 H3 lysine 56 methylation regulates DNA replication through its interaction with PCNA. *Molecular*
26 *cell* **46**(1): 7-17.
- 27
- 28

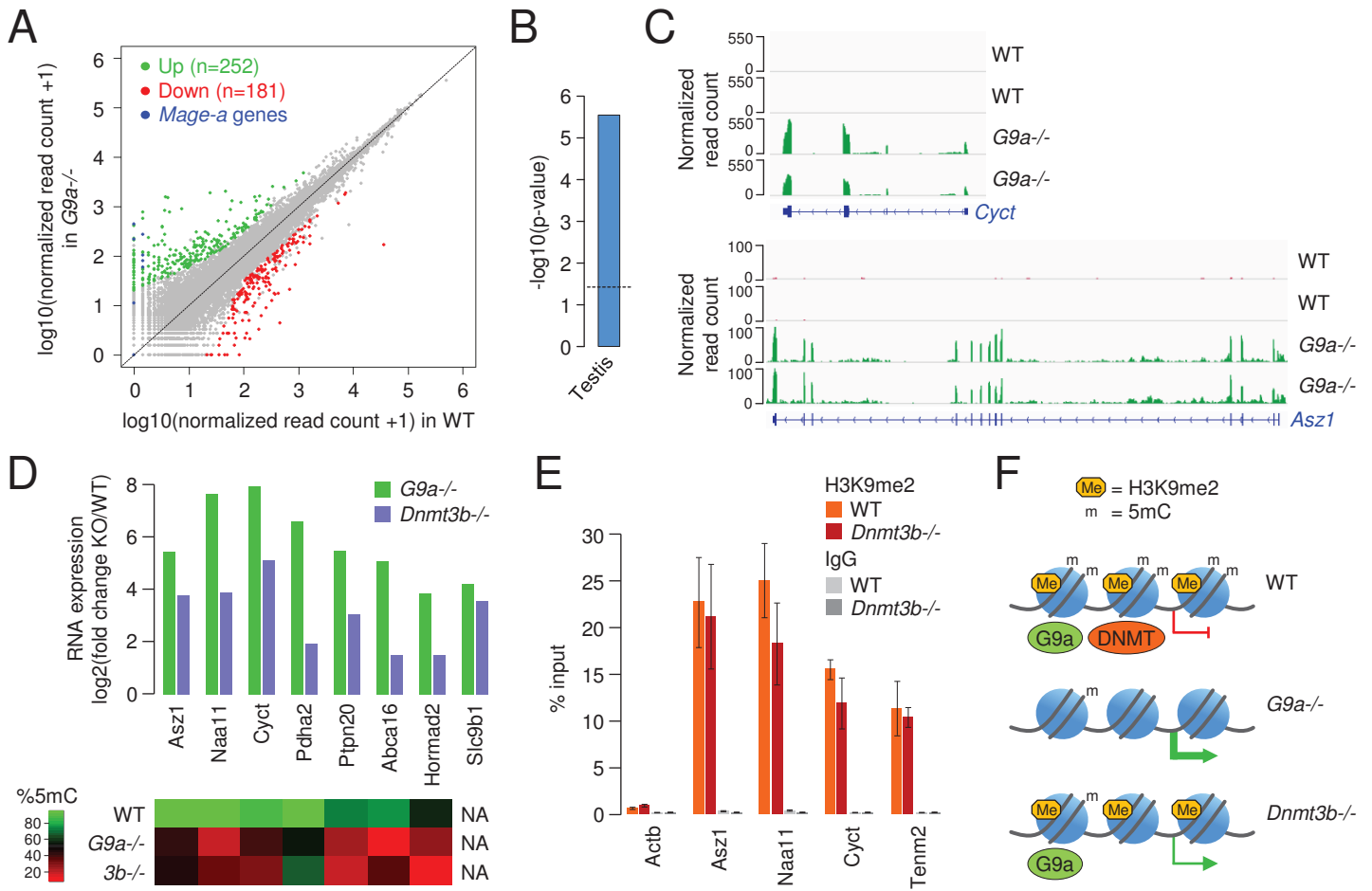












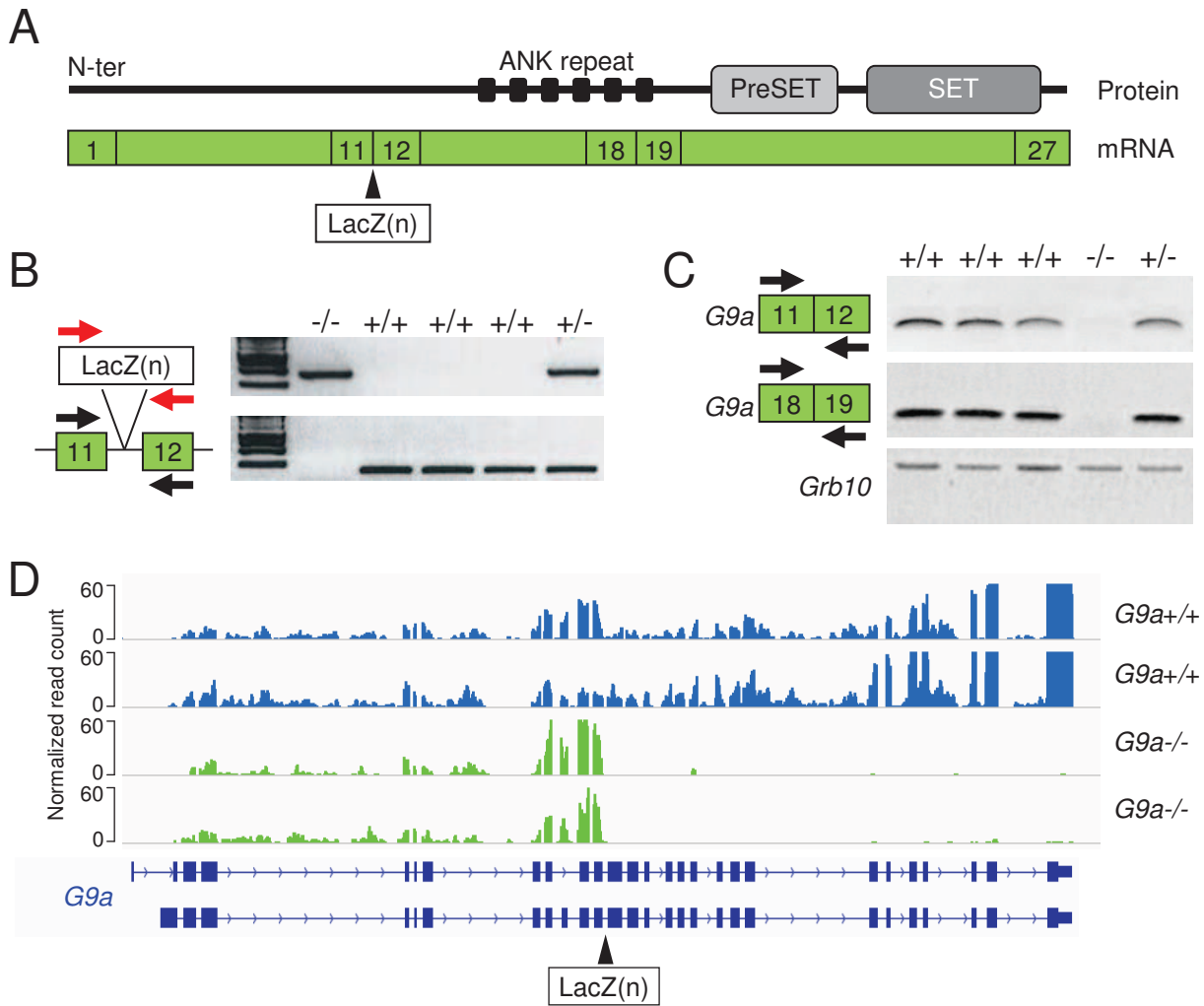


Figure S1. Characterization of G9a knockout embryos. **A.** Schematic representation of the mouse G9a protein (top) and transcript (bottom) with the site of gene trap insertion between the exons 11 and 12 (Wagschal et al., 2008). **B.** Examples of PCR genotyping using two primer pairs in the LacZ insertion cassette (red arrows) and in the exons 11-12 spanning the site of insertion (black arrows). **C.** Expression analysis by RT-PCR with primers spanning exons 11-12 and exons 18-19 (black arrows) indicates that the transcription of the *G9a* gene is abolished after the site of LacZ insertion in *G9a*^{-/-} animals. *Grb10* was amplified as a control. **D.** RNA-Seq in E8.5 embryos confirm that the transcription of the *G9a* gene is abolished after the site of LacZ insertion in knock-out embryos.

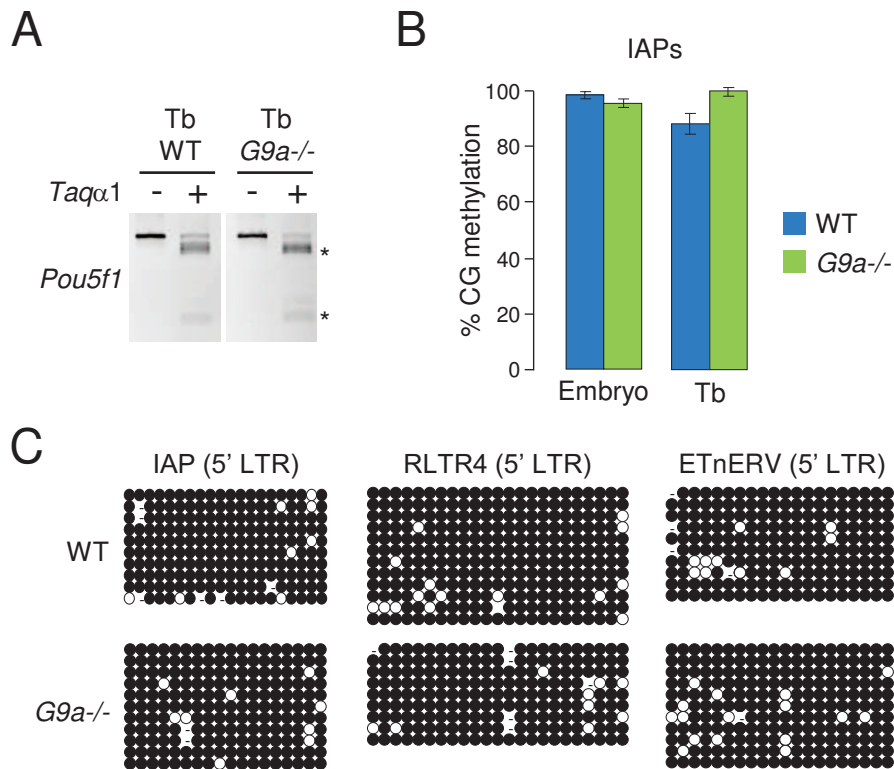
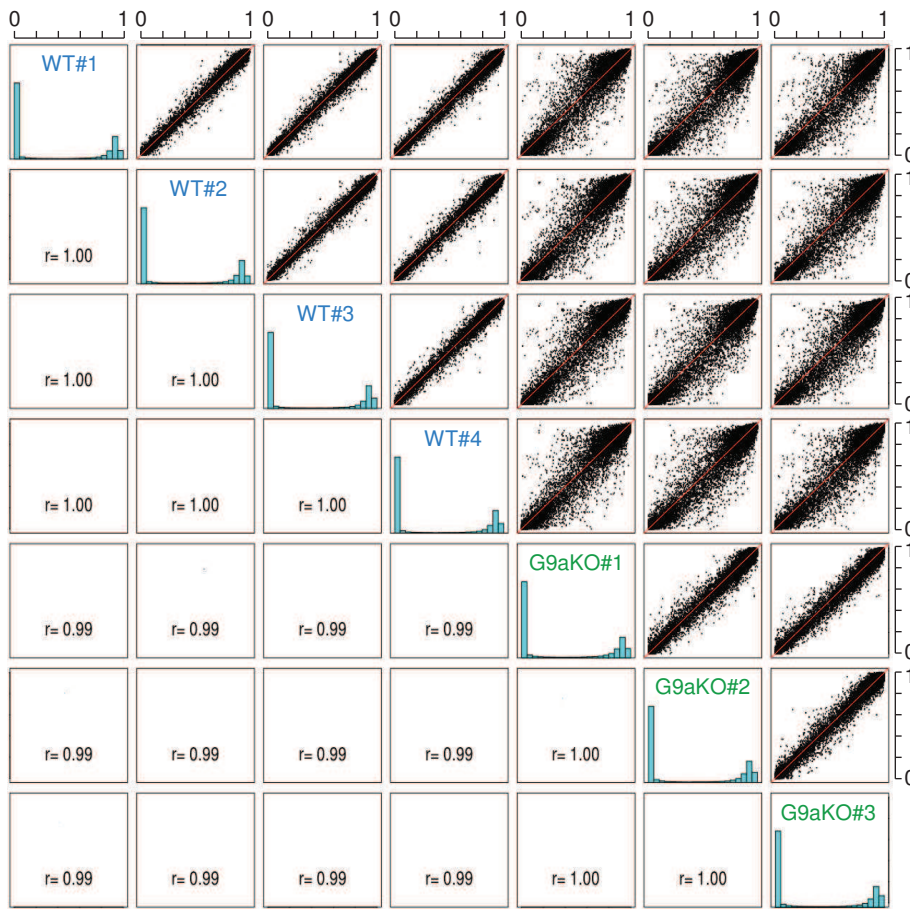


Figure S2. Methylation analysis in *G9a*^{-/-} embryos and extra-embryonic tissues. **A.** COBRA showing that the *Pou5f1* promoter is methylated at similar levels in trophoblast (Tb) cells from WT and *G9a*^{-/-} E9.5 conceptus. **B.** Quantification of DNA methylation in IAP retrotransposons by *Mcr*BC digestion followed by qPCR in WT and *G9a*^{-/-} E9.5 embryos and trophoblast (Tb) (mean \pm SD, n=3). **C.** Bisulfite cloning and sequencing at three retrotransposons in WT and *G9a*^{-/-} E9.5 embryos. Circles represent methylated (black) and unmethylated (white) CpGs, small lines indicate undefined methylation status, empty spaces indicate a deleted cytosine. Each horizontal row represents one sequenced clone.

A

Sample	Paired-end reads	Mapping efficiency (%)	Reads with unique hits (%)	Conversion rate (%)	Sequencing depth (median)	CpGs 8x
WT#1	32 547 365	93.98	75.24	99.67	62x	1 313 494
WT#2	47 789 978	93.21	74.44	99.76	91x	1 299 216
WT#3	39 181 062	94.31	75.92	99.78	79x	1 379 871
WT#4	43 058 637	88.20	70.08	99.58	67x	1 365 044
G9aKO#1	24 574 882	93.09	74.89	99.65	45x	1 336 445
G9aKO#2	30 063 968	91.56	74.35	99.81	57x	1 224 131
G9aKO#3	23 032 908	90.40	72.88	99.71	44x	1 243 754

B



C

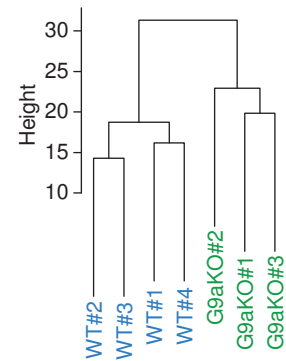


Figure S3. Features of RRBS methylomes in WT and *G9a*^{-/-} embryos. **A.** Table summarizing the number of sequenced reads, mapped reads, the bisulfite conversion rate and the number of CpGs sequenced at least 8x. **B.** Correlation matrix of RRBS methylation scores between WT and *G9a*^{-/-} replicate embryos. The graphs show the correlation of methylation scores in 50,000 random genomic tiles (400bp) and the histograms show the distribution of methylation in all genomic tiles (400bp). The Pearson's correlation coefficients (r) are indicated. **C.** Dendrogram representing the hierarchical clustering of the RRBS methylation scores measured in genomic tiles (400bp).

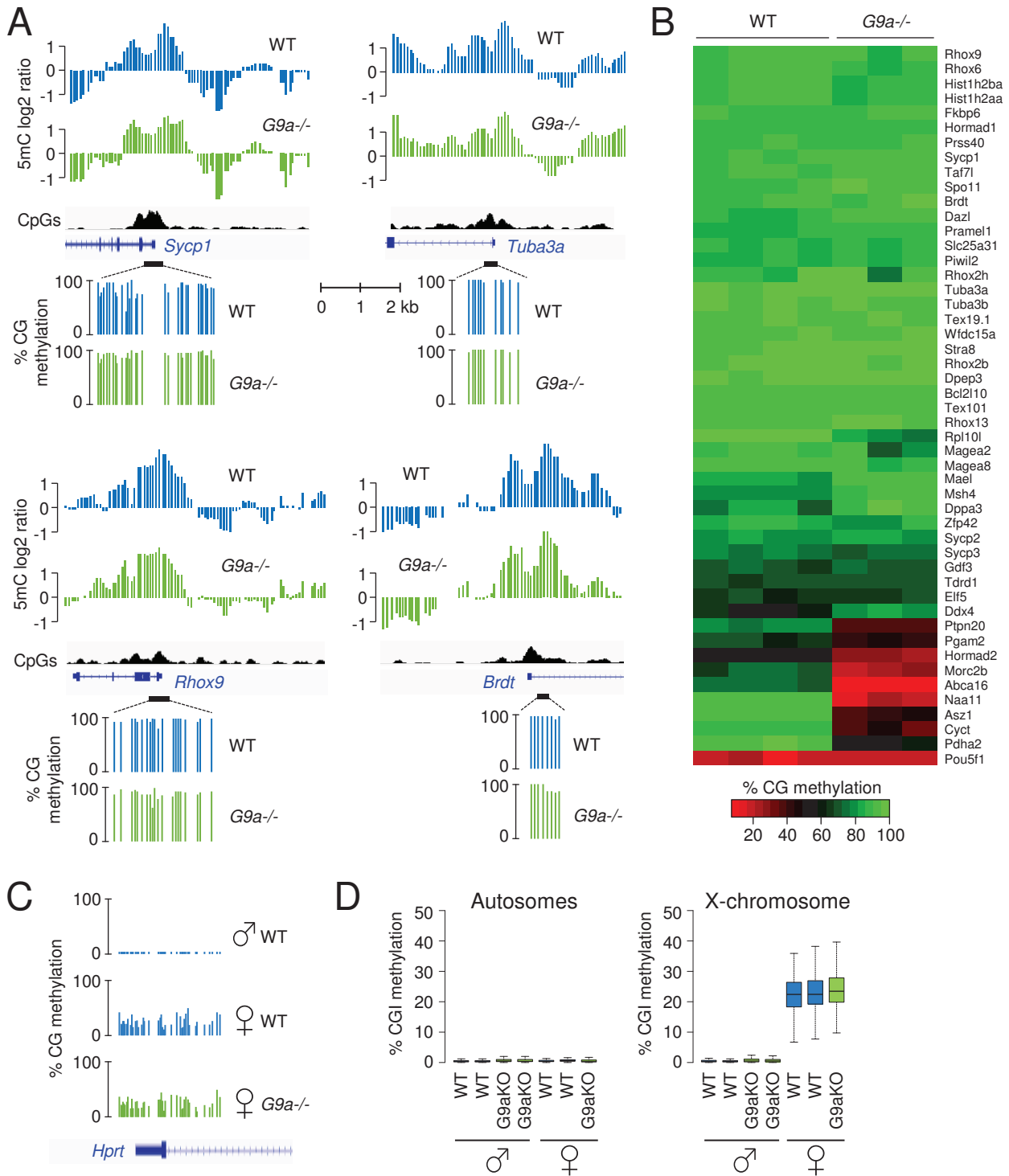


Figure S4. Gene promoter methylation on autosomes and the X-chromosome in *G9a*^{-/-} embryos. **A.** MeDIP and RRBS profiles in the promoters of germline and placenta-specific genes. **B.** Heatmap representation of the RRBS methylation scores in gene promoters (-1000 to +1000 bp from the TSS) of germline and pluripotency genes. **C.** RRBS methylation scores in the CpG island promoter of the X-linked *Hprt* gene. **D.** Box plots of the methylation of autosomal and X-linked CpG islands in male and female embryos, indicating that *G9a* is dispensable for CpG island methylation on the inactive X chromosome in females.

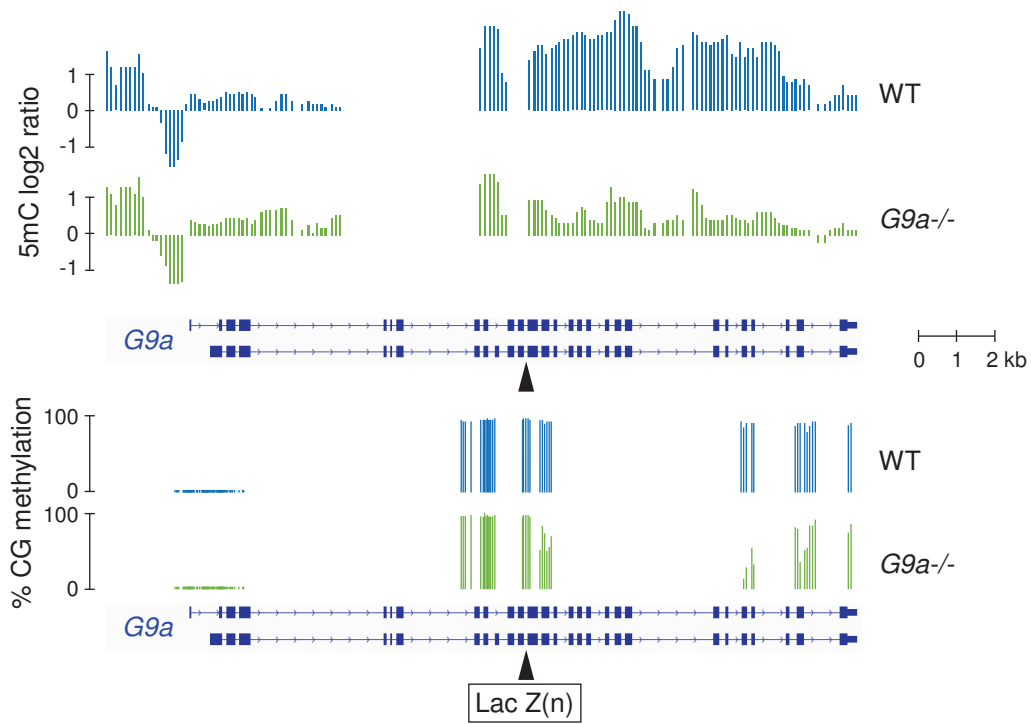


Figure S5. Hypomethylation of the *G9a* gene body in *G9a*^{-/-} embryos. The upper tracks depict smoothed MeDIP log₂ ratios of individual oligonucleotides in the *G9a* gene, and the graphs on the bottom depict methylation levels at individual CpGs covered by RRBS. The hypomethylation is detected by MeDIP and RRBS after the site of LacZ insertion (black arrow), beyond which transcription is aborted.

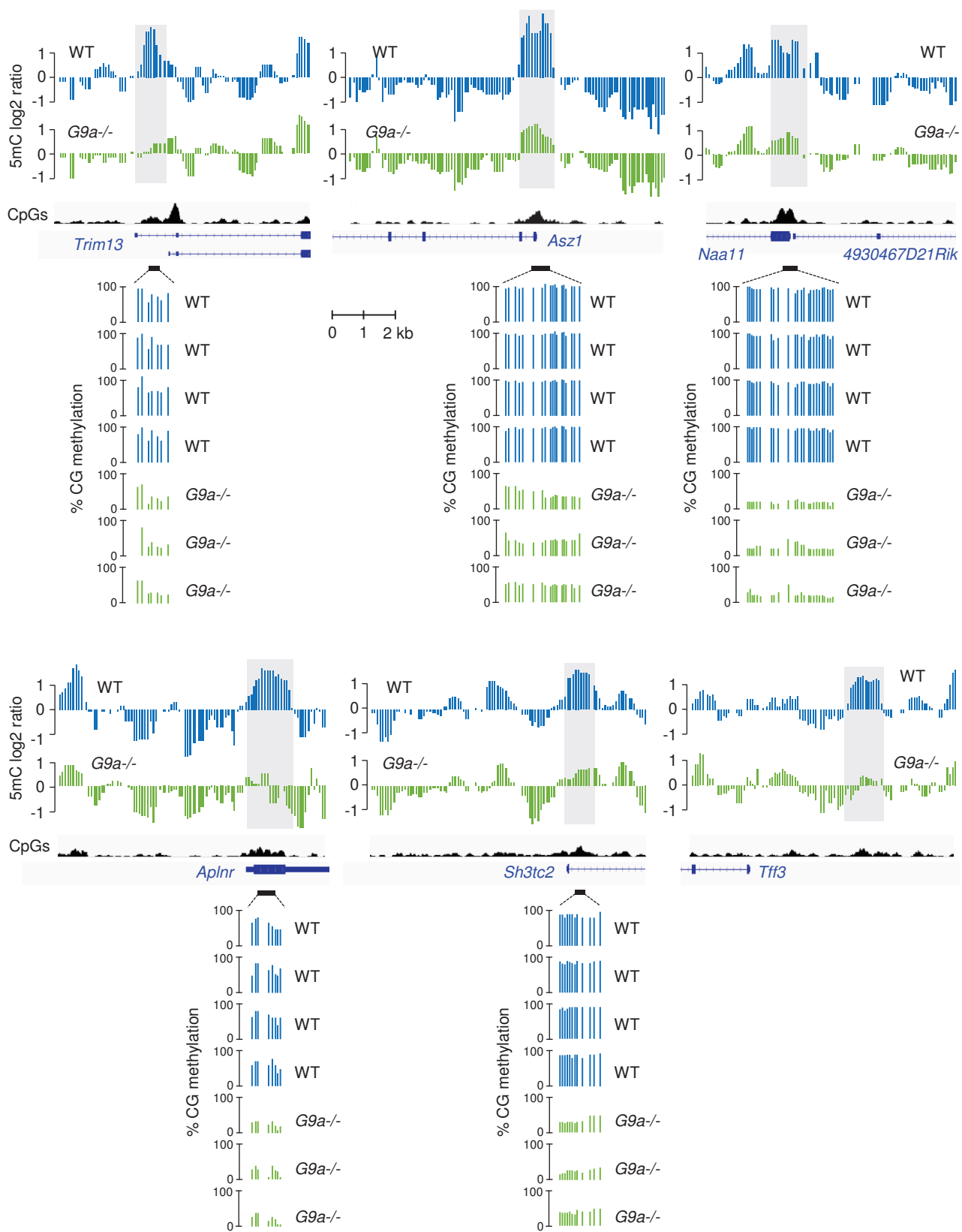


Figure S7. Examples of hypomethylated regions (HMRs) in *G9a*^{-/-} embryos. The upper tracks depict smoothed MeDIP log₂ ratios of individual oligonucleotides, and the graphs on the bottom depict RRBS methylation scores of individual CpGs in four WT and three *G9a*^{-/-} embryos. In each case, the difference measured by RRBS is confirmed by MeDIP (grey boxes).

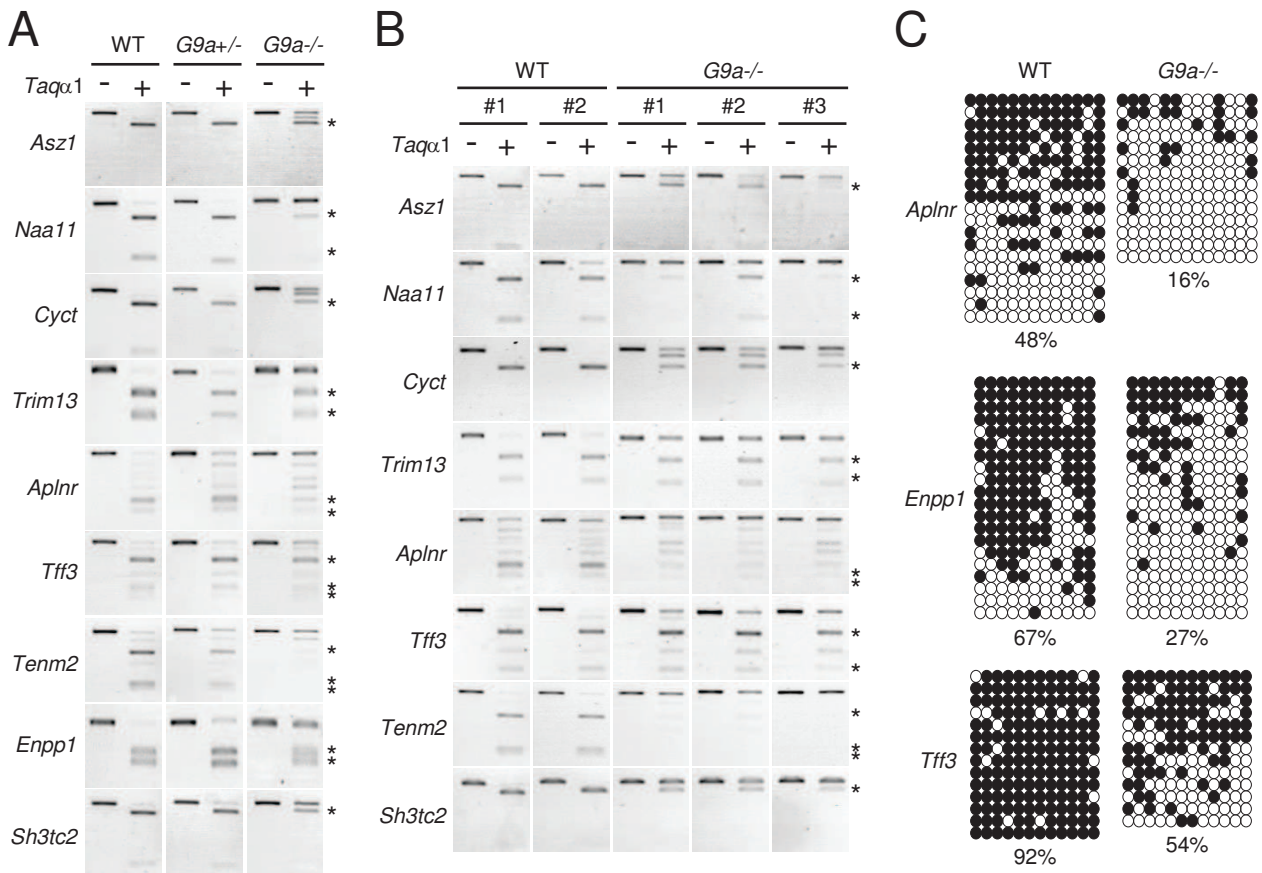


Figure S8. Validations at sequences hypomethylated in *G9a-/-* embryos. **A.** COBRA validations at candidate genes in wild type (WT), heterozygous (*G9a+/-*) and homozygous mutant (*G9a-/-*) littermate embryos collected at E9.5. The restriction fragments marked with asterisks are the end products of the digestion and indicate methylation. **B.** COBRA validations at eight target genes in two WT and three *G9a-/-* embryos collected at E8.5. **C.** Bisulfite sequencing at three targets in WT and *G9a-/-* E9.5 embryos. The percentages of CpG methylation per amplicon are indicated.

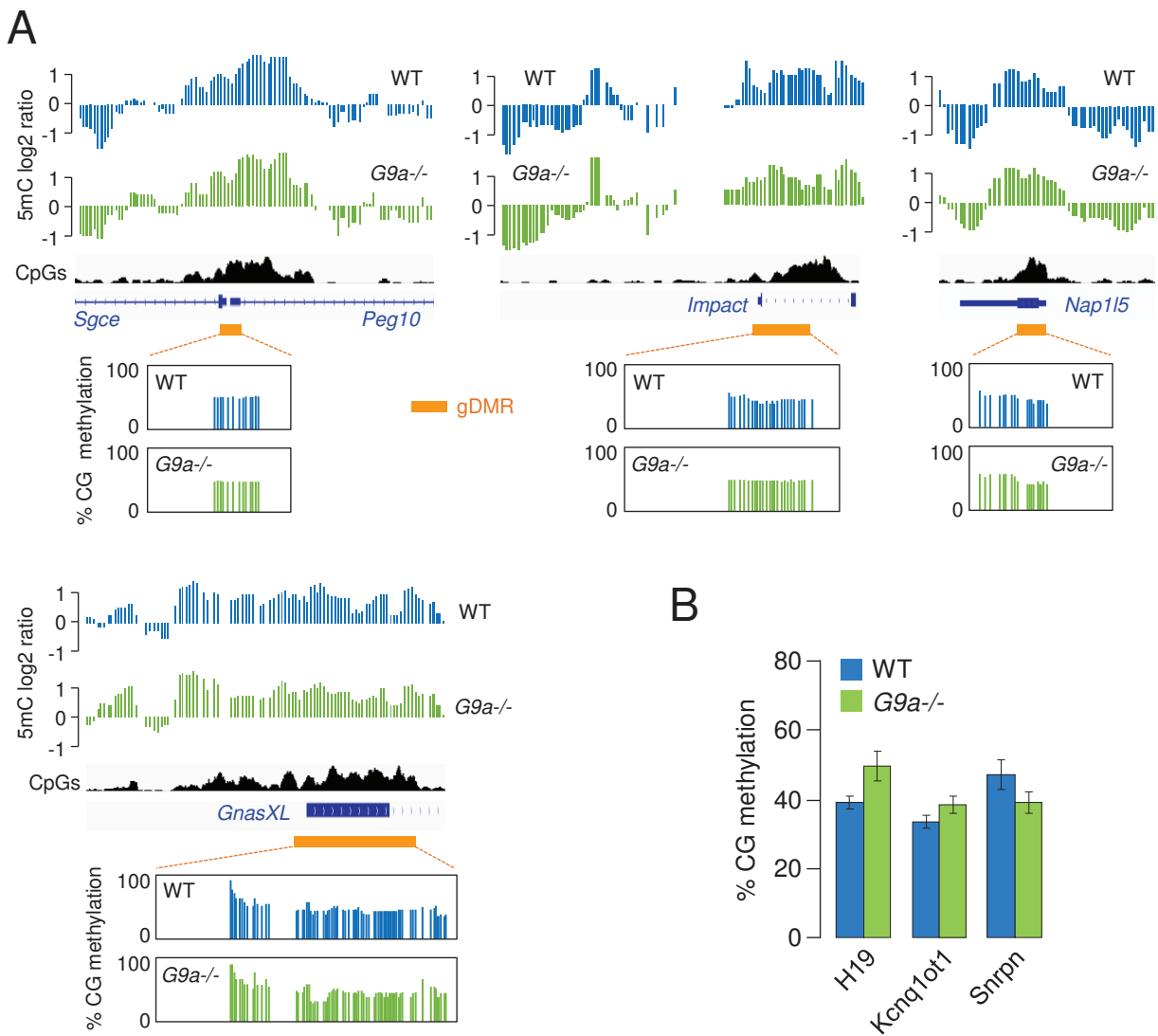


Figure S9. DNA methylation of imprinted germline DMRs (gDMRs) in *G9a*^{-/-} embryos. **A.** Examples of DNA methylation profiles by MeDIP and RRBS at gDMRs. The orange bars mark the position of the gDMRs. **B.** Methylation analysis of three gDMRs by *Mcr*BC digestion followed by qPCR in E9.5 embryos (mean \pm SD, n=3). Both paternal (*H19*) and maternal (*Kcnq1ot1*, *Snrpn*) gDMRs exhibit comparable methylation levels close to 50% in WT and *G9a*^{-/-} embryos.

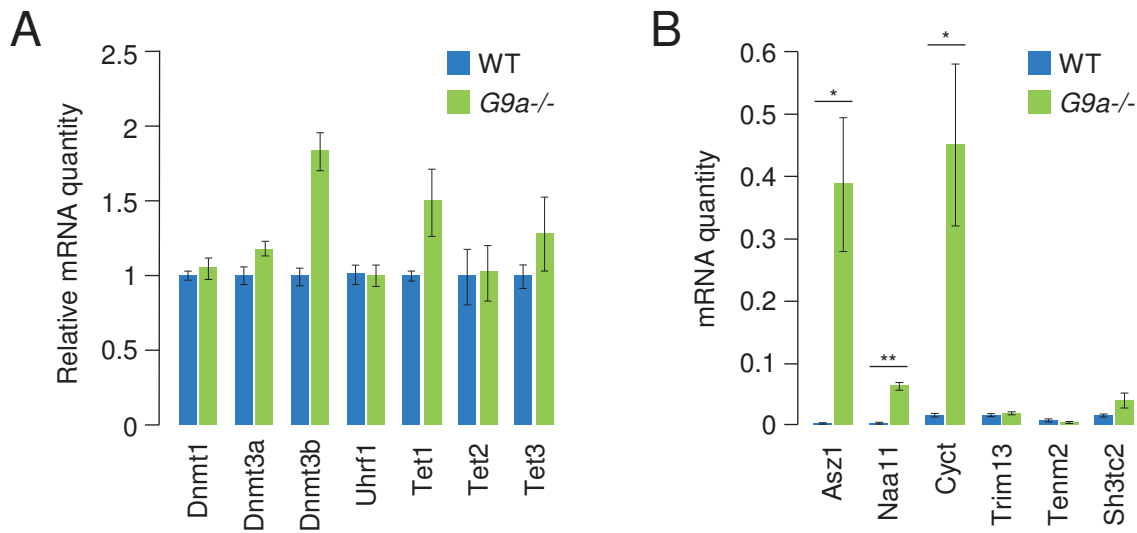


Figure S10. Quantification of gene expression by RT-qPCR in *G9a*^{-/-} embryos. **A.** Expression of genes encoding components of the DNA methylation machinery in WT and *G9a*^{-/-} E9.5 embryos. Expression was normalized to the expression of two housekeeping genes (*Actb* and *Rpl13a*) and then the values in WT embryos were set at 1 (mean \pm SEM, n=2 embryos). **B.** Expression of genes with a promoter-proximal HMR in WT and *G9a*^{-/-} E9.5 embryos, normalized to the expression of two housekeeping genes (*Actb* and *Rpl13a*) (mean \pm SEM, n=3 embryos). * = p<0.05, ** = p<0.01 (t-test).

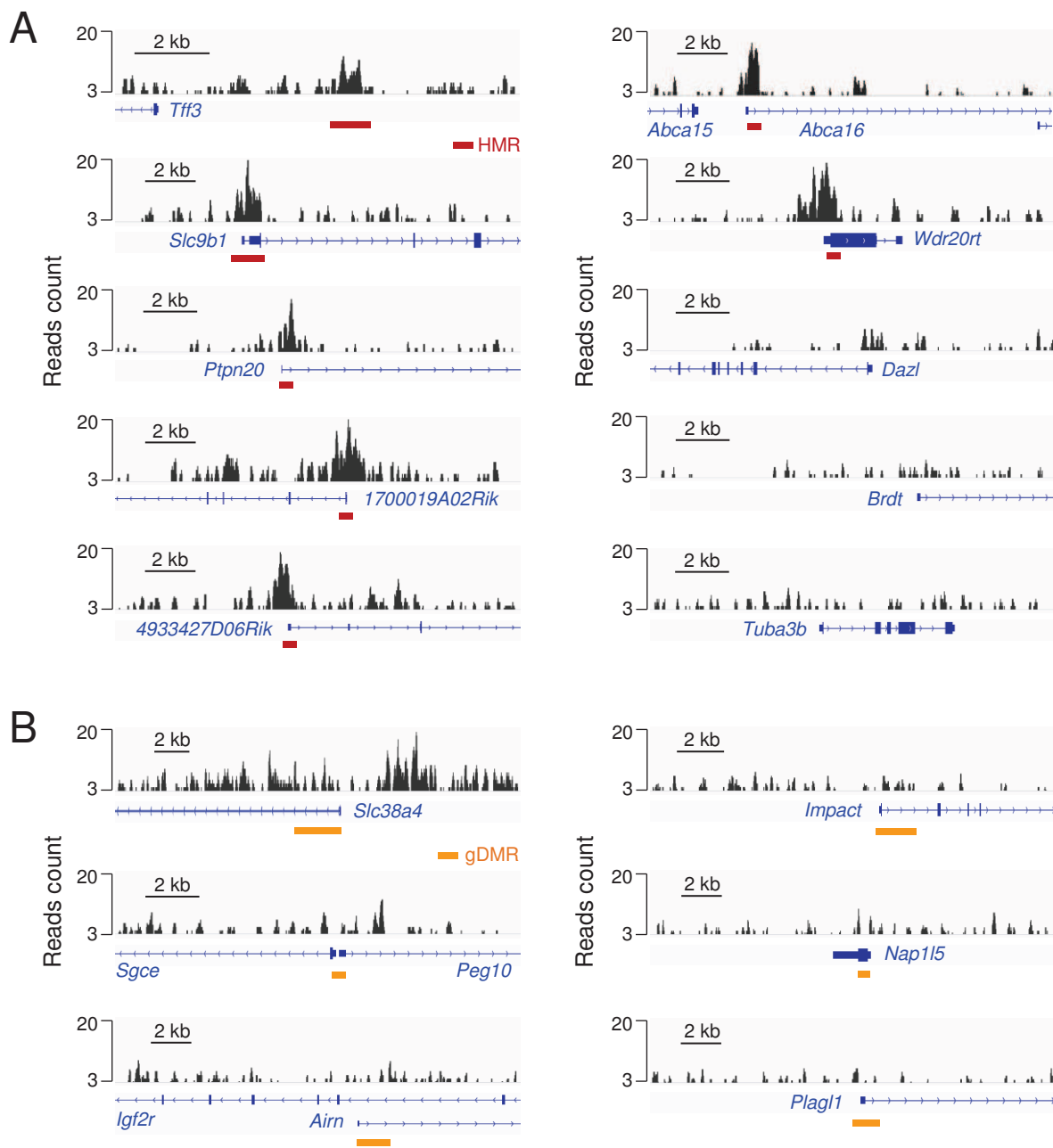


Figure S11. G9a ChIP-Seq profiles in mESCs at promoter-proximal HMRs and imprinted gDMRs. A. Browser views of G9a ChIP-Seq profiles in WT TT2 m ESCs (Mozzetta et al., 2013) at several germline genes with reduced promoter DNA methylation in *G9a*^{-/-} embryos. In all cases, a peak of G9a binding co-localizes with the HMR (red bars). In comparison, germline gene promoters with unchanged DNA methylation in *G9a*^{-/-} embryos (*Dazl*, *Brdt*, *Tuba3b*) are not enriched for G9a. **B.** G9a ChIP-Seq profiles in WT TT2 ESCs at imprinted gDMRs (orange bars). Enriched G9a binding is detected in the vicinity of the *Slc38a4* gDMR but not at other gDMRs.

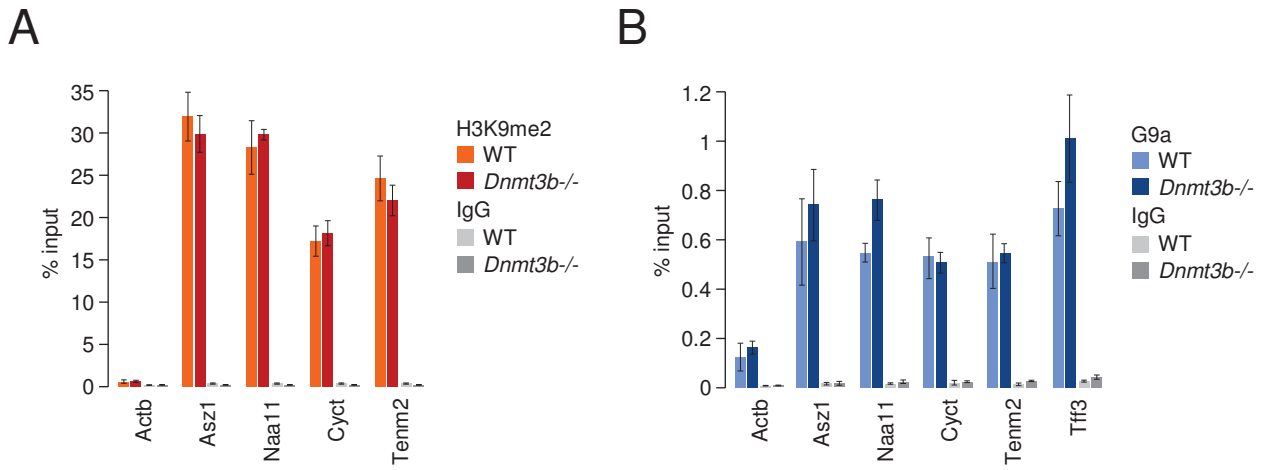


Figure S12. ChIP-qPCR analysis of H3K9me2 and G9a in WT and *Dnmt3b*^{-/-} MEFs. A. ChIP-qPCR analysis of H3K9me2 at HMRs in primary MEFs derived from WT and *Dnmt3b*^{-/-} embryos, represented as the percentage of input (mean \pm SEM, n=4 for WT, n=3 for *Dnmt3b*^{-/-}). The promoter of the housekeeping genes *Actb* served as a negative control. **B.** ChIP-qPCR analysis of G9a binding at HMRs in primary MEFs derived from WT and *Dnmt3b*^{-/-} embryos, represented as the percentage of input (mean \pm SEM, n=2 for WT, n=4 for *Dnmt3b*^{-/-}).

CONCLUSIONS

Les conclusions principales de ce travail sont :

- *G9a* est exprimé au moment de l'implantation de l'embryon et au cours de la mise en place de la méthylation de l'ADN entre 4.5 et 8.5 jpc.
- *G9a* a un impact global sur le méthylome des cellules ES en culture, mais est dispensable pour la méthylation globale du génome durant l'embryogenèse *in vivo*.
- *G9a* joue un rôle indispensable pour le recrutement de la méthylation de l'ADN à des sites spécifiques, incluant des promoteurs à îlots CpG de gènes spécifiques de la lignée germinale.
- *G9a* joue un rôle dans la maintenance de la méthylation de la gDMR du locus soumis à empreinte *Slc38a4 in vivo*.
- *G9a* cible les régions qui se déméthylent dans les mutants *G9a*^{-/-}, ce qui suggère une action directe sur la méthylation de l'ADN *in cis*.
- La marque répressive H3K9m2 couvre une très grande partie du génome *in vivo* et sa perte dans les embryons *G9a*^{-/-} ne corrèle pas avec les sites de perte de méthylation de l'ADN. Cela suggère que *G9a* influence la méthylation de l'ADN indépendamment de son activité de méthylase d'histones.
- *G9a* a besoin de la méthylation de l'ADN pour réprimer ses gènes cibles dans l'embryon.

RESULTATS PROJET 3 :

Identification du rôle d'E2F6 dans le recrutement de la méthylation de l'ADN durant le développement embryonnaire murin.

INTRODUCTION

Nos résultats obtenus dans les souris *G9a*^{-/-} nous ont permis d'éclaircir partiellement les mécanismes de recrutement de la méthylation de l'ADN vers certains gènes méiotiques. Cependant, les modes de ciblage de la machinerie de méthylation spécifiquement sur la famille des gènes méiotiques sont encore loin d'être élucidés.

En parallèle de *G9a*, nous nous sommes intéressés à la protéine E2F6. Comme décrit dans mon introduction, E2F6 est un facteur de transcription qui interagit avec DNMT3B et les souris *E2F6*^{-/-} ont le même phénotype que les souris porteuses d'une mutation hypomorphe de *Dnmt3b*, c'est à dire des transformations homéotiques (Courel et al., 2008 ; Velasco et al., 2010). Quelques études suggèrent qu'E2F6 aurait un rôle dans le recrutement de la méthylation de l'ADN au niveau de promoteurs de la lignée germinale (Kehoe et al., 2008; Storre et al., 2005; Velasco et al., 2010; Laseva et al., 2013 ; Storre et al., 2005). En effet, le site d'interaction d'E2F6 à l'ADN est enrichi à ces promoteurs, et les cellules (MEFs et ES) déficientes pour E2F6 montrent une réactivation de quatre gènes méiotiques, *Stag3*, *Smc1β*, *Tuba3a* et *Slc25a31*.

E2F6 semble donc être un candidat pour recruter la méthylation de l'ADN au niveau de gènes de la lignée germinale pendant le développement embryonnaire. De plus, il a été montré qu'E2F6 fait partie d'un complexe répresseur contenant *G9a*, ce qui suggère une interaction possible avec *G9a* dans le contrôle de la méthylation de l'ADN (Ogawa et al., 2002; Mozzetta et al., 2014).

L'objectif de ce projet était d'évaluer la contribution du facteur E2F6 à la mise en place de la méthylation de l'ADN au cours de l'embryogenèse de la souris. En

utilisant des approches de cartographie du méthylome par RRBS, nous avons identifié des régions différentiellement méthylées dans les embryons *E2f6*^{-/-} comparés aux embryons sauvages à 8.5 jpc, et étudié si l'absence d'E2F6 entraîne une réactivation des gènes différentiellement méthylés.

Je présente ici les résultats préliminaires de ce projet que j'ai pu obtenir avec l'aide d'un étudiant en Master 2 Thomas Dahlet.

TRAVAUX REALISES

Les travaux réalisés pour ce projet sont les suivants :

1-Mesurer l'expression du gène *E2f6* par RT-qPCR aux stades de développement embryonnaire 4.5 jpc jusqu'à 8.5 jpc.

2-Disséquer des embryons et préparer des échantillons d'ADN et d'ARN d'embryons murins sauvages et *E2f6*^{-/-} au stade 8.5 jpc.

3-Préparer des librairies de séquençage RRBS sur l'ADN des embryons *E2f6*^{-/-} et WT obtenus à 8.5 jpc.

4-Réaliser des expériences de RNA-seq sur l'ARN des embryons *E2f6*^{-/-} et WT obtenus à 8.5 jpc.

5-Valider les régions hypométhylées dans les embryons *G9a*^{-/-} par traitement au bisulfite de sodium suivi de digestions enzymatiques (COBRA).

6-Participer aux analyses bio-informatiques des résultats. Ce travail a majoritairement été réalisé par Sylvain Guibert, Michaël Dumas et Michaël Weber.

RESULTATS

I) Expression d'E2F6 dans l'embryogénèse et rôle dans la méthylation de gènes candidats

Tout comme pour *G9a*, nous nous sommes d'abord intéressés à l'expression d'*E2f6* au cours du développement embryonnaire, afin de savoir s'il pouvait avoir un rôle au moment de la phase de méthylation du génome. Pour cela, nous avons extrait l'ARN d'embryons entre et 4.5 et 8.5 jpc puis nous avons analysé l'expression d'*E2f6* par RT-qPCR (**Figure 1**). Ces résultats montrent que l'expression d'*E2f6* augmente au cours du temps, pour atteindre un niveau cinq fois supérieur à 8.5 jpc comparé à 4.5 jpc. Cette augmentation suit la mise en place globale de la méthylation de l'ADN entre 4.5 et 8.5 jpc aux îlots CpG (**Article 1**). Ce profil d'expression est compatible avec un rôle d'E2F6 dans le recrutement de la méthylation de l'ADN au cours de l'embryogénèse.

Afin d'étudier le rôle d'E2F6 dans la méthylation de l'ADN, nous avons utilisé une lignée de souris *E2f6*^{-/-}. Cette lignée nous a été fournie par le laboratoire de Jacqueline Lees du Massachusetts Institute of Technology à Cambridge, USA. Dans ces mutants, un codon stop a été inséré au niveau de l'exon 4, ce qui a pour conséquence de tronquer la protéine de ses domaines d'interaction à l'ADN et de dimérisation (**Figure 2**). Les souris mutantes *E2f6*^{-/-} sont viables et fertiles mais présentent un phénotype léger de transformation homéotique. Pour nos expériences, seules les souris *E2f6*^{-/-} provenant de parents hétérozygotes *E2f6*^{+/-} ont été utilisées.

Tout d'abord, nous avons analysé la méthylation de l'ADN à des gènes candidats par COBRA dans des embryons à 8.5 jpc. Nous nous sommes intéressés aux promoteurs des gènes méiotiques *Slc25a31*, *Tuba3a* et *Smc1β* qui ont été décrits précédemment comme étant déméthylés dans des cellules MEFs et des cellules ES *E2F6*^{-/-} (Kehoe et al., 2008; Storre et al., 2005; Velasco et al., 2010; Laseva et al., 2013 ; Storre et al., 2005). Les résultats montrent une hypométhylation de ces promoteurs dans les embryons *E2f6*^{-/-} au stade 8.5 jpc (**Figure 3**). Ces résultats valident les observations faites précédemment dans des modèles cellulaires

in vitro et suggèrent un rôle important d'E2F6 dans le recrutement de la méthylation de l'ADN à des promoteurs à îlots CpG au cours de l'embryogenèse.

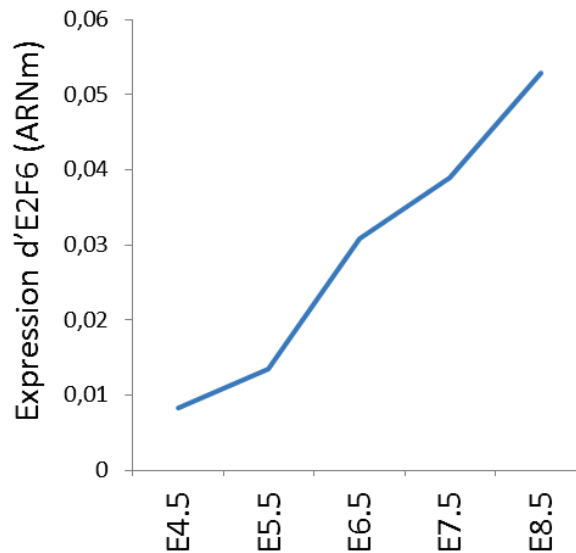


Figure 1 : Expression du gène *E2f6* mesurée par RT-qPCR dans les embryons.

L'expression a été mesurée par RT-qPCR sur un ensemble de 5 embryons aux stades 7.5 jpc et 8.5 jpc. A 4.5 jpc, 5.5 jpc et 6.5 jpc, 10 embryons ont été utilisés. L'expression est normalisée à celles de deux gènes de ménage (*actine β*, et *rpl13a*). Les oligonucléotides utilisés ciblent le dernier exon du gène pour amplifier toutes les isoformes possibles.

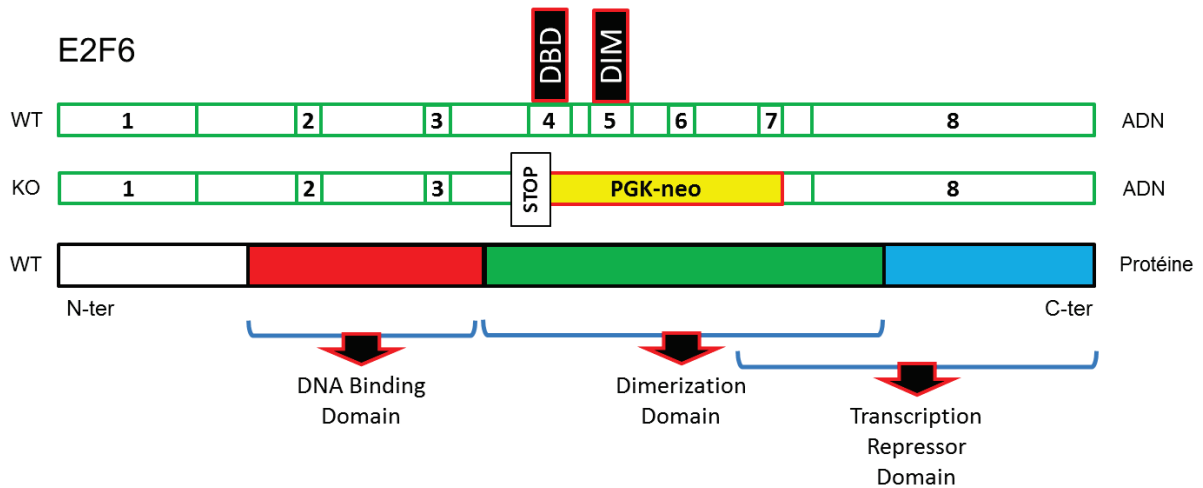


Figure 2 : Représentation schématique de la protéine E2F6 et de l'allèle mutant du gène *E2f6*.

Une cassette PGK-neo, codant pour une résistance à la néomycine, et un codon STOP sont insérés à la place des exons 4 à 6 du gène *E2f6*. Grâce à cette insertion, la protéine E2F6 est tronquée de ses domaines d'interaction à l'ADN, de dimérisation et de répression de la transcription. DBD : DNA Binding Domain. DIM : Dimerization Domain.

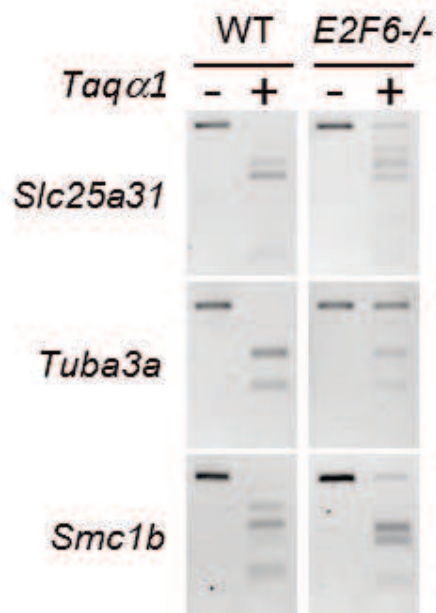


Figure 3 : Validations par COBRA de promoteurs de gènes hypométhylés dans les embryons *E2f6*^{-/-} à 8.5 jpc.

II) Le méthylome des embryons *E2F6*^{-/-} révèle une conservation globale de la méthylation de l'ADN et quelques régions hypométhylées

Afin de trouver d'autres gènes cibles d'E2F6 *in vivo* pour le recrutement de la méthylation de l'ADN, nous avons généré des cartographies du méthylome par RRBS dans des embryons sauvages et mutants *E2f6*^{-/-} (**Figure 4.A**). Les données de RRBS n'ont révélé aucun signe d'hypométhylation globale au niveau du génome (**Figure 4.A-C**) et des éléments transposables (**Figure 4.D**) dans les embryons déficients en *E2f6*, indiquant qu'E2F6 n'a pas un impact global sur la méthylation de l'ADN. Cependant, dans ces embryons mutants, on trouve tout de même 118 régions perdants plus de 20% de méthylation et 56 régions gagnants plus de 20% de méthylation (**Figure 5.A**). Les régions hypométhylées se trouvent à la fois dans des promoteurs, des exons, introns, régions inter-géniques et éléments transposables (**Figure 5.B**).

Nous nous sommes plus particulièrement intéressés aux régions promotrices perdant au moins 20% de la méthylation d'ADN (**Figure 6**).

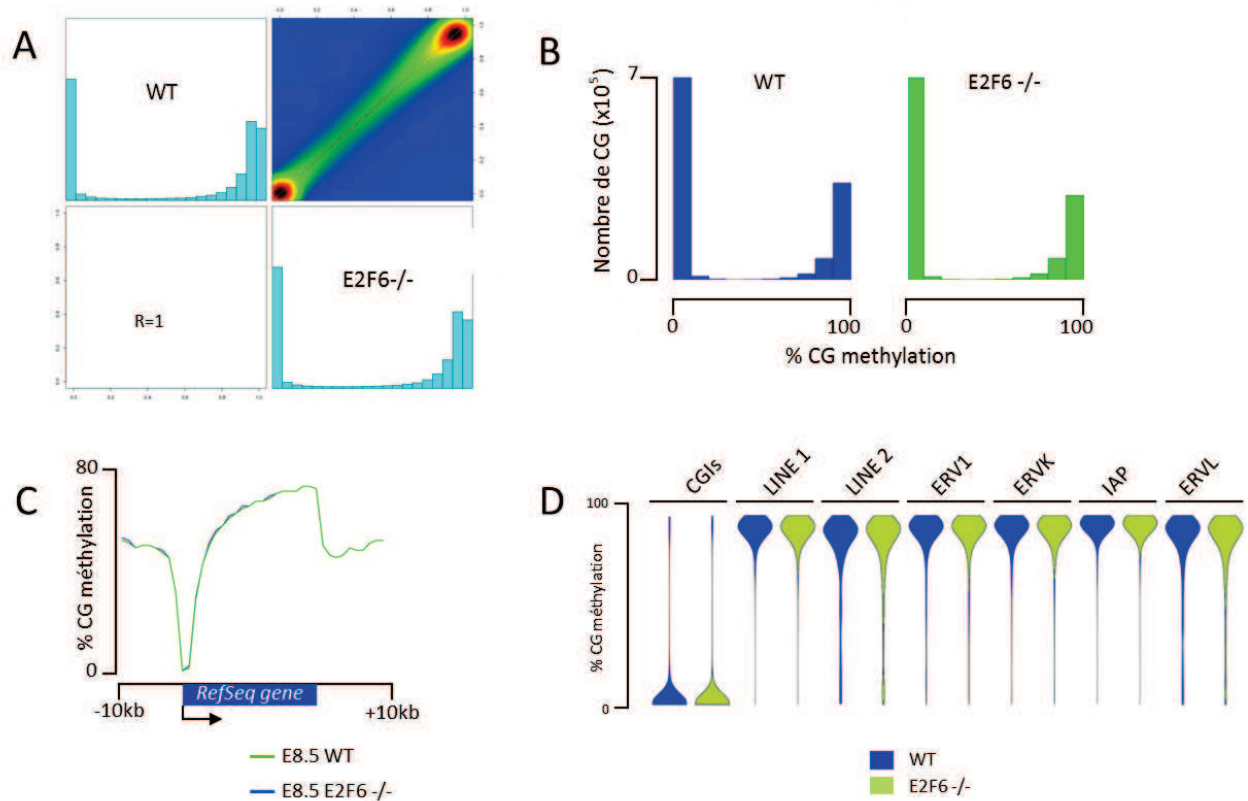


Figure 4: Analyse du niveau de méthylation de l'ADN obtenu par RRBS dans des embryons WT et *E2f6*^{-/-} à 8.5 jpc.

A : Le graphique montre la corrélation de méthylation dans 50.000 fenêtres génomiques prises au hasard. Les histogrammes représentent la distribution de la méthylation de l'ADN dans toutes les fenêtres. Le coefficient de corrélation de Pearson (r) est indiqué. **B :** Histogrammes de densité de la méthylation de l'ADN dans des embryons WT et *E2f6*^{-/-} à 8.5 jpc. **C :** Distribution de la méthylation moyenne de l'ADN au niveau des gènes RefSeq flanqués de 10kb dans les embryons WT et *E2f6*^{-/-} à 8.5 jpc. **D :** Graphique en violon de la méthylation mesurée dans les îlots CpG (CGIs) et les rétrotransposons dans les embryons WT et *E2f6*^{-/-} à 8.5 jpc.

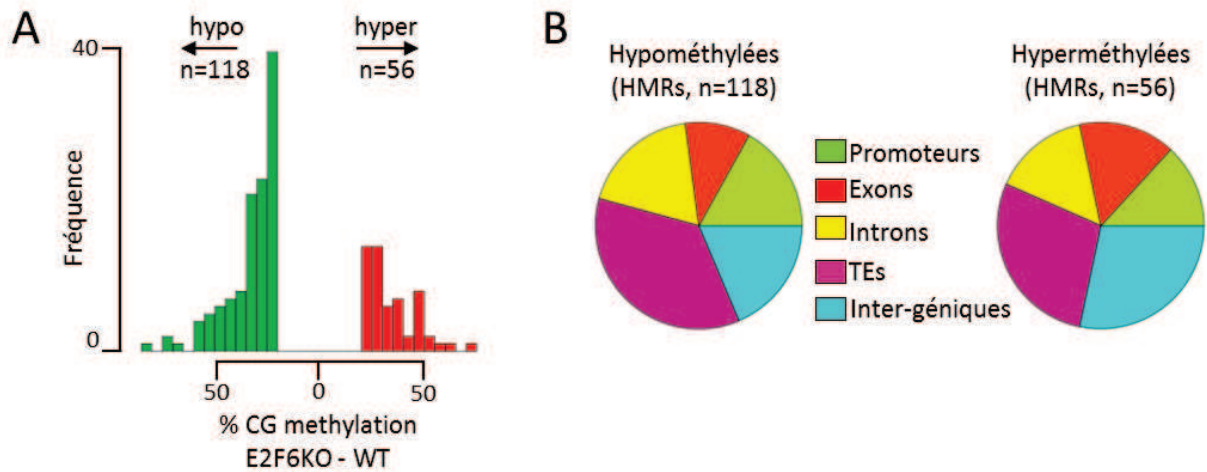


Figure 5 : Analyse des séquences hyperméthylées et hypométhylées, trouvées par RRBS dans des embryons WT et *E2f6*^{-/-} à 8.5 jpc.

A : Histogramme du nombre de séquences acquérant et perdant plus de 20% de méthylation dans les embryons *E2f6*^{-/-} par rapport aux embryons WT. **B** : Histogrammes en camembert du pourcentage de régions hypométhylées et hyperméthylées se trouvant dans des promoteurs, exons, introns, régions inter-géniques et éléments transposables (TEs).

Gene	WT	<i>E2f6</i> ^{-/-}
<i>Chn1os3</i>	0.478	0.278
<i>Boll</i>	0.429	0.228
<i>Ctnnb1</i>	0.707	0.504
<i>Ttc34</i>	0.7	0.497
<i>Slc44a4</i>	0.62	0.417
<i>Mir184</i>	0.681	0.477
<i>Ept1</i>	0.349	0.141
<i>Zfhx2os</i>	0.672	0.461
<i>Nkpd1</i>	0.559	0.347
<i>E2f6</i>	0.223	0.01
<i>Aebp2</i>	0.556	0.343
<i>2410012M07Rik</i>	0.512	0.298
<i>Chrna3</i>	0.846	0.631
<i>Rab35</i>	0.589	0.374
<i>Mael</i>	0.748	0.531
<i>Mettl7a1</i>	0.297	0.078
<i>Kif16b</i>	0.292	0.072
<i>Snx18</i>	0.565	0.345
<i>Slc25a31</i>	0.84	0.618
<i>Henmt1</i>	0.486	0.26
<i>Khdrbs1</i>	0.228	0
<i>Fas</i>	0.786	0.557
<i>Atp5g1</i>	0.424	0.194
<i>Pcbp3</i>	0.328	0.096
<i>Tspyl5</i>	0.256	0.024
<i>4930483I18Rik</i>	0.409	0.169
<i>Htra3</i>	0.426	0.185
<i>Thbs1</i>	0.654	0.412
<i>Ccnj</i>	0.589	0.336

Gene	WT	<i>E2f6</i> ^{-/-}
<i>Stard10</i>	0.966	0.702
<i>Lcn6</i>	0.63	0.362
<i>Rnf185</i>	0.718	0.448
<i>Mfsd6l</i>	0.318	0.045
<i>Lipe</i>	0.342	0.055
<i>Thpo</i>	0.888	0.598
<i>Gpr26</i>	0.349	0.047
<i>A730043L09Rik</i>	0.367	0.061
<i>C330024D21Rik</i>	0.938	0.622
<i>Olfr368</i>	0.572	0.252
<i>Phyh</i>	0.693	0.339
<i>Xlr3b</i>	0.842	0.488
<i>Flad1</i>	0.696	0.34
<i>Gm1564</i>	0.366	0.006
<i>Ppp2r2a</i>	0.812	0.445
<i>1700048O20Rik</i>	0.67	0.277
<i>Asic1</i>	0.495	0.084
<i>Tnfrsf18</i>	0.987	0.555
<i>Smc1b</i>	0.475	0.004
<i>Hormad1</i>	0.847	0.375
<i>Stag3</i>	0.74	0.188
<i>Rbm44</i>	0.585	0.026
<i>1700123I01Rik</i>	0.755	0.176
<i>Hsf2bp</i>	0.867	0.269
<i>Ept1</i>	0.758	0.157
<i>Gm2762</i>	0.7	0.015
<i>Gpat2</i>	0.941	0.164
<i>Tuba3a</i>	0.97	0.117

Figure 6 : Régions promotrices hypométhylées, identifiées par RRBS dans des embryons WT et *E2f6*^{-/-} à 8.5 jpc.

Le tableau présente le niveau de méthylation mesuré dans les embryons WT et *E2f6*^{-/-} (moyenne de deux embryons). Les régions présentées ici perdent au moins 20% de méthylation dans les embryons *E2f6*^{-/-} comparés aux embryons sauvages. Il s'agit de régions promotrices distantes de moins de 1000pb du TSS.

III) L'influence d'E2F6 sur la méthylation de l'ADN semble directe.

Nous avons ensuite regardé par quels mécanismes E2F6 contrôle la méthylation de l'ADN. Les RNA-seq ont été effectués dans trois embryons *E2f6*^{-/-} et trois embryons WT à 8.5 jpc (**Figure 7**). Les analyses montrent que l'expression des gènes codants pour les DNMTs, UHRF1 et TETs ne sont pas perturbés dans les mutants *E2f6*^{-/-} (**Figure 8.A**). Cela suggère qu'E2F6 régule directement la méthylation de l'ADN en *cis*, ce qui devra être confirmé par des expériences d'immunoprécipitation de chromatine.

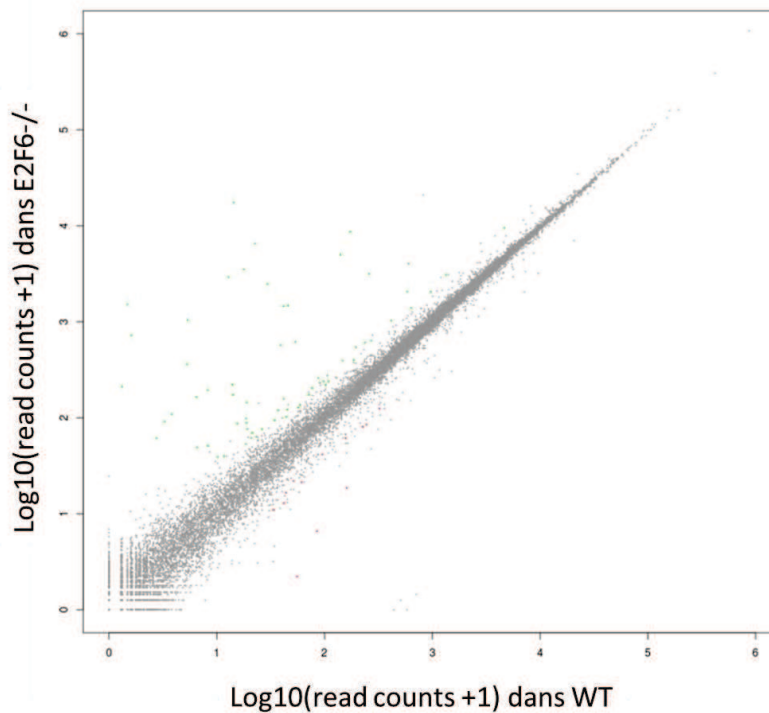


Figure 7 : Comparaison du niveau d'expression, obtenu par RNA-seq, des gènes RefSeq entre des embryons WT et *E2f6*^{-/-} à 8.5 jpc.

En verts, les gènes surexprimés, et en rouges, les gènes sous-exprimés dans les embryons *E2f6*^{-/-}.

IV) E2F6 réprime des gènes méiotiques

Afin d'étudier si les changements de méthylation en absence de E2F6 sont associés à des changements d'expression de gènes, nous avons réalisé des expériences de transcriptome par RNA-Seq dans trois embryons *E2f6*^{-/-} et trois embryons WT à 8.5 jpc (**Figure 7**). Ainsi, parmi les 118 régions hypométhylées identifiées par RRBS (au moins 20% de différence de méthylation), 17 correspondent à des promoteurs de gènes qui sont surexprimés (rapport fpm KO/WT > 2) dans les embryons mutants (**Figure 8.B**). Parmi ces gènes, 14 d'entre eux sont des gènes spécifiques de la lignée germinale (**Figure 8.B**). Cela indique que E2F6 est nécessaire pour méthyler et réprimer une partie des gènes de la lignée germinale dans l'embryon. La plupart de ces gènes est également surexprimée dans les embryons *Dnmt3b*^{-/-} (**Figure 8.B**), ce qui indique que la méthylation de l'ADN est nécessaire pour la fonction répressive de E2F6.

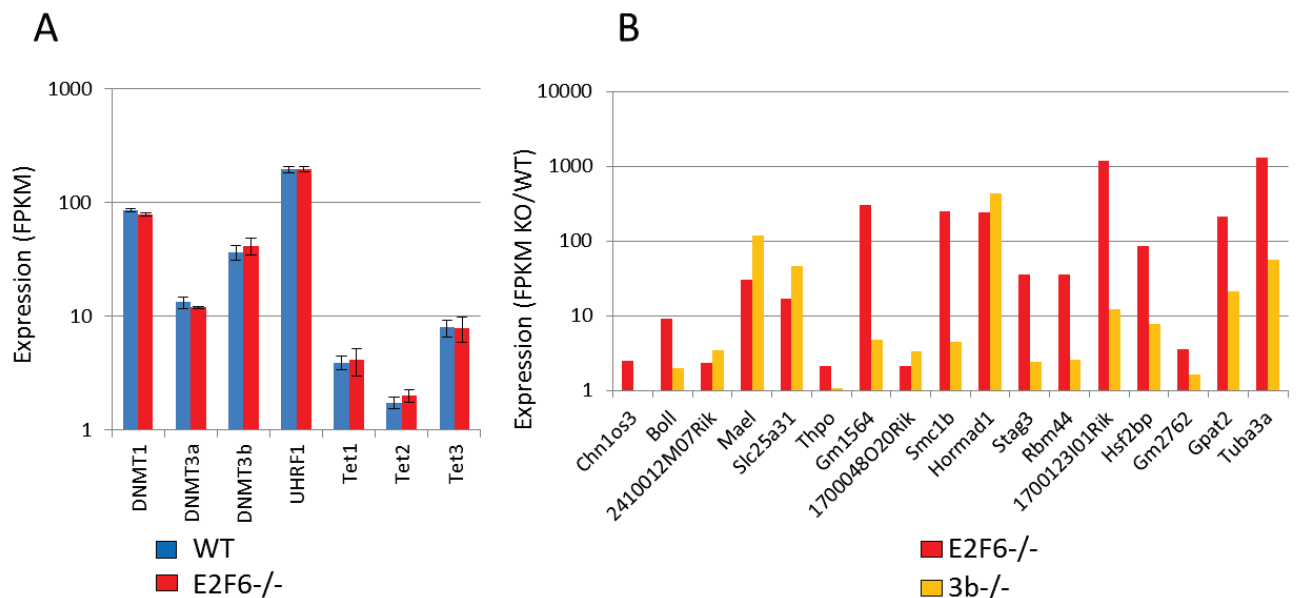


Figure 8 : Analyse des RNAseq obtenus dans des embryons WT et *E2F6*^{-/-} à 8.5 jpc.

A : Quantification de l'expression des gènes codants pour des protéines de la machinerie de méthylation de l'ADN (n=3 embryons). **B** : Comparaison de la surexpression de gènes dans des embryons WT, *E2f6*^{-/-} et *Dnmt3b*^{-/-} à 8.5 jpc. Parmi ces gènes, seuls *Thpo*, *Chn1os3* et *Prr23a3* ne sont pas de la lignée germinale.

V) E2F6 et G9a ne régulent pas la méthylation des mêmes séquences

Compte tenu du lien possible entre E2F6 et G9a, nous avons étudié si les régions hypométhylées dans les mutants *E2f6*^{-/-} sont les mêmes que celles hypométhylées dans les mutants *G9a*^{-/-}. Une analyse globale des régions méthylées par E2F6 montre qu'elles sont hypométhylées dans les embryons *E2f6*^{-/-} ainsi que *Dnmt3b*^{-/-}, mais pas dans les embryons *G9a*^{-/-} (**Figure 9.A**). Quand on analyse les promoteurs des gènes de la lignée germinale hypométhylés dans les mutants E2f6, il apparaît qu'aucun d'entre eux n'est commun avec ceux hypométhylés dans les embryons mutants *G9a*^{-/-} (**Figure 9.B**). G9a et E2F6 ont donc des cibles différentes pour le contrôle de la méthylation de l'ADN.

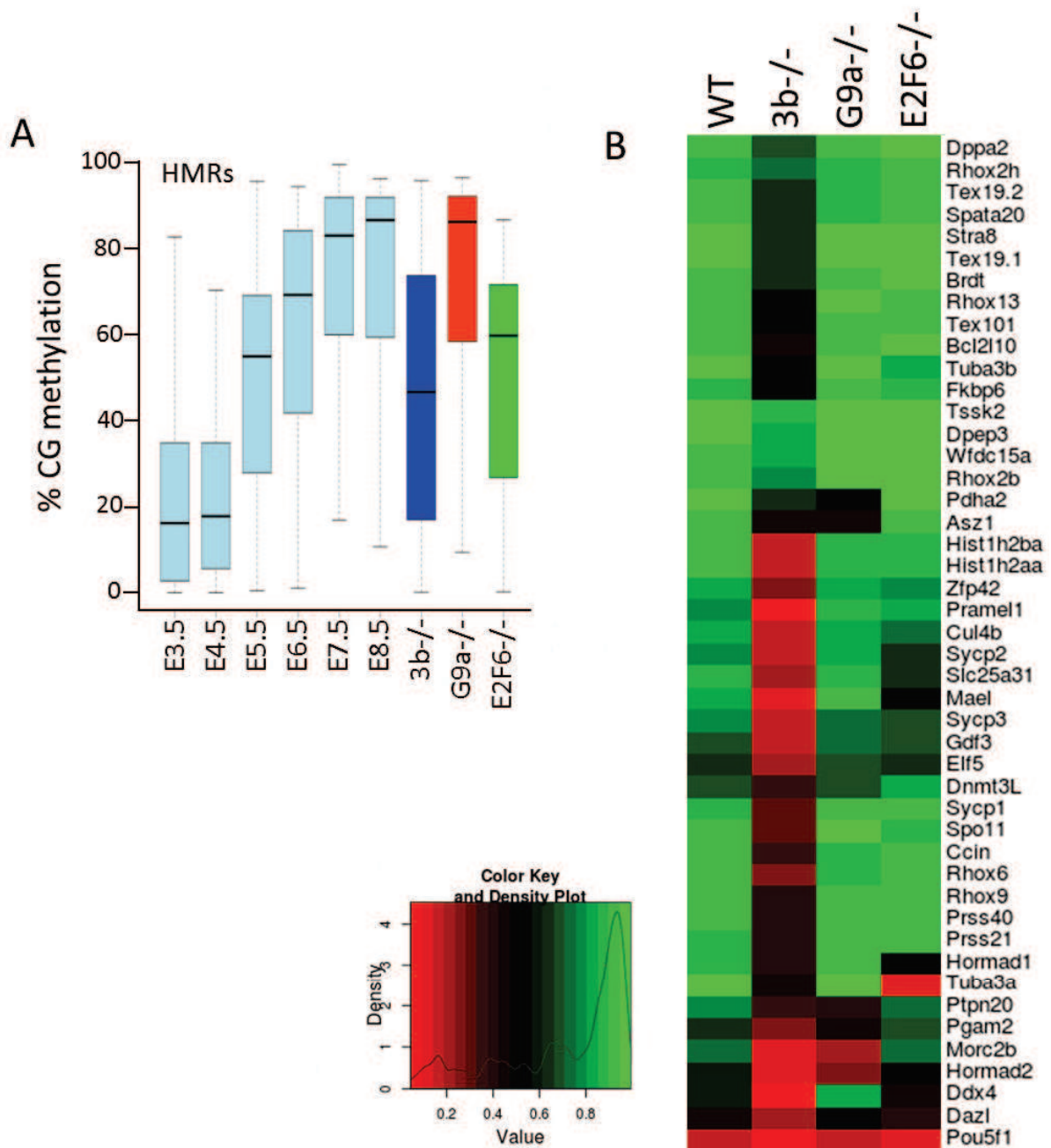


Figure 9 : Comparaison des régions hypométhylées dans les mutants *E2f6*^{-/-} et *G9a*^{-/-}.

A : Graphique montrant la dynamique d'acquisition de la méthylation sur les régions hypométhylées (perdant plus de 20% de la méthylation dans les embryons *E2f6*^{-/-}) dans l'embryogénèse (bleu clair) et des embryons 8.5 jpc *Dnmt3b*^{-/-} (bleu foncé), *G9a*^{-/-} (rouge) et *E2f6*^{-/-} (vert). **B** : Graphique représentant le niveau de méthylation de l'ADN mesuré sur les promoteurs (entre -1000 et +1000 pb du TSS) de gènes de la lignée germinale et de pluripotence.

CONCLUSIONS

Les conclusions principales de ce travail sont :

- E2F6 est exprimé après l'implantation de l'embryon au moment de la mise en place de la méthylation de l'ADN entre 4.5 et 8.5 jpc.
- E2F6 est impliqué dans le recrutement de la méthylation de l'ADN à des sites spécifiques *in vivo*, incluant des promoteurs à îlots CpG de plusieurs gènes méiotiques.
- G9a et E2F6 ne contrôlent pas la méthylation de l'ADN des mêmes régions génomiques *in vivo*.

DISCUSSION

La méthylation de l'ADN s'acquière de manière uniforme sur le génome entre 4.5 jpc et 8.5 jpc.

En utilisant des méthodes de cartographie à haute résolution, nous avons pu mettre en évidence la cinétique d'acquisition de la méthylation de l'ADN sur le génome au cours du développement embryonnaire. Ces données viennent grandement compléter les études déjà réalisées (Borgel et al., 2010; Mohn et al., 2008; Smith et al., 2012). De plus, elles permettent d'étudier quand, où et à quelle vitesse les gènes acquièrent ou perdent la méthylation de l'ADN durant l'embryogénèse entre 3.5 jpc et 8.5 jpc. Ces études ont toutefois été réalisées dans des embryons entiers entre 3.5 et 8.5 jpc. Il est donc possible que l'acquisition soit quelque peu différente d'une cellule à l'autre dans l'embryon, ce qui pourrait être étudié par des approches de cartographie sur cellules uniques qui sont actuellement développées (Smallwood et al., 2014).

Grâce à ces données, nous avons pu observer la mise en place de la méthylation de l'ADN sur certains éléments répétés, îlots CpG, différentes régions intra- et inter-géniques et les régions soumises à l'empreinte. Nous observons que le gain de la méthylation de l'ADN se fait de manière rapide entre les stades 4.5 jpc et 6.5 jpc. Cette acquisition *de novo* se fait de manière homogène sur tout le génome et semble donc se mettre en place par défaut. Ceci suggère que les DNMTs sont recrutées par défaut sur toute la chromatine grâce à leur domaine PWWP, et non pas à certaines régions par des facteurs de transcription.

La mise en place de la méthylation de l'ADN sur les îlots CpG (CGis) ne poursuit pas la même cinétique que le reste du génome. En effet, l'acquisition est plus lente et progressive entre 4.5 jpc et 8.5 jpc. Seulement 5% des îlots CpG gagnent plus de 50% de méthylation dans les embryons à 8.5 jpc. Il est d'ailleurs intéressant de remarquer que la méthylation des CGis n'atteint que très rarement un taux supérieur à 90% comme sur le reste du génome. Si la théorie selon laquelle la méthylation de l'ADN se met en place par défaut sur tout le génome est juste, alors il semblerait que les îlots soient spécifiquement protégés de cette acquisition. Quelles

sont les hypothèses capables d'expliquer la protection des îlots CpG contre la méthylation de l'ADN ? Le facteur SP1, par exemple, pourrait maintenir certains promoteurs de gènes non méthylés (Brandeis et al., 1994). La protéine CFP1 (CXXC Finger Protein 1) qui possède un domaine CXXC de liaison aux CpG non méthylés pourrait également contribuer à maintenir les îlots CpG dans un état hypométhylé (Deaton et Bird, 2011). On peut aussi imaginer que les îlots CpG soient appauvris en histones ou certaines marques d'histone, ce qui défavoriserait le recrutement des protéines DNMT3 par leur domaine PWWP. Une autre possibilité est que la méthylation de l'ADN soit constamment retirée de ces îlots CpG. Dans les ES, il a été montré que l'absence des enzymes de déméthylation TET, conduit à une hyperméthylation de certains CGIs (Dawlaty et al., 2011, 2014).

D'autre part, nous avons découvert une nouvelle classe d'îlots CpG dont la méthylation de l'ADN atteint un niveau ne dépassant pas 60% de méthylation à 8.5 jpc et dans les tissus adultes. Nous avons appelé cette nouvelle classe pmCGIs (partially methylated CGIs). Ceux-ci sont principalement localisés sur des gènes de la lignée germinale. Contrairement aux DMRs de gènes soumis à empreinte, cette méthylation partielle reflète une méthylation partielle acquise sur les deux allèles. Nous n'avons à ce jour pas d'hypothèse pour expliquer la méthylation partielle de ces séquences.

Pour éclaircir la protection des îlots CpG contre la méthylation et le recrutement partiel de la méthylation sur les pmCGI, il serait intéressant de cartographier plusieurs marques d'histones par CHIP-Seq dans les embryons entre 4.5 et 8.5 jpc afin de déterminer si on peut corrélérer l'acquisition et le niveau de méthylation des îlots CpG avec certaines combinaisons de marques d'histones. Ces expériences sont réalistes grâce au protocole de CHIP de marques d'histones sur de faibles quantités de cellules embryonnaires que j'ai validé au cours de ma thèse.

La méthylation de l'ADN, une marque épigénétique vraiment répressive ?

Nous avons remarqué que la méthylation de l'ADN se met en place préférentiellement aux régions intra-géniques et inter-géniques. C'est aussi le cas pour les îlots CpG qui sont beaucoup plus fréquemment méthylés dans les corps de gènes que dans les promoteurs. Cette méthylation des îlots intra-géniques se retrouve préférentiellement dans des gènes exprimés et impliqués dans des fonctions développementales et de la morphogénèse. De plus, nos expériences de RNA-Seq dans les embryons *Dnmt3b*^{-/-} et dKO *Dnmt3a*^{-/-} *Dnmt3b*^{-/-} montrent que l'absence de méthylation de l'ADN n'affecte l'expression que de très peu de gènes dans les embryons à 8.5 jpc. Ce sont surtout les gènes à promoteurs à îlots CpG fortement méthylés tels que les gènes méiotiques qui sont réactivés. D'autres analyses de transcriptome dans des cellules différenciées globalement déméthylées ont également conclu à un impact minimal sur le transcriptome (Ramos et al., 2015). Globalement, ces résultats remettent en cause la vision classique de la méthylation de l'ADN comme une marque épigénétique essentiellement répressive.

Il est possible que d'autres systèmes épigénétiques parallèles compensent l'absence de méthylation de l'ADN dans l'embryon. Cela est certainement dû à la mise en place de modifications d'histones en amont de la méthylation de l'ADN. La méthylation de l'ADN pourrait avoir un rôle dans la régulation plus fine de l'expression des gènes, par exemple en influençant l'expression de transcrits alternatifs ou l'épissage de certains exons (Lev Maor et al., 2015 ; Shukla et al., 2011). Nous n'avons à ce jour pas encore exploré en détails cette piste. Pour cela nous devons faire une analyse plus poussée des données de RNA-seq pour étudier si l'absence de méthylation influence les isoformes ou les profils d'épissages de certains transcrits.

DNMT3A et DNMT3B ont des fonctions spécifiques et redondantes.

Grâce au RRBS, nous avons aussi étudié les rôles des enzymes DNMT3A et DNMT3B sur la méthylation de l'ADN dans un contexte *in vivo* et à l'échelle génomique. Jusqu'alors, leurs rôles respectifs dans la mise en place de la méthylation durant l'embryogénèse étaient peu caractérisés car étudiés à l'échelle de gènes candidats ou dans les cellules ES en culture. Nous avons ainsi pu montrer que ces enzymes ont des fonctions spécifiques et redondantes *in vivo*.

Tout d'abord nous avons regardé comment s'expriment les gènes de chacune de ces enzymes durant le développement embryonnaire. Le gène *Dnmt3b* se surexprime entre 4.5 jpc et 8.5 jpc, avec un pic d'expression dans les épiblastes à 6.5 jpc. Ce pic correspond avec le moment durant lequel la méthylation de l'ADN se met en place sur le génome. L'expression du gène *Dnmt3a* est plus faible mais s'active également au même moment que celle de *Dnmt3b*. Ces résultats sont cohérents avec les niveaux de protéines évalués par marquages immunologiques (Watanabe et al., 2002 ; Okano et al., 1999).

Lorsqu'on analyse des embryons déficients pour l'une ou l'autre de ces deux enzymes, on n'observe qu'une déméthylation partielle du génome. A l'inverse l'inactivation simultanée des deux enzymes engendre une perte totale de méthylation du génome. Cela indique que ces enzymes ont des fonctions redondantes car elles sont donc capables de compenser partiellement l'absence de l'autre pour mettre en place la méthylation de l'ADN. Il est d'ailleurs intéressant de remarquer que très peu de régions perdent la totalité de la méthylation dans les embryons simples mutants *Dnmt3a*^{-/-} ou *Dnmt3b*^{-/-}, ce qui confirme le caractère redondants de ces enzymes. Ce travail de concert existe certainement pour assurer une mise en place efficace de la méthylation de l'ADN sur tout le génome. En effet, on sait que l'absence de méthylation conduit à une mort embryonnaire très précoce (Okano et al., 1999).

DNMT3B a cependant des fonctions spécifiques. L'absence de DNMT3B touche principalement les îlots CpG et en particulier ceux qui acquièrent lentement de la méthylation entre 4.5 et 6.5 jpc. Parmi ces régions se trouvent des promoteurs de gènes soumis à l'empreinte génomique, de gènes de la lignée germinale et les promoteurs du chromosome X inactif. Les pmCGis sont eux aussi déméthylés dans

les embryons déficients en DNMT3B, mais très peu dans les embryons déficients en DNMT3A. Quels mécanismes recrutent DNMT3B plutôt que DNMT3A à ces cibles ? Il est probable que DNMT3B soit recrutée par des facteurs protéiques ou des ARN non codants qui interagissent préférentiellement avec DNMT3B plutôt que DNMT3A. L'identification différentielle des partenaires protéiques de DNMT3B et DNMT3A par spectrométrie de masse pourrait nous aider à éclaircir ces mécanismes.

Les souris déficientes en DNMT3A meurent 4 semaines après la naissance (Okano et al., 1999). A 8.5 jpc, la méthylation du génome est très peu affectée par son absence. Ces résultats confirment que DNMT3A joue certainement des rôles importants plus tard dans le développement des souris pour la mise en place de profils de méthylation spécifiques dans les lignages différenciés. Pour le tester, il serait intéressant de lancer des analyses de méthylome dans plusieurs tissus de souris *Dnmt3a*^{-/-} âgées de 3 semaines. Les souris déficientes en DNMT3B, quant à elles, meurent à 14.5 jpc (Okano et al., 1999). La méthylation de l'ADN dans les embryons *Dnmt3b*^{-/-} à 8.5 jpc est plus touchée que dans les embryons *dnmt3a*^{-/-}, et de nombreuses régions sont déméthylées, ce qui peut expliquer le phénotype plus sévère des embryons *Dnmt3b*^{-/-} par rapport aux embryons *Dnmt3a*^{-/-}. Cela corrèle également avec l'expression plus forte de *Dnmt3b* par rapport à *Dnmt3a* durant la mise en place de la méthylation.

DNMT3A et DNMT3B ont des fonctions de maintien de la méthylation *in vivo*.

En complément de ce travail, nous avons étudié des embryons à 8.5 jpc déficients pour les deux enzymes DNMT3A et DNMT3B à la fois. Nos analyses nous ont permis de mettre en évidence qu'en l'absence de ces deux enzymes, le génome est globalement déméthylé, soulignant le caractère redondant de ces deux protéines. Étonnamment, on observe aussi une absence totale de méthylation au niveau des gDMRs de loci soumis à empreinte parentale dans les embryons dKO mais pas les simples mutants. Ces résultats sont une surprise car ils sont en désaccord avec certains résultats publiés par les équipes de H. Sasaki et M. Okano montrant que la présence de DNMT1 seule est suffisante pour maintenir la méthylation des gDMRs

de *H19*, *Dlk1* et *Igf2r* dans les embryons à 9.5 jpc (Oda et al., 2006 ; Hirasawa et al., 2008). Nos données de RRBS ne couvrent pas les gDMRs des gènes *H19* et *Dlk1* qui sont relativement pauvres en CpG, donc nos résultats sont en réalité discordant uniquement pour le locus *Igf2r*. Nous ne pouvons également pas exclure un effet indirect sur l'action de DNMT1 ou UHRF1 dans les embryons dKO. A ce stade je peux préciser que les données de RNA-Seq n'indiquent pas de changement d'expression des gènes *Dnmt1*, *Uhrf1* et *Zfp57* dans les embryons dKO. Cependant cela doit être validé au niveau des protéines par immunofluorescence par exemple.

Cela signifie-t-il que DNMT3A et DNMT3B ont également une fonction redondante de maintien de la méthylation dans le génome entier ? Les données obtenues dans les cellules ES suggèrent que DNMT1, DNMT3A et DNMT3B pourraient agir de concert pour maintenir la méthylation de l'ADN sur certaines régions, en particulier des éléments répétés, mais cela n'a jamais été démontré dans un contexte physiologique *in vivo* (Kim et al., 2002 ; Chen et al., 2003; Liang et al., 2002 ; Arand et al., 2012). L'étude de nos embryons dKO ne permet pas de répondre à cette question car nous ne pouvons pas distinguer l'effet de maintien de la méthylation par rapport à l'absence de méthylation *de novo*. Pour répondre à cette question, il faudrait générer des souris avec une inactivation conditionnelle des deux enzymes après que la méthylation globale du génome ait été acquise. Cela peut être envisagé grâce à l'utilisation de nos allèles 2lox en combinaison avec une souris exprimant le recombinaise Cre dans un lignage différencié.

Ces résultats apportent des éclaircissements sur la mise en place de la méthylation qui semble se faire par défaut sur tout le génome, et sur les rôles des enzymes DNMTs dans cette acquisition. Il nous faut tout de même confirmer certains résultats par des expériences complémentaires, car ils sont en désaccord avec des résultats déjà publiés. Aussi, il reste encore à clarifier comment les DNMTs ne sont pas recrutées au niveau de certains îlots CpG ou pourquoi le sont-elles spécifiquement à certains plutôt qu'à d'autres. C'est ce que nous avons essayé de comprendre en nous intéressant à G9a et E2F6, deux facteurs potentiellement impliqués dans le ciblage de la méthylation de l'ADN.

G9a contrôle la méthylation de l'ADN sur des régions spécifiques du génome

Le contrôle de la méthylation de l'ADN est encore très mal compris à ce jour. Nous donnons ici quelques éléments de réponse après étude de l'histone méthyltransférase G9a *in vivo*.

Dans les cellules ES, Il a été montré que G9a peut interagir avec DNMT3A et DNMT3B et qu'elle peut recruter la méthylation de l'ADN à de nombreuses régions riches en CpG tels que des ICRs, des rétrotransposons, des gènes de la pluripotente, du développement et de la lignée germinale (Dong et al., 2008 ; Tachibana et al., 2008 ; Myant et al., 2011 ; Ikegami et al., 2007 ; Xin et al., 2003). Grâce à des méthodes de cartographie à haute résolution, nous avons étudié l'impact de l'inactivation de G9a sur le méthylome des embryons *G9a*^{-/-} et des cellules ES *G9a*^{-/-}. On remarque que la méthylation globale du génome des embryons *G9a*^{-/-} n'est pas affectée contrairement à celle des cellules ES *G9a*^{-/-}. Cette différence entre les embryons et les cellules ES *G9a*^{-/-} pourrait s'expliquer par le fait que l'absence de G9a est compensée par d'autres mécanismes dans les embryons mais pas les cellules ES. Les cellules ES pourraient aussi être plus sensibles à un effet mineur de G9a car elles subissent un grand nombre de divisions cellulaires en culture.

Ainsi seul un nombre restreint d'ilots CpG est hypométhylé dans les embryons *G9a*^{-/-}, incluant des promoteurs de gènes spécifiques de la lignée germinale. G9a joue-t-il un rôle direct pour le recrutement de la méthylation de l'ADN à ces régions spécifiques du génome? Pour répondre à cette question, nous avons réalisé des ChIP afin de montrer que G9a se lie à proximité des régions hypométhylées dans les mutants *G9a*^{-/-}. De plus, l'absence de G9a n'affecte pas l'expression des autres gènes de la machinerie de méthylation de l'ADN telles que les *Dnmts*, *Uhrf1* ou les *Tets*. Seul *Dnmt3b* se surexprime un peu dans les embryons *G9a*^{-/-}, ce qui est certainement une conséquence du retard de développement des embryons mutants.

Hypothèses sur les mécanismes de contrôle de la méthylation de l'ADN par G9a

A ce stade, nous ne savons pas si l'hypométhylation de certaines régions dans les embryons *G9a*^{-/-} est due à un problème de mise en place *de novo* ou de maintien de la méthylation de l'ADN. Afin de répondre à cette question, nous avons voulu regarder l'état de la méthylation plus en amont au stade 6.5 jpc. Cependant, les embryons ont un retard de développement qui ne leur permet pas de s'implanter et nous n'avons donc pas réussi à obtenir d'embryons à ce stade. Pour différencier un effet sur la mise en place *de novo* ou le maintien de la méthylation de l'ADN, il serait nécessaire de générer des souris avec une inactivation conditionnelle de G9a par la technologie Lox-Cre après que la méthylation de l'ADN du génome soit mise en place.

G9a contrôle-t-elle la méthylation de l'ADN via la modification H3K9me2 ? Nos expériences de ChIP-Seq dans des embryons à 8.5 jpc ou réalisées dans les cellules ES (Liu et al., 2015) montrent que cette modification d'histones est présente à nos régions cibles, mais aussi sur la quasi-totalité du génome. Les résultats obtenus précédemment dans les cellules ES et les embryons montrent que G9a est responsable de la majorité de la marque H3K9me2 dans le génome, ce qui implique que l'inactivation de G9a doit induire une perte globale de H3K9me2 sur tout le génome des embryons *G9a*^{-/-}. Un ChIP-seq de la modification H3K9me2 dans les embryons *G9a*^{-/-} nous permettrait de confirmer si l'absence de G9a affecte H3K9me2 sur tout le génome ou uniquement à des régions spécifiques du génome. Si l'effet sur H3K9me2 est global, cela signifie que la perte de H3K9me2 ne corrèle pas avec les pertes de méthylation de l'ADN qui sont limitées à certaines séquences spécifiques. Nos résultats tendent à favoriser cette hypothèse car nous avons mis en évidence que d'autres promoteurs de gènes comme *Tex101* et *Sycp2* perdent la modification d'histone H3K9me2 dans les embryons *G9a*^{-/-} mais sans effet sur la méthylation de l'ADN.

Ces résultats suggèrent que l'impact de G9a sur la méthylation de l'ADN n'est pas dépendant de son activité catalytique sur H3K9. Des études de complémentation avec des mutants catalytiques de G9a ont été effectuées dans des cellules ES *G9a*^{-/-} et ont montré que l'activité catalytique n'est pas nécessaire pour rétablir la

méthylation de l'ADN (Dong et al., 2008; Epsztejn-Litman et al., 2008; Tachibana et al., 2008). Cependant, compte tenu de l'effet potentiellement indirect de G9a sur le méthylome global des cellules ES, il est probable que ces résultats ne peuvent pas être extrapolés aux embryons. Une étude d'embryons mutants pour le site catalytique de G9a, généré par la technologie Crisp-Cas9, permettrait de répondre à cette question dans le contexte physiologique *in vivo*.

Il serait également intéressant d'étudier l'impact de GLP, le partenaire de G9a, sur la méthylation de l'ADN. Il est possible que GLP compense l'absence de G9a à de nombreuses cibles génomiques, ce qui pourrait expliquer les effets restreints sur la méthylation de l'ADN. L'étude d'embryon *Glp*^{-/-} et double mutants *G9a*^{-/-} *Glp*^{-/-} pourrait apporter des réponses à ces questions et nous permettre de savoir si GLP a des fonctions spécifiques ou redondantes avec G9a dans le contrôle de la méthylation de l'ADN *in vivo*.

Autre possibilité de contrôle de la méthylation de l'ADN par G9a

G9a pourrait également contrôler la méthylation via d'autres mécanismes. En effet, G9a peut méthyler d'autres résidus d'histones tels que H3K27, H3K56 ou des variants d'H1. Ces modifications à des endroits spécifiques pourraient être responsables du recrutement de la machinerie de méthylation de l'ADN à certaines régions plutôt qu'à d'autres. Des études ont montré que la modification H3K9me2 pouvait également servir de substrat à d'autres histones méthyltransférase comme SETDB1 qui peut trimétyler H3K9. La nouvelle modification épigénétique, H3K9m3, pourrait à son tour recruter la méthylation de l'ADN. Des études par CHIP du niveau de H3K9m3 dans les embryons *G9a*^{-/-} permettraient d'évaluer si cette modification d'histone est impliquée dans le ciblage de la méthylation de l'ADN. Enfin il a été montré que G9a interagit directement avec DNMT3A et DNMT3B, ce qui suggère que G9a pourrait favoriser la méthylation de l'ADN via un recrutement direct des DNMTs.

E2F6 contrôle la méthylation de l'ADN sur des régions spécifiques du génome

Nos résultats dans les embryons déficients en G9a nous montrent que cette enzyme est indispensable pour la modification des histones H3K9m2 et le recrutement de la méthylation de l'ADN à des régions spécifiques du génome, incluant des promoteurs de gènes de la lignée germinale. Cependant, le système permettant de recruter la méthylation et/ou de maintenir G9a au niveau de certaines régions plutôt qu'à d'autres n'est pas encore élucidé. Nous nous sommes tournés vers l'hypothèse d'un facteur de transcription se liant à G9a. E2F6 a été montré comme se trouvant dans un même complexe répressif que G9a, qui pourrait être capable de recruter la méthylation de l'ADN (Ogawa et al., 2002; Mozzetta et al., 2014).

E2F6 est un facteur de transcription exprimé de manière croissante durant l'embryogénèse de 4.5 jpc à 8.5 jpc. Des données sporadiques ont montré que l'absence d'E2F6 dans des cellules ES et des MEFs engendre une réactivation de quatre gènes de la lignée germinale ainsi qu'une perte de recrutement de DNMT3B au niveau de ces mêmes gènes (Kehoe et al., 2008; Storre et al., 2005; Velasco et al., 2010; Laseva et al., 2013 ; Storre et al., 2005) . Afin d'évaluer le rôle de ce facteur à l'échelle génomique et dans les conditions physiologiques du développement embryonnaire, nous avons cartographié la méthylation de l'ADN dans des embryons déficients en E2F6 à 8.5 jpc. Les résultats montrent qu'E2F6 joue un rôle indispensable dans le recrutement de la méthylation de l'ADN à des sites spécifiques, incluant les promoteurs à îlots CpG de plusieurs gènes de la lignée germinale. Des RNA-seq réalisés dans des embryons mutants *E2f6*^{-/-} montrent que les gènes perdant de la méthylation de l'ADN au niveau de leurs promoteurs sont fortement réactivés. Ces mêmes expériences de RNA-Seq indiquent que l'expression des autres gènes de la machinerie de méthylation de l'ADN que sont *Uhrf1*, les *Dnmts* et les *Tet* n'est pas perturbée, suggérant un effet direct d'E2F6 sur la méthylation de l'ADN en *cis*. Cette hypothèse devra être vérifiée par des expériences de ChIP et ChIP-Seq. Pour comprendre comment E2F6 recrute la méthylation de l'ADN, nous pouvons tester l'existence d'interactions directes entre E2F6 et les

protéines DNMT par co-immunoprécipitation ou, d'une manière plus générale, identifier les partenaires protéiques d'E2F6 par immunoprécipitation suivie de spectrométrie de masse. On sait d'autre part qu'E2F6 interagit en complexe avec des protéines Polycomb dont EZH2 qui catalyse le dépôt de la marque d'histone H3K27me3 (Attwooll et al., 2005). Une étude approfondie des marques H3K27m2 et H3K27m3 par CHIP aux régions cibles d'E2F6 pourrait aider à comprendre si ces modifications sont en lien avec le rôle d'E2F6 dans la méthylation de l'ADN.

Au vu des résultats précédent montrant que G9a et E2F6 peuvent être isolées dans le même complexe protéique (Ogawa et al., 2002), nous avons émis l'hypothèse selon laquelle E2F6 pourrait recruter G9a afin d'établir la méthylation de l'ADN à des régions spécifiques. Cependant, les régions déméthylées dans les embryons *E2f6*^{-/-} ne correspondent pas aux régions déméthylées dans les embryons *G9a*^{-/-}. Ces deux facteurs semblent donc avoir des rôles distincts dans la méthylation de différentes régions de l'ADN. Toutefois cela n'exclue pas qu'E2F6 puisse jouer un rôle dans le recrutement de G9a à certaines séquences. Dans le cadre de cette hypothèse, nous essayerons de voir si l'absence d'E2F6 peut affecter la modification d'histone H3K9m2 au niveau des cibles d'E2F6 et de G9a.

Plus de la moitié des gènes dépendants d'E2F6 sont de la lignée germinale. Cependant une grande partie des gènes de cette famille qui sont méthylés dans l'embryogenèse ne sont pas affectés dans les mutants *E2f6*^{-/-}. Il y a donc certainement d'autres facteurs de transcription qui sont impliqués dans le contrôle de la méthylation de ces gènes.

Pourquoi l'absence de méthylation de l'ADN est-elle létale dans l'embryon ?

Les cellules ES triple knock-out pour *Dnmt1*, *Dnmt3a* et *Dnmt3b* sont viables et prolifèrent. Pourtant l'absence de DNMT3B seule suffit à induire une létalité dans l'embryon. Nos expériences de transcriptome indiquent que les embryons *Dnmt3b*^{-/-} ou dKO *Dnmt3a*^{-/-} *Dnmt3b*^{-/-} réactivent un petit nombre de gènes, Parmi ceux-ci, on

retrouve des surexpressions légères de gènes de pluripotence, de développement mais surtout une forte surexpression de gènes spécifiques de la lignée germinale. Actuellement, on ne sait pas quels sont les gènes parmi ceux-ci qui ont un réel caractère létal pour l'embryon de souris.

Certains gènes sont surexprimés en commun dans les embryons mutants *G9a*^{-/-} et *Dnmt3b*^{-/-} qui meurent tous les deux entre 10.5 et 14.5 jpc. Il se pourrait donc que les gènes responsables de la mort des embryons fassent partie des gènes communs entre les individus *G9a*^{-/-} et *Dnmt3b*^{-/-}. Ces gènes font tous partie de la famille des gènes spécifiques de la lignée germinales. Les souris déficientes en DNMT3B ne réactivent pas certains de ces gènes hypométhylés aussi fortement que le font les souris mutantes *G9a*^{-/-}, ce qui pourrait expliquer la létalité plus tardive des embryons *Dnmt3b*^{-/-} vers 14.5 jpc. Des données préliminaires à 6.5 jpc, 8.5 jpc et 11.5 jpc, non présentées ici, semblent montrer que certains gènes réactivés dans les mutants *G9a*^{-/-}, se réactivent de plus en plus au cours du temps dans les embryons *Dnmt3b*^{-/-}. Il se pourrait que l'expression de ces gènes atteignent le même niveau d'expression que dans les mutants *G9a*^{-/-} et soient la cause de la létalité embryonnaire tardive des animaux *Dnmt3b*^{-/-}. A l'inverse, seul un petit nombre de gènes méiotiques est hypométhylé et réactivé en commun dans les embryons déficients en E2F6 ou DNMT3B. Puisque les souris *E2f6*^{-/-} sont viables, nous pouvons conclure que la surexpression de ces quelques gènes en particulier ne participe pas à la létalité des embryons *Dnmt3b*^{-/-}.

Les données sur l'acquisition de la méthylation de l'ADN au cours du développement embryonnaire montrent que celle-ci se fait entre 4.5 jpc et 8.5 jpc. Son rôle ne doit devenir indispensable qu'après 8.5 jpc puisqu'elle n'est pas encore en place avant. D'autres systèmes tels que les modifications d'histones et/ou des facteurs de transcription sont certainement mis en avant durant ces étapes afin de réprimer les gènes en attendant que la méthylation de l'ADN soit correctement en place, sinon les embryons mourraient certainement plus tôt.

BIBLIOGRAPHIE

- Alvarez, M.C., Santos, J.C., Maniezzo, N., Ladeira, M.S., da Silva, A.L.C., Scaletsky, I.C.A., Pedrazzoli, J., and Ribeiro, M.L. (2013). MGMT and MLH1 methylation in *Helicobacter pylori*-infected children and adults. *World J. Gastroenterol. WJG* 19, 3043–3051.
- 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.
- Antelo, M., Balaguer, F., Shia, J., Shen, Y., Hur, K., Moreira, L., Cuatrecasas, M., Bujanda, L., Giraldez, M.D., Takahashi, M., et al. (2012). A high degree of LINE-1 hypomethylation is a unique feature of early-onset colorectal cancer. *PLoS One* 7, e45357.
- Arand, J., Spieler, D., Karius, T., Branco, M.R., Meilinger, D., Meissner, A., Jenuwein, T., Xu, G., Leonhardt, H., Wolf, V., et al. (2012). In Vivo Control of CpG and Non-CpG DNA Methylation by DNA Methyltransferases. *PLoS Genet* 8, e1002750.
- Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol. Cell* 31, 785–799.
- Arima, T., Hata, K., Tanaka, S., Kusumi, M., Li, E., Kato, K., Shiota, K., Sasaki, H., and Wake, N. (2006). Loss of the maternal imprint in *Dnmt3Lmat^{-/-}* mice leads to a differentiation defect in the extraembryonic tissue. *Dev. Biol.* 297, 361–373.
- Arnaud, P. (2010). Genomic imprinting in germ cells: imprints are under control. *Reprod. Camb. Engl.* 140, 411–423.
- Attwooll, C., Oddi, S., Cartwright, P., Prosperini, E., Agger, K., Steensgaard, P., Wagener, C., Sardet, C., Moroni, M.C., and Helin, K. (2005). A Novel Repressive E2F6 Complex Containing the Polycomb Group Protein, EPC1, That Interacts with EZH2 in a Proliferation-specific Manner. *J. Biol. Chem.* 280, 1199–1208.
- Auclair, G., and Weber, M. (2012). Mechanisms of DNA methylation and demethylation in mammals. *Biochimie* 94, 2202–2211.
- Auer-Grumbach, M., Mauko, B., Auer-Grumbach, P., and Pieber, T.R. (2012). Molecular genetics of hereditary sensory neuropathies. *NeuroMolecular Med.* 8, 147–158.
- 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.
- Balassiano, K., Lima, S., Jenab, M., Overvad, K., Tjonneland, A., Boutron-Ruault, M.C., Clavel-Chapelon, F., Canzian, F., Kaaks, R., Boeing, H., et al. (2011). Aberrant DNA methylation of cancer-associated genes in gastric cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC-EURGAST). *Cancer Lett.* 311, 85–95.

- Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S.C., Mann, M., and Kouzarides, T. (2010). Nucleosome-Interacting Proteins Regulated by DNA and Histone Methylation. *Cell* 143, 470–484.
- Bartolomei, M.S., Zemel, S., and Tilghman, S.M. (1991). Parental imprinting of the mouse H19 gene. *Nature* 351, 153–155.
- Baubec, T., Ivánek, R., Lienert, F., and Schübeler, D. (2013). Methylation-Dependent and -Independent Genomic Targeting Principles of the MBD Protein Family. *Cell* 153, 480–492.
- Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schübeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* 520, 243–247.
- Berger, S.L., Kouzarides, T., Shiekhhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. *Genes Dev.* 23, 781–783.
- Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* 125, 315–326.
- Bian, C., Chen, Q., and Yu, X. (2015). The zinc finger proteins ZNF644 and WIZ regulate the G9a/GLP complex for gene repression. *eLife* 4.
- Bian, C., Chen, Q., and Yu, X. The zinc finger proteins ZNF644 and WIZ regulate the G9a/GLP complex for gene repression. *eLife* 4.
- Blackledge, N.P., Zhou, J.C., Tolstorukov, M.Y., Farcas, A.M., Park, P.J., and Klose, R.J. (2010). CpG islands recruit a histone H3 lysine 36 demethylase. *Mol. Cell* 38, 179–190.
- Booth, M.J. [b1] (analytic), Branco, M.R. [b2 b3] (analytic), Ficz, G. [b2] (analytic), Oxley, D. [b4] (analytic), Krueger, F. [b5] (analytic), Reik, W. [b2 b3] (analytic), and Balasubramanian, S. [b1 b6 b7] (analytic) (2012). Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution (English). *Sci. Wash DC* 336, 934–937.
- Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. *Nat. Genet.* 42, 1093–1100.
- 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* 1760.
- 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. (2001a). Dnmt3L and the establishment of maternal genomic imprints. *Science* 294, 2536–2539.
- Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., and Bestor, T.H. (2001b). Dnmt3L and the establishment of maternal genomic imprints. *Science* 294, 2536–2539.

- Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A., and Cedar, H. (1994). Sp1 elements protect a CpG island from de novo methylation. *Nature* 371, 435–438.
- Carmell, M.A., Girard, A., van de Kant, H.J.G., Bourc’his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. (2007). MIWI2 Is Essential for Spermatogenesis and Repression of Transposons in the Mouse Male Germline. *Dev. Cell* 12, 503–514.
- Caspary, T., Cleary, M.A., Baker, C.C., Guan, X.J., and Tilghman, S.M. (1998). Multiple mechanisms regulate imprinting of the mouse distal chromosome 7 gene cluster. *Mol. Cell. Biol.* 18, 3466–3474.
- Chan, S., and Majeti, R. (2013). Role of DNMT3A, TET2, and IDH1/2 mutations in pre-leukemic stem cells in acute myeloid leukemia. *Int. J. Hematol.* 98, 648–657.
- Chédin, 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, L., Chen, K., Lavery, L.A., Baker, S.A., Shaw, C.A., Li, W., and Zoghbi, H.Y. (2015). MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. *Proc. Natl. Acad. Sci.* 112, 5509–5514.
- 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.
- Chinwalla, A.T., Cook, L.L., Delehaunty, K.D., Fewell, G.A., Fulton, L.A., Fulton, R.S., Graves, T.A., Hillier, L.W., Mardis, E.R., McPherson, J.D., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562.
- 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.
- Ciccone, D.N., Su, H., Hevi, S., Gay, F., Lei, H., Bajko, J., Xu, G., Li, E., and Chen, T. (2009). KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature* 461, 415–418.
- Clouaire, T., and Stancheva, I. (2008). Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin? *Cell. Mol. Life Sci. CMLS* 65, 1509–1522.
- Clouaire, T., Webb, S., Skene, P., Illingworth, R., Kerr, A., Andrews, R., Lee, J.-H., Skalnik, D., and Bird, A. (2012). Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes Dev.* 26, 1714–1728.
- 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.
- 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.

Couronné, 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.

Daniel, J.M., and Reynolds, A.B. (1999). The Catenin p120ctn Interacts with Kaiso, a Novel BTB/POZ Domain Zinc Finger Transcription Factor. *Mol. Cell. Biol.* 19, 3614–3623.

Davies, M.N., Volta, M., Pidsley, R., Lunnon, K., Dixit, A., Lovestone, S., Coarfa, C., Harris, R.A., Milosavljevic, A., Troakes, C., et al. (2012). Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome Biol.* 13, R43.

Dawlaty, M.M., Ganz, K., Powell, B.E., Hu, Y.-C., Markoulaki, S., Cheng, A.W., Gao, Q., Kim, J., Choi, S.-W., Page, D.C., et al. (2011). Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* 9, 166–175.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. *Genes Dev.* 25, 1010–1022.

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.

de Greef, J.C., Wang, J., Balog, J., den Dunnen, J.T., Frants, R.R., Straasheijm, K.R., Aytekin, C., van der Burg, M., Duprez, L., Ferster, A., et al. (2011). Mutations in ZBTB24 Are Associated with Immunodeficiency, Centromeric Instability, and Facial Anomalies Syndrome Type 2. *Am. J. Hum. Genet.* 88, 796–804.

Delgado, S., Gómez, M., Bird, A., and Antequera, F. (1998). Initiation of DNA replication at CpG islands in mammalian chromosomes. *EMBO J.* 17, 2426–2435.

Delhommeau, F., Dupont, S., Valle, V.D., James, C., Trannoy, S., Massé, A., Kosmider, O., Le Couedic, J.-P., Robert, F., Alberdi, A., et al. (2009). Mutation in TET2 in Myeloid Cancers. *N. Engl. J. Med.* 360, 2289–2301.

Dennis, K., Fan, T., Geiman, T., Yan, Q., and Muegge, K. (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes Dev.* 15, 2940–2944.

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.

Divanlian, K., Tsanev, D., Franke, L., and Bozhkova, K. (1989). [An in vitro micromethod for studying phagocytic and bacteriocidal activity of granulocytes from peripheral blood]. *Eksp. Med. Morfol.* 28, 49–53.

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

- Dong, K.B., Maksakova, I.A., Mohn, F., Leung, D., Appanah, R., Lee, S., Yang, H.W., Lam, L.L., Mager, D.L., Schübeler, D., et al. (2008). DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *EMBO J.* *27*, 2691–2701.
- Edmunds, J.W., Mahadevan, L.C., and Clayton, A.L. (2008). Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J.* *27*, 406–420.
- Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., Deplus, R., Fuks, F., Shinkai, Y., Cedar, H., et al. (2008). De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nat. Struct. Mol. Biol.* *15*, 1176–1183.
- Estève, P.-O., Chin, H.G., Smallwood, A., Feehery, G.R., Gangisetty, O., Karpf, A.R., Carey, M.F., and Pradhan, S. (2006). Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes Dev.* *20*, 3089–3103.
- 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.
- Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H., and Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat. Cell Biol.* *8*, 188–194.
- Feng, Q., and Zhang, Y. (2001). The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. *Genes Dev.* *15*, 827–832.
- Feng, S., Cokus, S.J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., et al. (2010). Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 8689–8694.
- Ferguson-Smith, A.C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. *Nat. Rev. Genet.* *12*, 565–575.
- Filion, G.J.P., Zhenilo, S., Salozhin, S., Yamada, D., Prokhortchouk, E., and Defossez, P.-A. (2006). A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Mol. Cell. Biol.* *26*, 169–181.
- Fraineau, S., Palii, C.G., Allan, D.S., and Brand, M. (2015). Epigenetic regulation of endothelial-cell-mediated vascular repair. *FEBS J.* *282*, 1605–1629.
- 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.
- Fujita, N., Watanabe, S., Ichimura, T., Tsuruzoe, S., Shinkai, Y., Tachibana, M., Chiba, T., and Nakao, M. (2003). Methyl-CpG Binding Domain 1 (MBD1) Interacts with the Suv39h1-HP1 Heterochromatic Complex for DNA Methylation-based Transcriptional Repression. *J. Biol. Chem.* *278*, 24132–24138.
- Fuks, F., Hurd, P.J., Wolf, D., Nan, X., Bird, A.P., and Kouzarides, T. (2003a). The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J. Biol. Chem.* *278*, 4035–4040.
- Fuks, F., Hurd, P.J., Deplus, R., and Kouzarides, T. (2003b). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res.* *31*, 2305–2312.

Gabel, H.W., Kinde, B., Stroud, H., Gilbert, C.S., Harmin, D.A., Kastan, N.R., Hemberg, M., Ebert, D.H., and Greenberg, M.E. (2015). Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* *522*, 89–93.

Gama-Sosa, M.A., Midgett, R.M., Slagel, V.A., Githens, S., Kuo, K.C., Gehrke, C.W., and Ehrlich, M. (1983). Tissue-specific differences in DNA methylation in various mammals. *Biochim. Biophys. Acta BBA - Gene Struct. Expr.* *740*, 212–219.

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.

Gheldof, N., Tabuchi, T.M., and Dekker, J. (2006). The active FMR1 promoter is associated with a large domain of altered chromatin conformation with embedded local histone modifications. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 12463–12468.

Globisch, D., Münzel, M., Müller, M., Michalakakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M., and Carell, T. (2010). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* *5*, e15367.

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^{Asp} by the DNA methyltransferase homolog Dnmt2. *Science* *311*, 395–398.

Greer, E.L., and Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* *13*, 343–357.

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.

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. (2011). 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. *Development* *140*, 562–572.

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, 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.

- Hahn, M.A., Wu, X., Li, A.X., Hahn, T., and Pfeifer, G.P. (2011). Relationship between Gene Body DNA Methylation and Intragenic H3K9me3 and H3K36me3 Chromatin Marks. *PLoS ONE* 6.
- Hajkova, P., Ancelin, K., Waldmann, T., Lacoste, N., Lange, U.C., Cesari, F., Lee, C., Almouzni, G., Schneider, R., and Surani, M.A. (2008). Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* 452, 877–881.
- 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.R., 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.
- 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. *Dev. Camb. Engl.* 129, 1983–1993.
- 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.
- Hellman, A., and Chess, A. (2007). Gene body-specific methylation on the active X chromosome. *Science* 315, 1141–1143.
- 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.
- Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* 18, 6538–6547.
- Hermann, A., Gowher, H., and Jeltsch, A. (2004). Biochemistry and biology of mammalian DNA methyltransferases. *Cell. Mol. Life Sci. CMLS* 61, 2571–2587.
- 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.
- Holliday, R., and Pugh, J.E. (1975). DNA modification mechanisms and gene activity during development. *Science* 187, 226–232.
- Holz-Schietinger, C., and Reich, N.O. (2010). The Inherent Processivity of the Human de Novo Methyltransferase 3A (DNMT3A) Is Enhanced by DNMT3L. *J. Biol. Chem.* 285, 29091–29100.
- Hon, G.C., Rajagopal, N., Shen, Y., McCleary, D.F., Yue, F., Dang, M.D., and Ren, B. (2013a). Adult tissue methylomes harbor epigenetic memory at embryonic enhancers. *Nat. Genet.* 45, 1198–1206.
- Hon, G.C., Rajagopal, N., Shen, Y., McCleary, D.F., Yue, F., Dang, M.D., and Ren, B. (2013b). Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat. Genet.* 45, 1198–1206.

- Hotchkiss, R.D. (1948). The Quantitative Separation of Purines, Pyrimidines, and Nucleosides by Paper Chromatography. *J. Biol. Chem.* *175*, 315–332.
- Houlden, H., King, R., Blake, J., Groves, M., Love, S., Woodward, C., Hammans, S., Nicoll, J., Lennox, G., O'Donovan, D.G., et al. (2006). Clinical, pathological and genetic characterization of hereditary sensory and autonomic neuropathy type 1 (HSAN I). *Brain* *129*, 411–425.
- Howlett, S.K., and Reik, W. (1991). Methylation levels of maternal and paternal genomes during preimplantation. *Development* *113*, 119–127.
- Hu, X., Zhang, L., Mao, S.-Q., Li, Z., Chen, J., Zhang, R.-R., Wu, H.-P., Gao, J., Guo, F., Liu, W., et al. (2014). Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. *Cell Stem Cell* *14*, 512–522.
- Hudler, P. (2012). Genetic aspects of gastric cancer instability. *ScientificWorldJournal* *2012*, 761909.
- Ikegami, K., Iwatani, M., Suzuki, M., Tachibana, M., Shinkai, Y., Tanaka, S., Grealley, J.M., Yagi, S., Hattori, N., and Shiota, K. (2007). Genome-wide and locus-specific DNA hypomethylation in G9a deficient mouse embryonic stem cells. *Genes Cells Devoted Mol. Cell. Mech.* *12*, 1–11.
- Illingworth, R., Kerr, A., DeSousa, D., Jørgensen, 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*.
- Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R.W., 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*, e1001134.
- Inoue, A., Shen, L., Matoba, S., and Zhang, Y. (2015). Haploinsufficiency, but Not Defective Paternal 5mC Oxidation, Accounts for the Developmental Defects of Maternal Tet3 Knockouts. *Cell Rep.* *10*, 463–470.
- Iqbal, K., Jin, S.-G., Pfeifer, G.P., and Szabó, P.E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc. Natl. Acad. Sci.* *108*, 3642–3647.
- Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., et al. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* *41*, 178–186.
- Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* *416*, 556–560.
- 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.
- Johnson, T.B., and Coghill, R.D. (1925). RESEARCHES ON PYRIMIDINES. C111. THE DISCOVERY OF 5-METHYL-CYTOSINE IN TUBERCULINIC ACID, THE NUCLEIC ACID OF THE TUBERCLE BACILLUS1. *J. Am. Chem. Soc.* *47*, 2838–2844.
- Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* *13*, 484–492.

- Jones, P.A., and Liang, G. (2009). Rethinking how DNA Methylation Patterns are Maintained. *Nat. Rev. Genet.* *10*, 805–811.
- Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., and Wolffe, A.P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* *19*, 187–191.
- Kagiwada, 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, 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.
- 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.
- Kerkel, K., Spadola, A., Yuan, E., Kosek, J., Jiang, L., Hod, E., Li, K., Murty, V.V., Schupf, N., Vilain, E., et al. (2008). Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat. Genet.* *40*, 904–908.
- 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.
- 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.
- Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* *31*, 89–97.
- 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*.
- Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X.S., and Ahringer, J. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* *41*, 376–381.
- Kouzarides, T. (2002). Histone methylation in transcriptional control. *Curr. Opin. Genet. Dev.* *12*, 198–209.
- 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*, e04094.
- Kremer, E.J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S.T., Schlessinger, D., Sutherland, G.R., and Richards, R.I. (1991). Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)_n. *Science* *252*, 1711–1714.

- Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930.
- Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008). Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. *PLoS Genet.* 4.
- Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W., et al. (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev.* 22, 908–917.
- 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. *Genes. N. Y. N* 2000 35, 88–93.
- Langemeijer, S.M.C., Jansen, J.H., Hooijer, J., van Hoogen, P., Stevens-Linders, E., Massop, M., Waanders, E., van Reijmersdal, S.V., Stevens-Kroef, M.J.P.L., Zwaan, C.M., et al. (2011). TET2 mutations in childhood. *Leukemia* 25, 189–192.
- Lao, V.V., and Grady, W.M. (2011). Epigenetics and colorectal cancer. *Nat. Rev. Gastroenterol. Hepatol.* 8, 686–700.
- Lees-Murdock, D.J., and Walsh, C.P. (2008). DNA methylation reprogramming in the germ line. *Epigenetics* 3, 5–13.
- Lees-Murdock, D.J., De Felici, M., and Walsh, C.P. (2003). Methylation dynamics of repetitive DNA elements in the mouse germ cell lineage. *Genomics* 82, 230–237.
- Lehnertz, B., Ueda, Y., Derijck, A.A.H.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H.F.M. (2003). Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. *Curr. Biol.* 13, 1192–1200.
- 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.
- Lev Maor, G., Yearim, A., and Ast, G. (2015). The alternative role of DNA methylation in splicing regulation. *Trends Genet.* 31, 274–280.
- 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.-Y., Lees-Murdock, D.J., Xu, G.-L., and Walsh, C.P. (2004). Timing of establishment of paternal methylation imprints in the mouse. *Genomics* 84, 952–960.
- 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.
- Liang, G., Chan, M.F., Tomigahara, Y., Tsai, Y.C., Gonzales, F.A., Li, E., Laird, P.W., and Jones, P.A. (2002). Cooperativity between DNA Methyltransferases in the Maintenance Methylation of Repetitive Elements. *Mol. Cell. Biol.* 22, 480–491.

- Lindroth, A.M., Shultis, D., Jasencakova, Z., Fuchs, J., Johnson, L., Schubert, D., Patnaik, D., Pradhan, S., Goodrich, J., Schubert, I., et al. (2004). Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. *EMBO J.* *23*, 4286–4296.
- 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*, 523–536.
- Lister, R., Pelizzola, M., Dowen, 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.
- Liu, H., Kim, J.-M., and Aoki, F. (2004). Regulation of histone H3 lysine 9 methylation in oocytes and early pre-implantation embryos. *Dev. Camb. Engl.* *131*, 2269–2280.
- Liu, X., Gao, Q., Li, P., Zhao, Q., Zhang, J., Li, J., Koseki, H., and Wong, J. (2013). UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nat. Commun.* *4*, 1563.
- Loh, M., Liem, N., Vaithilingam, A., Lim, P.L., Sapari, N.S., Elahi, E., Mok, Z.Y., Cheng, C.L., Yan, B., Pang, B., et al. (2014). DNA methylation subgroups and the CpG island methylator phenotype in gastric cancer: a comprehensive profiling approach. *BMC Gastroenterol.* *14*, 55.
- 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*.
- Lu, H., and Hallstrom, T.C. (2013). The Nuclear Protein UHRF2 Is a Direct Target of the Transcription Factor E2F1 in the Induction of Apoptosis. *J. Biol. Chem.* *288*, 23833–23843.
- Lu, F., Liu, Y., Jiang, L., Yamaguchi, S., and Zhang, Y. (2014). Role of Tet proteins in enhancer activity and telomere elongation. *Genes Dev.* *28*, 2103–2119.
- Lucifero, D., Mann, M.R.W., Bartolomei, M.S., and Trasler, J.M. (2004). Gene-specific timing and epigenetic memory in oocyte imprinting. *Hum. Mol. Genet.* *13*, 839–849.
- 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., Portela, A., Liz, J., Melo, S.A., Rossi, S., Spizzo, R., Croce, C.M., Calin, G.A., and Esteller, M. (2010). CpG island hypermethylation-associated silencing of non-coding RNAs transcribed from ultraconserved regions in human cancer. *Oncogene* *29*, 6390–6401.
- Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., and Maleszka, R. (2010). The Honey Bee Epigenomes: Differential Methylation of Brain DNA in Queens and Workers. *PLoS Biol.* *8*.
- Magnúsdóttir, E., Dietmann, S., Murakami, K., Günesdogan, U., Tang, F., Bao, S., Diamanti, E., Lao, K., Gottgens, B., and Surani, M.A. (2013). A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nat. Cell Biol.* *15*, 905–915.

- 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.
- Mancini-DiNardo, D., Steele, S.J.S., Levorse, J.M., Ingram, R.S., and Tilghman, S.M. (2006). Elongation of the *Kcnq1ot1* transcript is required for genomic imprinting of neighboring genes. *Genes Dev.* *20*, 1268–1282.
- Maresca, A., Zaffagnini, M., Caporali, L., Carelli, V., and Zanna, C. (2015). DNA methyltransferase 1 mutations and mitochondrial pathology: is mtDNA methylated? *Front. Genet.* *6*, 1–15.
- Marin, M., Karis, A., Visser, P., Grosveld, F., and Philipsen, S. (1997). Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. *Cell* *89*, 619–628.
- 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.
- Matzke, M.A., and Mosher, R.A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* *15*, 394–408.
- Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., et al. (2010). Conserved Role of Intragenic DNA Methylation in Regulating Alternative Promoters. *Nature* *466*, 253–257.
- McGrath, J., and Solter, D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* *37*, 179–183.
- Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* *33*, 5868–5877.
- 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.
- Métivier, R., Gallais, R., Tiffoche, C., Le Péron, C., Jurkowska, R.Z., Carmouche, R.P., Ibberson, D., Barath, P., Demay, F., Reid, G., et al. (2008). Cyclical DNA methylation of a transcriptionally active promoter. *Nature* *452*, 45–50.
- Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* *448*, 553–560.
- Mitchell, S.M., Ross, J.P., Drew, H.R., Ho, T., Brown, G.S., Saunders, N.F.W., Duesing, K.R., Buckley, M.J., Dunne, R., Beetson, I., et al. (2014). A panel of genes methylated with high frequency in colorectal cancer. *BMC Cancer* *14*, 54.
- 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.

- 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.
- Mozzetta, C., Pontis, J., Fritsch, L., Robin, P., Portoso, M., Proux, C., Margueron, R., and Ait-Si-Ali, S. (2014). The Histone H3 Lysine 9 Methyltransferases G9a and GLP Regulate Polycomb Repressive Complex 2-Mediated Gene Silencing. *Mol. Cell* 53, 277–289.
- Myant, K., Termanis, A., Sundaram, A.Y.M., Boe, T., Li, C., Merusi, C., Burrage, J., de Las Heras, J.I., and Stancheva, I. (2011). LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. *Genome Res.* 21, 83–94.
- Naito, T., Noshio, K., Ito, M., Igarashi, H., Mitsushashi, K., Yoshii, S., Aoki, H., Nomura, M., Sukawa, Y., Yamamoto, E., et al. (2014). IGF2 differentially methylated region hypomethylation in relation to pathological and molecular features of serrated lesions. *World J. Gastroenterol.* WJG 20, 10050–10061.
- 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.
- Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Pagnani, A., Zecchina, R., Parlato, C., and Oliviero, S. (2013). Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells. *Genome Biol.* 14, R91.
- Neumann, B., and Barlow, D.P. (1996). Multiple roles for DNA methylation in gametic imprinting. *Curr. Opin. Genet. Dev.* 6, 159–163.
- 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.
- Nitta, H., Unoki, M., Ichiyanagi, K., Kosho, T., Shigemura, T., Takahashi, H., Velasco, G., Francastel, C., Picard, C., Kubota, T., et al. (2013). Three novel ZBTB24 mutations identified in Japanese and Cape Verdean type 2 ICF syndrome patients. *J. Hum. Genet.* 58, 455–460.
- Oda, M., Yamagiwa, A., Yamamoto, S., Nakayama, T., Tsumura, A., Sasaki, H., Nakao, K., Li, E., and Okano, M. (2006). DNA methylation regulates long-range gene silencing of an X-linked homeobox gene cluster in a lineage-specific manner. *Genes Dev.* 20, 3382–3394.
- Ogawa, H., Ishiguro, K.-I., Gaubatz, S., Livingston, D.M., and Nakatani, Y. (2002). A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science* 296, 1132–1136.
- Ohinata, Y., Payer, B., O’Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S.C., Obukhanych, T., Nussenzweig, M., Tarakhovskiy, A., et al. (2005). Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* 436, 207–213.

- Ohki, I., Shimotake, N., Fujita, N., Jee, J.-G., Ikegami, T., Nakao, M., and Shirakawa, M. (2001). Solution Structure of the Methyl-CpG Binding Domain of Human MBD1 in Complex with Methylated DNA. *Cell* *105*, 487–497.
- Ohno, R., Nakayama, M., Naruse, C., Okashita, N., Takano, O., Tachibana, M., Asano, M., Saitou, M., and Seki, Y. (2013). A replication-dependent passive mechanism modulates DNA demethylation in mouse primordial germ cells. *Development* *140*, 2892–2903.
- Okano, M., Xie, S., and Li, E. (1998). 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., 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.
- 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.* *CB 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.
- Papait, R., Pistore, C., Negri, D., Pecoraro, D., Cantarini, L., and Bonapace, I.M. (2007). Np95 Is Implicated in Pericentromeric Heterochromatin Replication and in Major Satellite Silencing. *Mol. Biol. Cell* *18*, 1098–1106.
- Park, J.K., Ryu, J.K., Yoon, W.J., Lee, S.H., Lee, G.Y., Jeong, K.-S., Kim, Y.-T., and Yoon, Y.B. (2012). The role of quantitative NPTX2 hypermethylation as a novel serum diagnostic marker in pancreatic cancer. *Pancreas* *41*, 95–101.
- Patnaik, D., Chin, H.G., Estève, P.-O., Benner, J., Jacobsen, S.E., and Pradhan, S. (2004). Substrate Specificity and Kinetic Mechanism of Mammalian G9a Histone H3 Methyltransferase. *J. Biol. Chem.* *279*, 53248–53258.
- Peat, J.R., Dean, W., Clark, S.J., Krueger, F., Smallwood, S.A., Ficiz, 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.
- Penn, N.W., Suwalski, R., O’Riley, C., Bojanowski, K., and Yura, R. (1972). The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem. J.* *126*, 781–790.
- Peters, A.H.F.M., O’Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schöfer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A., et al. (2001). Loss of the Suv39h Histone Methyltransferases Impairs Mammalian Heterochromatin and Genome Stability. *Cell* *107*, 323–337.
- Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. *Nat. Biotechnol.* *28*, 1057–1068.
- Prissette, 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.

- 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.
- Raddatz, G., Guzzardo, P.M., Olova, N., Fantappiè, M.R., Rampp, M., Schaefer, M., Reik, W., Hannon, G.J., and Lyko, F. (2013). Dnmt2-dependent methylomes lack defined DNA methylation patterns. *Proc. Natl. Acad. Sci. U. S. A.* 110, 8627–8631.
- Radpour, R., Kohler, C., Haghghi, M.M., Fan, A.X.C., Holzgreve, W., and Zhong, X.Y. (2009). Methylation profiles of 22 candidate genes in breast cancer using high-throughput MALDI-TOF mass array. *Oncogene* 28, 2969–2978.
- Ratel, D., Ravanat, J.-L., Berger, F., and Wion, D. (2006). N6-methyladenine: the other methylated base of DNA. *Bioessays* 28, 309–315.
- 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.
- Reik, W., and Walter, J. (2001). Genomic imprinting: parental influence on the genome. *Nat. Rev. Genet.* 2, 21–32.
- Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Non-Coding RNAs. *Cell* 129, 1311–1323.
- Rothbart, S.B., Krajewski, K., Nady, N., Tempel, W., Xue, S., Badeaux, A.I., Baryste-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.
- 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. *Development* 131, 975–982.
- Saitou, M. (2009). Germ cell specification in mice. *Curr. Opin. Genet. Dev.* 19, 386–395.
- Saitou, M., Kagiwada, S., and Kurimoto, K. (2012). Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. *Dev. Camb. Engl.* 139, 15–31.
- 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.
- Sapari, N.S., Loh, M., Vaithilingam, A., and Soong, R. (2012). Clinical Potential of DNA Methylation in Gastric Cancer: A Meta-Analysis. *PLoS ONE* 7.
- Sasaki, H., and Matsui, Y. (2008). Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat. Rev. Genet.* 9, 129–140.
- Sato, M., Kimura, T., Kurokawa, K., Fujita, Y., Abe, K., Masuhara, M., Yasunaga, T., Ryo, A., Yamamoto, M., and Nakano, T. (2002). Identification of PGC7, a new gene expressed specifically in preimplantation embryos and germ cells. *Mech. Dev.* 113, 91–94.

- 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.
- Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., and Rauscher, F.J. (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev.* *16*, 919–932.
- 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.
- 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. *Dev. Camb. Engl.* *134*, 2627–2638.
- Sekimata, M., and Homma, Y. (2004). Sequence-specific transcriptional repression by an MBD2-interacting zinc finger protein MIZF. *Nucleic Acids Res.* *32*, 590–597.
- Sequeira-Mendes, J., Díaz-Uriarte, R., Apedaile, A., Huntley, D., Brockdorff, N., and Gómez, M. (2009). Transcription Initiation Activity Sets Replication Origin Efficiency in Mammalian Cells. *PLoS Genet.* *5*.
- Sérandour, A.A., Avner, S., Percevault, F., Demay, F., Bizot, M., Lucchetti-Miganeh, C., Barloy-Hubler, F., Brown, M., Lupien, M., Métivier, R., et al. (2011). Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res.* *21*, 555–565.
- 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, A., 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.
- Shinkai, Y., and Tachibana, M. (2011). H3K9 methyltransferase G9a and the related molecule GLP. *Genes Dev.* *25*, 781–788.
- 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*.

- Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479, 74–79.
- Skinner, M.K. (2008). What is an Epigenetic Transgenerational Phenotype? F3 or F2. *Reprod. Toxicol. Elmsford N* 25, 2–6.
- Sleutels, F., Zwart, R., and Barlow, D.P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* 415, 810–813.
- Smallwood, A., Estève, P.-O., Pradhan, S., and Carey, M. (2007). Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes Dev.* 21, 1169–1178.
- 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.
- Smeets, D.F., Moog, U., Weemaes, C.M., Vaes-Peeters, G., Merckx, G.F., Niehof, J.P., and Hamers, G. (1994). ICF syndrome: a new case and review of the literature. *Hum. Genet.* 94, 240–246.
- Smith, E.Y., Futtner, C.R., Chamberlain, S.J., Johnstone, K.A., and Resnick, J.L. (2011). Transcription is required to establish maternal imprinting at the Prader-Willi syndrome and Angelman syndrome locus. *PLoS Genet.* 7, e1002422.
- Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* 484, 339–344.
- Song, J., Rechkoblit, O., Bestor, T.H., and Patel, D.J. (2011). Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* 331, 1036–1040.
- Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W.T.C., Bauer, C., Münzel, M., Wagner, M., Müller, 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.
- Statham, A.L., Robinson, M.D., Song, J.Z., Coolen, M.W., Stirzaker, C., and Clark, S.J. (2012). Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. *Genome Res.* 22, 1120–1127.
- Storre, J., Schäfer, A., Reichert, N., Barbero, J.L., Hauser, S., Eilers, M., and Gaubatz, S. (2005). Silencing of the Meiotic Genes SMC1 β and STAG3 in Somatic Cells by E2F6. *J. Biol. Chem.* 280, 41380–41386.

- Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., and Tajima, S. (2004). DNMT3L Stimulates the DNA Methylation Activity of Dnmt3a and Dnmt3b through a Direct Interaction. *J. Biol. Chem.* *279*, 27816–27823.
- Sun, Z., Wu, Y., Ordog, T., Baheti, S., Nie, J., Duan, X., Hojo, K., Kocher, J.-P., Dyck, P.J., and Klein, C.J. (2014). Aberrant signature methylome by DNMT1 hot spot mutation in hereditary sensory and autonomic neuropathy 1E. *Epigenetics* *9*, 1184–1193.
- 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.
- Tachibana, M., Sugimoto, K., Fukushima, T., and Shinkai, Y. (2001). SET Domain-containing Protein, G9a, Is a Novel Lysine-preferring Mammalian Histone Methyltransferase with Hyperactivity and Specific Selectivity to Lysines 9 and 27 of Histone H3. *J. Biol. Chem.* *276*, 25309–25317.
- Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H., et al. (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev.* *16*, 1779–1791.
- Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T., and Shinkai, Y. (2005). Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes Dev.* *19*, 815–826.
- Tachibana, M., Matsumura, Y., Fukuda, M., Kimura, H., and Shinkai, Y. (2008). G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription. *EMBO J.* *27*, 2681–2690.
- 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.
- Tamaru, H., and Selker, E.U. (2001). A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. *Nature* *414*, 277–283.
- Tamaru, H., Zhang, X., McMillen, D., Singh, P.B., Nakayama, J., Grewal, S.I., Allis, C.D., Cheng, X., and Selker, E.U. (2003). Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in *Neurospora crassa*. *Nat. Genet.* *34*, 75–79.
- Thijssen, P.E., Ito, Y., Grillo, G., Wang, J., Velasco, G., Nitta, H., Unoki, M., Yoshihara, M., Suyama, M., Sun, Y., et al. (2015). Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome. *Nat. Commun.* *6*.
- Timp, W., and Feinberg, A.P. (2013). Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. *Nat. Rev. Cancer* *13*, 497–510.
- 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.* *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.* e201591382.
- Ueda, J., Tachibana, M., Ikura, T., and Shinkai, Y. (2006). Zinc Finger Protein Wiz Links G9a/GLP Histone Methyltransferases to the Co-repressor Molecule CtBP. *J. Biol. Chem.* *281*, 20120–20128.
- Valinluck, V., and Sowers, L.C. (2007). Endogenous Cytosine Damage Products Alter the Site Selectivity of Human DNA Maintenance Methyltransferase DNMT1. *Cancer Res.* *67*, 946–950.
- Varley, K.E., Gertz, J., Bowling, K.M., Parker, S.L., Reddy, T.E., Pauli-Behn, F., Cross, M.K., Williams, B.A., Stamatoyannopoulos, J.A., Crawford, G.E., et al. (2013). Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res.* *23*, 555–567.
- Velasco, G., Hubé, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-Péquignot, E., and Francastel, C. (2010). Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 9281–9286.
- Vilkaitis, G., Suetake, I., Klimašauskas, S., and Tajima, S. (2005). Processive Methylation of Hemimethylated CpG Sites by Mouse Dnmt1 DNA Methyltransferase. *J. Biol. Chem.* *280*, 64–72.
- Viré, E.C.R.L.M.C.L.E.A.D.J.-M.M.M.C.L.L.Y.F. (2006). The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* *439*, 871–874.
- Waddington, C.H. (2012). The epigenotype. 1942. *Int. J. Epidemiol.* *41*, 10–13.
- Wagschal, A., Sutherland, H.G., Woodfine, K., Henckel, A., Chebli, K., Schulz, R., Oakey, R.J., Bickmore, W.A., and Feil, R. (2008). G9a Histone Methyltransferase Contributes to Imprinting in the Mouse Placenta. *Mol. Cell. Biol.* *28*, 1104–1113.
- Walter, M.J., Ding, L., Shen, D., Shao, J., Grillot, M., McLellan, M., Fulton, R., Schmidt, H., Kalicki-Veizer, J., O’Laughlin, M., et al. (2011). Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. *Leukemia* *25*, 1153–1158.
- Wan, M., Lee, S.S., Zhang, X., Houwink-Manville, I., Song, H.R., Amir, R.E., Budden, S., Naidu, S., Pereira, J.L., Lo, I.F., et al. (1999). Rett syndrome and beyond: recurrent spontaneous and familial MECP2 mutations at CpG hotspots. *Am. J. Hum. Genet.* *65*, 1520–1529.
- Waterland, R.A., Kellermayer, R., Rached, M.-T., Tatevian, N., Gomes, M.V., Zhang, J., Zhang, L., Chakravarty, A., Zhu, W., Laritsky, E., et al. (2009). Epigenomic profiling indicates a role for DNA methylation in early postnatal liver development. *Hum. Mol. Genet.* *18*, 3026–3038.
- Weber, A., Marquardt, J., Elzi, D., Forster, N., Starke, S., Glaum, A., Yamada, D., Defossez, P.-A., Delrow, J., Eisenman, R.N., et al. (2008). Zbtb4 represses transcription of P21CIP1 and controls the cellular response to p53 activation. *EMBO J.* *27*, 1563–1574.

- Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schübeler, 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., Pääbo, S., Rebhan, M., and Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* *39*, 457–466.
- Willard, H.F., and Hendrich, B.D. (1999). Breaking the silence in Rett syndrome. *Nat. Genet.* *23*, 127–128.
- Wong, E., Yang, K., Kuraguchi, M., Werling, U., Avdievich, E., Fan, K., Fazzari, M., Jin, B., Brown, A.M.C., Lipkin, M., et al. (2002). Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 14937–14942.
- Wong, E.M., Southey, M.C., Fox, S.B., Brown, M.A., Dowty, J.G., Jenkins, M.A., Giles, G.G., Hopper, J.L., and Dobrovic, A. (2011). Constitutional methylation of the BRCA1 promoter is specifically associated with BRCA1 mutation-associated pathology in early-onset breast cancer. *Cancer Prev. Res. Phila. Pa* *4*, 23–33.
- Woodcock, C.L., and Ghosh, R.P. (2010). Chromatin higher-order structure and dynamics. *Cold Spring Harb. Perspect. Biol.* *2*, a000596.
- Xie, W., Barr, C.L., Kim, A., Yue, F., Lee, A.Y., Eubanks, J., Dempster, E.L., and Ren, B. (2012). Base-resolution Analyses of Sequence and Parent-of-Origin Dependent DNA Methylation in the Mouse Genome. *Cell* *148*, 816–831.
- Xin, Z., Tachibana, M., Guggiari, M., Heard, E., Shinkai, Y., and Wagstaff, J. (2003). Role of Histone Methyltransferase G9a in CpG Methylation of the Prader-Willi Syndrome Imprinting Center. *J. Biol. Chem.* *278*, 14996–15000.
- 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-Péquignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* *402*, 187–191.
- 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.
- Yamaji, M., Seki, Y., Kurimoto, K., Yabuta, Y., Yuasa, M., Shigeta, M., Yamanaka, K., Ohinata, Y., and Saitou, M. (2008). Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nat. Genet.* *40*, 1016–1022.
- 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.
- Yoon, H.-G., Chan, D.W., Reynolds, A.B., Qin, J., and Wong, J. (2003). N-CoR Mediates DNA Methylation-Dependent Repression through a Methyl CpG Binding Protein Kaiso. *Mol. Cell* *12*, 723–734.

Yu, M., Hon, G.C., Szulwach, K.E., Song, C.-X., Zhang, L., Kim, A., Li, X., Dai, Q., Park, B., Min, J.-H., et al. (2012). Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome. *Cell* *149*, 1368–1380.

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, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev.* *13*, 1924–1935.

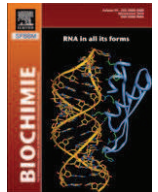
Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W., et al. (2010). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Res.* *38*, 4246–4253.

Ziller, M.J., Müller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., Boyle, P., Epstein, C.B., Bernstein, B.E., Lengauer, T., et al. (2011). Genomic Distribution and Inter-Sample Variation of Non-CpG Methylation across Human Cell Types. *PLoS Genet.* *7*.



Contents lists available at SciVerse ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi

Review

Mechanisms of DNA methylation and demethylation in mammals

Ghislain Auclair, Michael Weber*

UMR 7242, Biotechnology and Cell Signalling, Université de Strasbourg, CNRS, ESBS, Bd Sébastien Brant, BP 10413, 67412 Illkirch Cedex, France

ARTICLE INFO

Article history:

Received 15 February 2012

Accepted 15 May 2012

Available online xxx

Keywords:

DNA methylation

Demethylation

TET

DNMT

Reprogramming

CpG island

ABSTRACT

Cytosine methylation is an epigenetically propagated DNA modification that can modify how the DNA molecule is recognized and expressed. DNA methylation undergoes extensive reprogramming during mammalian embryogenesis and is directly linked to the regulation of pluripotency and cellular identity. Studying its regulation is also important for a better understanding of the many diseases that show epigenetic deregulations, in particular, cancer. In the recent years, a lot of progress has been made to characterize the profiles of DNA methylation at the genome level, which revealed that patterns of DNA methylation are highly dynamic between cell types. Here, we discuss the importance of DNA methylation for genome regulation and the mechanisms that remodel the DNA methylome during mammalian development, in particular the involvement of the rediscovered modified base 5-hydroxymethylcytosine.

© 2012 Published by Elsevier Masson SAS.

1. Introduction

The regulation of complex eukaryote genomes entails not only sequence-specific DNA-binding factors, but also additional levels of regulation such as DNA modifications, histone post-translational modifications and chromatin remodeling. These modifications are often referred to as being “epigenetic” although some of them have not been shown to fulfill the strict definition of epigenetics, which implies a heritability through mitosis or meiosis [1]. During development and cellular differentiation, these epigenetic marks undergo dynamic changes that ultimately contribute to produce and maintain distinct cell types of an organism [2]. Elucidating how these epigenetic marks participate in the regulation of cellular identity is crucial to better understand embryonic development and the etiology of many diseases.

Methylation is the most abundant epigenetic modification that directly affects the DNA molecule in eukaryotes. It consists in the addition of a methyl group on the carbon 5 of the cytosine, thereby creating 5-methylcytosine (5mC), and is catalyzed by enzymes of the DNA methyltransferase (DNMT) family. In mammals, it almost exclusively occurs in the context of CG dinucleotides, which are infrequent in the genome except at short DNA sequences termed ‘CpG islands’ that typically contain around 5–10 CpGs per 100 bp and often colocalize with gene promoters and regulatory regions. Of note, cytosine methylation mapping experiments at base pair

resolution revealed that methylation can also occur at non-CpG sites (predominantly CpA methylation) in oocytes [3–5] and adult brain [6], as well as ES cells [7–9], albeit at much lower frequencies than CpG methylation. It is currently unknown why this non-CpG methylation is more frequent in oocytes and brain, and if it plays any functional role.

Cytosine methylation is generally viewed as a repressive mark that inhibits transcriptional initiation, either by preventing the binding of certain transcription factors, or by recruiting methyl-binding proteins (MBPs) and generating a repressed chromatin environment. These patterns of DNA methylation can be stably propagated during cell division, which makes it a paradigm for epigenetic regulation that can mediate long lasting changes in gene expression even when the initial triggering signal has disappeared. In addition, profiles of DNA methylation are perturbed in many diseases, in particular all types of cancer that show both genome-wide hypomethylation and aberrant hypermethylation of tumor-suppressor genes or non-coding RNAs [10–12]. An illustration of the key roles of DNA methylation in cancer is that mutations in some of the enzymes that methylate (DNMT3A) and hydroxymethylate (TET2) DNA are frequent in leukemia [13,14].

Cytosine methylation is found in many species of the plant, fungi and animal kingdoms, but has apparently undergone a very complex evolutionary history. In fungi for example, cytosine methylation is absent in *Saccharomyces cerevisiae* but present in other species such as *Neurospora crassa*. Similarly, in insects, cytosine methylation is absent in *Drosophila melanogaster* but present in the honey bee *Apis mellifera* [15]. In addition, the distribution of methylated cytosines within the genome also varies between

* Corresponding author. Tel.: +33 3 68 85 44 35.

E-mail address: michael.weber@unistra.fr (M. Weber).

organisms. Data from many organisms indicate that there is a gradual transition from partial to global genome methylation around the time of emergence of vertebrates [16,17]. In invertebrates, fungi and plants, the most common pattern is methylation-free domains interspersed with heavily methylated sequences that correspond to repetitive mobile elements and gene bodies, whereas all vertebrates show globally high levels of CpG methylation in all genome compartments except CpG islands. In this review, we will focus on the distribution and regulation of DNA methylation in mammalian genomes, with particular emphasis on recent discoveries that help to clarify the mechanisms that establish or erase cytosine methylation during development and cellular differentiation.

2. DNA methylation in mammalian genomes

2.1. Genome-wide distribution of 5mC

A critical step toward understanding the role of DNA methylation is to determine where 5mC is in the genome. This is now feasible with the recent development of methods for genome-wide mapping of 5mC using microarrays or high-throughput sequencing [18]. Some methods isolate methylated or hypomethylated DNA by digestion with methylation-sensitive or resistant enzymes, with the caveat that they are biased toward certain restriction motifs. Another approach is to sonicate DNA and enrich for methylated fragments by immunoprecipitation with a monoclonal anti-5 methylcytosine antibody (MeDIP for "Methylated DNA Immunoprecipitation") [19], which can be combined with high-throughput sequencing for genome-wide maps at moderate resolution (~250 bp). The gold standard approach relies on chemical treatment with sodium bisulfite, which converts unmethylated cytosines into uracils while leaving methylated cytosines unchanged. After PCR replacement of uracils by thymines, the methylation status of single cytosines can be deduced from alignment against the reference sequence. This technique has the advantage of generating information at a single-nucleotide resolution, but one limitation is that it requires tremendous sequencing efforts if applied genome-wide on large mammalian genomes.

These large-scale methylome data indicate that cytosine methylation is present throughout mammalian genomes (in intergenic regions, coding regions, mobile elements and certain promoters) and revealed an inverse correlation between cytosine methylation and CpG density: CpG-poor DNA, which comprises most of the genome, shows high levels of cytosine methylation, whereas CpG islands remain mostly unmethylated [7,20]. There are however rare methylated CpG islands in somatic cells. Some are associated with allele-specific gene silencing in the context of parental genomic imprinting or on the inactive copy of the X chromosome in female cells. Recent mapping experiments of promoter DNA methylation also revealed rare methylated CpG islands (representing around 3% of the total number of promoter CpG islands) in the promoters of important developmental genes such as germline, pluripotency and *Hox* genes [21–24], suggesting that promoter DNA methylation restricts pluripotency of differentiated cells. This hypothesis is further supported by the observation that dedifferentiation of somatic cells into pluripotent cells entails demethylation of pluripotency genes and can be improved by interfering with DNA methylation [25,26].

A role of DNA methylation at non-promoter sequences is also emerging. Indeed, several studies reported that cell type-specific DNA methylation profiles vary more frequently at intergenic sequences compared to annotated gene promoters. These sites of differential methylation might regulate the activity of distant

enhancers [9,27,28] or the transcription of non-coding RNAs and uncharacterized transcripts [12,29]. Another intriguing observation is that DNA methylation levels within gene bodies are also dynamic in relation to gene expression. In general, higher levels of intragenic methylation correlate with higher expression levels [30,31], even though it is not yet clear if there is a causal relationship. One proposed model is that intragenic DNA methylation could modulate alternative transcription initiation sites within gene bodies [32]. Another possible role of intragenic DNA methylation is to modulate splicing events. Several studies have observed a possible crosstalk between DNA methylation and splicing in mammalian cells [33,34], as well as in the honey bee [35]. In addition, it has been observed that DNA methylation in mammals is increased at exons relative to introns [31,36], leading to the provocative hypothesis that the differential epigenetic marking of DNA participates in the process of exon definition [37]. In terms of mechanisms, we do not yet have a good understanding of how DNA methylation could impact post-transcriptional RNA processing, but there is indeed increasing evidence that transcription and RNA processing can be tightly coupled [37]. One proposed model is that DNA methylation could modulate the binding of DNA-binding factors such as CTCF, which could induce a local pausing of the RNA polymerase and favor the co-transcriptional assembly of the spliceosome at splice sites [34].

2.2. The DNA methyltransferase family

Cytosine methylation is catalyzed by enzymes of the DNA methyltransferase (DNMT) family that comprises three active members in mammals sharing a conserved catalytic domain. The founding member, DNMT1, 0 is responsible for maintaining DNA methylation during DNA replication. Replication creates sites of hemimethylated DNA, which are the preferential substrates for DNMT1 that copies patterns of CG methylation onto the newly synthesized DNA strand. This is made possible because CG sites are palindromic and provides an example of epigenetic mark being faithfully copied during cell division. Recent structural data revealed how DNMT1 prefers hemimethylated over unmethylated DNA by showing that binding of DNMT1 to hemimethylated CpGs makes DNA accessible for the catalytic domains, whereas binding to unmethylated CpGs protects the nascent strand from methylation by an autoinhibitory conformational change mediated by the CXXC domain [38,39]. DNMT1 localizes at sites of DNA replication during S-phase via interactions with PCNA, as well as UHRF1 (also known as NP95), a factor that binds hemimethylated CGs and is required to recruit DNMT1 at sites of hemimethylation [40,41]. The role of DNMT1 in the maintenance of DNA methylation *in vivo* is evidenced by the fact that a reduction of DNMT1 in mouse preimplantation embryos leads to a failure to maintain gamete-derived DNA methylation at imprinted loci [42,43].

DNMT3A and DNMT3B enzymes are responsible for *de novo* DNA methylation during development [44]. Both enzymes are highly homologous but possess distinct target specificities and expression patterns [45,46]. DNMT3B is more prevalent in early embryonic stages and is the main enzyme responsible for the acquisition of DNA methylation during implantation [24], whereas DNMT3A is expressed in later embryonic stages and differentiated cells. In particular, DNMT3A is responsible for setting up DNA methylation patterns in maturing gametes [45,47,48] and has also been implicated in the epigenetic control of postnatal somatic stem cells [49,50]. In accordance with their function at distinct times of development, mouse knockout models have shown that *Dnmt3b*^{-/-} mice die during early embryonic development whereas *Dnmt3a*^{-/-} mice survive until 4 weeks after birth [44]. Besides their action as *de novo* methyltransferases, it has also been

proposed that DNMT3A and DNMT3B participate in the maintenance by methylated sites that were missed by DNMT1 [51].

Another protein, DNMT3L, is related in sequence to DNMT3A/B but lacks a functional catalytic domain. DNMT3L is abundant in early embryos and germ cells and its absence, despite the lack of a functional methyltransferase domain, leads to a failure to establish DNA methylation in both the male and female germline at most genomic targets, including the imprinting control regions (ICRs) and several repetitive elements [45,48,52–55]. Biochemical and structural studies have shown that DNMT3L is in fact a cofactor that physically interacts with the DNMT3 enzymes, stimulates their *de novo* methylation activity and promotes their recruitment to chromatin [56,57].

3. Reprogramming of DNA methylation during development

3.1. DNA methylation reprogramming in preimplantation embryos

Genomic DNA methylation is markedly reprogrammed during early embryonic development in mammals. Following fertilization in mouse, levels of cytosine methylation globally decrease from the first cleavage stages up to the early blastocyst, before being re-acquired during and after implantation [58] (Fig. 1). Immunofluorescence studies have shown that this global demethylation occurs asymmetrically on both parental genomes: the paternal DNA rapidly loses 5mC signal in the zygote before DNA replication, probably by a process involving oxidation by the TET proteins (see below), whereas the maternal DNA is thought to be demethylated over several cell divisions by a replication-dependent passive demethylation (reviewed in Ref. [58]). Recently, the first comprehensive maps of DNA methylation by MeDIP and bisulfite sequencing in mouse preimplantation embryos confirmed that blastocysts have a globally demethylated genome [4,24,48,55]. Notable exceptions are imprinting control regions (ICRs) that faithfully maintain allele-specific patterns of DNA methylation after fertilization and direct the monoallelic expression of associated imprinted genes in embryonic development. Interestingly, a recent study carefully examined the profiles of allelic methylation at imprinted loci in gametes and preimplantation embryos and revealed that the regions of differential methylation can vary

between gametes and early embryos [3], indicating that certain gametic ICRs are not entirely protected from epigenetic reprogramming during early embryonic development. Certain classes of mobile elements such as intracisternal A-particles (IAPs) are also partially resistant to methylation erasure in preimplantation embryos [59], which might help to prevent their transient activation.

What is the functional importance of global demethylation after fertilization? One proposed model is that it may facilitate the activation of the pluripotency program in the embryo [60], which is supported by the observation that improper paternal demethylation (in a model of embryos lacking the TET3 protein, see Section 4.4) is associated with a delay in the activation of certain pluripotency genes and an increased incidence of developmental failures [61]. Yet major points of uncertainty remain, among which the fact that there are still debates on the conservation of the rapid demethylation in the male pronucleus immediately following fertilization in certain other mammalian species such as sheep or rabbit [62–65]. Recent studies also lead to reevaluate the extent of methylation erasure in mouse preimplantation embryos. Indeed, the first maps of DNA methylation in mouse blastocysts revealed an unanticipated high number of non-imprinted sequences, including CpG islands and gene promoters, that maintain substantial levels of gametic methylation in preimplantation embryos [4,24,48,55], indicating that resistance to demethylation is common after fertilization. This is also supported by a recent study suggesting that epitope masking might overestimate the loss of 5mC signal by immunofluorescence in mouse preimplantation embryos [66]. Most of the identified CpG islands and gene promoters inherit DNA methylation from the oocyte, which is reminiscent of the fact that most ICRs are of maternal origin and indicates that the maternal genome is more prone to transmit functional gene methylation states to the embryo than the paternal genome. The role of this persistent methylation is not elucidated but one can speculate that it might modulate gene expression states and lineage decisions in embryos, in particular in the most early phases of development as oocyte-contributed DNA methylation tends to be resolved to a hypomethylated state in postimplantation embryos [4]. The mechanisms that confer resistance to global demethylation are also unclear but might involve DNA-binding factors that specifically

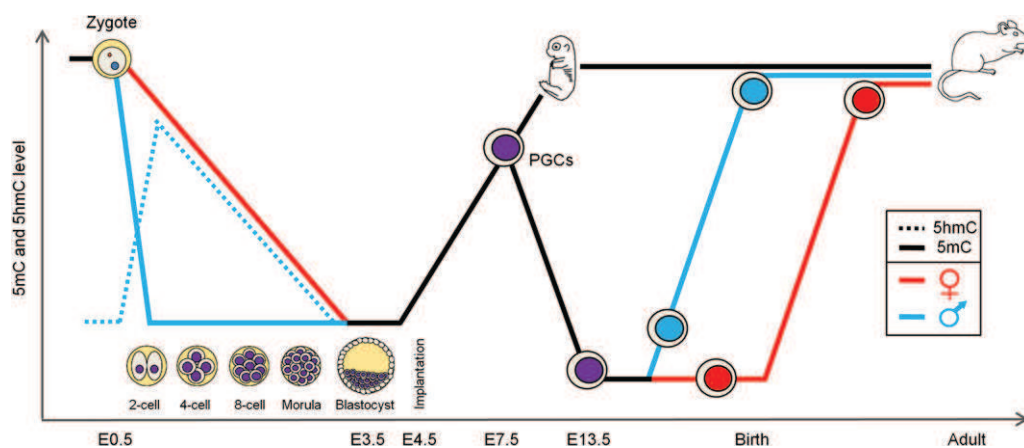


Fig. 1. Schematic representation of the waves of DNA methylation reprogramming during mouse development. After fertilization, the levels of genomic DNA methylation globally decrease to reach a low point in preimplantation blastocysts. The paternal genome rapidly loses 5-methylcytosine (5mC) in the zygote, whereas the maternal genome undergoes a replication-dependent loss of methylation during the first cleavage stages. The rapid erasure of 5mC on the paternal genome is concomitant with an appearance of 5-hydroxymethylcytosine (5hmC, represented by the blue dotted line), which is then lost by replication-dependent dilution or DNA repair. In contrast, immunostaining do not detect high levels of 5hmC on the maternal genome in preimplantation embryos. Note that many sequences partially escape methylation erasure after fertilization (see text). New patterns of DNA methylation are re-established in postimplantation embryos by the *de novo* methyltransferases DNMT3A and DNMT3B. After implantation, the cells of the epiblast generate all the cells of the embryo, including the primordial germ cells (PGCs). PGCs undergo a second wave of methylation erasure that is completed by embryonic day 13.5 (E13.5). When epigenetic reprogramming is completed, PGCs enter meiosis in females and mitotic arrest in males.

protect some sequences against demethylation, as shown recently with the zinc finger protein ZFP57 that protects ICRs from demethylation in preimplantation embryos and ES cells via binding to a hexanucleotide motif [67,68].

3.2. DNA methylation erasure in primordial germ cells

A second wave of DNA methylation reprogramming occurs in germ cells [58] (Fig. 1). After implantation, the epiblast becomes the source of all embryonic lineages, including the precursors of germ cells that are called primordial germ cells (PGCs). PGCs form a cluster of around 40 cells at embryonic day 7.25 (E7.25) and then proliferate and migrate to colonize the genital ridges by E10.5, where they continue to proliferate until E13.5 and enter meiotic prophase in females and mitotic arrest in males. PGC specification occurs in the epiblast at E6.0 in response to bone morphogenetic protein (BMP) signaling, which induces the expression of early PGC markers such as *Prdm1* (also known as *Blimp1*) at E6.25, *Prdm14* at E6.5 and *Dppa3* (also known as *PGC7* and *Stella*) at E7.25 [69–71]. These markers initiate a cascade of events in PGCs that include repression of somatic genes, reacquisition of pluripotency and initiation of epigenetic reprogramming. While PGCs initially bear an epigenetic state similar to the one of surrounding somatic cells, they undergo exchange of histone variants, loss of histone modifications, and erasure of DNA methylation during their proliferation and migration [58,72,73]. Genome-wide demethylation in PGCs is evidenced by a global loss of 5mC signal in immunofluorescence and is completed in both sexes by E13.5, soon after sex determination has taken place. Sequence-specific analyses initially revealed that it results in an erasure of methylation at imprinted loci, the *Xist* promoter and many transposable elements [59,74,75]. Sparse evidence also indicated that demethylation affects other sequences but this needs to be confirmed genome-wide. Recently, bisulfite sequencing was used to show that male and female E13.5 PGCs indeed show a marked hypomethylation compared to other tissues in all genome compartments [76]. This is corroborated by another methylation mapping study in E13.5 PGCs that shows a global erasure of DNA methylation at all gene loci [77]. The kinetics of this demethylation appear to vary depending on the target sequence. Whereas several studies reported a rapid demethylation of imprinted sequences between E10.5 and E12.5 [74,78], many other genes initiate progressive erasure of DNA methylation in earlier stages [77], which coincides with a loss of 5mC immunostaining observed when PGCs start migrating from E8.0 onwards [73]. These complex kinetics of demethylation might reflect the use of a combination of passive and active processes to achieve full demethylation in PGCs. A consequence of this global methylation erasure is that it limits the potential for transgenerational transmission of epimutations, although some regions escape methylation reprogramming such as IAPs and LTR-ERV1 mobile element, as well as rare single-copy regions [59,77].

3.3. DNA demethylation in other contexts

DNA demethylation was long thought to occur only during specific developmental windows in zygotes and PGCs, however a number of recent studies reported DNA demethylation in response to various stimuli in other cellular contexts. First, the progresses in reprogramming technologies have indicated that somatic cells have the ability to demethylate endogenous pluripotency genes during the course of reprogramming, although very inefficiently unless the cells are fused with embryonic stem (ES) cells [26,79]. Second, localized rapid demethylation events have been observed in cultured cells at gene promoters following transcriptional activation by hormones [80,81], as well as at distal

regulatory regions such as enhancers and insulators [9,28,82]. Third, demethylation events are also observed during tissue-specific differentiation of somatic cells. This is the case in the hematopoietic system where several groups reported demethylation events, including at gene promoters, during the differentiation of human or mouse hematopoietic stem cells into terminally differentiated hematopoietic cells [24,83–85]. A recent report using reduced representation bisulfite sequencing (RRBS) even showed that there is a surprising genome-scale reduction in cytosine methylation in around three cell divisions during mouse erythropoiesis *in vivo* [86]. Active demethylation has also been observed in postmitotic hematopoietic cells such as human monocytes differentiating into dendritic cells, which clearly demonstrates that demethylation can occur independently of DNA replication [87]. Besides hematopoietic cells, rapid DNA demethylation events at specific targets have also been observed at myogenic genes during muscle differentiation 20 years ago [88–90], in liver [91,92] and in neurons [93]. A striking example is the synaptic activity-dependent demethylation of the promoters of the *Bdnf* and *Fgf1* genes in postmitotic neurons [94–96]. A constant in all these studies is that demethylation seems to mostly affect sequences with moderate CpG richness, suggesting that only methylation of CpG-poor regions is susceptible to rapid erasure in somatic cells whereas methylation of CpG islands is probably only reversible in zygotes or PGCs. Altogether, these data indicate that DNA methylation patterns in mammalian genomes are much more plastic than initially anticipated and suggest that DNA demethylation can mediate rapid responses to external stimuli.

4. Mechanisms that shape mammalian DNA methylomes

A key question is to elucidate the mechanisms responsible for specifying the methylation state of a given sequence at a given time in the genome. Addressing these questions is not only important for understanding the remodeling of DNA methylation in relation to cellular identity, but also to get insights into abnormal regulation of DNA methylation in disease. The mechanisms that attract DNA methylation to specific sequences have been well dissected in plants and fungi where DNA methylation is dependent on methylation of the lysine 9 of histone H3 (H3K9) and RNAi pathways [97]. However much less is known in mammals where many of the mechanisms identified in lower organisms do not seem to play equivalent functions. Recent data indicate that in mammals, the methylation of specific targets could be achieved either by directing the recruitment of the DNA methylation machinery to specific regions, or by directing the displacement of factors that protect against the default methylation state. Progress has also been made to elucidate the long debated mechanisms of cytosine demethylation with the involvement of 5-hydroxymethylcytosine.

4.1. Role of sequence-specific DNA-binding factors

It is becoming increasingly evident that the DNA sequence itself is a major determinant of DNA methylation states. This is supported by the observation that DNA methylation states in the human genome are frequently modulated by single-nucleotide polymorphisms (SNPs) in *cis* [98,99]. Similarly in mouse, a recent comprehensive base-resolution mapping of cytosine methylation in different genetic backgrounds identified numerous cytosine methylation sites that depend on the immediate DNA sequence [6]. Recently, the influence of the underlying DNA sequence on DNA methylation patterns was functionally demonstrated in mouse ES cells. In this study, the authors repeatedly inserted DNA fragments of approximately 1 kb in the same inert genomic location, which has the advantage of eliminating the variability conferred by the

site of integration. Strikingly, they observed that most of the integrated DNA sequences mimicked the DNA methylation pattern of the endogenous sequence, demonstrating that the information that specifies DNA methylation lies in the local DNA sequence [100]. Two models can account for this dependency between the genetic and epigenetic information. The first model is that the DNMT enzymes themselves have a preference for certain DNA sequences. In line with this idea, a recent study revealed that DNMT3A and DNMT3B have preferences for certain DNA sequences flanking the CpG site [101]. However it is unlikely that these intrinsic sequence preferences alone can explain the complexity of DNA methylation patterns. The second model is that DNA sequence recruits sequence-specific DNA binding factors that in turn influence DNA methylation. A perfect example is provided by ZFP57, whose binding to the methylated hexanucleotide motif TGCCGC confers resistance to demethylation in ES cells [67,68]. Another possible scenario for a crosstalk between genetics and epigenetics is that sequence-specific transcription factors physically interact with the DNMT3 enzymes and recruit them to their target sites, as shown for example for E2F6 [102,103] (Fig. 2A). Alternatively, binding of sequence-specific factors could also protect from DNA methylation (Fig. 2B). This was suggested by early work using transgenic mice that showed that Sp1 transcription factor binding sites are necessary to maintain certain gene promoters in an unmethylated state [104], and was recently confirmed in mouse ES cells where promoters bearing mutations in several known binding sites are unable to maintain a hypomethylated state [100]. This suggests that displacement of protecting factors could be a way to achieve *de novo* DNA methylation during development. Finally, several recent studies also indicate that, unexpectedly, DNA-binding factors such as hormone receptors [80,81], FOXA1 [28], REST or CTCF [9] can lead to local demethylation of the surrounding DNA at gene promoters or enhancers (Fig. 2C), which could serve to potentiate the binding of the pioneer factor or other factors. Altogether these data indicate that DNA-binding factors play a major role in shaping the DNA methylome in mammalian cells.

4.2. Crosstalk between DNA methylation and histone modifications

Histone post-translational modifications are also candidate regulators of DNA methylation states. This is not mutually exclusive from a role of DNA-binding factors as these could function by recruiting cofactors that modify the local chromatin marks. While histone modifications have been identified as key determinants of DNA methylation in plants and fungi, the picture is more complex in mammals. First, there seem to be only a limited dependence between cytosine methylation and repressive histone modifications in mammalian genomes. Histone H3 Lysine 9 (H3K9) and Lysine 27 (H3K27) methylation have been proposed to act as signals that recruit DNA methylation in mammalian cells, however this could not yet be validated at many targets. G9a, a H3K9 methyltransferase, has been implicated in the recruitment of DNA methylation at certain genes in ES cells, but it appears independent of its catalytic activity [105–107]. Other studies suggest a possible link between Histone H3 Lysine 36 (H3K36) methylation and cytosine methylation. Indeed, both marks co-localize in mammalian genomes, predominantly in the body of genes [108,109]. In addition, it has been shown that DNMT3A possesses a domain that binds to H3K36me3, which provides a potential functional link between both marks [110]. The best evidence linking chromatin marks to DNA methylation in mammals is the inverse correlation between DNA methylation and Histone H3 Lysine 4 (H3K4) methylation. The presence of H3K4 methylation is a hallmark of unmethylated CpG islands and is the best predictor of DNA hypomethylation at the genome level [20,23]. The role of H3K4 methylation is corroborated by studies showing that DNMT3L/DNMT3A/B complexes preferentially bind chromatin when it is unmethylated on H3K4 [56,111], and that the action of the histone H3K4 demethylase KDM1B is required to establish DNA methylation at some imprinting control regions in oocytes [112]. Recent studies uncovered other factors involved in establishing a specific chromatin feature at unmethylated CpG islands. One example is CFP1, a protein that interacts with non-methylated CpGs and the

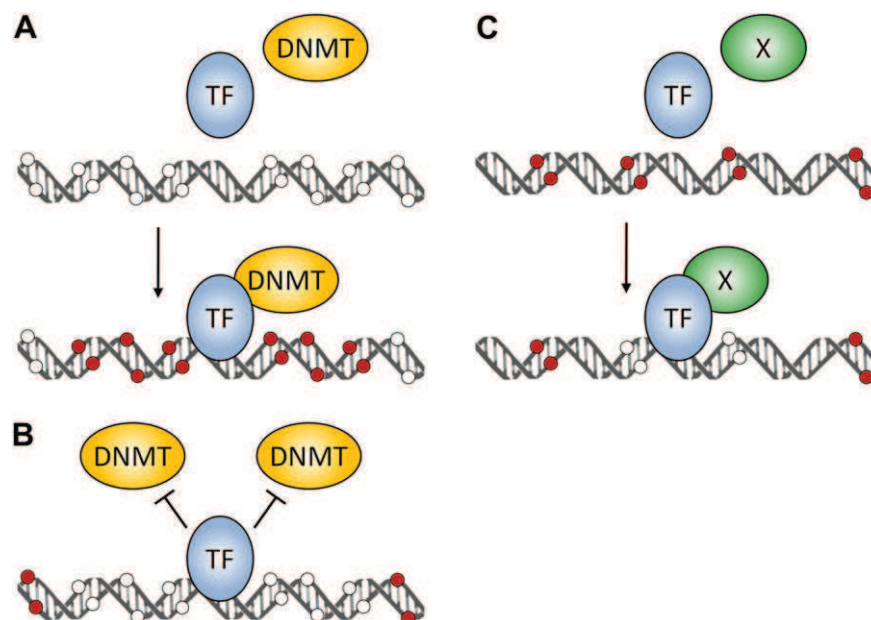


Fig. 2. Influence of the DNA sequence on DNA methylation patterns *in cis*. A. Certain sequence-specific transcription factors (TF) have been shown to interact with DNA methyltransferases (DNMT) and recruit them to their DNA binding sites. B. Alternatively, the recruitment of DNA-binding transcription factors can help to protect from DNA methylation and maintain a hypomethylated state. The mechanism of exclusion from DNA methylation is unknown but could be explained by a reduced accessibility of DNMTs to the local DNA. C. Several DNA binding factors, such as hormone receptors or CTCF, induce a local demethylation of the surrounding DNA when they bind to their target sites in the genome, which could help to favor the recruitment of other factors to the same site. This demethylation could be achieved either by recruiting unknown demethylating factors (X) or by interfering with DNA methylation maintenance during replication. Open circles represent unmethylated CpGs, red circles represent methylated CpGs.

H3K4 methyltransferase Setd1, and is required for maintaining high levels of H3K4 methylation at CpG islands [113]. Altogether, these data suggest that H3K4 methylation plays a prominent role in protecting DNA against *de novo* DNA methylation, and that its removal could act as a signal to target DNA methylation in the genome.

4.3. Role of non-coding RNAs and transcription

A link between DNA methylation and small non-coding RNAs is emerging in mammalian germ cells. Piwi-interacting RNAs (piRNAs) are a class of 24 to 32-nucleotides small RNAs abundant in germ cells where they bind to PIWI proteins, a subfamily of Argonaute proteins. These piRNA complexes have been linked to the silencing of retrotransposons during spermatogenesis. Interestingly, it has been shown that DNA methylation of several classes of retrotransposons is not properly established in the male germline of mutant mice lacking the PIWI proteins MILI and MIWI2 [114,115], suggesting that piRNAs participate in the recruitment of the DNA methylation machinery to retrotransposons. This model has been further corroborated by a study showing that the piRNA pathway is required to establish DNA methylation at the imprinted *Rasgrf1* locus that contains a retrotransposon sequence recognized by piRNAs [116]. Future studies will help to clarify the mechanisms by which the piRNAs recruit DNA methylation to their target sequences, and if piRNAs also regulate DNA methylation in somatic cells as recently observed in the model organism *Aplysia* [117].

In addition to trans-acting small RNAs, cis-acting RNAs of longer size have also been implicated in epigenetic gene regulation. While these long non-coding RNAs have so far mostly been implicated in pathways of chromatin modifications, it is conceivable that they might also regulate DNA methylation. Several studies have shown that non-coding antisense transcripts, such as *Air* or *Kcnq1ot1*, are required for proper DNA methylation at somatic differentially methylated regions (DMRs) in imprinted loci, although the mechanisms involved could be indirect and remain to be established [118]. In certain cases, a possibility is that the process of transcription itself is involved in DNA methylation. Indeed, it has been shown that transcription of protein-coding RNAs across imprinted DMRs is required for their DNA methylation in oocytes at the *Gnas* and *Snrpn* loci [119,120]. The importance of transcription in gametes is further corroborated by genome-scale studies showing that sites of DNA methylation in oocytes are preferentially located within active transcription units [48,55]. Altogether this suggests that transcription could be one factor that facilitates the acquisition of DNA methylation in germ cells, perhaps by exposing the DNA to the DNA methylation machinery or by inducing a favorable chromatin signature [121].

4.4. Mechanisms of DNA demethylation: is 5hmC the key?

The mechanisms of cytosine demethylation have been a matter of debate for many years. One simple way for achieving demethylation is a lack of methylation maintenance during DNA replication, which results in a progressive dilution of DNA methylation over several rounds of DNA replication (Fig. 3). This mechanism is believed to be responsible for the erasure of DNA methylation on the maternal genome after fertilization, which is supported by the fact that the maintenance enzyme DNMT1 is excluded from the nucleus during preimplantation development [42]. DNMT1 is however required to maintain DNA methylation at certain sequences such as ICRs during this period, indicating that a low level of active DNMT1 must remain in the nucleus [42]. Yet rapid replication-independent erasure of cytosine methylation has been reported in zygotes and PGCs, as well as an increasing number of

other biological contexts (see above), indicating that mechanisms for active removal of 5mC exist. In the last four years, a tremendous amount of studies have contributed to elucidate potential pathways for active DNA demethylation involving DNA repair, which are summarized in detail in some recent reviews [122–124].

Considerable excitement was created with the discovery that 5-hydroxymethylcytosine (5hmC) is a prominent epigenetic mark of DNA in mammals that could act as an intermediate in the cytosine demethylation reaction. 5hmC has initially been detected in DNA from various animals many years ago [125] and is particularly abundant in cells from early embryos and the nervous system [126,127]. The key finding that brought 5hmC back into the spotlights is the discovery that members of the ten–eleven translocation (TET) family oxidize 5mC into 5hmC (Fig. 3) [128]. These proteins are conserved in metazoans and three TET proteins exist in vertebrates (TET1, TET2, TET3). All three TET proteins show a broad expression pattern in various tissues, with TET1 being particularly abundant in ES cells, TET2 in hematopoietic cells and TET3 in oocytes and zygotes. The TET proteins have now been implicated in various biological processes such as early embryogenesis, stem cell differentiation and hematopoiesis (for detailed reviews see Refs. [129,130]). It was later found that TET can further catalyze the oxidation of 5hmC into 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) [131,132], which are detectable in genomic DNA of various cells and could be additional intermediates of demethylation (Fig. 3). TET-mediated oxidation of 5mC has been implicated in the erasure of DNA methylation on the paternal genome after fertilization because immunofluorescence studies revealed that the loss of 5mC signal coincides with the appearance of 5hmC [133,134] (Fig. 1), as well as 5fC and 5caC [135]. This was confirmed by an elegant genetic study showing that loss of maternal TET3 impairs methylation erasure on the paternal genome in mouse preimplantation embryos [61].

One possible scenario for the TET-mediated demethylation is that the oxidation derivatives lead to passive demethylation because they are not properly recognized by the methylation maintenance machinery. This is supported by the observations that DNMT1 is not active on hemi-hydroxymethylated DNA [136]. This scenario is plausible during preimplantation development *in vivo* because 5hmC, 5caC and 5fC signals on the paternal genome persist in the cleavage stage embryos and are gradually diluted during replication [134,135,137]. Another scenario is that 5hmC, 5caC or 5fC trigger removal by DNA glycosylases such as thymine DNA glycosylase (TDG), followed by base excision repair (BER) (Fig. 3). Although no glycosylase has yet been identified to act on 5hmC, it has been shown that TDG can efficiently remove 5caC and 5fC from DNA [132,138]. The important role of BER glycosylases is supported by the observation that TDG deficiency leads to embryonic lethality and hypermethylation of developmental genes in mice [139,140]. The role of BER is further supported by the observations that in addition to enzymes that initiate BER, BER components PARP1, APE1 and XRCC1 are detected in the paternal pronucleus [141] and in E11.5 PGCs [142] at the time of demethylation. In addition, inhibition of BER with PARP or APE1 inhibitors results in an increase of DNA methylation on the zygotic paternal genome [142].

Other possible players in the reaction of demethylation are deaminases of the AID/APOBEC family that could trigger demethylation by BER. These deaminases could directly deaminate 5mC, which creates T:G mismatches that can be repaired by DNA glycosylases such as TDG and MBD4 followed by BER [140] (Fig. 3). Alternatively, they could require prior oxidation of 5mC by the TET proteins and deaminate 5hmC, thereby creating 5hmU that could be removed by the DNA glycosylases TDG or single-strand-selective monofunctional uracil-DNA glycosylase 1 (SMUG1), followed by BER (Fig. 3). This latter model is supported by the fact that AID/

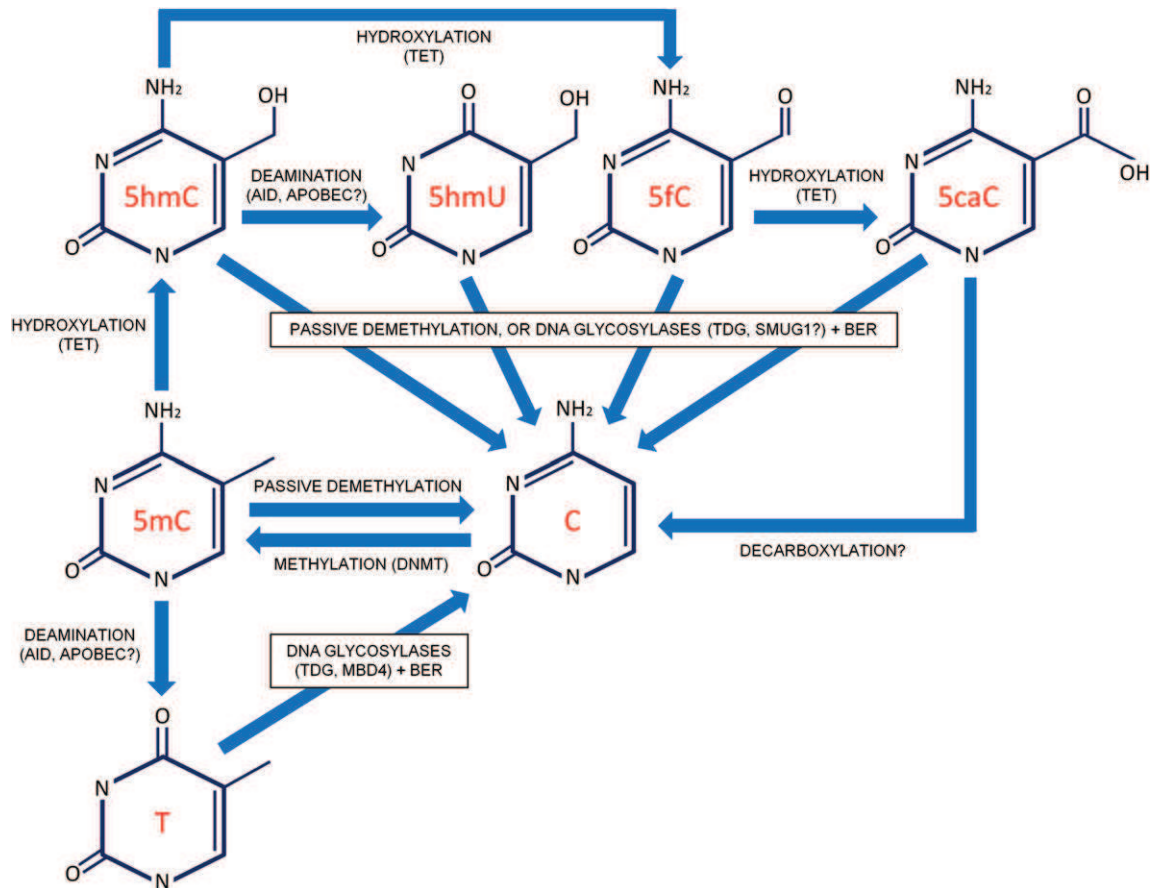


Fig. 3. Putative chemical pathways for cytosine demethylation. 5-methylcytosine (5mC) is generated and maintained during replication by DNA methyltransferases (DNMT). Erasure of cytosine methylation can be achieved by lack of maintenance leading to passive dilution during replication. One potential mechanism for active replication-independent demethylation is deamination of 5mC into thymine (T) by cytosine deaminases such as AID or APOBEC, followed by excision of the thymine by the DNA glycosylases TDG or MBD4 and repair by the base excision repair (BER) pathway. 5mC can also be hydroxylated into 5-hydroxymethylcytosine (5hmC) by TET enzymes, which can further be oxidated into 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), or deaminated into 5-hydroxymethyluracil (5hmU). All these modified bases could lead to replication-dependent passive demethylation, or active demethylation via excision by DNA glycosylases (TDG, SMUG1) followed by BER. In addition, a putative decarboxylase could lead to decarboxylation of 5caC directly back to cytosine.

APOBEC deaminases promote 5hmC-dependant demethylation in combination with BER enzymes in cultured cells and brain, and that TDG and SMUG1 have a glycosylase activity on 5hmU [94,140]. In support of a role of deamination, AID has been shown to participate in the active demethylation of pluripotency genes during reprogramming of non-dividing heterokaryons [79]. AID has also been proposed to contribute to the global methylation erasure in developing germ cells, as PGCs isolated from *Aid*^{-/-} E13.5 embryos show a ~20% median CpG methylation level compared to ~10% in wild type PGCs [76]. Yet the substantial demethylation in *Aid*^{-/-} PGCs and the lack of profound phenotype in *Aid*^{-/-} mice suggest that other deaminases or pathways participate in methylation erasure in this context. In addition, it remains to determine if deamination and BER in PGCs act directly on 5mC or require prior oxidation of 5mC by TET proteins. One possible argument for the role of 5mC oxidation in PGCs is the fact that TET1 deficient mice show a decreased fertility [143], yet this requires further studies as DNA methylation patterns have not yet been analyzed in PGCs of TET-deficient mice.

In summary, a large variety of demethylation pathways have been proposed that still need to be verified experimentally *in vivo*. This diversity of mechanisms might reflect the fact that the organism utilizes different pathways of demethylation depending on the cell type and the genomic target, or sometimes even in

parallel to ensure robust epigenetic reprogramming [144]. We are only beginning to address the biological functions of cytosine hydroxymethylation, and whether it plays additional roles besides being an intermediate of demethylation [129,130]. The recent development of a novel method for genome-wide mapping of 5-hydroxymethylcytosine at base-pair resolution will certainly pave the way for future studies that will clarify the picture [145]. These will unravel the fundamental role of dynamic DNA methylation in genome regulation and cellular identity, which has important implications not only for development but also for cellular reprogramming technologies.

Acknowledgments

Our team is supported by the Centre National de la Recherche Scientifique (CNRS), the University of Strasbourg, the ATIP-AVENIR program, the Agence Nationale de la Recherche (Grant No. ANR-09-BLAN-0261-03), and the CEFIC Long Research Initiative (Grant No. LRI-EMSG49-CNRS-08).

References

- [1] S.L. Berger, T. Kouzarides, R. Shiekhattar, A. Shilatifard, An operational definition of epigenetics, *Genes Dev.* 23 (2009) 781–783.

- [2] A. Meissner, Epigenetic modifications in pluripotent and differentiated cells, *Nat. Biotechnol.* 28 (2010) 1079–1088.
- [3] S. Tomizawa, H. Kobayashi, T. Watanabe, S. Andrews, K. Hata, G. Kelsey, H. Sasaki, Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes, *Development* 138 (2011) 811–820.
- [4] Z.D. Smith, M.M. Chan, T.S. Mikkelsen, H. Gu, A. Gnirke, A. Regev, A. Meissner, A unique regulatory phase of DNA methylation in the early mammalian embryo, *Nature* 484 (2012) 339–344.
- [5] T.R. Haines, D.I. Rodenhiser, P.J. Ainsworth, Allele-specific non-CpG methylation of the *Nf1* gene during early mouse development, *Dev. Biol.* 240 (2001) 585–598.
- [6] W. Xie, C.L. Barr, A. Kim, F. Yue, A.Y. Lee, J. Eubanks, E.L. Dempster, B. Ren, Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome, *Cell* 148 (2012) 816–831.
- [7] R. Lister, M. Pelizzola, R.H. Dowen, R.D. Hawkins, G. Hon, J. Tonti-Filippini, J.R. Nery, L. Lee, Z. Ye, Q.M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A.H. Millar, J.A. Thomson, B. Ren, J.R. Ecker, Human DNA methylomes at base resolution show widespread epigenomic differences, *Nature* 462 (2009) 315–322.
- [8] M.J. Ziller, F. Muller, J. Liao, Y. Zhang, H. Gu, C. Bock, P. Boyle, C.B. Epstein, B.E. Bernstein, T. Lengauer, A. Gnirke, A. Meissner, Genomic distribution and inter-sample variation of non-CpG methylation across human cell types, *PLoS Genet.* 7 (2011) e1002389.
- [9] M.B. Stadler, R. Murr, L. Burger, R. Ivanek, F. Lienert, A. Scholer, C. Wirbelauer, E.J. Oakeley, D. Gaidatzis, V.K. Tiwari, D. Schubeler, DNA-binding factors shape the mouse methylome at distal regulatory regions, *Nature* 480 (2011) 490–495.
- [10] M. Esteller, Cancer epigenomics: DNA methylomes and histone-modification maps, *Nat. Rev. Genet.* 8 (2007) 286–298.
- [11] G.C. Hon, R.D. Hawkins, O.L. Caballero, C. Lo, R. Lister, M. Pelizzola, A. Valsesia, Z. Ye, S. Kuan, L.E. Edsall, A.A. Camargo, B.J. Stevenson, J.R. Ecker, V. Bafna, R.L. Strausberg, A.J. Simpson, B. Ren, Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer, *Genome Res.* (2011).
- [12] A. Lujambio, A. Portela, J. Liz, S.A. Melo, S. Rossi, R. Spizzo, C.M. Croce, G.A. Calin, M. Esteller, CpG island hypermethylation-associated silencing of non-coding RNAs transcribed from ultraconserved regions in human cancer, *Oncogene* 29 (2010) 6390–6401.
- [13] F. Delhommeau, S. Dupont, V. Della Valle, C. James, S. Trannoy, A. Masse, O. Kosmider, J.P. Le Couedic, F. Robert, A. Alberdi, Y. Lecluse, I. Plo, F.J. Dreyfus, C. Marzac, N. Casadevall, C. Lacombe, S.P. Romana, P. Dessen, J. Soulier, F. Viguie, M. Fontenay, W. Vainchenker, O.A. Bernard, Mutation in *TET2* in myeloid cancers, *N. Engl. J. Med.* 360 (2009) 2289–2301.
- [14] X.J. Yan, J. Xu, Z.H. Gu, C.M. Pan, G. Lu, Y. Shen, J.Y. Shi, Y.M. Zhu, L. Tang, X.W. Zhang, W.X. Liang, J.Q. Mi, H.D. Song, K.Q. Li, Z. Chen, S.J. Chen, Exome sequencing identifies somatic mutations of DNA methyltransferase gene *DNMT3A* in acute monocytic leukemia, *Nat. Genet.* 43 (2011) 309–315.
- [15] Y. Wang, M. Jorda, P.L. Jones, R. Maleszka, X. Ling, H.M. Robertson, C.A. Mizzen, M.A. Peinado, G.E. Robinson, Functional CpG methylation system in a social insect, *Science* 314 (2006) 645–647.
- [16] A. Zemach, I.E. McDaniel, P. Silva, D. Zilberman, Genome-wide evolutionary analysis of eukaryotic DNA methylation, *Science* (2010).
- [17] K. Okamura, K.A. Matsumoto, K. Nakai, Gradual transition from mosaic to global DNA methylation patterns during deuterostome evolution, *BMC Bioinf.* 11 (Suppl. 7) (2010) S2.
- [18] P.W. Laird, Principles and challenges of genomewide DNA methylation analysis, *Nat. Rev. Genet.* 11 (2010) 191–203.
- [19] M. Weber, J.J. Davies, D. Wittig, E.J. Oakeley, M. Haase, W.L. Lam, D. Schubeler, Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells, *Nat. Genet.* 37 (2005) 853–862.
- [20] A. Meissner, T.S. Mikkelsen, H. Gu, M. Wernig, J. Hanna, A. Sivachenko, X. Zhang, B.E. Bernstein, C. Nusbaum, D.B. Jaffe, A. Gnirke, R. Jaenisch, E.S. Lander, Genome-scale DNA methylation maps of pluripotent and differentiated cells, *Nature* 454 (2008) 766–770.
- [21] R. Illingworth, A. Kerr, D. Desousa, H. Jorgensen, P. Ellis, J. Stalker, D. Jackson, C. Clee, R. Plumb, J. Rogers, S. Humphray, T. Cox, C. Langford, A. Bird, A novel CpG island set identifies tissue-specific methylation at developmental gene loci, *PLoS Biol.* 6 (2008) e22.
- [22] F. Mohn, M. Weber, M. Rebhan, T.C. Roloff, J. Richter, M.B. Stadler, M. Bibel, D. Schubeler, Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors, *Mol. Cell* 30 (2008) 755–766.
- [23] M. Weber, I. Hellmann, M.B. Stadler, L. Ramos, S. Paabo, M. Rebhan, D. Schubeler, Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome, *Nat. Genet.* 39 (2007) 457–466.
- [24] J. Borgel, S. Guibert, Y. Li, H. Chiba, D. Schubeler, H. Sasaki, T. Forne, M. Weber, Targets and dynamics of promoter DNA methylation during early mouse development, *Nat. Genet.* 42 (2010) 1093–1100.
- [25] N. Bhutani, J.J. Brady, M. Damian, A. Sacco, S.Y. Corbel, H.M. Blau, Reprogramming towards pluripotency requires AID-dependent DNA demethylation, *Nature* (2009).
- [26] T.S. Mikkelsen, J. Hanna, X. Zhang, M. Ku, M. Wernig, P. Schorderet, B.E. Bernstein, R. Jaenisch, E.S. Lander, A. Meissner, Dissecting direct reprogramming through integrative genomic analysis, *Nature* 454 (2008) 49–55.
- [27] C. Schmidl, M. Klug, T.J. Boeld, R. Andreesen, P. Hoffmann, M. Edinger, M. Rehli, Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity, *Genome Res.* (2009).
- [28] A.A. Serandour, S. Avner, F. Percevault, F. Demay, M. Bizot, C. Lucchetti-Miganeh, F. Barloy-Hubler, M. Brown, M. Lupien, R. Metivier, G. Salbert, J. Eeckhoutte, Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers, *Genome Res.* 21 (2011) 555–565.
- [29] R.S. Illingworth, U. Gruenewald-Schneider, S. Webb, A.R. Kerr, K.D. James, D.J. Turner, C. Smith, D.J. Harrison, R. Andrews, A.P. Bird, Orphan CpG islands identify numerous conserved promoters in the mammalian genome, *PLoS Genet.* 6 (2010).
- [30] A. Hellman, A. Chess, Gene body-specific methylation on the active X chromosome, *Science* 315 (2007) 1141–1143.
- [31] L. Laurent, E. Wong, G. Li, T. Huynh, A. Tsiganos, C.T. Ong, H.M. Low, K.W. Kin Sung, I. Rigoutsos, J. Loring, C.L. Wei, Dynamic changes in the human methylome during differentiation, *Genome Res.* 20 (2010) 320–331.
- [32] A.K. Maunakea, R.P. Nagarajan, M. Bilienky, T.J. Ballinger, C. D'Souza, S.D. Fouse, B.E. Johnson, C. Hong, C. Nielsen, Y. Zhao, G. Turecki, A. Delaney, R. Varhol, N. Thiessen, K. Shchors, V.M. Heine, D.H. Rowitch, X. Xing, C. Fiore, M. Schillebeeckx, S.J. Jones, D. Haussler, M.A. Marra, M. Hirst, T. Wang, J.F. Costello, Conserved role of intragenic DNA methylation in regulating alternative promoters, *Nature* 466 (2010) 253–257.
- [33] J.I. Young, E.P. Hong, J.C. Castle, J. Crespo-Barreto, A.B. Bowman, M.F. Rose, D. Kang, R. Richman, J.M. Johnson, S. Berget, H.Y. Zoghbi, Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005) 17551–17558.
- [34] S. Shukla, E. Kavak, M. Gregory, M. Imashimizu, B. Shutinoski, M. Kashlev, P. Oberdoerffer, R. Sandberg, S. Oberdoerffer, CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing, *Nature* 479 (2011) 74–79.
- [35] F. Lyko, S. Foret, R. Kucharski, S. Wolf, C. Falckenhayn, R. Maleszka, The honey bee epigenomes: differential methylation of brain DNA in queens and workers, *PLoS Biol.* 8 (2010) e1000506.
- [36] E. Hodges, A.D. Smith, J. Kendall, Z. Xuan, K. Ravi, M. Rooks, M.Q. Zhang, K. Ye, A. Bhattacharjee, L. Brizuela, W.R. McCombie, M. Wigler, G.J. Hannon, J.B. Hicks, High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing, *Genome Res.* 19 (2009) 1593–1605.
- [37] R.F. Luco, M. Allo, I.E. Schor, A.R. Kornblihtt, T. Misteli, Epigenetics in alternative pre-mRNA splicing, *Cell* 144 (2011) 16–26.
- [38] J. Song, O. Rechko, T.H. Bestor, D.J. Patel, Structure of DNMT1–DNA complex reveals a role for autoinhibition in maintenance DNA methylation, *Science* 331 (2011) 1036–1040.
- [39] J. Song, M. Teplova, S. Ishibe-Murakami, D.J. Patel, Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation, *Science* 335 (2012) 709–712.
- [40] M. Bostick, J.K. Kim, P.O. Esteve, A. Clark, S. Pradhan, S.E. Jacobsen, UHRF1 plays a role in maintaining DNA methylation in mammalian cells, *Science* 317 (2007) 1760–1764.
- [41] J. Sharif, M. Muto, S. Takebayashi, I. Suetake, A. Iwamatsu, T.A. Endo, J. Shinga, Y. Mizutani-Koseki, T. Toyoda, K. Okamura, S. Tajima, K. Mitsuya, M. Okano, H. Koseki, The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA, *Nature* 450 (2007) 908–912.
- [42] R. Hirasawa, H. Chiba, M. Kaneda, S. Tajima, E. Li, R. Jaenisch, H. Sasaki, Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development, *Genes Dev.* 22 (2008) 1607–1616.
- [43] Y. Kurihara, Y. Kawamura, Y. Uchijima, T. Amamo, H. Kobayashi, T. Asano, H. Kurihara, Maintenance of genomic methylation patterns during preimplantation development requires the somatic form of DNA methyltransferase 1, *Dev. Biol.* 313 (2008) 335–346.
- [44] M. Okano, D.W. Bell, D.A. Haber, E. Li, DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development, *Cell* 99 (1999) 247–257.
- [45] Y. Kato, M. Kaneda, K. Hata, K. Kumaki, M. Hisano, Y. Kohara, M. Okano, E. Li, M. Nozaki, H. Sasaki, 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 (2007) 2272–2280.
- [46] D. Watanabe, I. Suetake, T. Tada, S. Tajima, Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis, *Mech. Dev.* 118 (2002) 187–190.
- [47] M. Kaneda, M. Okano, K. Hata, T. Sado, N. Tsujimoto, E. Li, H. Sasaki, Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting, *Nature* 429 (2004) 900–903.
- [48] S.A. Smallwood, S.I. Tomizawa, F. Krueger, N. Ruf, N. Carli, A. Segonds-Pichon, S. Sato, K. Hata, S.R. Andrews, G. Kelsey, Dynamic CpG island methylation landscape in oocytes and preimplantation embryos, *Nat. Genet.* (2011).
- [49] G.A. Challen, D. Sun, M. Jeong, M. Luo, J. Jelinek, J.S. Berg, C. Bock, A. Vasanthakumar, H. Gu, Y. Xi, S. Liang, Y. Lu, G.J. Darlington, A. Meissner,

- J.P. Issa, L.A. Godley, W. Li, M.A. Goodell, Dnmt3a is essential for hematopoietic stem cell differentiation, *Nat. Genet.* 44 (2011) 23–31.
- [50] H. Wu, V. Coskun, J. Tao, W. Xie, W. Ge, K. Yoshikawa, E. Li, Y. Zhang, Y.E. Sun, Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes, *Science* 329 (2010) 444–448.
- [51] P.A. Jones, G. Liang, Rethinking how DNA methylation patterns are maintained, *Nat. Rev. Genet.* 10 (2009) 805–811.
- [52] D. Bourc'his, T.H. Bestor, Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L, *Nature* 431 (2004) 96–99.
- [53] D. Bourc'his, G.L. Xu, C.S. Lin, B. Bollman, T.H. Bestor, Dnmt3L and the establishment of maternal genomic imprints, *Science* 294 (2001) 2536–2539.
- [54] K. Hata, M. Okano, H. Lei, E. Li, Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice, *Development* 129 (2002) 1983–1993.
- [55] H. Kobayashi, T. Sakurai, M. Imai, N. Takahashi, A. Fukuda, O. Yayoi, S. Sato, K. Nakabayashi, K. Hata, Y. Sotomaru, Y. Suzuki, T. Kono, Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks, *PLoS Genet.* 8 (2012) e1002440.
- [56] S.K. Ooi, C. Qiu, E. Bernstein, K. Li, D. Jia, Z. Yang, H. Erdjument-Bromage, P. Tempst, S.P. Lin, C.D. Allis, X. Cheng, T.H. Bestor, DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA, *Nature* 448 (2007) 714–717.
- [57] D. Jia, R.Z. Jurkowska, X. Zhang, A. Jeltsch, X. Cheng, Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation, *Nature* 449 (2007) 248–251.
- [58] M. Saitou, S. Kagiwada, K. Kurimoto, Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells, *Development* 139 (2012) 15–31.
- [59] N. Lane, W. Dean, S. Erhardt, P. Hajkova, A. Surani, J. Walter, W. Reik, Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse, *Genesis* 35 (2003) 88–93.
- [60] C.R. Farthing, G. Ficiz, R.K. Ng, C.F. Chan, S. Andrews, W. Dean, M. Hemberger, W. Reik, Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes, *PLoS Genet.* 4 (2008) e1000116.
- [61] T.P. Gu, F. Guo, H. Yang, H.P. Wu, G.F. Xu, W. Liu, Z.G. Xie, L. Shi, X. He, S.G. Jin, K. Iqbal, Y.G. Shi, Z. Deng, P.E. Szabo, G.P. Pfeifer, J. Li, G.L. Xu, The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes, *Nature* 477 (2011) 606–610.
- [62] K. Lepikhov, V. Zakhartchenko, R. Hao, F. Yang, C. Wrenzycki, H. Niemann, E. Wolf, J. Walter, Evidence for conserved DNA and histone H3 methylation reprogramming in mouse, bovine and rabbit zygotes, *Epigenet. Chromatin* 1 (2008) 8.
- [63] A.R. Reis Silva, P. Adenot, N. Daniel, C. Archilla, N. Peynot, C.M. Lucci, N. Beaujean, V. Duranthon, Dynamics of DNA methylation levels in maternal and paternal rabbit genomes after fertilization, *Epigenetics* 6 (2011) 987–993.
- [64] W. Dean, F. Santos, M. Stojkovic, V. Zakhartchenko, J. Walter, E. Wolf, W. Reik, Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos, *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 13734–13738.
- [65] N. Beaujean, G. Hartshorne, J. Cavilla, J. Taylor, J. Gardner, I. Wilmut, R. Meehan, L. Young, Non-conservation of mammalian preimplantation methylation dynamics, *Curr. Biol.* 14 (2004) R266–R267.
- [66] Y. Li, C. O'Neill, Persistence of cytosine methylation of DNA following fertilisation in the mouse, *PLoS One* 7 (2012) e30687.
- [67] X. Li, M. Ito, F. Zhou, N. Youngson, X. Zuo, P. Leder, A.C. Ferguson-Smith, A maternal-zygotic effect gene, *Zfp57*, maintains both maternal and paternal imprints, *Dev. Cell* 15 (2008) 547–557.
- [68] S. Quenneville, G. Verde, A. Corsinotti, A. Kapopoulou, J. Jakobsson, S. Offner, I. Baglivo, P.V. Pedone, G. Grimaldi, A. Riccio, D. Trono, In embryonic stem cells, *ZFP57/KAP1* recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions, *Mol. Cell* 44 (2011) 361–372.
- [69] Y. Ohinata, B. Payer, D. O'Carroll, K. Ancelin, Y. Ono, M. Sano, S.C. Barton, T. Obukhanych, M. Nussenzweig, A. Tarakhovskiy, M. Saitou, M.A. Surani, *Blimp1* is a critical determinant of the germ cell lineage in mice, *Nature* 436 (2005) 207–213.
- [70] M. Sato, T. Kimura, K. Kurokawa, Y. Fujita, K. Abe, M. Masuhara, T. Yasunaga, A. Ryo, M. Yamamoto, T. Nakano, Identification of *PGC7*, a new gene expressed specifically in preimplantation embryos and germ cells, *Mech. Dev.* 113 (2002) 91–94.
- [71] M. Yamaji, Y. Seki, K. Kurimoto, Y. Yabuta, M. Yuasa, M. Shigeta, K. Yamanaka, Y. Ohinata, M. Saitou, Critical function of *Prdm14* for the establishment of the germ cell lineage in mice, *Nat. Genet.* 40 (2008) 1016–1022.
- [72] P. Hajkova, K. Ancelin, T. Waldmann, N. Lacoste, U.C. Lange, F. Cesari, C. Lee, G. Almouzni, R. Schneider, M.A. Surani, Chromatin dynamics during epigenetic reprogramming in the mouse germ line, *Nature* 452 (2008) 877–881.
- [73] Y. Seki, K. Hayashi, K. Itoh, M. Mizugaki, M. Saitou, Y. Matsui, Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice, *Dev. Biol.* 278 (2005) 440–458.
- [74] P. Hajkova, S. Erhardt, N. Lane, T. Haaf, O. El-Maarri, W. Reik, J. Walter, M.A. Surani, Epigenetic reprogramming in mouse primordial germ cells, *Mech. Dev.* 117 (2002) 15–23.
- [75] D.J. Lees-Murdock, M. De Felici, C.P. Walsh, Methylation dynamics of repetitive DNA elements in the mouse germ cell lineage, *Genomics* 82 (2003) 230–237.
- [76] C. Popp, W. Dean, S. Feng, S.J. Cokus, S. Andrews, M. Pellegrini, S.E. Jacobsen, W. Reik, Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency, *Nature* 463 (2010) 1101–1105.
- [77] S. Guibert, T. Forné, M. Weber, Global profiling of DNA methylation erasure in mouse primordial germ cells, *Genome Res.* 22 (2012) 633–641.
- [78] J.Y. Li, D.J. Lees-Murdock, G.L. Xu, C.P. Walsh, Timing of establishment of paternal methylation imprints in the mouse, *Genomics* 84 (2004) 952–960.
- [79] N. Bhutani, J.J. Brady, M. Damian, A. Sacco, S.Y. Corbel, H.M. Blau, Reprogramming towards pluripotency requires AID-dependent DNA demethylation, *Nature* 463 (2010) 1042–1047.
- [80] M.S. Kim, T. Kondo, I. Takada, M.Y. Youn, Y. Yamamoto, S. Takahashi, T. Matsumoto, S. Fujiyama, Y. Shirode, I. Yamaoka, H. Kitagawa, K. Takeyama, H. Shibuya, F. Ohtake, S. Kato, DNA demethylation in hormone-induced transcriptional derepression, *Nature* 461 (2009) 1007–1012.
- [81] R. Metivier, R. Gallais, C. Tiffoche, C. Le Peron, R.Z. Jurkowska, R.P. Carmouche, D. Ibberson, P. Barath, F. Demay, G. Reid, V. Benes, A. Jeltsch, F. Gannon, G. Salbert, Cyclical DNA methylation of a transcriptionally active promoter, *Nature* 452 (2008) 45–50.
- [82] A. Colaneri, N. Staffa, D.C. Fargo, Y. Gao, T. Wang, S.D. Peddada, L. Birnbaumer, Expanded methyl-sensitive cut counting reveals hypomethylation as an epigenetic state that highlights functional sequences of the genome, *Proc. Natl. Acad. Sci. U. S. A.* 108 (2011) 9715–9720.
- [83] H. Ji, L.L. Ehrlich, J. Seita, P. Murakami, A. Doi, P. Lindau, H. Lee, M.J. Arvey, R.A. Irizarry, K. Kim, D.J. Rossi, M.A. Inlay, T. Serwold, H. Karsunky, L. Ho, G.Q. Daley, I.L. Weissman, A.P. Feinberg, Comprehensive methylome map of lineage commitment from haematopoietic progenitors, *Nature* 467 (2010) 338–342.
- [84] M.T. Bocker, I. Hellwig, A. Breiling, V. Eckstein, A.D. Ho, F. Lyko, Genome-wide promoter DNA methylation dynamics of human hematopoietic progenitor cells during differentiation and aging, *Blood* 117 (2011) e182–e189.
- [85] V. Calvanese, A.F. Fernandez, R.G. Urduinguio, B. Suarez-Alvarez, C. Mangas, V. Perez-Garcia, C. Bueno, R. Montes, V. Ramos-Mejia, P. Martinez-Cambor, C. Ferrero, Y. Assenov, C. Bock, P. Menendez, A.C. Carrera, C. Lopez-Larrea, M.F. Fraga, A promoter DNA demethylation landscape of human hematopoietic differentiation, *Nucleic Acids Res.* 40 (2012) 116–131.
- [86] J.R. Shearstone, R. Pop, C. Bock, P. Boyle, A. Meissner, M. Socolovsky, Global DNA demethylation during mouse erythropoiesis in vivo, *Science* 334 (2011) 799–802.
- [87] M. Klug, S. Heinz, C. Gebhard, L. Schwarzfischer, S.W. Krause, R. Andreesen, M. Rehli, Active DNA demethylation in human postmitotic cells correlates with activating histone modifications, but not transcription levels, *Genome Biol.* 11 (2010) R63.
- [88] M. Szyf, J. Rouleau, J. Theberge, V. Bozovic, Induction of myogenic differentiation by an expression vector encoding the DNA methyltransferase cDNA sequence in the antisense orientation, *J. Biol. Chem.* 267 (1992) 12831–12836.
- [89] B.P. Brunk, D.J. Goldhamer, C.P. Emerson Jr., Regulated demethylation of the myoD distal enhancer during skeletal myogenesis, *Dev. Biol.* 177 (1996) 490–503.
- [90] M. Lucarelli, A. Fuso, R. Strom, S. Scarpa, The dynamics of myogenin site-specific demethylation is strongly correlated with its expression and with muscle differentiation, *J. Biol. Chem.* 276 (2001) 7500–7506.
- [91] G. Nagae, T. Isagawa, N. Shiraki, T. Fujita, S. Yamamoto, S. Tsutsumi, A. Nonaka, S. Yoshida, K. Matsusaka, Y. Midorikawa, S. Ishikawa, H. Soejima, M. Fukayama, H. Suemori, N. Nakatsuji, S. Kume, H. Aburatani, Tissue-specific demethylation in CpG-poor promoters during cellular differentiation, *Hum. Mol. Genet.* 20 (2011) 2710–2721.
- [92] R.A. Waterland, R. Keller Mayer, M.T. Rached, N. Tatevian, M.V. Gomes, J. Zhang, L. Zhang, A. Chakravarty, W. Zhu, E. Laritsky, W. Zhang, X. Wang, L. Shen, Epigenomic profiling indicates a role for DNA methylation in early postnatal liver development, *Hum. Mol. Genet.* 18 (2009) 3026–3038.
- [93] J.U. Guo, D.K. Ma, H. Mo, M.P. Ball, M.H. Jang, M.A. Bonaguidi, J.A. Balazer, H.L. Eaves, B. Xie, E. Ford, K. Zhang, G.L. Ming, Y. Gao, H. Song, Neuronal activity modifies the DNA methylation landscape in the adult brain, *Nat. Neurosci.* 14 (2011) 1345–1351.
- [94] J.U. Guo, Y. Su, C. Zhong, G.L. Ming, H. Song, Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain, *Cell* 145 (2011) 423–434.
- [95] D.K. Ma, M.H. Jang, J.U. Guo, Y. Kitabatake, M.L. Chang, N. Pow-Anpongkul, R.A. Flavell, B. Lu, G.L. Ming, H. Song, Neuronal activity-induced *Gadd45b* promotes epigenetic DNA demethylation and adult neurogenesis, *Science* 323 (2009) 1074–1077.
- [96] K. Martinowich, D. Hattori, H. Wu, S. Fouse, F. He, Y. Hu, G. Fan, Y.E. Sun, DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation, *Science* 302 (2003) 890–893.
- [97] M. Freitag, E.U. Selker, Controlling DNA methylation: many roads to one modification, *Curr. Opin. Genet. Dev.* 15 (2005) 191–199.
- [98] K. Kerker, A. Spadola, E. Yuan, J. Kosek, L. Jiang, E. Hod, K. Li, V.V. Murty, N. Schupf, E. Vilain, M. Morris, F. Haghghi, B. Tycko, Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation, *Nat. Genet.* 40 (2008) 904–908.
- [99] J. Gertz, K.E. Varley, T.E. Reddy, K.M. Bowling, F. Pauli, S.L. Parker, K.S. Kucera, H.F. Willard, R.M. Myers, Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation, *PLoS Genet.* 7 (2011) e1002228.

- [100] F. Lienert, C. Wirbelaer, I. Som, A. Dean, F. Mohn, D. Schubeler, Identification of genetic elements that autonomously determine DNA methylation states, *Nat. Genet.* 43 (2011) 1091–1097.
- [101] B.L. Wienholz, M.S. Kareta, A.H. Moarefi, C.A. Gordon, P.A. Ginno, F. Chedin, DNMT3L modulates significant and distinct flanking sequence preference for DNA methylation by DNMT3A and DNMT3B in vivo, *PLoS Genet.* 6 (2010).
- [102] E. Hervouet, F.M. Vallette, P.F. Cartron, Dnmt3/transcription factor interactions as crucial players in targeted DNA methylation, *Epigenetics* 4 (2009) 487–499.
- [103] G. Velasco, F. Hube, J. Rollin, D. Neuillet, C. Philippe, H. Bouzinba-Segard, A. Galvani, E. Viegas-Pequignot, C. Francastel, Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues, *Proc. Natl. Acad. Sci. U. S. A.* 107 (2010) 9281–9286.
- [104] M. Brandeis, D. Frank, I. Keshet, Z. Siegfried, M. Mendelsohn, A. Nemes, V. Temper, A. Razin, H. Cedar, Sp1 elements protect a CpG island from de novo methylation, *Nature* 371 (1994) 435–438.
- [105] K.B. Dong, I.A. Maksakova, F. Mohn, D. Leung, R. Appanah, S. Lee, H.W. Yang, L.L. Lam, D.L. Mager, D. Schubeler, M. Tachibana, Y. Shinkai, M.C. Lorincz, DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity, *EMBO J.* 27 (2008) 2691–2701.
- [106] S. Epsztejn-Litman, N. Feldman, M. Abu-Remaileh, Y. Shufaro, A. Gerson, J. Ueda, R. Deplus, F. Fuks, Y. Shinkai, H. Cedar, Y. Bergman, De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes, *Nat. Struct. Mol. Biol.* 15 (2008) 1176–1183.
- [107] M. Tachibana, Y. Matsumura, M. Fukuda, H. Kimura, Y. Shinkai, G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription, *EMBO J.* 27 (2008) 2681–2690.
- [108] N.P. Blackledge, J.C. Zhou, M.Y. Tolstorukov, A.M. Farcas, P.J. Park, R.J. Klose, CpG islands recruit a histone H3 lysine 36 demethylase, *Mol. Cell* 38 (2010) 179–190.
- [109] P. Kolasinska-Zwierz, T. Down, I. Latorre, T. Liu, X.S. Liu, J. Ahringer, Differential chromatin marking of introns and expressed exons by H3K36me3, *Nat. Genet.* 41 (2009) 376–381.
- [110] A. Dhayalan, A. Rajavelu, P. Rathert, R. Tamas, R.Z. Jurkowska, S. Ragozin, A. Jeltsch, The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation, *J. Biol. Chem.* 285 (2010) 26114–26120.
- [111] Y. Zhang, R. Jurkowska, S. Soeroes, A. Rajavelu, A. Dhayalan, I. Bock, P. Rathert, O. Brandt, R. Reinhardt, W. Fischle, A. Jeltsch, 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 (2010) 4246–4253.
- [112] D.N. Ciccone, H. Su, S. Hevi, F. Gay, H. Lei, J. Bajko, G. Xu, E. Li, T. Chen, KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints, *Nature* 461 (2009) 415–418.
- [113] J.P. Thomson, P.J. Skene, J. Selfridge, T. Clouaire, J. Guy, S. Webb, A.R. Kerr, A. Deaton, R. Andrews, K.D. James, D.J. Turner, R. Illingworth, A. Bird, CpG islands influence chromatin structure via the CpG-binding protein Cfp1, *Nature* 464 (2010) 1082–1086.
- [114] A.A. Aravin, R. Sachidanandam, D. Bourc'his, C. Schaefer, D. Pezic, K.F. Toth, T. Bestor, G.J. Hannon, A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice, *Mol. Cell* 31 (2008) 785–799.
- [115] S. Kuramochi-Miyagawa, T. Watanabe, K. Gotoh, Y. Totoki, A. Toyoda, M. Ikawa, N. Asada, K. Kojima, Y. Yamaguchi, T.W. Ijiri, K. Hata, E. Li, Y. Matsuda, T. Kimura, M. Okabe, Y. Sakaki, H. Sasaki, T. Nakano, DNA methylation of retrotransposon genes is regulated by Piwi family members MIL1 and MIWI2 in murine fetal testes, *Genes Dev.* 22 (2008) 908–917.
- [116] T. Watanabe, S. Tomizawa, K. Mitsuya, Y. Totoki, Y. Yamamoto, S. Kuramochi-Miyagawa, N. Iida, Y. Hoki, P.J. Murphy, A. Toyoda, K. Gotoh, H. Hiura, T. Arima, A. Fujiyama, T. Sado, T. Shibata, T. Nakano, H. Lin, K. Ichiyani, P.D. Soloway, H. Sasaki, Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse *Rasgrf1* locus, *Science* 332 (2011) 848–852.
- [117] P. Rajasethupathy, I. Antonov, R. Sheridan, S. Frey, C. Sander, T. Tuschl, E.R. Kandel, A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity, *Cell* 149 (2012) 693–707.
- [118] D.P. Barlow, Genomic imprinting: a mammalian epigenetic discovery model, *Annu. Rev. Genet.* 45 (2011) 379–403.
- [119] M. Chotalia, S.A. Smallwood, N. Ruf, C. Dawson, D. Lucifero, M. Frontera, K. James, W. Dean, G. Kelsey, Transcription is required for establishment of germline methylation marks at imprinted genes, *Genes Dev.* 23 (2009) 105–117.
- [120] E.Y. Smith, C.R. Futtner, S.J. Chamberlain, K.A. Johnstone, J.L. Resnick, Transcription is required to establish maternal imprinting at the Prader–Willi syndrome and Angelman syndrome locus, *PLoS Genet.* 7 (2011) e1002422.
- [121] S.A. Smallwood, G. Kelsey, De novo DNA methylation: a germ cell perspective, *Trends Genet.* 28 (2012) 33–42.
- [122] N. Bhutani, D.M. Burns, H.M. Blau, DNA demethylation dynamics, *Cell* 146 (2011) 866–872.
- [123] S.C. Wu, Y. Zhang, Active DNA demethylation: many roads lead to Rome, *Nat. Rev. Mol. Cell Biol.* 11 (2010) 607–620.
- [124] M. Teperek-Tkacz, V. Pasque, G. Gentsch, A.C. Ferguson-Smith, Epigenetic reprogramming: is deamination key to active DNA demethylation? *Reproduction* 142 (2011) 621–632.
- [125] N.W. Penn, R. Suwalski, C. O'Riley, K. Bojanowski, R. Yura, The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid, *Biochem. J.* 126 (1972) 781–790.
- [126] S. Kiaucionis, N. Heintz, The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain, *Science* 324 (2009) 929–930.
- [127] A. Ruzov, Y. Tsenkina, A. Serio, T. Dudnakova, J. Fletcher, Y. Bai, T. Chebotareva, S. Pells, Z. Hannoun, G. Sullivan, S. Chandran, D.C. Hay, M. Bradley, I. Wilmot, P. De Sousa, Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development, *Cell Res.* 21 (2011) 1332–1342.
- [128] M. Tahiliani, K.P. Koh, Y. Shen, W.A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L.M. Iyer, D.R. Liu, L. Aravind, A. Rao, Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1, *Science* 324 (2009) 930–935.
- [129] M.R. Branco, G. Ficz, W. Reik, Uncovering the role of 5-hydroxymethylcytosine in the epigenome, *Nat. Rev. Genet.* 13 (2011) 7–13.
- [130] H. Wu, Y. Zhang, Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation, *Genes Dev.* 25 (2011) 2436–2452.
- [131] S. Ito, L. Shen, Q. Dai, S.C. Wu, L.B. Collins, J.A. Swenberg, C. He, Y. Zhang, Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine, *Science* 333 (2011) 1300–1303.
- [132] Y.F. He, B.Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C.X. Song, K. Zhang, C. He, G.L. Xu, TET-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA, *Science* 333 (2011) 1303–1307.
- [133] M. Wossidlo, T. Nakamura, K. Lepikhov, C.J. Marques, V. Zakhartchenko, M. Boiani, J. Arand, T. Nakano, W. Reik, J. Walter, 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming, *Nat. Commun.* 2 (2011) 241.
- [134] K. Iqbal, S.G. Jin, G.P. Pfeifer, P.E. Szabo, Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine, *Proc. Natl. Acad. Sci. U. S. A.* 108 (2011) 3642–3647.
- [135] A. Inoue, L. Shen, Q. Dai, C. He, Y. Zhang, Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development, *Cell Res.* 21 (2011) 1670–1676.
- [136] V. Valinluck, L.C. Sowers, Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1, *Cancer Res.* 67 (2007) 946–950.
- [137] A. Inoue, Y. Zhang, Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos, *Science* 334 (2011) 194.
- [138] A. Maiti, A.C. Drohat, Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites, *J. Biol. Chem.* 286 (2011) 35334–35338.
- [139] D. Cortazar, C. Kunz, J. Selfridge, T. Lettieri, Y. Saito, E. MacDougall, A. Wirz, D. Schuermann, A.L. Jacobs, F. Siegrist, R. Steinacher, J. Jiricny, A. Bird, P. Schar, Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability, *Nature* 470 (2011) 419–423.
- [140] S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, L.K. Abramowitz, M.S. Bartolomei, F. Rambow, M.R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A.J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, A. Bellacosa, Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair, *Cell* 146 (2011) 67–79.
- [141] M. Wossidlo, J. Arand, V. Sebastiano, K. Lepikhov, M. Boiani, R. Reinhardt, H. Scholer, J. Walter, Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes, *EMBO J.* 29 (2010) 1877–1888.
- [142] P. Hajkova, S.J. Jeffries, C. Lee, N. Miller, S.P. Jackson, M.A. Surani, Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway, *Science* 329 (2010) 78–82.
- [143] M.M. Dawlaty, K. Ganz, B.E. Powell, Y.C. Hu, S. Markoulaki, A.W. Cheng, Q. Gao, J. Kim, S.W. Choi, D.C. Page, R. Jaenisch, TET1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development, *Cell Stem Cell* 9 (2011) 166–175.
- [144] J.A. Hackett, J.J. Zyllicz, M.A. Surani, Parallel mechanisms of epigenetic reprogramming in the germline, *Trends Genet.* 28 (2012) 164–174.
- [145] M.J. Booth, M.R. Branco, G. Ficz, D. Oxley, F. Krueger, W. Reik, S. Balasubramanian, Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution, *Science* (2012).

Identification de cibles et régulateurs de la méthylation de l'ADN chez la souris

La méthylation de l'ADN est une modification épigénétique qui prend place durant le développement embryonnaire sur le génome des Mammifères. Durant ma thèse, j'ai déterminé les cinétiques de mise en place de la méthylation de l'ADN sur le génome murin au cours de l'embryogénèse précoce. J'ai identifié les rôles spécifiques et redondants des ADN méthyltransférases DNMT3a et DNMT3b dans ce processus. J'ai également étudié le rôle de deux facteurs dans la mise en place de la méthylation de l'ADN dans l'embryon. Premièrement, j'ai déterminé que l'enzyme G9a joue un rôle essentiel pour la répression et le recrutement de la méthylation de l'ADN à des sites spécifiques du génome, incluant en particulier des promoteurs à îlots CpG de gènes méiotiques. Deuxièmement, l'étude du facteur E2F6 m'a permis de montrer que cette protéine est elle aussi impliquée dans le recrutement de la méthylation de l'ADN, et ce à des promoteurs de gènes méiotiques distincts de ceux régulés par G9a.

Méthylation de l'ADN, îlot CpG, développement embryonnaire, expression génique, G9a, E2F6, DNMT3a, DNMT3b.

DNA methylation is an epigenetic modification which is established during embryonic development on the mammalian genome. In my thesis, I determined the kinetics of DNA methylation acquisition on the mouse genome during early embryogenesis, and determined the specific and redundant roles of the DNA methyltransferases DNMT3a and DNMT3b in this process. I also studied the roles of two factors involved in setting up DNA methylation in embryos. First, I determined that the G9a enzyme plays an essential role for the *in vivo* repression and DNA methylation of specific genomic sites, including in particular the CpG island promoters of germline genes. Second, the study of the E2F6 factor allowed me to show that this protein is also involved in recruiting DNA methylation at a set of germline gene promoters than are distinct from those regulated by G9a.

DNA methylation, CpG island, embryonic development, gene expression, G9a, E2F6, DNMT3a, DNMT3b.