

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

UMR7242 : Biotechnologie et Signalisation Cellulaire

THÈSE présentée par :

Thomas DAHLET

soutenue le : **30 septembre 2019**

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

Discipline/ Spécialité : **Aspects moléculaires et cellulaires de la biologie**

**Méthylation de l'ADN :
Fonctions et ciblage au cours du
développement chez la souris**

THÈSE dirigée par :
Dr **WEBER Michaël**

Directeur de recherche, UMR7242 BSC, Université de Strasbourg

RAPPORTEURS :
Dr **BOURC'HIS Déborah**
Dr **DEJARDIN Jérôme**

Directeur de recherche, Institut Curie, Paris
Directeur de recherche, Institut de Génétique Humaine, Montpellier

EXAMINATEUR :
Dr **TORA Laszlo**

Directeur de recherche, Institut de Génétique et de Biologie
Moléculaire et Cellulaire

REMERCIEMENTS :

Voilà, il est temps pour moi de clore cette aventure et de remercier toutes les personnes qui y ont contribué de près ou de loin.

Je voudrais commencer par remercier les membres de mon jury, Déborah Bour'chis, Jérôme Déjardin et Laszlo Tora pour le temps que vous prendrez pour évaluer mon travail de thèse.

Ensuite, j'aimerais remercier Michaël (chef). Tout d'abord je devrais te remercier pour avoir répondu à mon (deuxième !) mail et accepté ma candidature de stage dans ton laboratoire. La deuxième tentative aura été la bonne ! Plus sérieusement, merci pour ta patience et ta confiance ainsi que pour ton aide dans la préparation du concours et la rédaction de cette thèse. J'ai particulièrement apprécié le fait que tu aies toujours considéré mes idées (même si elles sont souvent farfelues) et que tu m'as laissé la chance d'en explorer certaines, ça m'a vraiment permis de m'épanouir durant toutes ces années. Donc merci pour tout.

J'aimerais ensuite remercier mon ex-mentor, Ghislain. Etre encadré par quelqu'un en dernière année de thèse c'est pas toujours facile, pas le temps de souffler ! Merci de m'avoir initié à la paillasse (aux contaminations COBRA) et à l'épigénétique. T'as toujours été un exemple pour moi et j'espère avoir été digne de ta succession. Bonne chance à toi et à Rose pour votre start up !

Ensuite la personne avec qui j'ai tout partagé : Manon, ma bretonne préférée ! On a commencé ensemble en tant que « baby scientists » et on a vécu les galères à l'animalerie, les galères de WB, les mi-thèses à l'arrache mais aussi plein de bonnes choses, comme ... les galères à l'animalerie, les galères de WB ou les mi thèses à l'arrache. Parce que finalement des galères avec les bonnes personnes c'est pas vraiment des galères. Ça m'a fait énormément plaisir de partager ces années avec toi, elle aurait eu moins de saveur sinon. Tu es une personne hyper bienveillante et tu mérites le meilleur. Je serai là pour t'encourager dans la dernière ligne droite ;)

Ensuite j'aimerais remercier la #teamREIC. J'ai eu la chance d'évoluer dans une ambiance fantastique.

Un grand merci à Hala Al Adhami (à défaut de ne pas savoir prononcer ton prénom j'espère l'avoir bien écrit !) Merci pour toute ton aide et le temps que tu m'as consacré sur ces quatre années. J'ai beaucoup apprécié ces discussions passionnées de philosophie, de politique, mais surtout de psychologie dans le bureau. Et je pense que l'épisode de Paulette restera dans les annales du laboratoire pour un bout de temps (J'espère ne pas avoir fait de fautes d'accords...)

Andrea, tu es la première personne à qui j'ai parlé dans le labo, et tu m'as convaincu de venir. Merci pour tous ces moments, ta gentillesse, ta bonne humeur, et ta disponibilité !

Judith ... Merci énormément pour ton aide sur le projet E2F6, je crois que je n'aurai jamais réussi à finir le projet sans toi ... Merci aussi pour les dizaines (centaines ?! milli... néné faut pas déconner) de madeleines. Je pense que je ne sais toujours pas où est rangé l'IGEPAL dans le labo. Ta personne ferait également un très bon sujet de thèse Judith !

Ambre, AC/DC résonnera dans le labo encore pendant longtemps ! Plus sérieusement, je pense que tu es un exemple pour tout le monde. Déjà parce que Mendel te donne 10% de chance d'avoir les bons échantillons, que sur ces échantillons tu récoltes 10 cellules au FACS et que tu Madame fait du

RRBS sur du too low ! Et aussi, et surtout, parce que tu es la personne la plus optimiste que je connaisse.

Minimike, merci pour ton aide pour les analyses bioinformatiques. J'ai beaucoup apprécié ta disponibilité, notamment pour m'initier au monde de l'écran noir et ses formules compliquées !

Merci à Yannick (digne successeur de TopGun) et Anaïs pour ces discussions passionnées, les conseils ainsi que ces débats de statistiques en lab meeting !

Gott verdàmmi Eléa ! (bonjour internet pour l'orthographe) Tu as traversé le couloir pour nous rejoindre et me rappeler le pays en salle de culture. Quand est ce qu'on retourne grimper ?

Et Marion, miss références ! Garde la pêche et ta bonne humeur, ta thèse est bien partie !

Aux anciens Elouan, Claire, Patryk, Jitesh...

Enfin merci aux personnes de l'institut pour tous ces moments partagés notamment à midi et ses débats passionnants (ou complètement wtf...) Les soirées BBQ, karaoké sur BSC-plage (on a eu la table, il manque encore le sable fin et les palmiers, on y croit !). Je pense notamment à Leonel, Daisy, Jo-Ann, Nadja, Eric, ainsi qu'aux anciens, Carole, Kathline...

Enfin, à mes ex-collègues et amis du master BDCS (du plus petit au plus grand) : Anaïs, Mélanie, Maxime « Spitz » Diem et Thomas Kleiber.

Merci également à mes amis de tous les jours, mais surtout du jeudi soir ! Merci à Thomas Di Mattia « cet homme incroyable », Laure, camillenoll (en minuscule, sans espace), monsieur Husson, monsieur Schmitt, que de péripéties à vos côtés. Je pourrais également remercier les Berthoms, l'Exil, la « Cave », le 2... (je m'arrête là au risque de choquer certains...). En vrai, je suis très heureux d'avoir partagé ces moments avec vous, mais bon, vous le savez déjà (il y aurait tellement de choses à dire, et surtout la flemme après des mois de rédaction !). L'aventure continue !

Je voudrais enfin remercier mes parents, sans qui tout cela n'aurait pas été possible. Merci de m'avoir laissé l'opportunité de tracer ma voie et de m'avoir toujours soutenu dans mes projets professionnels durant toutes ces années. Cette thèse, c'est aussi un peu la vôtre. Merci également à Manon, Jan et Yanis. Tous ensemble vous êtes la base.

Et ... Maguelone... ma plus belle découverte. La méthylation de l'ADN a finalement servi à quelque chose.

Thomas DAHLET

LISTE DES ABBRÉVIATIONS :

5-AzadC	5-Aza-2'-Deoxycytidine
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5 hydroxymethylcytosine
5mC	5-methylcytosine
6mA	6-methyladenine
A	Adénine
ADD	Domaine protéique ATRX-DNMT3-DNMT3L
AML	Acute Myeloide Leukemia
ARNm	ARN messenger
ARNnc	ARN non codant
ARNt	ARN de transfert
ATP	Adénosine tri-phosphate
BAH	Domaine protéique bromo-adjacent homology
BER	Base Excision Repair
bp	base pair
C	Cytosine
CD	Catalytic Domain
cDNA	ADN complémentaire
CGI	CpG island
ChIP	Immunoprécipitation de la chromatine
chr	Chromosome
CMML	Chronic Myelomonocytic Leukemia
COBRA	Combined restriction bisulfite analysis
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeat
CXXC	Domaine protéique riche en cystéines (C)
DBD	DNA Binding Domain
dCas9	Dead Cas9
DHS	DNase hypersensitivity site
DKO	Double Knock Out
DMAPD	Domaine protéique d'interaction avec la protéine DMAP
DMC	Differentially Methylated Cytosine
DMR	Differentially Methylated Region
DNMT	DNA methyltransferase
ERV	Endogenous Retroviruses
ES	Cellule souche embryonnaire
F	Femelle
FACS	Fluorescence Activated Cell Sorting
FC	Fold change

FPKM	Fragments Per Kilobase of transcript per Million mapped reads
G	Guanine
Gb	Giga base
gDMR	germline Differentially Methylated Region
GFP	Green Fluorescent Protein
gRNA	ARN guide
GO	Gene Ontology
H	adénine/thymine/guanine
HCP	High CpG promoter
HEK	Human Embryonic Kidney
IAP	Intracisternal A Particules
ICF	Immunodeficiency, Centromeric instability, Facial anomalies
ICM	Inner Cell Mass
ICP	Intermediate CpG promoter
ICR	Imprinting Control Region
IGV	Integrative Genomics Viewer
iPSC	Cellule souche pluripotente induite
kb	Kilo base
kDa	kilo Dalton
KO	Knock Out
KRAB	Krüppel-associated Box
LCP	Low CpG promoter
LIF	Leukemia Inhibitor Factor
LINE	long interspersed nuclear elements
LMR	Low Methylated Region
LTR	Long Terminal Repeat
LZ	Leuzine-zipper
M	Mâle
MB	Marked-box
MBD	Methyl-CpG-binding domain proteins
MDS	Myelodysplastic Syndrome
MedIP	Methylated DNA immunoprecipitation
MEF	Mouse Embryonic Fibroblast
NLS	Nuclear Localization Signal
NPC	Neuronal Progenitor Cell
PBS	PCNA binding site
PcG	Polycomb group Protein
PCR	Polymerase Chain Reaction
PGC	Primordial Germ Cell
PMD	Partially Methylated Domain
PR-DUB	Polycomb Repressive Deubiquitination
PRC1	Polycomb Repressive Complex 1
PRC2	Polycomb Repressive Complex 2

PRE	Polycomb Responsive Element
PWWP	Domaine protéique riche en Proline (P) et Tryptophane (W)
RRBS	Reduced Representation Bisulfite Sequencing
SAH	S-adenosyl-L-homocysteine
SAM	S-adenosyl-L-methionine (AdoMet)
Scr	Scramble
shRNA	Small hairpin RNA
SINE	short interspersed nuclear elements
RB	Rétinoblastome
RMSK	Repeat Masker
RT-qPCR	Quantitative reverse transcription PCR
SEM	Erreur standard à la moyenne
T	Thymine
TAM	Tamoxifène
TDG	Thymine DNA Glycosylase
TE	Transposable Element
TET	Ten-Eleven Translocation
TKO	Triple Knock Out
TSS	Transcription Start Site
TTS	Transcription Termination Site
WGBS	Whole Genome Bisulfite Sequencing
WT	Wild Type

TABLE DES MATIÈRES :

INTRODUCTION	9
1. Introduction à l'épigénétique	11
1.1. Définition, historique.....	11
1.2. L'organisation nucléaire de l'ADN.....	12
1.3. La chromatine	13
1.4. Les modifications de l'ADN.....	15
2. La méthylation de l'ADN	17
2.1. Distribution génomique	17
2.2. Les ADN méthyltransférases	20
2.3. La déméthylation de l'ADN	25
2.4. Recrutement de la méthylation de l'ADN.....	27
2.5. Conséquences de la méthylation de l'ADN.....	30
2.6. Implication de la méthylation de l'ADN dans les pathologies	33
2.6.1. Les cancers	33
2.6.2. Syndromes.....	35
3. Rôle de la méthylation de l'ADN au cours du développement embryonnaire	37
3.1. Reprogrammation au cours du développement	37
3.1.1. Reprogrammation dans la lignée germinale	37
3.1.2. Reprogrammation de l'embryon précoce	39
3.2. Régulations des éléments transposables	44
3.3. Régulation des gènes soumis à l'empreinte.....	47
3.4. Régulation de l'identité cellulaire.....	48
4. Les complexes Polycomb	53
4.1. Composition protéique des complexes Polycomb	53
4.2. Recrutement des complexes Polycomb	55
4.3. Rôles des complexes Polycomb au cours du développement embryonnaire.....	57
4.4. Mécanismes de répression par les complexes Polycomb	59
4.5. Le complexe PRC1.6	60
4.5.1. Description du complexe PRC1.6	60
4.5.2. Rôles du complexe PRC1.6	66
4.6. Le facteur de transcription E2F6	68
OBJECTIFS.....	71
RÉSULTATS	75
1. Rôle des ADN méthyltransférases au cours du développement chez la souris.....	77
1.1. Contexte et objectifs	77
1.2. Stratégie expérimentale.....	78
1.3. Travaux réalisés	79
1.4. Article 1: Genome-wide analyses in the mouse embryo reveals the importance of DNA methylation for transcription integrity	80
2. Rôle du facteur de transcription E2F6 dans le ciblage épigénétique des gènes gamétiques dans l'embryon.....	121
2.1. Contexte et objectifs	121

2.2.	<i>Stratégie expérimentale</i>	122
2.3.	<i>Travaux réalisés</i>	123
2.4.	<i>Article 2: E2F6 initiates stable epigenetic silencing of the germline program during embryonic development</i>	124
DISCUSSION.....		159
1.	Rôle de la méthylation de l'ADN au cours du développement embryonnaire	161
1.1.	<i>Rôle des ADN méthyltransférases dans la méthylation du génome</i>	161
1.2.	<i>Rôle de la méthylation dans la répression du programme gamétique</i>	161
1.3.	<i>Rôle de la méthylation dans la répression des gènes 2-cellules</i>	162
1.4.	<i>Rôle de la méthylation dans la répression de gènes de cellules différenciées</i>	163
1.5.	<i>Rôle de la méthylation dans la répression des éléments transposables</i>	164
1.6.	<i>Causes de la létalité embryonnaire des embryons hypométhylés</i>	165
2.	Rôle du facteur de transcription E2F6 dans la répression des gènes gamétiques	166
2.1.	<i>E2F6 réprime un groupe de gènes gamétiques</i>	166
2.2.	<i>Mécanismes de répression des gènes gamétiques par PRC1.6</i>	166
2.3.	<i>Mécanismes de recrutement de la méthylation de l'ADN par E2F6</i>	168
2.4.	<i>E2F6 induit une répression durable par des mécanismes dépendants ou indépendants de la méthylation de l'ADN</i>	169
BIBLIOGRAPHIE.....		171
ANNEXES		203
1.	Rôle de la méthylation de l'ADN dans la régulation de l'intégrité transcriptomique	205
	<i>Annexe 1.1 : Statistiques de séquençage des librairies RRBS</i>	205
	<i>Annexe 1.2 : Statistiques de séquençage des librairies WGBS</i>	206
	<i>Annexe 1.3 : Statistiques de séquençage des librairies RNA-seq</i>	206
	<i>Annexe 1.4 : liste des gènes dérprimés dans les embryons DKO (top 40)</i>	207
	<i>Annexe 1.5 : Liste des éléments transposables dérprimés dans les embryons DKO et Dnmt1 KO</i>	208
	<i>Annexe 1.6 : Liste des gènes avec une initiation interne de la transcription</i>	210
	<i>Annexe 1.7 : Liste des oligos utilisés dans cette étude</i>	211
2.	Rôle du facteur de transcription E2F6 dans le ciblage épigénétique des gènes gamétiques dans l'embryon	212
	<i>Annexe 2.1 : Statistiques de séquençage des librairies RRBS</i>	212
	<i>Annexe 2.3 : Statistiques de séquençage des librairies WGBS</i>	213
	<i>Annexe 2.2 : Statistiques de séquençage des librairies RNA-seq</i>	213
	<i>Annexe 2.4 : Analyse ontologique des gènes régulés par E2F6</i>	213
	<i>Annexe 2.5 : Liste des gènes dérprimés en l'absence d'E2F6 dans les cellules ES</i>	214
	<i>Annexe 2.6 : liste des gènes dérprimés en l'absence de E2F6 dans l'embryon (E8.5)</i>	215
	<i>Annexe 2.7 : Liste des DMRs identifiées dans l'embryon E8.5 en l'absence de E2F6 (WGBS)</i>	216
	<i>Annexe 2.8 : Liste des oligos utilisés dans cette étude</i>	218

INTRODUCTION

1. Introduction à l'épigénétique

1.1. Définition, historique

L'ADN est une macromolécule nécessaire pour le développement, le fonctionnement et la reproduction de tout être vivant. Elle contient l'information génétique qui est représentée par une séquence composée de 4 bases différentes : l'adénine (A), la thymine (T), la cytosine (C) et la guanine (G). L'ensemble de cette information génétique, appelé génome, s'exprime par le biais des gènes qui sont définis comme les séquences codantes et leurs séquences promotrices. Les séquences codantes sont transcrites en ARNm qui seront ensuite traduites en protéines. Ces dernières sont des macromolécules ayant des propriétés biochimiques ou enzymatiques qui définissent les fonctions spécifiques de chaque cellule.

Le séquençage et l'annotation du génome murin et humain au début des années 2000 a permis d'observer que seule une petite proportion de celui-ci est composée de séquences codantes (environ 1%) (*Lander et al., 2001, Waterston et al., 2002*) (**figure 1A**). Le reste du génome est composé de séquences inter-géniques et de séquences répétées d'origine virale. Bien que les séquences inter-géniques aient été initialement décrites comme « ADN poubelle », elles ont un rôle primordial dans la régulation de l'expression des gènes. Cette régulation permet de conférer une identité et une fonction spécifique aux cellules qui expriment chacune des combinaisons de gènes différentes alors que toutes les cellules de l'organisme partagent le même génome. Une des disciplines qui étudie la régulation de l'expression des gènes est appelée « épigénétique ».

Le terme épigénétique est attribué à Conrad Waddington qui étudiait le lien entre génotype et phénotype dans le déterminisme cellulaire chez la drosophile. Il a ainsi nommé ce domaine de recherche « épigénétique », qui est issu des termes épigénèse et génétique (*Waddington, 2012*). Il a notamment représenté l'épigénétique avec son célèbre croquis d'une bille (qui représente une cellule) évoluant à travers un plateau en relief illustrant la régulation de sa différenciation en un type cellulaire donné (**figure 1B**). Une définition plus moderne de l'épigénétique a depuis été énoncée. Il s'agit de la discipline qui étudie les mécanismes régulant de manière réversible, transmissible et adaptative, l'expression des gènes indépendamment de la séquence d'ADN (*Berger et al., 2009*).

La régulation de l'expression des gènes peut s'effectuer sur différents niveaux qui peuvent s'influencer mutuellement : d'une part par l'organisation tridimensionnelle de l'ADN dans le noyau

cellulaire, d'autre part par la compaction locale de l'ADN au sein de la chromatine et enfin par la modification chimique de l'ADN (**figure 1C**).

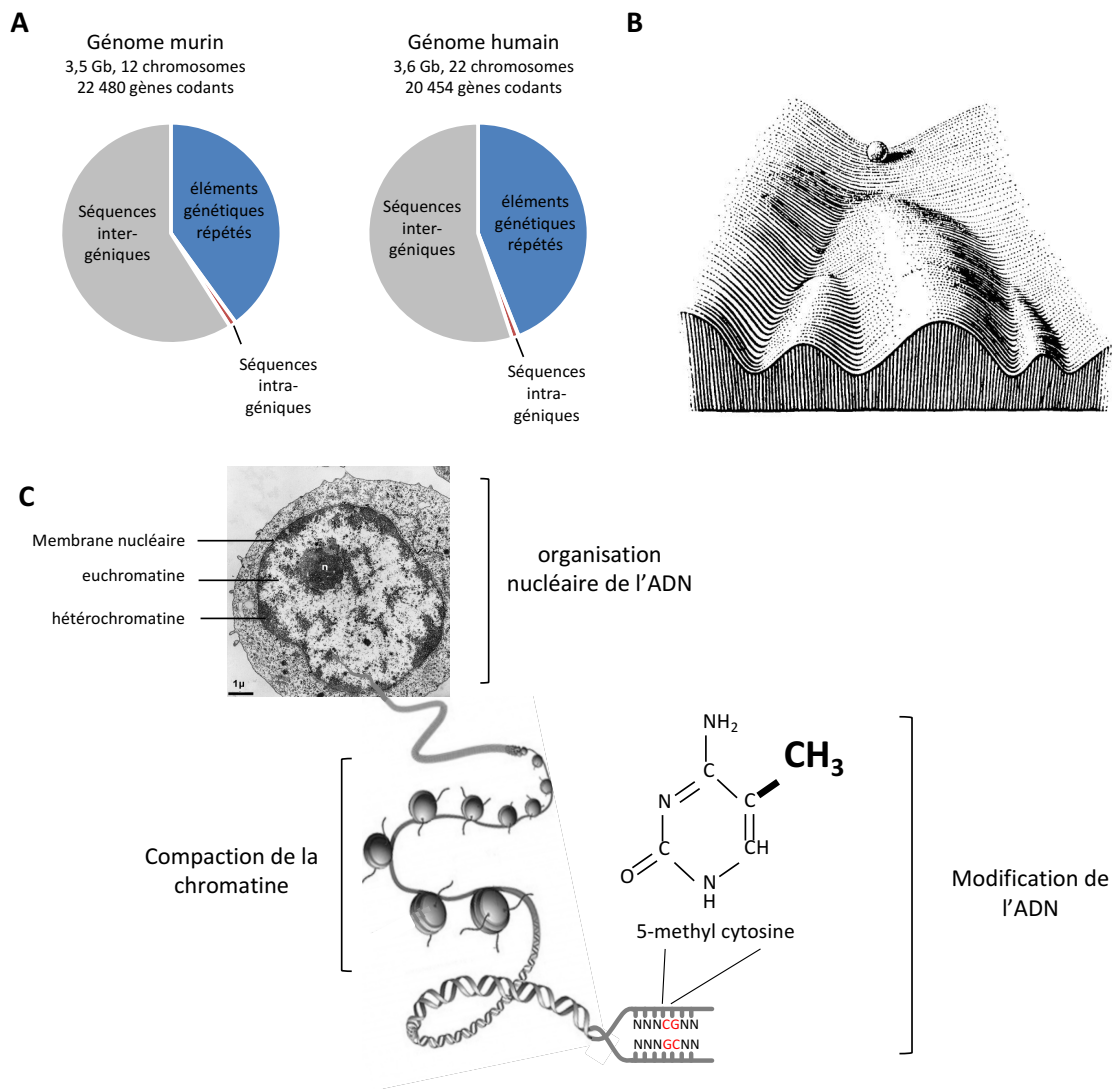


Figure 1 : L'épigénétique. A) Les génomes murins et humains possèdent globalement les mêmes caractéristiques génomiques : une petite proportion du génome correspond à des séquences codantes, et une grande proportion correspond aux séquences régulatrices. B) Illustration de l'épigénétique par Conrad Waddington. Adapté du livre « The strategy of genes » 1957. C) La régulation épigénétique s'effectue sur trois niveaux différents : l'architecture nucléaire, la compaction locale de la chromatine et la modification de l'ADN représentée par la 5 méthyl-cytosine.

1.2. L'organisation nucléaire de l'ADN

La longueur de l'ADN linéarisé d'une cellule est de 2 mètres chez la souris et l'homme. Celui-ci est contenu dans le noyau des cellules qui fait uniquement quelques micromètres de diamètre. L'organisation de l'ADN dans ce noyau est donc nécessaire afin de réguler son expression. Des

expériences de microscopie électronique ont initialement montré deux types de régions distinctes. D'une part, l'euchromatine correspond à des régions opaques aux électrons situées de manière diffuse dans le noyau. Elle témoigne d'une molécule d'ADN relâchée et permissive à la transcription. D'autre part, l'hétérochromatine correspond à des régions denses aux électrons situées sur la périphérie du noyau et au niveau de régions spécifiques comme le nucléole. Elle comprend une molécule d'ADN compactée et réfractaire à la transcription (**figure 1C**). On distingue également deux types d'hétérochromatine. L'hétérochromatine constitutive est identique dans tous les types cellulaires, et permet notamment la répression des éléments génétiques répétés. D'autre part, l'hétérochromatine facultative est variable entre les types cellulaires et permet la répression de gènes impliqués dans le déterminisme cellulaire.

L'interaction entre les différentes molécules d'ADN est donc importante pour organiser le noyau des cellules. Cette interaction peut être étudiée par des techniques moléculaires comme le Hi-C. Cette technologie est basée sur la ligation des brins d'ADN proches dans le noyau et le séquençage haut débit permettant la détermination de toutes les interactions entre les brins d'ADN dans le génome. Ces expériences ont mis en évidence un rôle des protéines CTCF et cohésine dans la délimitation de compartiments qui définissent des régions actives ou inactives pour la transcription des gènes (*Rowley and Corces, 2018*). Ces compartiments sont également enrichis par la présence de modifications d'histones caractéristiques.

1.3. La chromatine

La base moléculaire de la chromatine est le nucléosome. Celui-ci correspond à un complexe nucléoprotéique formé par l'interaction entre la charge positive des histones et la charge négative de l'ADN. Le nucléosome est composé d'un octamère d'histones comprenant 4 homodimères d'histones H2A, H2B, H3 et H4, ainsi que d'une séquence d'ADN de 146 paires de bases (bp) (**figure 2**). L'interaction entre plusieurs nucléosomes permet de compacter l'ADN localement et de créer un environnement répulsif pour la transcription des gènes.

Le positionnement des nucléosomes sur l'ADN peut être modulé par des complexes de remodelage de la chromatine dépendant de l'ATP comme le complexe SWI-SNF par exemple. Ces complexes ont la capacité de rapprocher ou d'incorporer des nucléosomes entre eux, entraînant une compaction locale de la chromatine. Ils permettent également l'espacement ou l'éviction de nucléosomes, entraînant un relâchement de la chromatine (*Narlikar et al., 2013*).

La compaction de la chromatine peut être régulée par des modifications post-traductionnelles des histones. Plusieurs dizaines de modifications ont été répertoriées telles que la méthylation, l'acétylation ou l'ubiquitination qui concernent généralement des résidus lysines de la queue de l'histone H3 (Kouzarides, 2007). Parmi les modifications les plus étudiées, et celles retrouvées dans ce manuscrit, on retrouve H3K9me2/3 (di ou tri-méthylation de la lysine en position 9 de l'histone H3), H3K27me3 (tri-méthylation de la lysine en position 27 de l'histone H3) ou H2AK119ub1 (mono-ubiquitination de la lysine en position 119 de l'histone H2A) qui correspondent à des modifications associées à la répression des gènes, et H3K4me3 (tri-méthylation de la lysine en position 4 de l'histone H3), H3K36me3 (tri-méthylation de la lysine 36 de l'histone H3), H3K9ac (acétylation de la lysine 9 de l'histone H3) ou H3K27a (acétylation de la lysine 27 de l'histone H3) qui correspondent à des modifications associées à l'activation des gènes (**Figure 2 ; Tableau 1**). La modification des histones est réversible et relativement dynamique au cours du déterminisme cellulaire. Elle est régulée par différents complexes multi-protéiques tels que les complexes Polycomb ou les complexes Trithorax par exemple.

Le rôle des modifications d'histones dans la régulation de la chromatine dépend du type de modification. Ainsi, l'acétylation des histones induit une neutralisation de la charge positive de la lysine diminuant l'interaction de l'histone avec l'ADN. A l'inverse, la méthylation n'affecte pas les propriétés biochimiques des histones mais peut servir de signal pour recruter d'autres effecteurs épigénétiques (Bannister and Kouzarides, 2011).

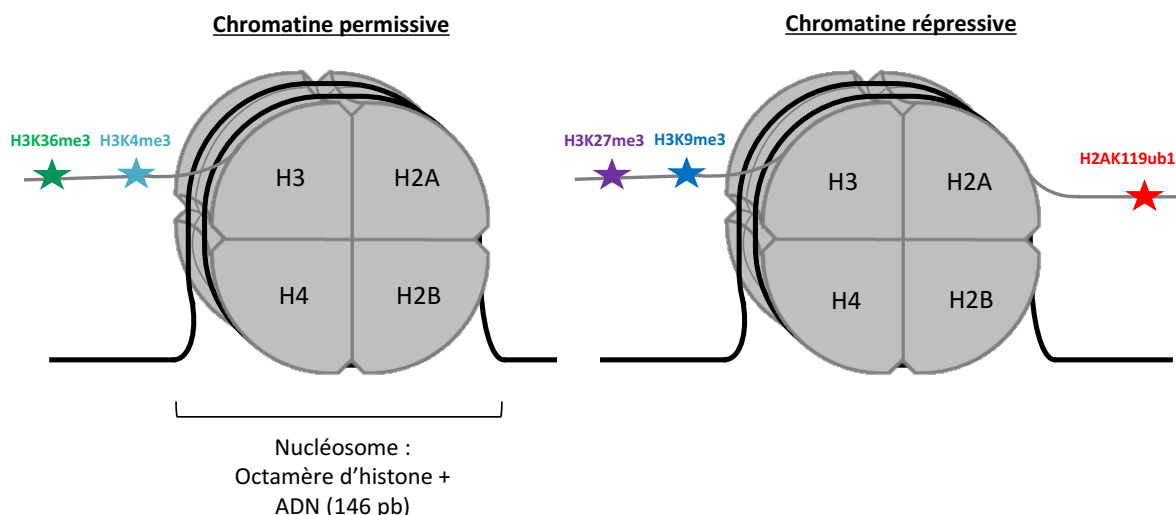


Figure 2 : Le nucléosome est la base moléculaire de la chromatine. Le nucléosome est un octamère d'histones comprenant quatre homodimères d'histone H2A, H2B, H3 et H4 et 146 paires de bases de l'ADN. Des modifications post-traductionnelles des histones peuvent être ajoutées sur la queue de l'histone H3 ou H2A créant un environnement permissif ou répressif pour l'expression des gènes.

marques permissives de la transcription

type de modification	localisation	enzymes impliquées dans la mise en place
H3K4me3	promoteur	SET1, MLL (complexe Trithorax)
H3K36me3	corps de gène	SETD2
H3K9ac	promoteur	GCN5/PCAF
H3K27ac	enhancer	p300/CBP

marques repressives de la transcription

type de modification	localisation	enzymes impliquées dans la mise en place
H3K9me2	promoteur	G9A/GLP
H3K9me3	promoteur	SETDB1, SUV39H1
H2AK119ub1	promoteur	RING1A/B (complexe PRC1)
H3K27me3	promoteur / enhancer	EZH1/2 (complexe PRC2)

Tableau 1 : Les modifications post-traductionnelles des histones. Les différentes modifications d'histones retrouvées dans ce manuscrit, ainsi que leurs localisations génomiques et les enzymes impliquées dans leurs mises en place. Les modifications d'histones peuvent influencer positivement ou négativement l'expression des gènes.

L'incorporation de variants d'histones permet également de réguler la compaction de la chromatine. Ces variants peuvent avoir différents rôles. Par exemple H2A.Z est impliqué dans la régulation de la transcription, tandis que H2A.X est incorporé au niveau des sites de cassures de l'ADN et permet le recrutement de la machinerie de réparation de l'ADN (Talbert and Henikoff, 2017).

1.4. Les modifications de l'ADN

L'expression des gènes peut également être régulée par certaines modifications chimiques de la molécule d'ADN. La modification de l'ADN la plus connue et étudiée est la méthylation du carbone 5 de la cytosine (5mC). Cette méthylation des cytosines est présente dans un contexte de dinucléotide CpG sur la séquence d'ADN et elle est considérée comme un répresseur de l'activité des gènes lorsqu'elle est présente au niveau des séquences régulatrices (Smith and Meissner, 2013).

Bien qu'elle soit relativement conservée au cours de l'évolution, la méthylation des cytosines est absente chez certains organismes modèles tels que *C. elegans*, *D. melanogaster* ou *S. pombe* (**figure 3**).

D'autres modifications chimiques de l'ADN ont été décrites dans le génome des êtres vivants. La méthylation des adénines sur le groupement azote 6 (6mA) a récemment été identifiée chez différentes espèces. Il est intéressant de constater qu'elle est particulièrement présente dans des

génomomes ne montrant pas de méthylation sur les cytosines. Elle a ainsi été retrouvée chez *C. elegans* (Greer et al., 2015), *D. melanogaster* (Zhang et al., 2015) mais également chez la souris (Wu et al., 2016) et chez l'homme (Xiao et al., 2018) (figure 3). De plus, des intermédiaires de la déméthylation de la 5mC comme l'hydroxy-méthylation du carbone 5 des cytosines (5hmC) ont potentiellement un rôle dans la régulation épigénétique des gènes (Shi et al., 2017).

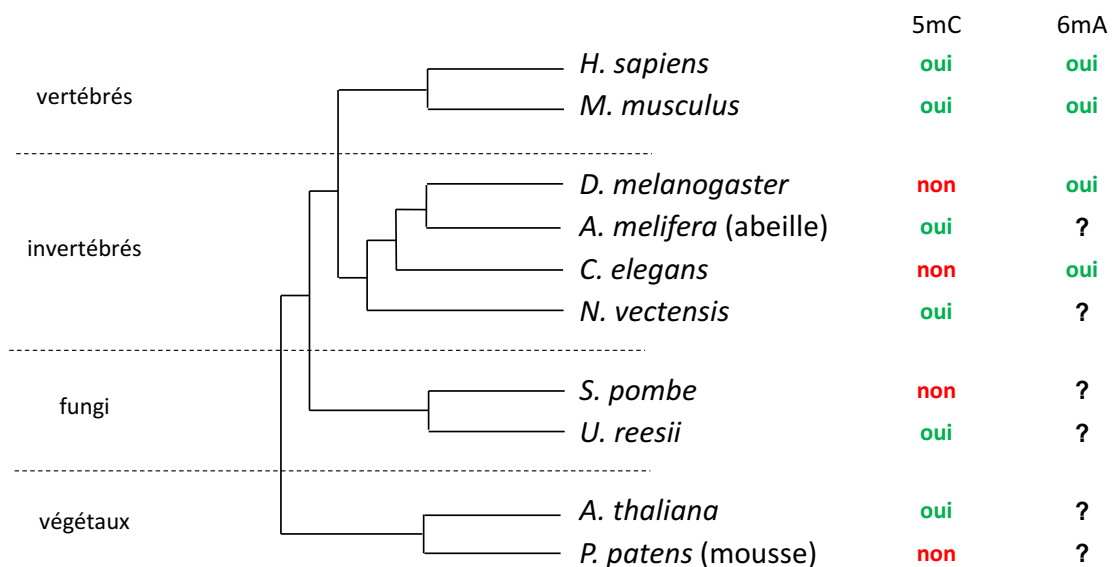


Figure 3 : Conservation de la méthylation de l'ADN au cours de l'évolution. La méthylation des cytosines est relativement conservée dans l'évolution. Celle-ci est néanmoins absente chez certains organismes modèles qui semblent avoir opté pour un mécanisme épigénétique impliquant la méthylation des adénines.

La méthylation du carbone 5 des cytosines de l'ADN est l'objet d'étude de ma thèse. C'est à cette modification que je ferai référence lorsque je discuterai de la méthylation de l'ADN dans la suite de ce manuscrit. Ma thèse est particulièrement focalisée sur le rôle de la méthylation de l'ADN au cours du développement embryonnaire chez la souris. Dans ce contexte, j'ai étudié son implication dans le maintien de l'intégrité transcriptomique, et les mécanismes permettant son ciblage au niveau du génome.

La suite de cette introduction s'attachera à introduire les concepts permettant la compréhension de ces problématiques. Tout d'abord, je présenterai les différents acteurs de la méthylation de l'ADN, ainsi que son rôle dans la régulation épigénétique des gènes. Ensuite, je discuterai la dynamique de la méthylation de l'ADN ainsi que sa fonction au cours du développement embryonnaire. Enfin, je présenterai les complexes Polycomb et leurs liens potentiels avec la méthylation de l'ADN.

2. La méthylation de l'ADN

2.1. Distribution génomique

La compréhension du rôle de la méthylation de l'ADN dans la régulation de l'expression des gènes nécessite de déterminer sa distribution dans le génome. Le développement de technologies permettant de cartographier la méthylation de l'ADN de manière génomique se sont alors développées.

L'utilisation d'anticorps spécifique contre les cytosines méthylées couplées à l'analyse haut débit a permis le développement du MedIP (*Weber et al., 2005*). Cette technique permet de déterminer les régions génomiques enrichies pour la méthylation de l'ADN mais elle ne permet pas de quantifier précisément le niveau de méthylation de chaque cytosine.

Des techniques plus résolutive ont été développées, en couplant le traitement de l'ADN au bisulfite de sodium avec le séquençage haut débit. Le bisulfite de sodium convertit les cytosines non méthylées en uracile puis en thymine après amplification par PCR, alors que les cytosines méthylées sont insensibles à ce traitement chimique (**figure 4A**). Le séquençage de l'ADN traité au bisulfite de sodium et la comparaison avec la séquence de référence du génome permet de déterminer le niveau de méthylation sur chaque cytosine du génome (**figure 4B**). Cette technique appelée WGBS (Whole Genome Bisulfite Sequencing) est la méthode de référence pour cartographier la méthylation du génome, mais est très coûteuse puisqu'elle nécessite de séquencer le génome entier (*Lister et al., 2009*) (**figure 4C**). Le développement d'un protocole alternatif permettant de s'intéresser uniquement aux îlots CpG (voir plus bas) permet de réduire les coûts du séquençage. Ce protocole nécessite de digérer le génome avec l'enzyme de restriction MspI. Celle-ci digère l'ADN au niveau du site de restriction CCGG qui est enrichi dans les îlots CpG. La sélection des fragments d'ADN de petites tailles permet ainsi de séquencer préférentiellement les îlots CpG. Cette technique est alors appelée RRBS (Reduced Representation Bisulfite Sequencing) (*Meissner et al., 2005*) (**figure 4C**).

Des protocoles permettant l'analyse du méthylome sur de faibles quantités d'ADN ont été développés. Les protocoles du RRBS et du WGBS ont été adaptés sur des cellules uniques mais présentent le désavantage de la dégradation de l'ADN par le bisulfite de sodium réduisant ainsi considérablement la couverture du génome (*Guo et al., 2013; Miura et al., 2012; Smallwood et al., 2014*).

Néanmoins, une méthode alternative ne convertissant pas l'ADN au bisulfite de sodium a récemment été développée. Cette technique appelée TAPS couple la déméthylation de l'ADN grâce à l'enzyme

TET et la conversion des intermédiaires de déméthylation en thymine (Liu et al., 2019). Cette méthode est aussi résolutive que le bisulfite de sodium mais présente l'avantage de ne pas dégrader l'ADN. La rendant très prometteuse pour son application dans le séquençage sur cellule unique. L'analyse de méthylome sur cellules uniques est utile pour étudier des phénomènes dans lesquels de faibles quantités de cellules sont disponibles, comme le développement embryonnaire ou des échantillons caractérisés par une grande hétérogénéité cellulaire comme les tumeurs. Enfin, des progrès sur cellules uniques plus récents permettent de combiner ces techniques avec l'analyse du transcriptome ou de l'accessibilité de la chromatine, permettant ainsi d'avoir un aperçu de l'état d'une cellule à un instant donné (Gu et al., 2019; Karemaker and Vermeulen, 2018; Li et al., 2019).

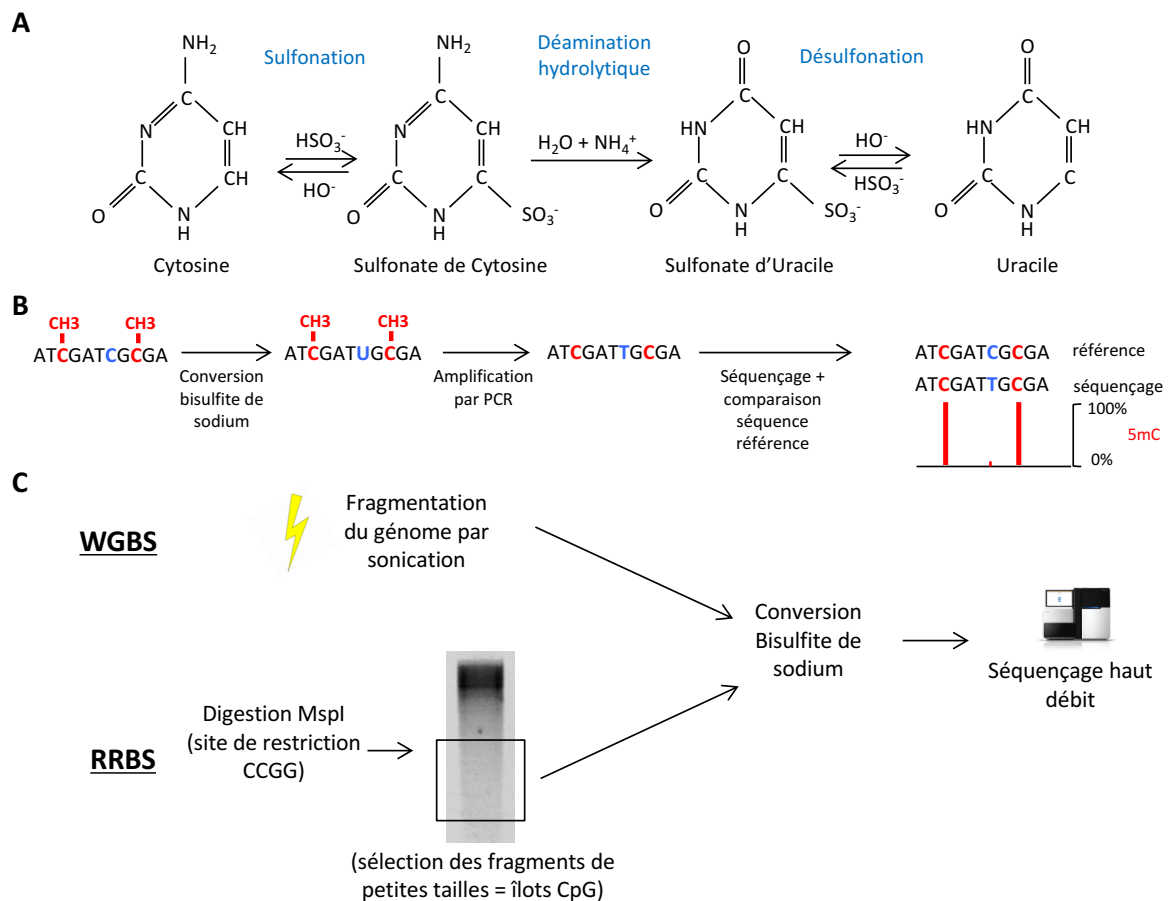


Figure 4 : Techniques permettant l'analyse de la méthylation de l'ADN à l'échelle génomique. A) Le traitement de l'ADN au bisulfite de sodium permet de convertir les cytosines non méthylées en uracile. B) Les cytosines non méthylées (en bleu) sont converties en uraciles, contrairement aux cytosines méthylées (en rouge). La comparaison de la séquence traitée au bisulfite de sodium avec une séquence référence permet de déterminer le niveau de méthylation sur chaque cytosine. C) Le RRBS permet de sélectionner uniquement les îlots CpG permettant de limiter les coûts du séquençage par rapport au WGBS.

L'analyse du méthylome de plusieurs types cellulaires a ainsi révélé que le génome est globalement hyperméthylé, avec un niveau de méthylation d'environ 80% (Lister et al., 2009; Stadler et al., 2011). La méthylation de l'ADN est notamment élevée au niveau des éléments transposables et plus généralement dans l'ensemble des régions intragéniques et intergéniques (**figure 5**). Ce niveau de méthylation est relativement stable entre les différents types cellulaires de l'organisme murin et humain. Les différences entre des cellules de tissus différents sont relativement restreintes et observées principalement dans les séquences régulatrices distantes appelées enhancers (Hon et al., 2013; Ziller et al., 2013).

La majorité de la méthylation est retrouvée dans un contexte de dinucléotide CpG. Néanmoins certaines cellules présentent une méthylation hors CpG dans des contextes CpHpG et CpHpH (où H peut être A, T ou G) (Lister et al., 2009). Cette méthylation est notamment retrouvée dans des cellules qui ne se divisent pas ou peu comme les cellules neuronales ou les ovocytes (Lister et al., 2013; Shirane et al., 2013), ou ayant une forte propension à acquérir la méthylation de l'ADN, comme les cellules souches embryonnaires (ES) (Lister et al., 2009).

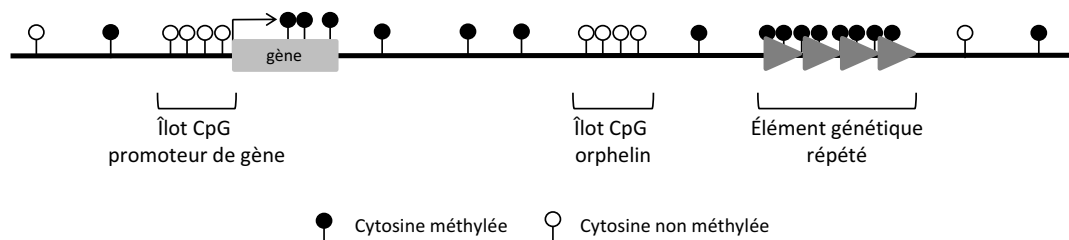


Figure 5 : Distribution génomique de la méthylation de l'ADN. Le génome des mammifères est globalement enrichi en méthylation de l'ADN, notamment au niveau des éléments génétiques répétés. Néanmoins, certaines régions riches en CpG, appelées îlots CpG, sont généralement dépourvues de méthylation de l'ADN.

Bien que le génome soit globalement hyperméthylé, certaines séquences sont dépourvues de méthylation. Celles-ci sont caractérisées par une forte concentration en dinucléotides CpG et ont été nommées îlots CpG (Bird, 1986) (**figure 5**). Les îlots CpG sont définis comme une séquence d'une longueur d'au moins 200 paires de bases, une concentration en CpG supérieure à 50% et un ratio de CpG observés par rapport à attendus supérieur à 0,6. Les mécanismes évolutifs permettant d'expliquer cette forte disparité des concentrations en CpG dans les génomes de mammifères ne sont pas encore pleinement élucidés. Cependant ils semblent être liés au fait que la méthylation de cytosines est mutagène. Ceci est causé par la déamination spontanée des cytosines méthylées en

thymines qui n'est pas réparée efficacement par la machinerie de réparation de l'ADN. Ainsi, les îlots CpG qui sont des régions qui restent généralement hypométhylées auraient résisté à ce phénomène de mutations des CpG au cours de l'évolution (Stadler et al., 2011; Weber et al., 2007).

La moitié des îlots CpG sont localisés au niveau des promoteurs de gènes chez la souris et chez l'homme. Ceux localisés au niveau des séquences intra ou inter-géniques sont appelés îlots CpGs orphelins (Illingworth et al., 2010) (**figure 5**). Les mécanismes impliqués dans l'état d'hypométhylation des îlots CpG semblent être conservés chez différentes espèces de mammifères. En effet, l'insertion d'un chromosome humain dans un cellule murine acquière un profil de méthylation identique par rapport au chromosome endogène humain, suggérant des mécanismes de protection identiques entre les deux espèces (Long et al., 2016). Ces mécanismes seront discutés plus en détail dans le chapitre 2.4 : recrutement de la méthylation de l'ADN.

En plus des îlots CpG, d'autres régions du génome sont hypométhylées. C'est le cas par exemple des « canyons » ou « vallées » qui sont de larges régions génomiques allant jusqu'au 20kb (Jeong et al., 2014; Xie et al., 2013). Ces régions sont enrichies en gènes du développement et par la modification d'histone H3K27me3 (Jeong et al., 2014).

La distribution de la méthylation de l'ADN n'est pas stable au cours de la vie d'un individu. En effet, on observe une large reprogrammation de la méthylation lors de l'embryogenèse précoce et au niveau des cellules germinales (Messerschmidt et al., 2014) (voir chapitre 3 : Rôle de la méthylation de l'ADN au cours du développement embryonnaire). Cette distribution est également altérée dans certaines pathologies comme le cancer.

2.2. Les ADN méthyltransférases

La méthylation des cytosines est catalysée par la famille des ADN méthyltransférases (DNMTs). Elle se compose de méthyltransférases de maintien (principalement DNMT1) qui copient la méthylation sur le brin néo-synthétisé après la réplication de l'ADN, et d'ADN méthyltransférases *de novo* (principalement DNMT3A ou DNMT3B) qui catalysent la méthylation de cytosines non méthylées indépendamment de la réplication de l'ADN (**figure 6A**). Les ADN méthyltransférases possèdent un domaine C-terminal catalytique conservé et un domaine N-terminal régulateur plus variable. Le domaine catalytique est composé de 10 motifs différents. Ils sont impliqués dans la réaction chimique permettant le transfert du groupement méthyl de la S-adenosyl-méthionine (SAM) sur le carbone 5 de la cytosine (Kumar et al., 1994) (**figure 6B-C**). La partie N-terminale régulatrice est

sensiblement différente entre l'enzyme DNMT1 et les enzymes DNMT3, ce qui pourrait contribuer à expliquer leurs rôles distincts dans la méthylation du génome (**figure 6C**).

DNMT1

L'enzyme DNMT1 est la première de la famille des DNMTs à avoir été clonée et étudiée (*Bestor et al., 1988*). Des études biochimiques ont montré qu'elle est capable de méthyler l'ADN hémi-méthylé avec une grande efficacité suggérant qu'elle est impliquée dans le maintien de la méthylation de l'ADN au cours des divisions cellulaires (*Hitt et al., 1988; Pradhan et al., 1999*). La fonction de DNMT1 nécessite également l'action de son cofacteur UHRF1 qui est capable d'interagir avec l'ADN hémi-méthylé par son domaine SRA et recrute DNMT1 aux sites de réplication (*Bashtrykov et al., 2014*). Le rôle de la protéine UHRF1 dans le recrutement de DNMT1 est majeur car des cellules ES mutantes pour *Uhrf1* ou *Dnmt1* présentent un génome globalement hypométhylé de manière équivalente (*Bostick et al., 2007; Sharif et al., 2007*). Récemment, le mécanisme de recrutement de DNMT1 par UHRF1 a été élucidé en montrant que UHRF1 catalyse l'ubiquitination de l'histone H3 qui est ensuite reconnue par le domaine RFT de DNMT1 (*Li et al., 2018a; Nishiyama et al., 2013*). Néanmoins DNMT1 est capable d'interagir avec d'autres protéines régulatrices. C'est le cas par exemple d'USP7 qui interagit avec le domaine BAH de DNMT1 et qui est responsable de sa stabilisation au niveau protéique (*Cheng et al., 2015; Qin et al., 2011*). De plus le domaine PBD de DNMT1 interagit avec la protéine de la fourche de réplication PCNA (*Schermelleh et al., 2007*). Néanmoins cette interaction ne semble pas nécessaire pour le maintien de la méthylation puisqu'une mutation de ce domaine PBD n'induit pas d'altération des profils de méthylation dans les cellules ES (*Garvilles et al., 2015*).

La protéine DNMT1 possède un domaine CXXC qui permet son auto-inhibition pour la méthylation *de novo*. En effet, des études structurales ont montré que l'interaction de DNMT1 avec l'ADN non méthylé provoque un changement de conformation de DNMT1 inhibant son activité catalytique (*Song et al., 2011*). Malgré cette propriété, certaines études suggèrent un rôle de DNMT1 dans la méthylation *de novo*. En effet, les études biochimiques ont montré que DNMT1 est capable de méthyler des brins d'ADN non méthylés, mais moins efficacement que les brins d'ADN hémi-méthylés (*Vilkaitis et al., 2005*). De plus, une modélisation mathématique a proposé un rôle de DNMT1 dans la méthylation *de novo* de certains éléments transposables dans les cellules ES (*Arand et al., 2012*). Les fonctions exactes de DNMT1 dans la méthylation *de novo* du génome sont donc encore à éclaircir. Une isoforme spécifique, DNMT1o, est exprimée durant le développement précoce de l'embryon (*Hirasawa et al., 2008*). Cette isoforme ne possède pas une portion de la partie N-terminale correspondant au domaine DMAP qui est impliqué dans l'interaction avec la protéine DMAP et

HDAC2 (Rountree et al., 2000) (figure 6C). Cette isoforme permet notamment le maintien de la méthylation sur les gènes soumis à l’empreinte dans l’embryon précoce (Cirio et al., 2008; Howell et al., 2001).

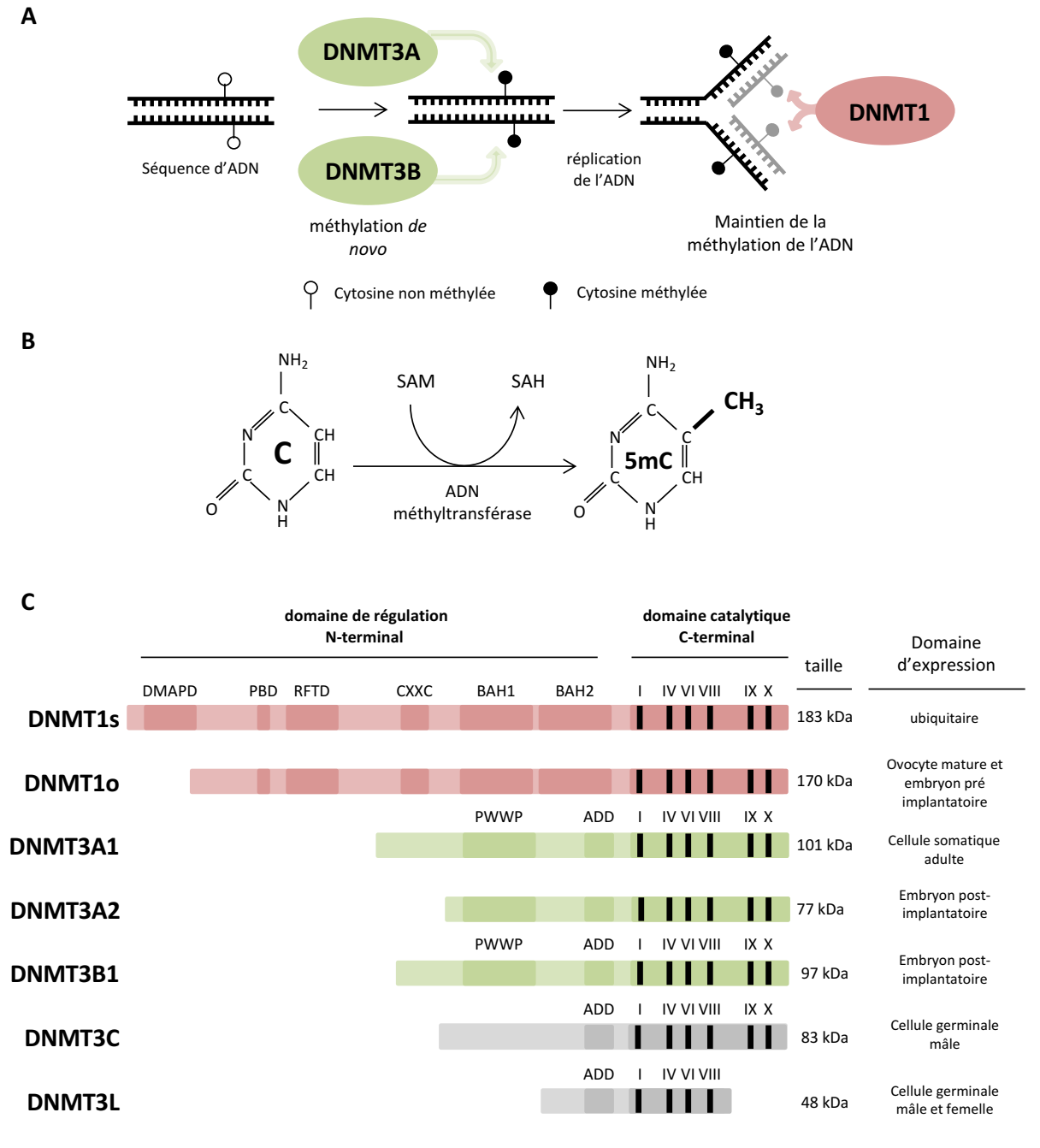


Figure 6 : La méthylation de l’ADN est catalysée par la famille des ADN méthyltransférases. A) Cette famille se décompose d’ADN méthyltransférases *de novo* représentées par DNMT3A et DNMT3B et d’une ADN méthyltransférase de maintien représentée par DNMT1. B) Les DNMTs transfèrent le groupement méthyl de la S-adenosyl-méthionine (SAM) sur le carbone 5 des cytosines. C) Structure des protéines DNMTs chez la souris. Elles possèdent un domaine catalytique conservé et un domaine de régulation N-terminal qui leur confèrent une fonction spécifique dans la méthylation du génome.

DNMT3A et DNMT3B

Les enzymes DNMT3 sont notamment composées de DNMT3A et DNMT3B qui possèdent une grande homologie de séquence d'acides aminés (Okano et al., 1998). Contrairement à l'enzyme DNMT1, elles ont la capacité de méthyler des brins d'ADN non méthylés avec une grande efficacité et sont considérées comme les principales méthyltransférases *de novo* de l'ADN (Okano et al., 1998). En effet, des études génétiques ont confirmé l'implication de ces enzymes dans la méthylation *de novo* de l'ADN en comparant des niveaux de méthylation entre des embryons déplétés pour *Dnmt3a* et *Dnmt3b* et des blastocystes (voir chapitre 3.1.2 : Reprogrammation de l'embryon précoce). Ces études ont montré une absence de méthylation *de novo* dans les embryons comparés au blastocyste sur certains locus spécifiques comme les gènes *Rhox* ou les rétrotransposons IAP (Oda et al., 2006; Okano et al., 1999). Plus récemment, les travaux du laboratoire ont montré que DNMT3B a une contribution plus importante que DNMT3A dans la méthylation du génome au cours de l'embryogenèse chez la souris (Auclair et al., 2014).

Néanmoins, comme pour DNMT1, des études suggèrent que la fonction de ces enzymes pourrait ne pas se restreindre à l'activité de méthylation *de novo*. En effet, de nombreuses études ont rapporté une hypométhylation du génome après l'inactivation de *Dnmt3a* et *Dnmt3b* dans les cellules ES (Chen et al., 2003; Leung et al., 2014a; Li et al., 2015; Thakur et al., 2016) et les fibroblastes embryonnaires murins (MEFs) (Dodge et al., 2005). Ces études suggèrent alors un rôle de DNMT3A et DNMT3B dans le maintien de la méthylation de l'ADN.

Le gène *Dnmt3b* est exprimé préférentiellement au cours du développement embryonnaire, avant d'être pas ou peu exprimé dans les cellules adultes. Plusieurs dizaines d'isoformes de DNMT3B ont été décrites dans la littérature, dont la majorité sont exprimées dans les cellules cancéreuses (Duymich et al., 2016). Ces isoformes sont générées grâce à diverses combinaisons de promoteurs et d'épissages alternatifs. Néanmoins, seules certaines isoformes possèdent un domaine catalytique fonctionnel. La principale isoforme est DNMT3B1, qui est exprimée préférentiellement dans les cellules ES et dans l'embryon (**figure 6C**). Certaines isoformes, inactives d'un point de vue catalytique, jouent un rôle de protéines accessoires. Elles régulent ainsi l'activité des enzymes DNMTs fonctionnelles, à l'image du rôle de DNMT3L (voir plus bas) (Duymich et al., 2016; Gordon et al., 2013).

Les animaux mutants pour *Dnmt3a* sont viables jusqu'à environ 4 semaines après la naissance, laissant supposer un rôle de DNMT3A préférentiellement dans les cellules différenciées (Okano et al.,

1999). On distingue une grande isoforme DNMT3A1 et une petite isoforme DNMT3A2 qui ne possède pas une portion de la partie N-terminale (Chen et al., 2002) (**figure 6C**). Cette dernière est exprimée durant le développement embryonnaire où elle participe à la reméthylation globale du génome avec DNMT3B (Auclair et al., 2014). L'isoforme DNMT3A1 est exprimée plus tardivement dans les cellules somatiques où elle est impliquée dans le déterminisme cellulaire, comme par exemple au sein de la lignée hématopoïétique (Challen et al., 2012; Watanabe et al., 2002).

Les enzymes DNMT3A et DNMT3B possèdent un domaine PWWP qui leur permettent d'interagir avec la modification d'histone H3K36me3 *in vitro* (Dhayalan et al., 2010). Ce domaine permet à l'enzyme DNMT3B d'être recrutée et de méthyler le corps de gènes exprimés dans les cellules ES. Néanmoins DNMT3A2 ne semble pas être impliquée dans cette fonction dans les cellules (Baubec et al., 2015).

Ces enzymes possèdent aussi un domaine ADD qui est conservé entre les enzymes DNMT3, DNMT3L et la protéine ATRX. Ce domaine permet également l'interaction avec la chromatine. Des études biochimiques et structurales ont montré que ce domaine interagit avec H3K4 non méthylé (Ooi et al., 2007; Zhang et al., 2010) et avec H3K9 méthylé (Dhayalan et al., 2011).

Les autres DNMTs

Les autres membres de la famille des ADN méthyltransférases sont DNMT3L et DNMT3C. La protéine DNMT3L est inactive d'un point de vue catalytique mais interagit avec les autres DNMTs stimulant leur activité (Chédin et al., 2002; Jia et al., 2007). Récemment, l'enzyme DNMT3C a été découverte chez la souris. Elle est exprimée exclusivement dans les cellules germinales où elle est impliquée dans la méthylation d'éléments transposables dans les cellules germinales mâles (Barau et al., 2016). Le gène *Dnmt3c* est issu d'une duplication du gène *Dnmt3b* et est présent chez la souris mais pas chez l'Homme. Enfin, l'enzyme DNMT2 a été découverte après DNMT1 car elle possède une forte homologie de séquence avec le domaine catalytique des autres DNMTs. Cependant, elle n'a aucun lien avec la méthylation de l'ADN mais méthyle les ARNt en position C38 et elle a été renommée TRDMT1 (Goll et al., 2006).

Les études contradictoires quant aux rôles des DNMTs dans la méthylation *de novo* et de maintien ont conduit la communauté scientifique à repenser le modèle initial de la méthylation du génome (Jeltsch et Jurkowska, 2014; Jones et Liang, 2009). Une coopération entre les différentes DNMTs

notamment dans le processus de maintien au cours de la réplication a alors été proposée. Néanmoins ce modèle reste à être vérifié *in vivo* et à l'échelle génomique.

2.3. La déméthylation de l'ADN

La méthylation de l'ADN est un processus réversible. La déméthylation des cytosines peut s'opérer de deux manières différentes. D'une part de manière passive, c'est-à-dire par l'absence de l'activité de maintien de la méthylation engendrant une déméthylation globale du génome au cours des divisions cellulaires. D'autre part de manière active, grâce à l'action des protéines TETs permettant la déméthylation ciblée de certaines séquences d'ADN indépendamment de la réplication de l'ADN.

La déméthylation passive

La réplication de l'ADN génère un brin néo-synthétisé qui est méthylé par l'action de DNMT1 et d'UHRF1. Dans le cas où ces enzymes ne réalisent pas cette fonction, une dilution progressive de la méthylation de l'ADN est observée au cours des divisions cellulaires. La déméthylation est alors globale puisqu'il est supposé que DNMT1 est responsable du maintien de la méthylation sur tout le génome (**figure 7**). On observe deux évènements dans lesquels l'ADN est globalement déméthylé de manière passive chez les mammifères : lors de l'embryogenèse précoce et lors de la maturation des cellules germinales primordiales (*Messerschmidt et al., 2014*). Dans les deux cas, DNMT1 et UHRF1 sont exclues du noyau des cellules empêchant le maintien de la méthylation de l'ADN (*Cardoso et Leonhardt, 1999; Seisenberger et al., 2012*).

La déméthylation active

Les mécanismes permettant la déméthylation active n'ont été élucidés que récemment. Chez les trypanosomes, les enzymes JBP1/2 sont capables d'hydroxyler la thymine méthylée permettant sa déméthylation. Une équipe a ainsi étudié la protéine TET1 qui présente une homologie de séquence avec les enzymes JBP1/2. L'expression ectopique de TET1 dans les cellules HEK conduit alors à la déméthylation du génome et à l'accumulation d'un dérivé de déméthylation : la 5-hydroxyméthylcytosine (5hmC) (*Kriaucionis et Heintz, 2009; Tahiliani et al., 2009*). Deux autres paralogues, TET2 et TET3 ont été ensuite caractérisés dans la déméthylation de l'ADN (*Ito et al., 2010*), ainsi que d'autres intermédiaires de déméthylation comme la 5-formylcytosine (5fC) ou la 5-carboxylcytosine (5caC) (*Ito et al., 2011*). Les protéines TETs sont ainsi capables de déméthylater la 5mC et de générer les intermédiaires de déméthylation 5hmC, la 5fC et 5caC par oxydations successives (*Wu et Zhang, 2017*)

(figure 7). Les intermédiaires de la déméthylation sont ensuite excisés du brin d'ADN par l'action des thymine ADN glycosylases (TDG). Cette action conduit à la formation d'un site abasique et une cytosine non méthylée est incorporée par la machinerie de réparation de l'ADN (He et al., 2011; Maiti et Drohat, 2011; Weber et al., 2016). L'inactivation de TDG dans les cellules ES conduit à l'accumulation de 5fC et 5caC mais pas de 5hmC, suggérant que ce dérivé n'est pas affecté par ce mécanisme (**figure 7**) (He et al., 2011; Shen et al., 2013). Les dérivés d'oxydation de 5mC pourraient également favoriser la déméthylation de manière passive car une étude biochimique a montré que DNMT1 a une affinité moindre pour l'ADN hémi-hydroxyméthylé comparé à l'ADN hémi-méthylé (Hashimoto et al., 2012).

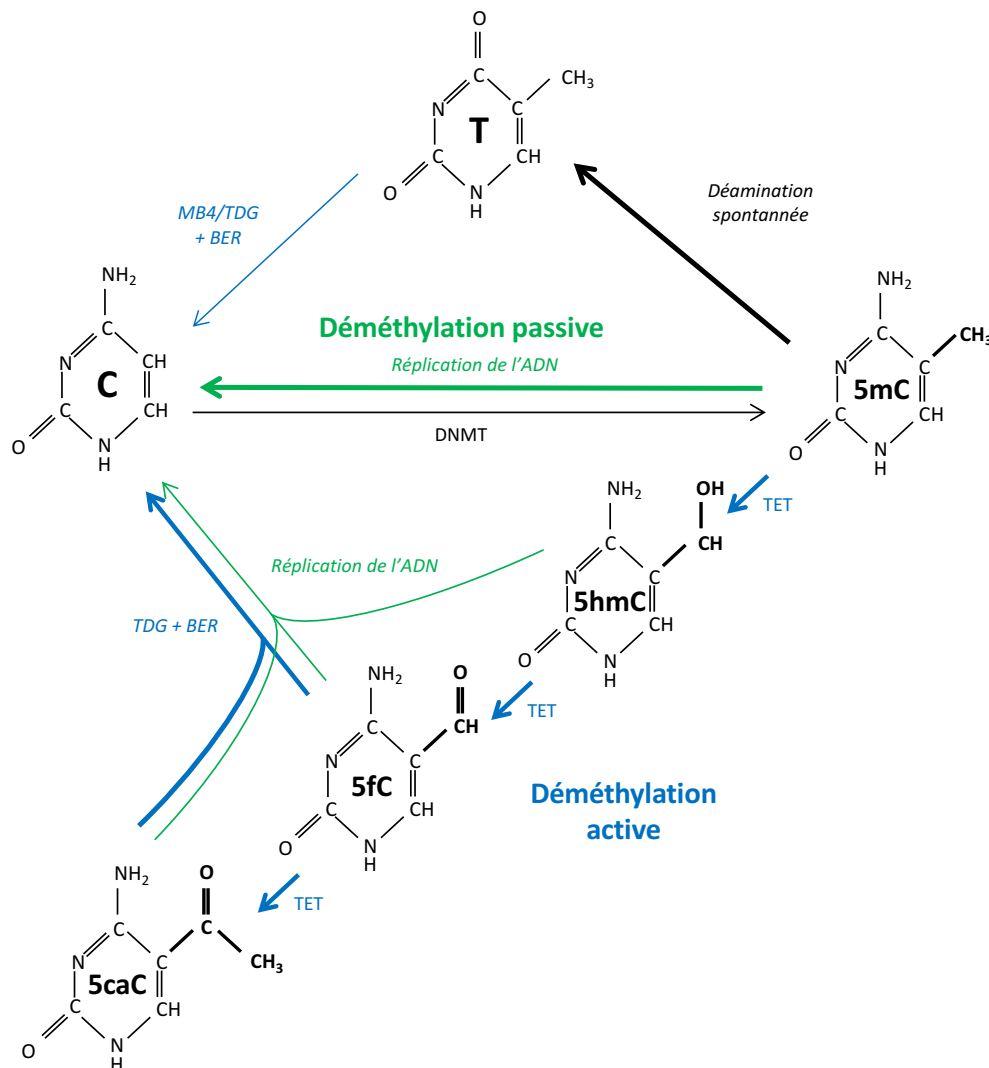


Figure 7 : Les mécanismes de déméthylation de l'ADN. L'ADN peut être déméthylé de manière passive à travers la répllication de l'ADN, ou de manière active via l'action des protéines TETs. Ces dernières génèrent des intermédiaires de déméthylation par oxydations successives qui seront ensuite excisés et remplacés par l'action de la machinerie TDG-BER ou dilués à travers la répllication de l'ADN. Enfin la déméthylation de l'ADN peut être médiée par la déamination spontanée des cytosines méthylées, et la réparation par la machinerie MB4/TDG-BER.

2.4. Recrutement de la méthylation de l'ADN

La distribution de la méthylation de l'ADN dans le génome est influencée d'une part par la nature de la séquence de l'ADN, mais également par l'environnement chromatinien. En effet, le recrutement des DNMTs au niveau de la chromatine peut être modulé par la présence de facteurs de transcription et la présence de modifications d'histones. Ces facteurs peuvent permettre le recrutement de la méthylation via une interaction directe avec les DNMTs, ou inversement peuvent contribuer à protéger certaines régions de la méthylation *de novo*. Plusieurs modèles de recrutement de la méthylation de l'ADN sont discutés ci-dessous.

Rôle de la séquence d'ADN et des facteurs de transcription

Le génome murin présente des îlots CpG principalement hypométhylés. Les rares exceptions dans les promoteurs de gènes sont préférentiellement localisés dans les promoteurs de gènes gamétiques (c'est à dire exprimés spécifiquement dans les cellules germinales) qui acquièrent une forte méthylation de l'ADN au cours du développement (Auclair *et al.*, 2014). Cet adressage spécifique pourrait être réalisé par l'activité du facteur de transcription E2F6 qui se lie au niveau de ces promoteurs et interagit avec DNMT3B (Velasco *et al.*, 2010) (**figure 8A**). D'autres facteurs de transcription ont potentiellement un rôle dans le recrutement de la méthylation de l'ADN sur des régions spécifiques du génome par interaction avec les DNMTs, par exemple MYC ou PU.1 (Brenner *et al.*, 2005; Suzuki *et al.*, 2006).

Inversement, la séquence d'ADN des îlots CpG semble jouer un rôle important dans leur protection contre la méthylation *de novo*. Un grand nombre de facteurs de transcription ont un dinucléotide CpG dans leur motif reconnu sur l'ADN et ont donc une occurrence de leurs motifs enrichi dans les îlots CpG. Certaines études ont montré que la mutation du motif de certains facteurs de transcription dans les cellules ES conduit à une plus forte méthylation des îlots CpG (Hartl *et al.*, 2019; Krebs *et al.*, 2014). Ceci suggère ainsi que la fixation de facteurs de transcription sur les îlots CpG contribue à les protéger de la méthylation de l'ADN. Ces résultats ont confirmé des données plus anciennes montrant par exemple que la mutation du motif reconnu par SP1 dans le promoteur du gène *Aprt* induit sa méthylation dans les cellules ES (Brandeis *et al.*, 1994) (**figure 8B**).

La densité en CpG des îlots semble également jouer un rôle dans leur protection contre la méthylation de l'ADN, puisqu'une anti corrélation entre la densité en CpG et la probabilité de méthylation a été observée. En effet, les îlots CpG de promoteurs de gènes ayant une densité élevée

en CpG sont généralement hypométhylés, tandis que ceux ayant une densité intermédiaire sont plus fréquemment méthylés dans les cellules mammifères (Weber et al., 2007). Les génomes de mammifères codent pour plusieurs protéines à domaine CXXC qui reconnaissent les CpGs non méthylés. C'est le cas notamment des protéines TETs qui sont enrichies préférentiellement aux niveaux des îlots CpG et contribuent à maintenir leur état d'hypométhylation (Dawlaty et al., 2014; Verma et al., 2018; Wu et al., 2011) (figure 8C). La protéine KDM2B (appelée également FBXL10) est une autre protéine à domaine CXXC qui se lie aux îlots CpG et son inactivation conduit à une hyperméthylation de nombreux îlots CpG dans les cellules ES murines (Boulard et al., 2015). D'autres protéines ayant un domaine CXXC sont enrichies au niveau des îlots CpG comme par exemple la protéine KDM1A qui induit la déméthylation de l'histone H3K36 (Blackledge et al., 2010) ou du facteur de transcription CFP1 qui fait partie d'un complexe contenant SET1 responsable de la méthylation de H3K4 (Thomson et al., 2010).

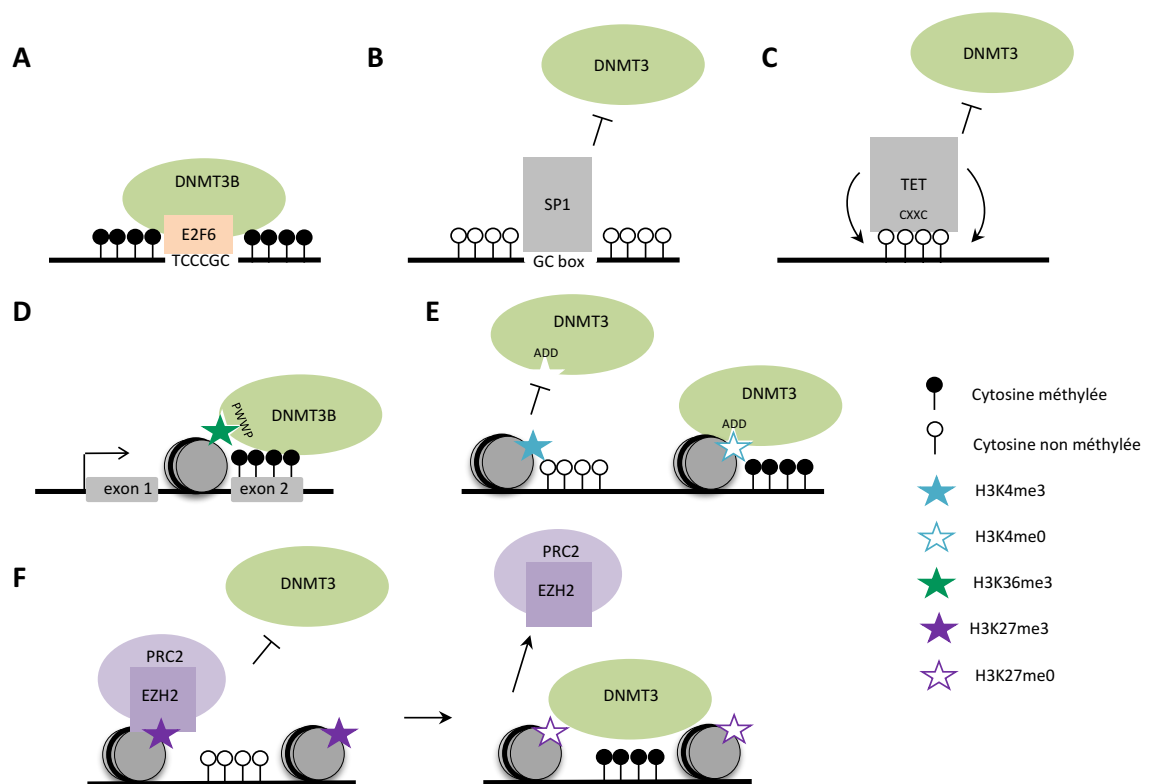


Figure 8 : Les mécanismes de recrutement de la méthylation de l'ADN. A-B-C) La séquence d'ADN est un paramètre important pour réguler la méthylation de l'ADN. Certains facteurs de transcription peuvent recruter *en cis* les DNMTs (A), ou empêcher le recrutement de la méthylation (B). Certaines protéines se liant au niveau des cytosines non méthylées par leur domaine CXXC peuvent également protéger contre la méthylation de l'ADN (C). D-E-F) La chromatine joue également un rôle dans la régulation de la méthylation. La marque d'histone H3K36me3 permet le recrutement direct de DNMT3B (D), tandis que les marques H3K4me3 (E) et H3K27me3 (F) protègent de l'acquisition de la méthylation de l'ADN.

Rôle de la chromatine

Les modifications post-traductionnelles des histones jouent également un rôle important dans la régulation de la méthylation de l'ADN. C'est le cas de la modification d'histone H3K36me3 qui est enrichie au niveau des corps de gènes exprimés. Le domaine PWWP de DNMT3B lui permet d'interagir avec cette marque et d'adresser la méthylation au niveau de corps de gènes exprimés dans les cellules ES (Baubec et al., 2015) (figure 8D). Des études biochimiques ont démontré que le domaine PWWP de DNMT3A a également la capacité d'interagir avec H3K36me3 (Dhayalan et al., 2010), néanmoins son recrutement par H3K36me3 n'a pas été validé dans les cellules ES, ni *in vivo* chez la souris (Baubec et al., 2015; Sendžikaitė et al., 2019).

La méthylation de l'ADN peut également être régulée par la marque d'histone H3K9me3. Cette marque d'hétérochromatine est enrichie au niveau des séquences génétiques répétées et au niveau de promoteurs de certains gènes réprimés. En effet, plusieurs études ont montré que l'inactivation de SETDB1 (appelée aussi ESET), une histone méthyltransférase impliquée dans la déposition de H3K9me3, altère le recrutement de la méthylation de l'ADN au niveau de rétrotransposons et de certaines séquences géniques dans les cellules ES et *in vivo* (Leung et al., 2014a; Liu et al., 2014; Rowe et al., 2013). Certaines études suggèrent également que la marque H3K9me3 est impliquée dans le maintien de la méthylation de l'ADN des gènes soumis à l'empreinte parentale au cours du développement (voir chapitre 3.3 : régulation des gènes soumis à l'empreinte). Bien que le lien entre H3K9me3 et la méthylation de l'ADN soit établi, les mécanismes moléculaires permettant leur interaction ne sont pas connus. Une étude biochimique a proposé que le domaine ADD des protéines DNMT3 interagisse préférentiellement avec l'histone H3 contenant la modification H3K9me3 (Dhayalan et al., 2011). Néanmoins l'importance fonctionnelle de cette interaction n'a pas été testée dans les cellules.

Par opposition, la présence de certaines marques d'histones est anti-corrélative de la méthylation de l'ADN. C'est le cas par exemple de H3K4me3 qui est spécifiquement enrichie au niveau des îlots CpG non méthylés (Meissner et al., 2008; Weber et al., 2007). Ceci suggère que la présence de cette modification d'histones contribue à la protection des îlots CpG contre la méthylation de l'ADN. Le mécanisme permettant l'exclusion des DNMTs par H3K4me3 fait intervenir le domaine ADD des protéines DNMT3s. En effet, plusieurs études ont montré que le domaine ADD interagit efficacement avec l'histone H3 seulement en absence de méthylation de H3K4, ce qui contribue à défavoriser l'action des DNMTs au niveau des régions génomiques fortement méthylées sur H3K4 (Ooi et al., 2007; Otani et al., 2009; Zhang et al., 2010) (figure 8E).

La présence de la modification H3K27me3 médiée par le complexe Polycomb PRC2, est également antagoniste de la méthylation de l'ADN (*Meissner et al., 2008; Mohn et al., 2008*). C'est particulièrement le cas au niveau des promoteurs de gènes riches en CpG dits "bivalents" qui sont marqués par des niveaux élevés de H3K4me3 et de H3K27me3 ainsi que par une absence de méthylation de l'ADN. Néanmoins le lien entre ces deux marques épigénétiques est encore partiellement ambigu, notamment dans le fait de savoir si c'est H3K27me3 qui exclut la méthylation de l'ADN ou l'inverse. L'inactivation de PRC2 n'induit pas d'hyperméthylation de l'ADN massive des îlots CpG dans les cellules ES (*Boulard et al., 2015*). Cependant, une autre équipe a observé qu'une mutation perte de fonction du complexe PRC2 dans les cellules ES, qui induit une perte globale de H3K27me3, conduit à l'accumulation de la méthylation de l'ADN dans les "canyons", qui sont des régions normalement hypométhylées (*Li et al., 2018b*). Ceci suggère ainsi un mécanisme de protection contre la méthylation de l'ADN par H3K27me3 dans ces « canyons » (**figure 8F**). Inversement, il a été montré que la perte de méthylation de l'ADN dans les cellules ES induit une redistribution des profils de méthylation de H3K27 et un gain de H3K27me3 dans de nombreuses régions génomiques, préférentiellement au niveau de régions déjà silencieuses (*Brinkman et al., 2012; Hagarman et al., 2013; King et al., 2016; Walter et al., 2016*). Le rôle négatif de la méthylation de l'ADN sur le recrutement de H3K27me3 a également été observé dans une autre étude visant à déterminer les séquences d'ADN minimales permettant le recrutement du complexe PRC2 dans les cellules ES. Cette étude a conclu que le recrutement de H3K27me3 nécessite des séquences riches en CpG et nécessairement protégées de la méthylation de l'ADN (*Jermann et al., 2014*). En résumé, ces études montrent dans leur ensemble que le lien entre la méthylation de l'ADN et la méthylation de H3K27 est complexe et probablement dépendent du contexte génomique.

2.5. Conséquences de la méthylation de l'ADN

La méthylation de l'ADN est une marque épigénétique associée à la répression de la transcription des gènes. Elle peut inhiber la fixation de facteurs de transcription ou servir également de signal permettant le recrutement d'autres effecteurs. Ces effecteurs peuvent être des protéines reconnaissant la méthylation des cytosines ou des facteurs de transcription reconnaissant un motif d'ADN méthylé.

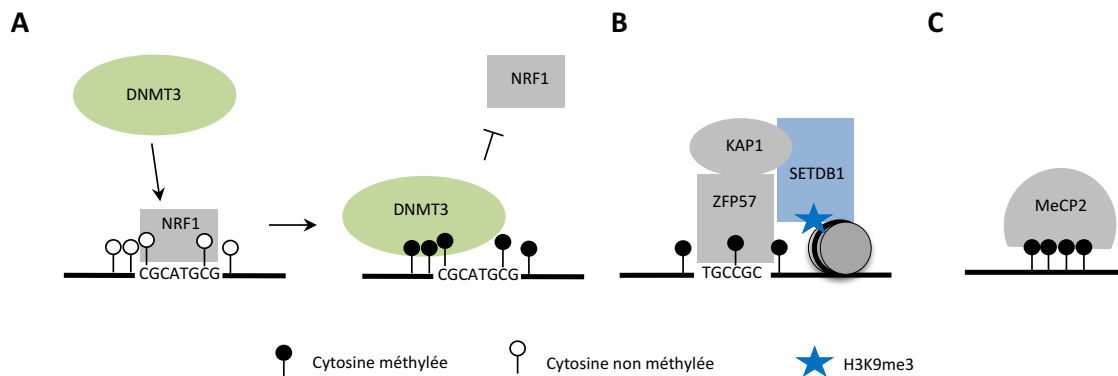


Figure 9 : Conséquences de la méthylation de l'ADN. L'acquisition de la méthylation permet de moduler la fixation de facteurs de transcription. Par exemple, NRF1 qui se lie à un motif non méthylé est évincé par la méthylation de ce motif (B), et ZFP57 se lie exclusivement sur un motif d'ADN méthylé (C). Certaines protéines se liant aux cytosines méthylées sont également recrutées comme MeCP2 (D).

Régulation de la fixation de facteurs de transcription

Il a été proposé depuis de nombreuses années que la méthylation de l'ADN inhibe la fixation de facteurs de transcription. Ainsi, de nombreuses études ont montré que la présence de 5mC inhibe la liaison de divers facteurs de transcription dans des tests d'affinité *in vitro*. Cependant, il reste dans la plupart des cas à démontrer si cela a une influence réelle sur la liaison des facteurs de transcription dans un contexte chromatinien. Par exemple, cela a été démontré dans le cas du facteur de transcription NRF1 dont la fixation est inhibée lorsque son motif est méthylé et qui se fixe à de nombreux sites de liaison *de novo* en absence de méthylation de l'ADN dans les cellules ES TKO (triple knock-out) (Domcke et al., 2015) (figure 9A).

A l'opposé, la méthylation de certains motifs d'ADN peut entraîner le recrutement de facteurs de transcriptions qui ont une affinité préférentielle pour l'ADN méthylé. C'est le cas par exemple de ZFP57 qui se lie à son motif TGCCGC préférentiellement lorsqu'il est méthylé et qui est impliqué dans le recrutement de KAP1 et le maintien de la méthylation de l'ADN des gènes soumis à l'empreinte parentale (Liu et al., 2014; Quenneville et al., 2011) (figure 9B).

La méthylation de l'ADN peut également servir de signal pour le recrutement de protéines de la famille des MBD (Methyl-CpG-binding domain proteins), tel que MeCP2 (Nan et al., 1998) (figure 9C). Plusieurs membres de cette famille de MBD ont été décrits et sont capables de lier la méthylation sur le génome des cellules ES (Baubec et al., 2013).

Rôle dans la régulation de l'intégrité transcriptomique

La méthylation de l'ADN joue un rôle important dans de nombreux processus de régulation transcriptomique tels que la répression des éléments transposables, la régulation des gènes soumis à l'empreinte parentale, l'inactivation du chromosome X ou la régulation de l'expression des gènes (voir le chapitre 3 : Rôle de la méthylation de l'ADN au cours du développement embryonnaire).

La présence de la méthylation des cytosines au niveau de promoteurs de gènes n'induit pas forcément une répression du promoteur. L'effet de la méthylation d'un promoteur d'un gène sur sa régulation dépend principalement de sa densité en CpG. Ainsi, une forte densité de cytosines méthylées dans un promoteur riche en CpG induit une répression forte et stable du promoteur, tandis que la méthylation d'un promoteur faible en CpG n'est généralement pas suffisante pour empêcher son activation (*Weber et al., 2007*).

La méthylation de l'ADN dans les régions inter-géniques peut également participer à la régulation de l'expression génique en modulant l'activité et la liaison de facteurs de transcription au niveau des enhancers. Les expériences de cartographie de méthylation ont démontré invariablement dans de nombreux types cellulaires que les séquences d'enhancers actifs sont hypométhylées (*Hon et al., 2013; Stadler et al., 2011; Ziller et al., 2013*). Une nouvelle fois, la question de la cause et la conséquence a été posée et il semble que dans la majorité des cas, la déméthylation des enhancers est simplement une conséquence de la liaison de facteurs de transcription (*Stadler et al., 2011*). Cependant cela n'exclut pas un rôle important de la méthylation de l'ADN pour réguler l'accessibilité des enhancers (*Mahé et al., 2017*).

Contrairement aux autres régions génomiques, la méthylation des séquences intra-géniques corrèle positivement avec la transcription. Cette méthylation est permise par le recrutement de DNMT3B par la marque active de la transcription H3K36me3 (*Baubec et al., 2015*). Le rôle de cette méthylation n'est pas encore élucidé. Certaines études suggèrent que la méthylation sert à empêcher l'activation de promoteurs cryptiques générant des protéines tronquées (*Neri et al., 2017*). D'autres ont proposé un rôle dans la régulation de l'épissage alternatif (*Shukla et al., 2011*).

2.6. Implication de la méthylation de l'ADN dans les pathologies

2.6.1. Les cancers

Des différences de méthylation de l'ADN ont été observées il y a longtemps entre des cellules cancéreuses et des cellules saines (*Feinberg et Vogelstein, 1983*). Néanmoins ce n'est qu'avec l'avènement des technologies de cartographie de la méthylation à haut débit que l'on a pu caractériser avec plus de précisions les perturbations des profils de méthylation dans les cellules cancéreuses. On a ainsi observé un phénomène d'hypométhylation globale et un phénomène d'hyperméthylation locale dans les cellules cancéreuses.

Profil de la méthylation de l'ADN dans les cancers

La première cartographie complète du méthylome d'une tumeur de colon par WGBS a révélé une hypométhylation globale des cellules cancéreuses couvrant une grande partie du génome (40%). Cette hypométhylation est répartie en larges domaines de 100kb à 1 Mb, qui ont été nommés domaines partiellement méthylés (PMDs) car ils comprennent un niveau de méthylation compris entre 20 et 60% (*Berman et al., 2011*). Ces régions interagissent avec la lamine nucléaire et sont répliquées de manière tardive au cours du cycle cellulaire. Par la suite, ces larges blocs d'hypométhylation ont été observés dans d'autres types de cancers et sont considérés comme une signature épigénétique cancéreuse (*Hon et al., 2012; Vidal et al., 2017*). L'hypométhylation de ces régions est progressive suivant l'évolution de la tumeur et peut donc servir de marqueur biologique de l'avancée de l'état tumoral (*Leygo et al., 2017; Vidal et al., 2017*). De manière intéressante on retrouve ce phénomène d'hypométhylation dans d'autres contextes biologiques comme par exemple lors de l'immortalisation de fibroblastes humains (*Lister et al., 2009*) ou du vieillissement cellulaire.

Une des conséquences de l'hypométhylation globale du génome est une possible activation d'éléments transposables qui peuvent se réinsérer aléatoirement dans le génome et ainsi favoriser l'instabilité génomique qui est un paramètre important du cancer (*Hanahan et Weinberg, 2011*). La méthylation de l'ADN étant connue pour réprimer les gènes gamétiques dans les cellules somatiques, l'hypométhylation du génome peut également favoriser la dérégulation de ces gènes dans les cellules cancéreuses (*Rousseaux et al., 2013*). Du fait de leur expression exclusive dans les cellules germinales ou les cancers, ces gènes sont aussi appelés gènes cancer-testis. Des anticorps dirigés contre les antigènes cancer-testis codés par ces gènes sont des outils de choix comme biomarqueurs

ou à des fins de ciblage thérapeutique (Gjerstorff et al., 2015). Etant donné l'importance grandissante de ces antigènes cancer-testis dans les stratégies thérapeutiques anti-cancéreuses, la compréhension des mécanismes de régulation de ces gènes par la méthylation de l'ADN est une question importante.

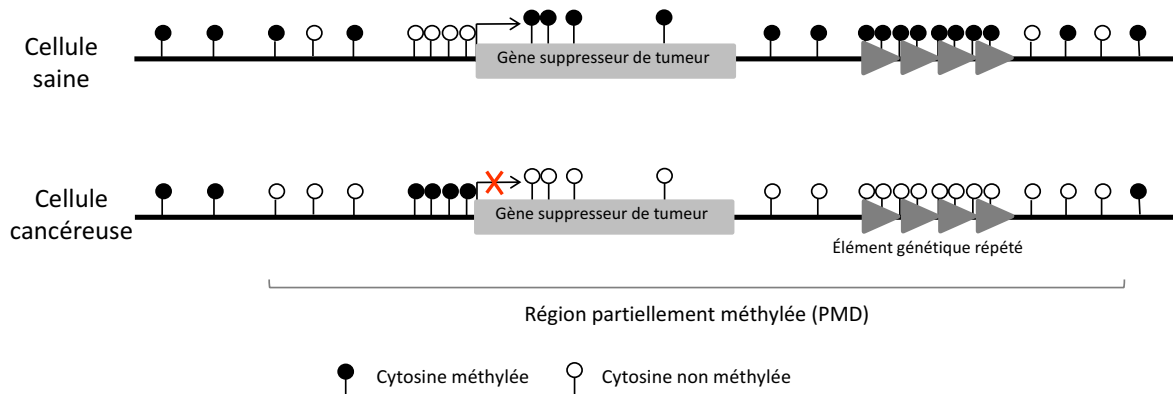


Figure 10 : Profil de la méthylation de l'ADN dans les cancers. Le profil de méthylation des cellules cancéreuses est caractérisé par une hypométhylation globale, conduisant à la dérégulation de gènes gamétiques et d'éléments transposables, et une hyperméthylation locale causant la répression de gènes suppresseurs de tumeurs. Ce profil de méthylation conduit ainsi à une instabilité génomique.

En parallèle, les cellules cancéreuses sont caractérisées par une fréquente hyperméthylation locale des îlots CpG. Cette hyperméthylation aberrante d'îlots CpG contribue à réprimer certains promoteurs de gènes suppresseurs de tumeurs qui ont des fonctions importantes dans la prolifération cellulaire, l'apoptose ou la stabilité génomique (**figure 10**). Cependant, ces hyperméthylations d'îlots CpG ont lieu préférentiellement sur des îlots CpG marqués par la marque d'histone H3K27me3 au cours du développement (Schlesinger et al., 2007), indiquant que ce phénomène concerne des gènes qui étaient en grande partie initialement silencieux. Un des objectifs à l'avenir est donc de discriminer les changements de méthylation de l'ADN "silencieux" de ceux qui ont un rôle causal dans la dérégulation de l'expression des gènes dans les cellules cancéreuses (Pfeifer, 2018).

Causes des défauts de méthylation dans les cancers

Les défauts de méthylation de l'ADN dans les cancers peuvent être causés par un manque de contrôle ou de maintien de la méthylation dû à une prolifération rapide des cellules, ou dans certains types de cancer, à des mutations de gènes impliqués dans le contrôle de la méthylation de l'ADN. C'est le cas des cancers hématologiques qui présentent des mutations fréquentes de gènes codant pour des effecteurs de la méthylation de l'ADN. Par exemple, *DNMT3A* est muté dans 25% des cas de cancers de leucémies myéloïdes aiguës (AML) (Ley et al., 2010). De plus *TET2* est également fréquemment muté dans 10-25% des cas des AMLs et syndromes myélodysplasiques (MDS) (Delhommeau et al., 2009), ainsi que près de 50% des cas de leucémies myéloïdes chroniques (CMML) (Kosmider et al., 2009). L'inactivation de *TET2* dans des cellules humaines est même capable d'initier le processus oncogénique (Rasmussen et Helin, 2016). Les mutations du gène *DNMT3A* jouent un rôle causal dans le développement cancéreux car elles sont présentes à des stades très précoces et même plusieurs années avant le déclenchement de la maladie (Desai et al., 2018; Shlush et al., 2014).

Le vieillissement cellulaire correspond à une perte progressive des fonctions cellulaires causées par un raccourcissement des chromosomes dû à l'absence des télomérases. Des changements de profil de méthylation de l'ADN surviennent alors au cours du vieillissement cellulaire et sont identiques à ceux observés dans le cancer (Kulis et al., 2015). On observe ainsi par exemple une hypométhylation globale de l'ADN au niveau de larges régions génomiques qui correspondent aux larges régions hypométhylées caractéristiques des cancers (Yuan et al., 2015). Cela suggère qu'une partie des changements de méthylation de l'ADN observés dans les cancers sont dus à une dérive de profils de méthylation au cours du vieillissement.

2.6.2. Syndromes

Le syndrome ICF

Le syndrome ICF est une maladie rare affectant une cinquantaine de personnes dans le monde qui est caractérisé par une immunodéficience, de l'instabilité génomique, et des anomalies de formation de la face. On distingue l'ICF de type 1 affectant la moitié des patients et qui est caractérisé par une mutation hypomorphe dans le gène *DNMT3B* (Xu et al., 1999). Les mutations se trouvent généralement au niveau du domaine catalytique et quelques fois dans le domaine PWWP (Ehrlich et al., 2008). L'ICF de type 2-4 est caractérisé par des mutations dans d'autres gènes, comme *ZBTB24* (de

Greef et al., 2011), *CDC7* ou *HELLS* (*Thijssen et al., 2015*). De manière intéressante, ces protéines sont des régulateurs potentiels de l'activité de DNMT3B (*Thompson et al., 2018; Yu et al., 2014*). Alors que tous les patients ICF présentent une hypométhylation caractéristique des régions péri-centromériques, les patients ICF1 présentent une hypométhylation des îlots CpG des gènes gamétiques qui les distinguent des patients ICF2-4 (*Velasco et al., 2018*).

Le syndrome de Rett

Le syndrome de Rett est une maladie neurologique dégénérative caractérisée par une mutation au niveau du gène *Mecp2* localisé sur le chromosome X (*Amir et al., 1999*). Du fait de sa localisation chromosomique, les garçons porteurs de la mutation ne survivent pas au-delà de deux ans d'âge, tandis que les filles sont mosaïques du fait de l'inactivation aléatoire du chromosome X, et survivent. La protéine MECP2 est capable d'interagir avec l'ADN méthylé et participe à la répression épigénétique de certains gènes (*Ehrhart et al., 2016*).

Le syndrome de l'X fragile

C'est une maladie neurologique caractérisée par un défaut d'expression du gène *Fmr1* impliqué dans le développement du cerveau. Le gène *Fmr1* est localisé au niveau d'une région télomérique ayant un motif CGG répété une cinquantaine de fois au niveau de son promoteur. Chez les individus porteurs de la maladie, ce motif CGG est répété 200 fois ou plus créant un îlot CpG, qui acquiert la méthylation de l'ADN et réprime le gène (*Verkerk et al., 1991*). Récemment une équipe a réussi à réactiver le gène *in vivo* dans un modèle murin en induisant une déméthylation ciblée avec la protéine TET1 fusionnée à la protéine Cas9 inactive (*Liu et al., 2018b*).

3. Rôle de la méthylation de l'ADN au cours du développement embryonnaire

3.1. Reprogrammation au cours du développement

Comme énoncé précédemment, la distribution de la méthylation de l'ADN n'est pas stable au cours du développement embryonnaire. Deux phases de reprogrammation épigénétique (incluant la méthylation de l'ADN) ont lieu au cours du développement dans les cellules germinales primordiales et lors de l'embryogenèse précoce. Le rôle de ces reprogrammations épigénétiques n'est pas entièrement élucidé, néanmoins il est probable qu'elles contribuent à restaurer l'état de pluripotence/totipotence cellulaire.

Dans un premier temps, je présenterai la dynamique de la méthylation de l'ADN au cours de ces phases de reprogrammation. Dans un deuxième temps, nous discuterons de son rôle dans la régulation de plusieurs éléments génomiques comme les éléments transposables ou l'empreinte génomique parentale, et enfin son rôle dans la régulation de l'identité cellulaire.

3.1.1. Reprogrammation dans la lignée germinale

Une importante reprogrammation épigénétique a lieu dans les cellules germinales. Celle-ci permet d'activer le programme gamétique et accompagne la différenciation des cellules germinales primordiales (PGCs). Les précurseurs de ces cellules les plus précoces ont été localisés dans l'épiblaste embryonnaire au stade embryonnaire E6.25. Ces cellules migrent ensuite jusqu'à l'emplacement des futures gonades du stade E8 à E11 (*Lesch et Page, 2012*). Au cours de leur migration, les cellules germinales primordiales subissent une déméthylation globale de leur génome passant de 70% à environ 5%. Après le déterminisme sexuel de l'embryon (au stade embryonnaire E12,5) les PGCs réacquièrent la méthylation de l'ADN à des stades différents selon le développement des spermatozoïdes ou des ovocytes (**figure 11**). Durant la phase de reméthylation, le spermatozoïde acquiert une forte méthylation du génome de l'ordre de 90% (*Popp et al., 2010*), alors que l'ovocyte acquiert une méthylation intermédiaire de l'ordre de 40% (*Smallwood et al., 2011*) (**figure 11**).

Déméthylation des cellules germinales primordiales

La déméthylation des cellules germinales primordiales se déroule entre les stades E7.5 et E12.5 atteignant un niveau de méthylation le plus bas de l'organisme (5%) (Kobayashi et al., 2013; Seisenberger et al., 2012). Cette déméthylation inclue toutes les régions génomiques y compris les gènes soumis à l'empreinte, alors que seuls certains éléments transposables résistent partiellement à la déméthylation (Guibert et al., 2012; Seisenberger et al., 2012). Le mécanisme de déméthylation global des PGCs semble être majoritairement passif. En effet les protéines impliquées dans le maintien de la méthylation comme DNMT1 ou UHRF1 sont faiblement actives car majoritairement exclues du noyau des cellules germinales (Kagiwada et al., 2013; Seisenberger et al., 2012). Une contribution des protéines TETs à la phase de déméthylation est suggérée par la forte expression des enzymes TET1 et TET2 tout comme la présence de l'intermédiaire de déméthylation 5hmC dans les cellules germinales (Hackett et al., 2013; Vincent et al., 2013). La déméthylation induite par les protéines TETs semble cependant avoir une contribution limitée puisqu'on observe pas de différences majeures de méthylation entre des PGC WT (wild type, sauvage) et KO (knock out) pour TET1 (Hill et al., 2018). Elle semble néanmoins nécessaire pour la déméthylation complète de certains gènes gamétiques (Yamaguchi et al., 2012) et des gènes soumis à l'empreinte parentale (Yamaguchi et al., 2013).

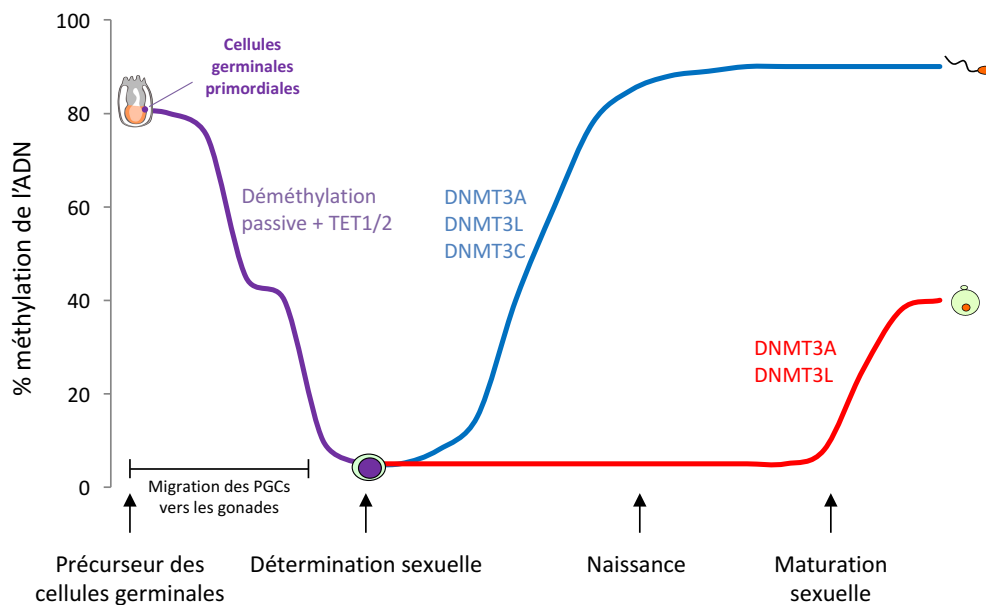


Figure 11 : Dynamique du méthylome des cellules germinales primordiales. Les précurseurs des cellules germinales subissent une déméthylation globale de la méthylation de l'ADN. Après la détermination sexuelle, les cellules germinales sont reméthylées à des stades différents entre les individus mâles et femelles.

Reméthylation des cellules germinales primordiales

La phase de reméthylation du génome des cellules germinales qui s'en suit a lieu à des âges différents chez les mâles et femelles. Les PGC mâles sont reméthylées rapidement à partir du stade E14.5 pour atteindre 90% de méthylation à 5-6 jours après la naissance (**Figure 11**) (Popp et al., 2010). Cette méthylation requiert l'activité de DNMT3A (Kaneda et al., 2004), de DNMT3L (Bourc'his et al., 2001) et de DNMT3C qui s'est spécialisée dans la méthylation de certains éléments transposables (Barau et al., 2016). La phase de reméthylation est cruciale pour la différenciation des spermatozoïdes. En effet, la mutation de chacune de ces enzymes dans les cellules germinales mâles conduit à des défauts méiotiques et une infertilité des animaux (Barau et al., 2016; Bourc'his et al., 2001; Kaneda et al., 2004).

Dans les cellules germinales femelles, la phase de méthylation est plus tardive et a lieu au moment de la phase de croissance de l'ovocyte après la puberté (**figure 11**) (Smallwood et al., 2011). Le méthylome des ovocytes se distingue notamment par une grande proportion de méthylation non CpG comme CpHpH ou CpHpG (Shirane et al., 2013). Ceci peut s'expliquer par le fait que ces cellules restent en phase de quiescence pendant une longue période sans division cellulaire. La méthylation *de novo* dans les ovocytes dépend surtout de l'activité de DNMT3A et DNMT3L mais ne semble pas impliquer DNMT3B (Shirane et al., 2013; Smallwood et al., 2011). Une grande proportion de la méthylation des ovocytes est acquise au niveau de corps de gènes exprimés, suggérant un lien entre la transcription et la méthylation *de novo*. Ce procédé permet notamment la méthylation des gènes soumis à l'empreinte maternelle (Veselovska et al., 2015). Les mécanismes de recrutement de la méthylation de l'ADN dans l'ovocyte sont partiellement élucidés et impliquent des interactions entre les modifications d'histones et la méthylation de l'ADN. En effet, la phase de méthylation *de novo* des ovocytes est précédée par un remodelage de certaines modifications d'histones comme H3K36me3 et H3K4me3 (Stewart et al., 2015). De plus, les enzymes KDM1A ou KDM1B qui déméthylent H3K4 sont nécessaires pour recruter la méthylation de l'ADN sur certains îlots CpG dans les ovocytes (Stewart et al., 2015).

3.1.2. Reprogrammation de l'embryon précoce

Une reprogrammation de la méthylation de l'ADN a également lieu au cours du développement embryonnaire précoce. Après la fécondation, l'ADN haploïde maternel et paternel est globalement déméthylé jusqu'au stade blastocyste pour ensuite réacquérir la méthylation après l'implantation de l'embryon (**figure 12**).

Déméthylation de l'embryon pré-implantatoire

Après la fécondation, l'ADN maternel et paternel forment deux pronoyaux distincts et ne fusionnent qu'après la première réplication du zygote. Des expériences d'immunofluorescences ont observé une cinétique de déméthylation de l'ADN différente entre ces pronoyaux (Mayer et al., 2000; Santos et al., 2002). Le pronoyau femelle est déméthylé progressivement jusqu'au stade blastocyste suggérant une déméthylation passive, tandis que le pronoyau mâle est déméthylé en quelques heures avant la première réplication du zygote suggérant une déméthylation active. Cependant, des études ultérieures par séquençage haut débit suggèrent que les deux génomes maternel et paternel subissent des processus de déméthylation à la fois actif et passif (Guo et al., 2014; Shen et al., 2014; Wang et al., 2014).

Cette phase de déméthylation post-zygotique fait intervenir la protéine TET3 qui est fortement exprimée durant le stade pré-implantatoire de l'embryon et colocalise principalement avec le pronoyau mâle (Gu et al., 2011; Wossidlo et al., 2011). L'activité de cette protéine induit l'enrichissement d'intermédiaires de la déméthylation de l'ADN tels que la 5hmC (Iqbal et al., 2011), la 5caC et la 5fC (Inoue et al., 2011). En conséquence, TET3 est nécessaire pour faciliter la déméthylation de certaines séquences dans le zygote (Guo et al., 2014; Peat et al., 2014; Shen et al., 2014). La déméthylation passive après la fécondation est également facilitée grâce à l'éviction de DNMT1 hors du noyau des cellules de l'embryon précoce (Cardoso and Leonhardt, 1999). L'isoforme DNMT1o d'origine ovocytaire, qui est exprimée spécifiquement durant cette phase, ne pénétrera dans le noyau des cellules qu'au stade embryonnaire 8-cellules (Howell et al., 2001).

La déméthylation globale du génome n'est pas totale puisqu'environ 20% de la méthylation subsistent dans les blastocystes (Wang et al., 2014). Les régions résistantes à la déméthylation correspondent aux gènes soumis à l'empreinte et à certains rétrotransposons (Smith et al., 2012; Wang et al., 2014). La méthylation de l'ADN est maintenue sur ces régions par l'activité de DNMT1, ce qui suggère que son exclusion du noyau n'est pas totale, et que des mécanismes de protection contre la déméthylation existent sur certaines régions au cours de la phase pré-implantatoire (Hirasawa et al., 2008).

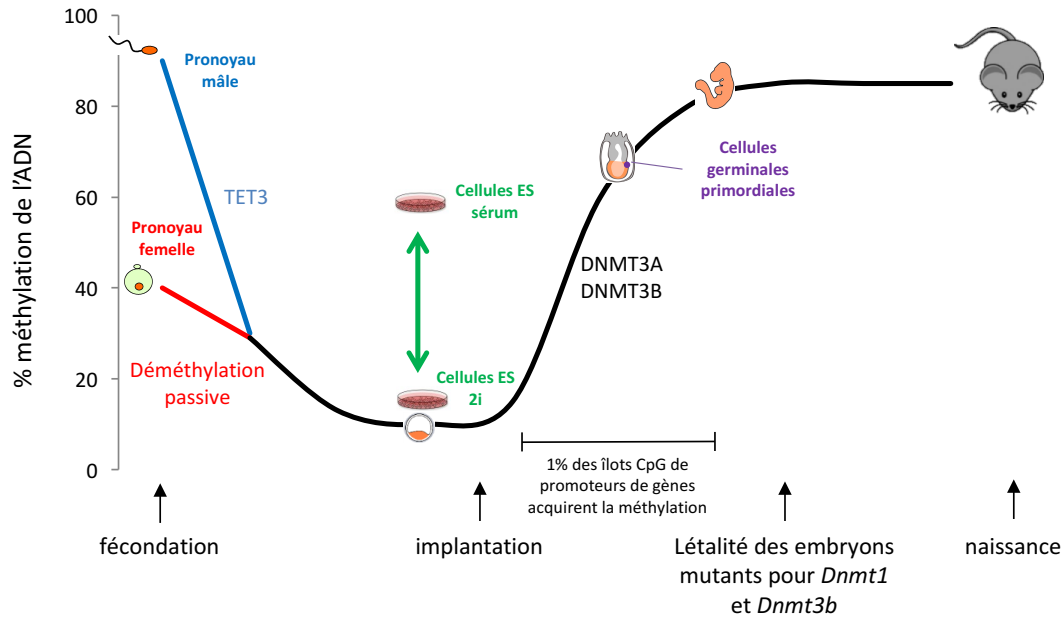


Figure 12 : Dynamisme du méthylome lors de l'embryogenèse précoce. Après la fécondation, les pronoyaux mâles et femelles sont déméthylés de manière active et passive respectivement. Après l'implantation du blastocyste, le génome va être reméthylé. Cette reméthylation concerne uniquement 1% des îlots CpG de promoteurs de gènes. Elle est nécessaire pour le bon développement embryonnaire étant donné la létalité des embryons mutants pour les *Dnmt3b* et *Dnmt1* après cette phase.

Méthylation *de novo* de l'embryon post-implantatoire

Après son implantation dans la paroi utérine, l'embryon globalement hypo-méthylé subit une vague de reméthylation jusqu'au stade embryonnaire E8.5 atteignant un niveau global de 80% (Smith et al., 2012). Il existe néanmoins des différences d'acquisition de la méthylation entre l'épiblaste et le tissu extra-embryonnaire, ainsi que de légères variations entre les cellules des différents feuillets embryonnaires (Zhang et al., 2018). Une étude du laboratoire a montré que la majorité de l'acquisition de la méthylation de l'ADN se fait entre le stade E4.5 et E6.5 (Auclair et al., 2014). De plus, la cinétique d'acquisition de méthylation est homogène dans la plupart des compartiments génomiques (Auclair et al., 2014). Durant cette phase de reméthylation, seuls 1% des îlots CpG des promoteurs de gènes acquièrent la méthylation de l'ADN. Ceci a lieu en particulier sur les promoteurs de gènes gamétiques, conduisant à leur répression (Auclair et al., 2014; Borgel et al., 2010). L'acquisition de la méthylation de l'ADN sur ces promoteurs suggèrent, soit un recrutement spécifique de la méthylation, soit l'absence d'un mécanisme de protection des îlots CpG.

La phase de reméthylation implique la coopération entre DNMT3A et DNMT3B dans la méthylation globale du génome (Auclair et al., 2014; Okano et al., 1999). Cependant DNMT3B a un rôle spécifique

dans la méthylation des îlots CpG (Auclair et al., 2014). Contrairement à la phase de méthylation dans les gamètes, la protéine DNMT3L a un rôle mineur durant la phase de reméthylation de l'embryon. Il a cependant été montré que les embryons mutants pour *Dnmt3l* semblent avoir un retard dans reméthylation sur certains gènes sans pour autant la compromettre (Guenatri et al., 2013).

Cette phase de reméthylation est vitale pour le développement de l'embryon. En effet les embryons mutants pour *Dnmt1* ou *Dnmt3b* ne se développent pas au-delà du stade E13.5 (Li et al., 1992; Okano et al., 1999). Néanmoins, les causes de cette létalité embryonnaire restent inconnues. Une étude transcriptomique sur des embryons globalement hypométhylés pourrait apporter de nouveaux éléments de réponse sur le rôle de la méthylation de l'ADN au cours du développement murin.

Modèle cellulaire d'étude de la méthylation embryonnaire

L'étude de la phase de reprogrammation de la méthylation du génome s'avère complexe compte tenu de la difficulté d'accessibilité des embryons et de la faible quantité de cellules. Les technologies permettant l'analyse à haut débit sur des cellules uniques se sont développées ces dernières années (Stuart and Satija, 2019), mais des études mécanistiques plus poussées semblent être difficiles à réaliser *in vivo*.

Les cellules ES peuvent être un bon modèle cellulaire d'analyse de cette reprogrammation de la méthylation de l'ADN. Elles sont dérivées de la masse cellulaire interne (ICM) des blastocystes et peuvent être différenciées en n'importe quel type cellulaire, hormis généralement le tissu extra-embryonnaire (Evans and Kaufman, 1981). Les cellules sont généralement cultivées en présence de sérum bovin fœtal et de LIF (leukemia inhibitor factor) permettant le maintien de leur état de pluripotence (Smith et al., 1988). Malgré le fait qu'elles soient issues de cellules hypométhylées, les cellules ES cultivées en sérum présentent un niveau de méthylation élevé, équivalent à celui d'un épiblaste embryonnaire au stade E6.5 (Ficz et al., 2013; Habibi et al., 2013) (**figure 12**). Ces cellules en culture sont hétérogènes sur plusieurs paramètres : d'un point de vue morphologique, d'expression de certains facteurs de pluripotence (Kalmar et al., 2009; Toyooka et al., 2008) et de niveau de méthylation de l'ADN (Hayashi et al., 2008). De plus, une petite proportion de cellules ES (environ 1%) adoptent un profil similaire aux cellules de l'embryon au stade 2-cellules qui sont hypométhylées (Eckersley-Maslin et al., 2016; Macfarlan et al., 2012).

Les cellules ES peuvent également être cultivées dans un milieu sans sérum et en présence de deux inhibiteurs PD0325901 et CHIR99021 (2i). Ces derniers inhibent des kinases de la voie de signalisation

MAPK et WNT (Ying et al., 2008). Ce changement de milieu de culture induit une modification importante du transcriptome, une reprogrammation de la distribution de certaines modifications d'histone comme H3K27me3 (Marks et al., 2012) et une déméthylation globale du génome. En effet, ces cellules ont un méthylome proche de celui d'un blastocyste avec une persistance de la méthylation au niveau de certains éléments transposables et des gènes soumis à l'empreinte (Ficz et al., 2013; Habibi et al., 2013). Cette déméthylation est causée en partie par l'expression de *Prdm14* dans les cellules ES en 2i qui réprime l'expression de *Dnmt3a* et *Dnmt3b* (Leitch et al., 2013; Yamaji et al., 2013), ainsi que par la déstabilisation de la protéine UHRF1 et l'inhibition de la voie de maintien de la méthylation globale (von Meyenn et al., 2016) (figure 13). Un milieu 2i supplémenté en vitamine C permet de réduire encore davantage le niveau de méthylation à environ 5% sur tout le génome (Walter et al., 2016). En effet la vitamine C potentialise l'activité des protéines TETs, impliquées dans la déméthylation de l'ADN (Blaschke et al., 2013).

Les cellules ES sont donc un bon modèle d'étude de la reprogrammation épigénétique du développement embryonnaire. Les cellules ES cultivées en milieu 2i avec un méthylome proche de celui d'un blastocyste peuvent être transférées en milieu sérum et atteindre un niveau de méthylation proche d'un épiblaste, récapitulant la phase de méthylation *de novo* (figure 12).

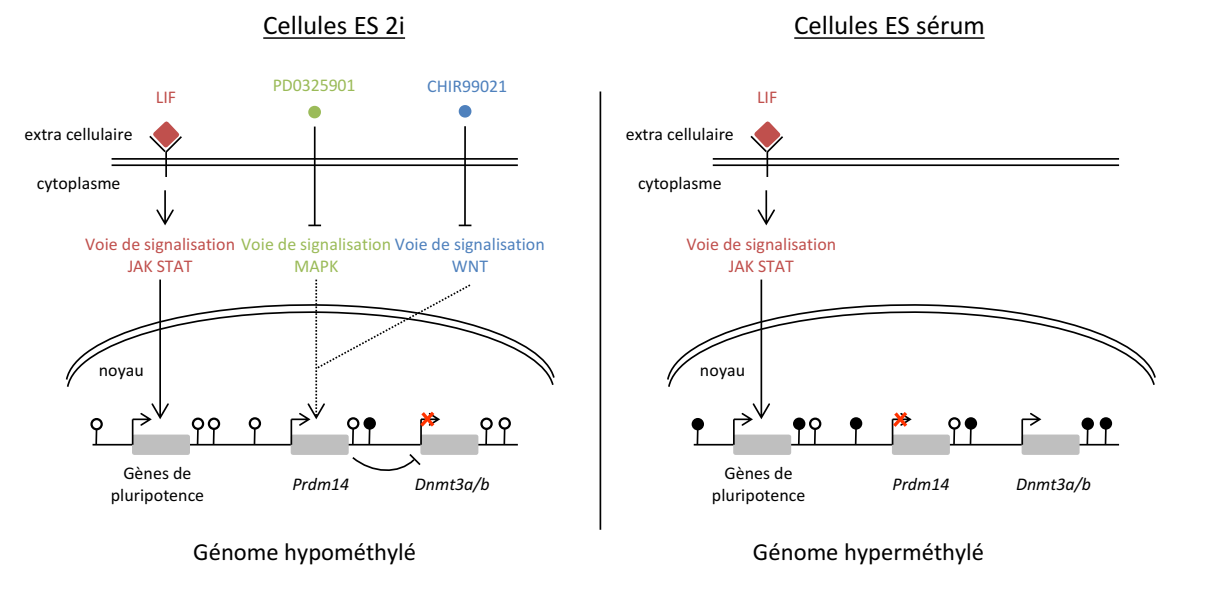


Figure 13 : Les cellules ES sont un modèle d'étude de la reprogrammation de la méthylation de l'ADN. Les cellules ES cultivées dans un milieu avec du sérum adoptent le méthylome d'un épiblaste embryonnaire. Les cellules ES dans un milieu en 2i induisent une cascade moléculaire qui réprime l'expression des *Dnmt3a/b* qui participe à la déméthylation des cellules. Le niveau de méthylation dans ces cellules est ainsi proche de celui d'un blastocyste.

3.2. Régulations des éléments transposables

Définitions

Les éléments répétés constituent 40% du génome murin. Ils sont constitués d'éléments répétés en tandem comme les régions satellites situées au niveau des régions péri-centromériques par exemple, et d'éléments transposables (TE). Certains de ces éléments transposables ont une origine virale (comme les ERVs – voir plus bas), tandis que les autres ont une origine inconnue du fait de leurs mutations progressives au cours de l'évolution. On distingue les transposons à ADN ou à ARN (appelés rétrotransposons) selon leurs modes de mobilisation. Les transposons à ADN ont une mobilisation de type « couper-coller » et les rétrotransposons « copier-coller ». Une classification plus précise a permis de les regrouper en plusieurs familles et sous familles en fonction notamment de la présence de séquences LTR (*Wicker et al., 2007*). Les éléments transposables les plus représentés dans le génome murin sont les rétrotransposons avec les éléments LTR (long terminal repeat), ERVs et les éléments non LTR, SINEs et LINEs. (2002) (figure 14).

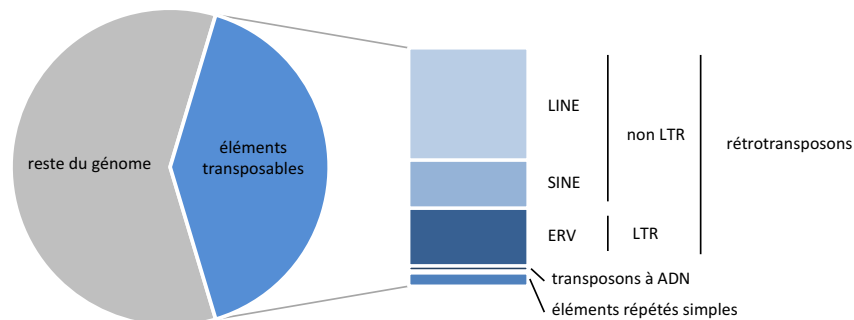


Figure 14 : Les éléments génétiques répétés. Les éléments transposables constituent une part importante du génome murin. La classification de ces éléments transposables dépend de leur mode de mobilisation (à travers des séquences d'ADN ou d'ARN). On distingue les transposons à ADN et les rétrotransposons. Ces derniers sont également classés en fonction de la présence de séquences LTR.

La plupart de ces éléments transposables sont inactifs dans le génome à cause de leurs mutations et troncations progressives au cours de l'évolution. Néanmoins, les rétrotransposons les plus jeunes, c'est à dire s'étant insérés le plus tard dans l'évolution sont toujours potentiellement actifs. L'activation de ceux-ci conduit à leurs transpositions et insertions aléatoires dans le génome générant des mutations et une instabilité génomique (*Goodier and Kazazian, 2008*). En effet, on considère que dans le génome murin 10% de mutations sont causées par l'activation de

rétrotransposons (*Maksakova et al., 2006*). Ces rétrotransposons doivent donc être réprimés afin de prévenir leurs activations.

Les rétrotransposons jouent néanmoins un rôle primordial dans le moteur de l'évolution. Ainsi plusieurs gènes impliqués dans des processus biologiques du développement sont régulés par des éléments transposables. C'est le cas par exemple du rétrotransposon MERVL qui régule l'expression des gènes 2-cellules (c'est à dire exprimés spécifiquement dans le stade embryonnaire 2-cellules) dans l'embryon précoce (*De Iaco et al., 2017; Hendrickson et al., 2017; Macfarlan et al., 2012; Whiddon et al., 2017*). De la même manière, l'expression du rétrotransposon LINE-1 joue un rôle dans l'embryogenèse précoce et il est considéré comme un marqueur de la totipotence (*Jachowicz et al., 2017*). L'évolution est ainsi témoin d'une « course à l'armement » entre l'activation des rétrotransposons responsables de mutations dans le génome, et les mécanismes moléculaires permettant leurs répressions.

Répression des éléments transposables au cours du développement embryonnaire

La reprogrammation de la méthylation de l'ADN dans les cellules germinales primordiales conduit à la déméthylation de la plupart des éléments transposables. Les rétrotransposons les plus « jeunes » d'un point de vue évolutif restent généralement méthylés tels que les éléments IAP (*Seisenberger et al., 2012*). Néanmoins, une étude a récemment mis à jour des rétrotransposons métastables très jeunes d'un point de vue évolutif et qui ne maintiennent pas la méthylation de l'ADN au cours de la reprogrammation germinale (*Kazachenka et al., 2018*). Les mécanismes permettant le maintien spécifique de la méthylation de l'ADN sur ces rétrotransposons restent néanmoins inconnus. Celui-ci pourrait faire intervenir SETDB1 qui catalyse le dépôt de H3K9me3 sur les rétrotransposons dans les PGCs et dont l'ablation conduit à la déméthylation ces rétrotransposons (*Liu et al., 2014*). Des mécanismes indépendants de la méthylation de l'ADN sont à l'œuvre pour réprimer les rétrotransposons déméthylés dans les PGCs. Ces mécanismes font intervenir les petits ARNs qui interagissent avec la protéine PIWI (piARN) et répriment les rétrotransposons au niveau post-transcriptionnel dans les PGCs (*Inoue et al., 2017*). Plus tard, au cours du développement des prospermatogonies, les piARN participent au ciblage de H3K9me3 et de la méthylation de l'ADN sur les rétrotransposons et vont ainsi contribuer à un basculement de mécanismes de répression post-transcriptionnel vers des mécanismes de répression transcriptionnel (*Aravin et al., 2008; Inoue et al., 2017; Kuramochi-Miyagawa et al., 2008*). La reméthylation des éléments transposables dans les PGCs mâles et femelles implique l'activité de DNMT3L et de DNMT3C, tandis que les rôles de DNMT3A et DNMT3B sont encore mal définis. Ainsi, l'ablation de DNMT3L et DNMT3C dans les cellules

germinales conduit à une hypométhylation et une dérégulation des rétrotransposons dans les spermatocytes (Barau et al., 2016; Bourc'his and Bestor, 2004). Ceci suggère ainsi un rôle important de la méthylation de l'ADN dans la répression des rétrotransposons dans les cellules germinales.

Comme dans les cellules germinales, les rétrotransposons doivent être réprimés durant la phase de reprogrammation de l'embryogenèse précoce. Cette répression implique la voie des facteurs de transcription en doigt de zinc ZFP, qui reconnaissent les séquences des rétrotransposons et recrutent le répresseur KAP1 (aussi appelé TRIM28). Ce dernier est indispensable pour la répression des rétrotransposons IAP et ERV dans les cellules ES et dans l'embryon précoce (Rowe et al., 2010). La voie ZFP/KAP1 fonctionne notamment en recrutant la marque d'histone H3K9me3 sur les rétrotransposons. En effet, l'inactivation de plusieurs histones méthyltransférases de H3K9 telles que *Setdb1* (Karimi et al., 2011; Matsui et al., 2010), *G9a* (Maksakova et al., 2013) ou *Suv39h* (Bulut-Karslioglu et al., 2014) conduisent à la dérégulation des rétrotransposons dans les cellules ES. La marque d'histone H3K9me3 pourrait alors jouer un rôle dans la protection des rétrotransposons contre la déméthylation. En effet l'inactivation de *Setdb1* dans les cellules ES conduit à une accumulation de 5hmC qui est médiée par TET2 sur les rétrotransposons (Coluccio et al., 2018; Deniz et al., 2018).

Au cours de la phase de reméthylation du génome, la voie ZFP/KAP1/SETDB1 favorise le dépôt de méthylation de l'ADN au niveau des rétrotransposons (Leung et al., 2014a; Rowe et al., 2013), ce qui pourrait initier un basculement vers des voies de répression des rétrotransposons dépendantes de la méthylation de l'ADN. Ainsi, l'ablation de DNMT1 ou d'UHRF1 induit une forte dérégulation des éléments IAPs dans les embryons post-implantatoires (Sharif et al., 2007; Walsh et al., 1998).

De manière intéressante, les cellules ES en culture récapitulent ce qui se passe au cours de la phase de reprogrammation épigénétique et adoptent des mécanismes de régulation des éléments transposables indépendants de la méthylation de l'ADN. En effet, les cellules ES *Dnmt* TKO qui sont totalement déméthylées ne montrent pas de réactivation massive d'éléments transposables (Karimi et al., 2011). Le passage de cellules ES d'un milieu sérum à un milieu en 2i+VitC a permis de disséquer le mécanisme de régulation des rétrotransposons dans des cellules durant la phase de déméthylation. Il a été observé que des marques répressives d'histones telles que H3K9me3 et H3K27me3 se substituent rapidement à la méthylation de l'ADN pour assurer la répression des rétrotransposons, passant d'un mécanisme dépendant à un mécanisme indépendant de la méthylation de l'ADN (Walter et al., 2016). Récemment plusieurs expériences de criblage ont identifié un grand nombre de nouvelles protéines impliquées dans la répression des rétrotransposons (Fukuda et al., 2018; Liu et al., 2018a; Robbez-Masson et al., 2018) Ces protéines pourraient également contribuer à assurer la répression fiable des rétrotransposons en présence ou en absence de la méthylation de l'ADN.

3.3. Régulation des gènes soumis à l’empreinte

Définitions

Les gènes soumis à l’empreinte ont la particularité d’être méthylés de manière mono-allélique en fonction de l’origine parentale des allèles. On distingue ainsi les gènes soumis à l’empreinte d’origine maternelle ou paternelle selon qu’ils héritent de la méthylation de l’ovocyte ou du spermatozoïde respectivement. Environ 150 gènes soumis à l’empreinte sont répertoriés chez la souris et sont généralement regroupés en clusters. Ces clusters sont régulés par l’acquisition de la méthylation de l’ADN sur des séquences régulatrices riches en CpG appelées ICRs (imprinted control region). Les ICRs sont donc méthylés de manière différentielle entre les deux allèles parentaux et sont appelés DMRs (differentially methylated region). Les ICRs peuvent se trouver au niveau du promoteur des gènes soumis à l’empreinte et ainsi directement réguler leur expression. Ils peuvent également se trouver sur des séquences intra ou inter-génique et réguler l’expression de plusieurs gènes *en cis* en agissant en tant qu’insulateur ou via l’expression d’ARNnc (ARN non codants) (Kelsey Gavin and Feil Robert, 2013) (figure 15).

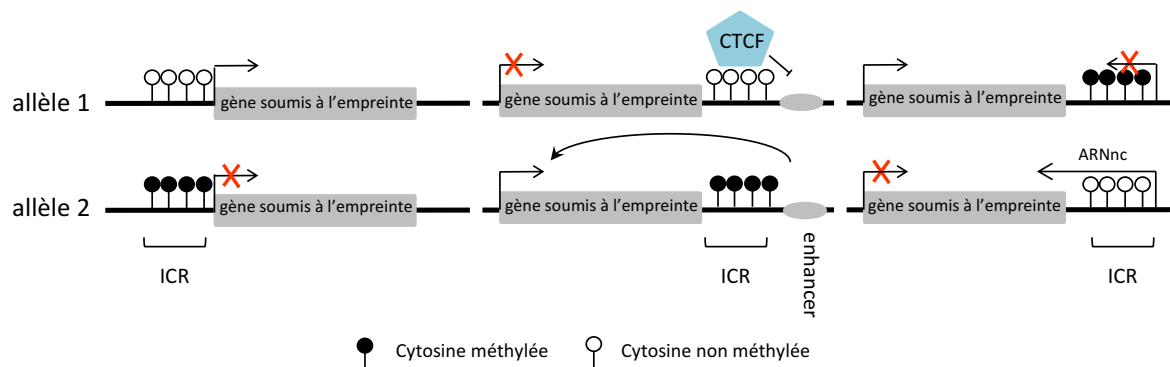


Figure 15 : Régulation des gènes soumis à l’empreinte. Les gènes soumis à l’empreinte sont contrôlés par des séquences riches en CpG appelées ICR. Ces ICRs permettent de contrôler l’expression mono-allélique de ces gènes en régulant leur promoteur. Ils peuvent également réguler leur expression *en cis* en agissant en tant qu’insulateur ou en exprimant un ARN non codant.

Régulation de l’empreinte au cours du développement embryonnaire

Les gènes soumis à l’empreinte sont protégés de la reprogrammation de la méthylation de l’ADN dans l’embryon précoce mais pas dans les cellules germinales. Lors de l’embryogenèse précoce, la

méthylation différentielle mise en place dans les gamètes parentaux au niveau des ICR résiste à la déméthylation de l'ADN. Le maintien de la méthylation des ICRs après la fécondation est réalisé par l'action de DNMT1 et UHRF1 (Hirasawa et al., 2008; Maenohara et al., 2017). La protection contre la déméthylation fait également intervenir des voies moléculaires similaires à celles qui ciblent des éléments transposables. Ainsi, les facteurs de transcription en doigts de zinc ZFP57 et ZFP445 interagissent avec des motifs d'ADN présents dans les ICRs et sont indispensables pour le maintien de la méthylation des ICRs au cours de l'embryogenèse (Li et al., 2008; Takahashi et al., 2019). Ceux-ci recrutent alors le répresseur KAP1 qui favorise le recrutement de H3K9me3, DNMT1 et UHRF1 sur les ICRs (Messerschmidt et al., 2012; Quenneville et al., 2011). Par ailleurs, la protection des ICRs contre la déméthylation après la fécondation nécessite également la protéine STELLA (Nakamura et al., 2007). Le génome des cellules germinales primordiales est globalement déméthylé y compris les ICRs contrôlant les gènes soumis à l'empreinte. La déméthylation des ICRs est permise grâce à l'éviction de la machinerie de maintien de la méthylation de l'ADN (Kagiwada et al., 2013) et par l'action de la protéine TET1 (Yamaguchi et al., 2013). Les ICRs sont ensuite méthylés durant les phases de reméthylation de l'ovocyte et du spermatozoïde via l'action de DNMT3A et DNMT3L (Bourc'his et al., 2001; Kaneda et al., 2004).

3.4. Régulation de l'identité cellulaire

Régulation de la pluripotence

La reprogrammation de la méthylation de l'ADN au cours du développement pourrait permettre la mise en place de la pluripotence. En effet, la déméthylation des promoteurs des gènes de pluripotence *Oct4* et *Nanog* permettent leur expression au stade blastocyste (Farthing et al., 2008; Gu et al., 2011). Ainsi, un délai de déméthylation de l'ADN dû à l'absence de TET1 induit également un délai de l'expression de *Oct4* au cours du développement (Gu et al., 2011).

Une petite population de cellules ES en culture adoptent un profil caractéristique de cellules totipotentes, caractérisé par l'expression de gènes 2-cellules ou de certains rétrotransposons comme MERVL (Macfarlan et al., 2012). Une étude transcriptomique sur cellules uniques a montré que l'absence de *Dnmt1* n'induit pas la dérèpression de ces éléments caractéristiques de la totipotence dans les cellules ES (Eckersley-Maslin et al., 2016). Ceci suggère que la méthylation de l'ADN ne participe pas à la mise en place de cet état cellulaire *in vitro*.

Régulation de la différenciation

Les cellules ES sont un bon modèle permettant d'étudier le rôle de la méthylation de l'ADN dans la différenciation cellulaire. L'absence des DNMTs conduit à une incapacité de différenciation des cellules ES à cause de la persistance de l'expression de marqueurs de la pluripotence (Jackson *et al.*, 2004; Panning and Jaenisch, 1996). De plus, des modèles de différenciation des cellules ES en cellules neuronales ont montré un changement important du méthylome entre les cellules souches et les cellules différenciées (Meissner *et al.*, 2008; Mohn *et al.*, 2008). Ces études suggèrent alors un rôle important de la méthylation de l'ADN dans la différenciation cellulaire *in vitro*.

Le développement des techniques de tri cellulaire par cytométrie en flux a permis l'exploration de la reprogrammation de la méthylation de l'ADN dans la différenciation cellulaire *in vivo*. Ainsi des variations importantes de méthylome ont été observées au cours de la différenciation de précurseurs de cellules du système nerveux (Lister *et al.*, 2013), du système hématopoïétique (Farlik *et al.*, 2016) ou encore des cardiomyocytes (Gilsbach *et al.*, 2014).

Ces études *in vitro* et *in vivo* ont permis l'observation du même phénomène : la différenciation cellulaire s'accompagne d'un recrutement de la méthylation de l'ADN au niveau de promoteurs de gènes de pluripotence ou embryonnaires, ainsi que la déméthylation des enhancers qui favorisent l'expression de gènes d'un lignage cellulaire spécifique (figure 16).

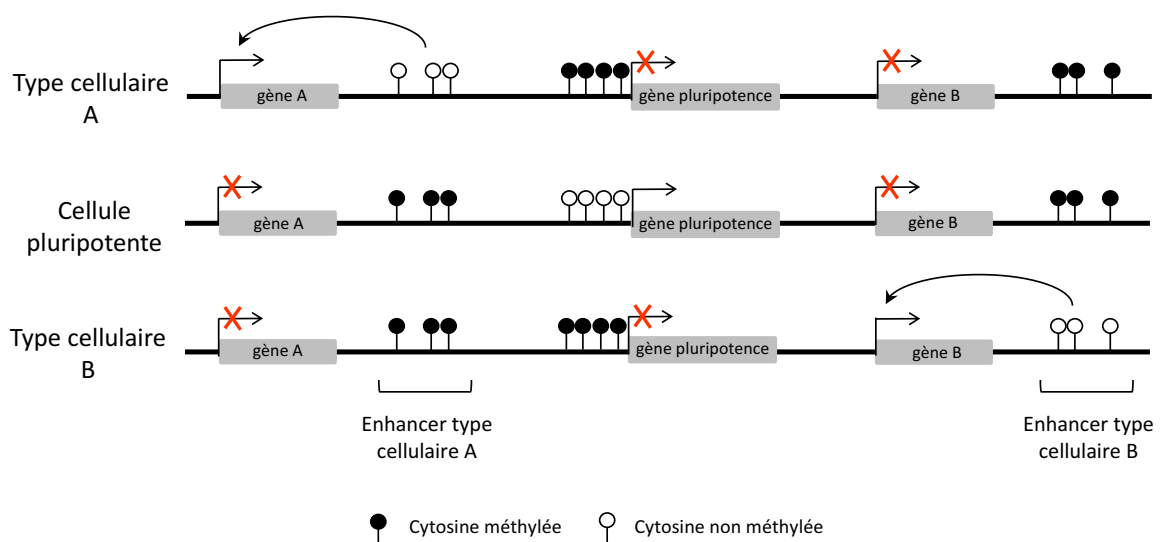


Figure 16 : Régulation de la différenciation cellulaire. La différenciation cellulaire s'accompagne d'une acquisition de la méthylation sur les gènes de pluripotence, ainsi que d'une déméthylation spécifique des enhancers contrôlant les gènes spécifiquement exprimés dans un type cellulaire donné.

La phase de reméthylation du génome lors de l'embryogenèse coïncide avec la différenciation des cellules souches pluripotentes afin de générer les différents tissus et organes de l'organisme. L'étude de la mise en place du plan de formation de l'organisme a révélé un rôle important de la déméthylation de l'ADN et des protéines TETs. En effet, ces protéines sont impliquées dans la déméthylation et l'activation d'enhancers (Bogdanović et al., 2016). Les embryons déplétés pour les 3 gènes *Tet1-2-3* ne montrent aucune différence morphologique par rapport aux embryons contrôles jusqu'au stade épiblaste où ils meurent dû à des défauts lors de la gastrulation (Dai et al., 2016). Ces résultats suggèrent alors que la méthylation de l'ADN est acquise uniformément dans l'embryon post-implantatoire, puis que les cellules subissent des événements de déméthylation spécifiques pour générer un type cellulaire donné.

Cas du programme gamétique

La reméthylation du génome à l'implantation conduit à la méthylation de seulement 1% des îlots CpG situés dans les promoteurs de gènes (Auclair et al., 2014). Cette méthylation a lieu spécifiquement dans les promoteurs de gènes gamétiques et contribue à leur répression dans l'embryon. En effet, les embryons déplétés pour *Dnmt3b* présentent une hypométhylation et une dérégulation de certains de ces gènes (Auclair et al., 2014; Borgel et al., 2010).

Comme énoncé plus haut, il est fort probable que toutes les cellules embryonnaires acquièrent la méthylation avant l'apparition des précurseurs de la lignée germinales au stade E6.5. Ainsi, la phase de déméthylation de l'ADN dans les PGCs pourrait permettre de coupler dans le temps la reprogrammation épigénétique avec l'activation du programme d'expression gamétique durant la différenciation de la lignée germinale (Hill et al., 2018; Yamaguchi et al., 2012). En effet, les cellules germinales mutées pour le gène *Dnmt1* activent les gènes gamétiques de manière trop précoce, conduisant à un hypogonadisme et une infertilité des animaux (Hargan-Calvopina et al., 2016).

Les gènes gamétiques sont également régulés par des mécanismes indépendants de la méthylation de l'ADN. Certaines protéines Polycomb appartenant au complexe PRC1.6 notamment ont un rôle dans la répression du programme gamétique (Endoh et al., 2017; Hisada et al., 2012; Qin et al., 2012; Suzuki et al., 2016). De plus, l'ablation de certaines histones méthyltransférases comme SETDB1 ou G9A ont conduit à la dérégulation des gènes gamétiques dans les cellules ES et dans l'embryon murin (Auclair et al., 2016; Karimi et al., 2011). Ces enzymes sont également impliquées dans la répression des éléments transposables en absence de la méthylation de l'ADN, (voir chapitre 3.2 : régulation des éléments transposables). Il est ainsi tentant de spéculer qu'un mécanisme de compensation,

similaire aux éléments transposables, est également mis en place pour réprimer les gènes gamétiques en absence de la méthylation de l'ADN lors de l'embryogenèse précoce.

Au cours de cette thèse, je me suis ainsi intéressé aux rôles des complexes Polycomb dans la régulation des gènes gamétiques.

4. Les complexes Polycomb

La protéine Polycomb (PC) a été initialement découverte chez la drosophile dont la perte de fonction conduit à la dérégulation des gènes Hox. Progressivement, d'autres gènes dont la mutation conduit au même phénotype ont été découverts conduisant à la définition du groupe des protéines Polycomb (PcG). Ces protéines sont impliquées dans la répression de l'expression des gènes et ont un effet antagoniste des protéines Trithorax (*Schuettengruber et al., 2017*). Des études protéomiques et biochimiques ont démontré que ces protéines s'organisent en complexes. On dénombre ainsi 2 complexes Polycomb principaux : PRC1 et PRC2. D'autres complexes Polycomb ont été récemment décrits comme PR-DUB (Polycomb repressive deubiquitination) mais ne seront pas décrits ici. Le complexe PRC1 est caractérisé par la présence d'enzymes qui déposent la marque d'histone H2AK119ub1, tandis que PRC2 est caractérisé par la présence d'enzymes qui déposent la marque d'histone H3K27me3. Depuis leurs découvertes initiales, les complexes Polycomb sont considérés comme des éléments centraux de la régulation de la transcription et sont impliqués dans de nombreux processus biologiques dont le développement embryonnaire. En effet, mis à part leurs rôles dans la régulation des gènes Hox, ils régulent la différenciation et la pluripotence des cellules souches, et sont impliqués dans les processus oncogéniques (*Schuettengruber et al., 2017*).

4.1. Composition protéique des complexes Polycomb

Le complexe Polycomb PRC2

L'activité principale du complexe PRC2 est la méthylation de H3K27 qui est réalisée par l'une des deux enzymes EZH1 ou EZH2. Ces enzymes sont capables de catalyser la méthylation successive de H3K27me1-2-3 via leurs domaines SET (*Cao et al., 2002*). L'activité catalytique de EZH1/2 requiert l'interaction avec les protéines EED et SUZ12 qui les rendent fonctionnelles (*Margueron et al., 2009; Pasini et al., 2004*). Ces 3 protéines EZH1/2, EED et SUZ12 sont ainsi les membres minimaux du complexe PRC2 fonctionnel. Des protéines accessoires peuvent également interagir avec ce complexe afin de réguler sa fonction. Les protéines PCL1/2/3 sont impliquées dans le recrutement du complexe (*Hunkapiller et al., 2012*) tandis que les protéines AEBP2 et JARD2 jouent un rôle dans son recrutement et stimulent l'activité de EZH2 en présence de la modification d'histone H2AK119ub1 (*Kalb et al., 2014; Li et al., 2010*). L'interaction mutuellement exclusive des protéines accessoires PCL1/2/3 et AEBP2/JARID2 avec le complexe minimal fonctionnel PRC2 définissent le complexe PRC2.1 et PRC2.2 respectivement (*Hauri et al., 2016*) (**figure 17A**).

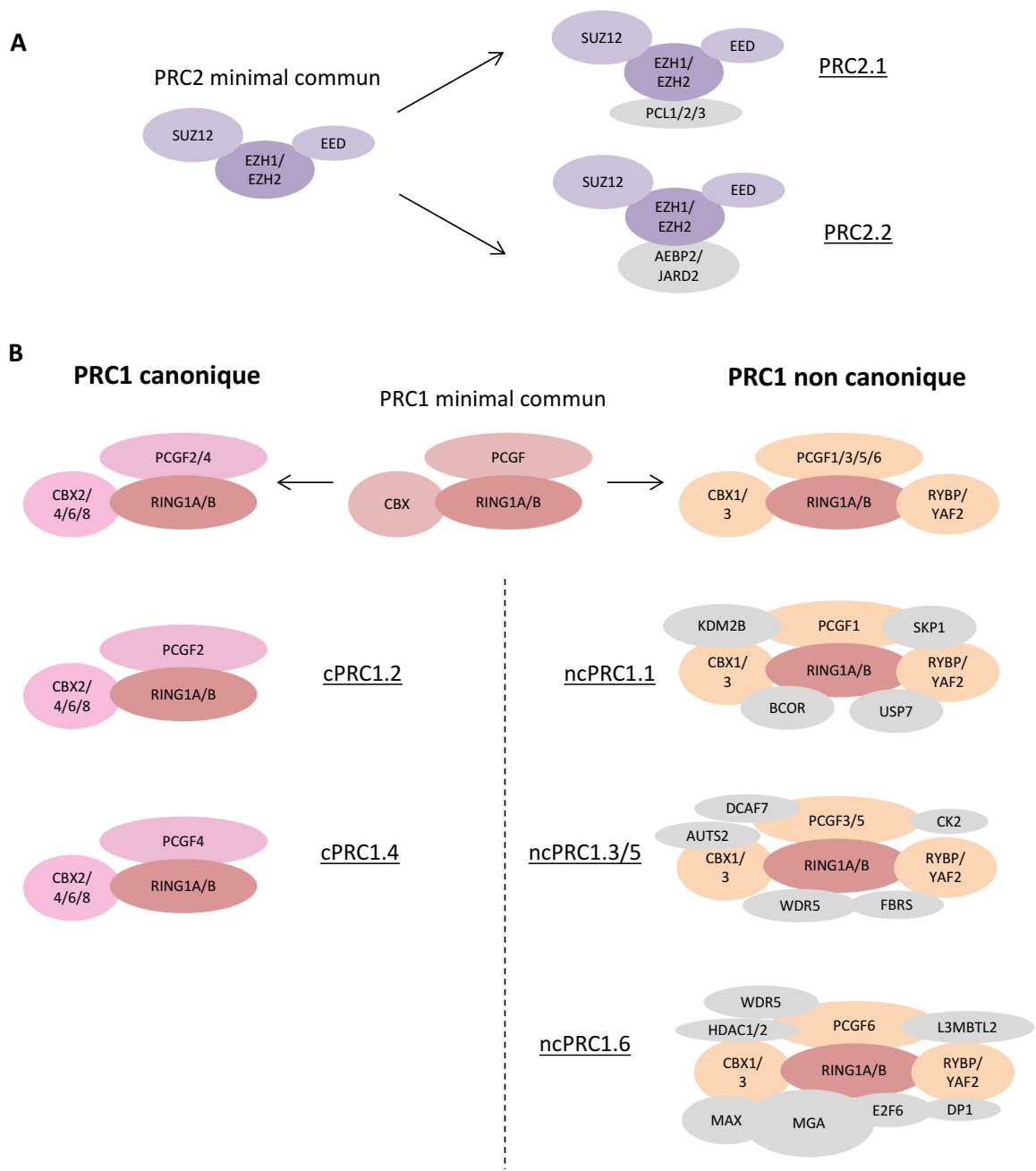


Figure 17 : Composition protéique des complexes Polycomb. A-B) On retrouve deux complexes Polycomb différents dans les cellules de mammifères : le complexe PRC2 (A) et le complexe PRC1 (B). A) L'interaction de protéines accessoires avec les protéines principales du complexe PRC2 permet de déterminer deux sous complexes PRC2.1 et PRC2.2. B) On retrouve deux types de complexes PRC1 en fonction de la composition protéique : le complexe PRC1 canonique comprenant CBX2/4/6/8 et les complexes PRC1 non canoniques comprenant les protéines CBX1/3 et RYBP/YAF2. La dénomination des sous complexes PRC1 est déterminée par la présence de la protéine PCGF.

Le complexe Polycomb PRC1

L'activité principale du complexe PRC1 est la déposition de la marque H2AK119ub1 qui est médiée par le domaine E3 ligase des enzymes RING1A/B (*Wang et al., 2004*). L'activité de RING1A/B requiert l'interaction avec les protéines PCGF régulatrices. On retrouve 6 différents paralogues de PCGF (PCGF1-6) qui interagissent avec RING1A/B de manière mutuellement exclusive. L'interaction d'un paraglogue PCGF (PCGF1-6) avec RING1A/B va alors définir un sous-complexe de PRC1 (PRC1.1 à PRC1.6) (*Gao et al., 2012*) (**figure 17B**). On distingue alors des complexes PRC1 canoniques et des complexes PRC1 non canoniques en fonction de leurs modes de recrutement au niveau de la chromatine (voir chapitre 4.2 : recrutement des complexes Polycomb). Ainsi, les complexes canoniques cPRC1.2 et cPRC1.4 possèdent les protéines CBX2/4/6/8 leur permettant d'interagir avec H3K27me3, tandis que les complexes ncPRC1.1/3/5/6 possèdent les protéines CBX1/3 qui interagissent avec H3K9me3 et sont recrutés de manière indépendante de H3K27me3. Ces complexes non canoniques sont également caractérisés par leur interaction spécifique avec RYBP/YAF2 (*Gao et al., 2012*) (**figure 17B**).

4.2. Recrutement des complexes Polycomb

Chez la drosophile, les complexes Polycomb sont recrutés à travers les PRE (Polycomb responsive elements) qui sont des motifs d'ADN reconnus par des facteurs de transcription. Néanmoins, aucun de ces facteurs de transcription conservés entre la drosophile et les mammifères ne permettent de recruter les complexes Polycomb chez les mammifères (*Entrevan et al., 2016*). Les modèles de recrutement des complexes Polycomb y sont donc plus divers et complexes. Les complexes PRC1 et PRC2 peuvent être recrutés individuellement mais également de manière synergique.

Certaines études ont montré que les complexes Polycomb PRC1 et PRC2 colocalisent dans les noyaux des cellules ES (*Boyer et al., 2006; Ku et al., 2008*) suggérant un mécanisme de recrutement entre ces deux complexes. Initialement, un modèle hiérarchique a été proposé dans lequel le complexe PRC2 permet le recrutement de PRC1 grâce aux protéines CBX2/4/6/7/8 qui permettent d'interagir avec H3K27me3 (**figure 18A**). En effet, des études biochimiques (*Cao et al., 2002; Fischle et al., 2003*) et structurales (*Min et al., 2003*) ont démontré l'interaction entre CBX2/4/6/7/8 et H3K27me3. De plus l'ablation génétique du complexe PRC2 empêche le recrutement de PRC1 au niveau des gènes Hox (*Boyer et al., 2006; Cao et al., 2002; Mujtaba et al., 2008*). Néanmoins il a récemment été montré que l'ablation génétique de PRC1 affecte également le recrutement de PRC2, contredisant ce modèle hiérarchique (*Blackledge et al., 2014*). Le mécanisme moléculaire permettant le recrutement de PRC2

par PRC1 peut être médié par l'interaction des protéines accessoires JARID-AEBP2 avec H2AK119ub1 (Cooper et al., 2016; Kalb et al., 2014) (**figure 18B**). Le recrutement des complexes Polycomb PRC1 et PRC2 semble donc être plutôt coopératif que hiérarchique comme il l'a été proposé chez la drosophile (Kahn et al., 2016). Ceci présuppose que les complexes PRC1 et PRC2 peuvent être recrutés de manière indépendante.

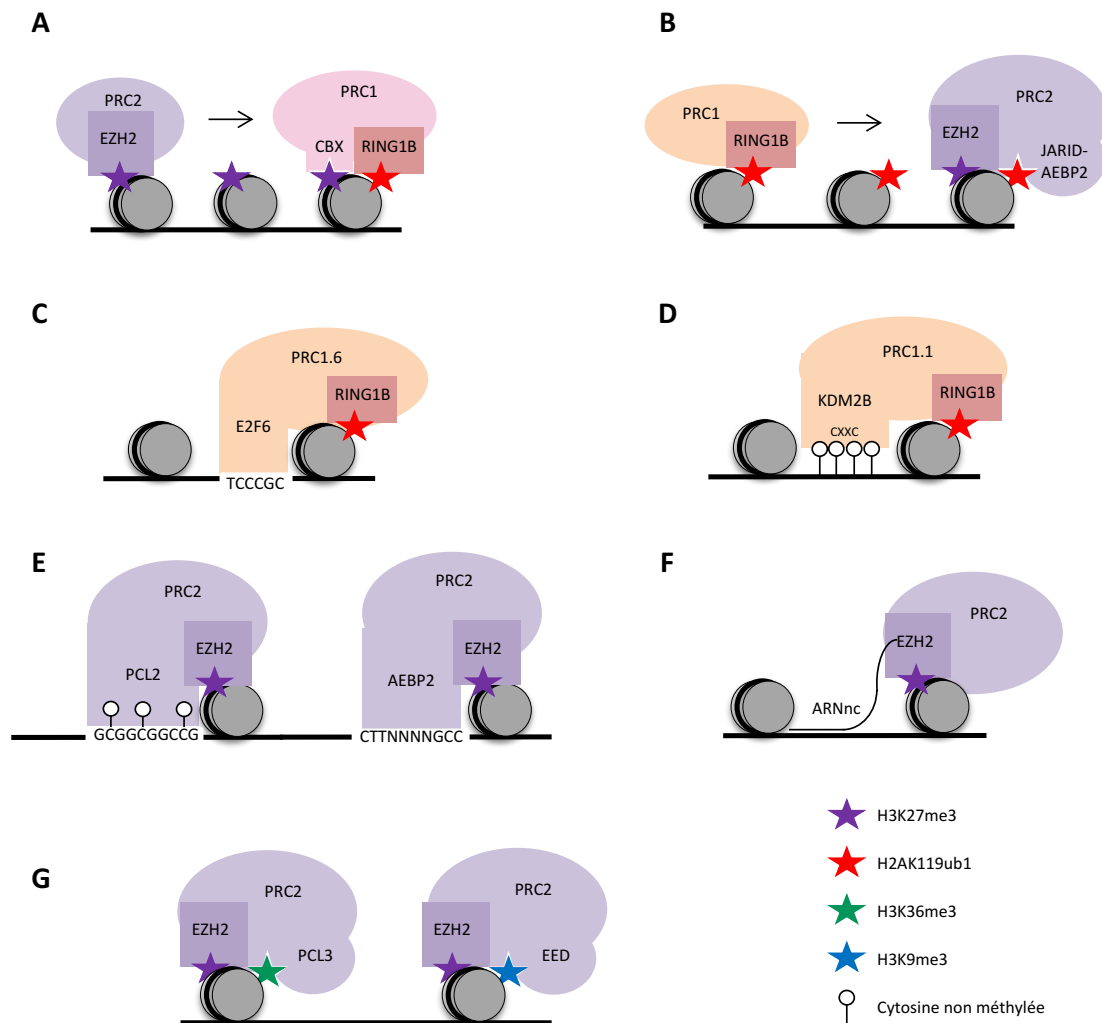


Figure 18 : Mécanismes de recrutement des complexes Polycomb. A-B) Un modèle coopératif des complexes PRC1 et PRC2 a été proposé. Le complexe PRC1 est recruté par H3K27me3 déposée par PRC2 (A) et le complexe PRC2 est recruté par H2AK119ub1 déposée par PRC1 (B). C-D) Le complexe PRC1 peut être recruté indépendamment du complexe PRC2 via des facteurs de transcriptions se liant à l'ADN (C), ou par des protéines reconnaissant l'ADN non méthylé (D). E-F-G) Le complexe PRC2 peut être recruté indépendamment du complexe PRC1 par des facteurs de transcription reconnaissant un motif d'ADN (E), par des ARN non codants (F) ou par des modifications d'histones (G)

Les complexes PRC1 non canoniques sont recrutés indépendamment du complexe PRC2 (Tavares et al., 2012) car ceux-ci ne possèdent pas les protéines CBX responsables de l'interaction avec H3K27me3 (Gao et al., 2012). Ces complexes possèdent des facteurs de transcription se liant à des motifs d'ADN comme MAX/MGA ou E2F6 (Stielow et al., 2018) (**figure 18C**). Il existe également un contexte nucléotidique favorable au recrutement de PRC1 non canonique, notamment au niveau des îlots CpG non méthylés qui sont reconnus par le domaine CXXC de KDM2B (Farcas et al., 2012; Wu et al., 2013) (**figure 18D**). Du fait de leurs mécanismes de recrutement différents, les complexes PRC1 canoniques et non canoniques ont des cibles différentes et donc des fonctions différentes dans les cellules ES (Morey et al., 2013).

Le complexe PRC2 est lui ciblé par différents mécanismes indépendamment du complexe PRC1. Tout d'abord, les protéines accessoires du complexe PRC2 jouent un rôle important dans son recrutement comme AEBP2 (Kim et al., 2009) et PCL2 qui ciblent préférentiellement les îlots CpG non méthylés (Li et al., 2017; Perino et al., 2018) (**figure 18E**). Ensuite, il a été montré que des ARNnc ont un rôle dans le ciblage du complexe notamment au niveau du chromosome X (Zhao et al., 2008, 2010) (**figure 18F**). Enfin, des modifications d'histones peuvent jouer un rôle dans le recrutement de PRC2, comme H3K36me3 qui interagit avec PCL3 (Ballaré et al., 2012) ou H3K9me3 qui est reconnue par EED *in vitro* (Margueron et al., 2009) (**figure 18G**). En effet une étude a montré un recrutement de PRC2 altéré dans les cellules KO pour l'histone méthyl transférase G9A (Mozzetta et al., 2014).

4.3. Rôles des complexes Polycomb au cours du développement embryonnaire

Les complexes Polycomb font partie des principaux complexes réprimant les gènes dans les cellules de mammifères. Ils sont ainsi impliqués dans un grand nombre de fonctions physiologiques, comme la régulation de la pluripotence, ou de la différenciation cellulaire. Ils sont largement impliqués dans le développement embryonnaire (Aloia et al., 2013) et dans le processus oncogénique (Comet et al., 2016).

Le rôle des complexes Polycomb dans le développement embryonnaire est démontré par la létalité précoce des embryons déplétés pour SUZ12 (PRC2) à E8.5 (Pasini et al., 2004) et RING1B (PRC1) à E10.5 (Voncken et al., 2003). A part leurs rôles dans la mise en place de l'axe antéro-postérieure en régulant les gènes *Hox*, les complexes Polycomb jouent un rôle prépondérant dans la régulation de l'auto-renouveau des cellules souches et dans leur différenciation (Schuettengruber et al., 2017). Les gènes impliqués dans ces processus sont caractérisés par un promoteur bivalent possédant à la fois la marque d'histones répressive H3K27me3 médiée par Polycomb, et la marque d'histone activatrice

H3K4me3 médiée par le complexe Trithorax (*Mikkelsen et al., 2007*). Le processus de différenciation cellulaire nécessite ainsi un remodelage de l'activité de Polycomb avec son éviction du promoteur de gènes impliqués dans la différenciation cellulaire.

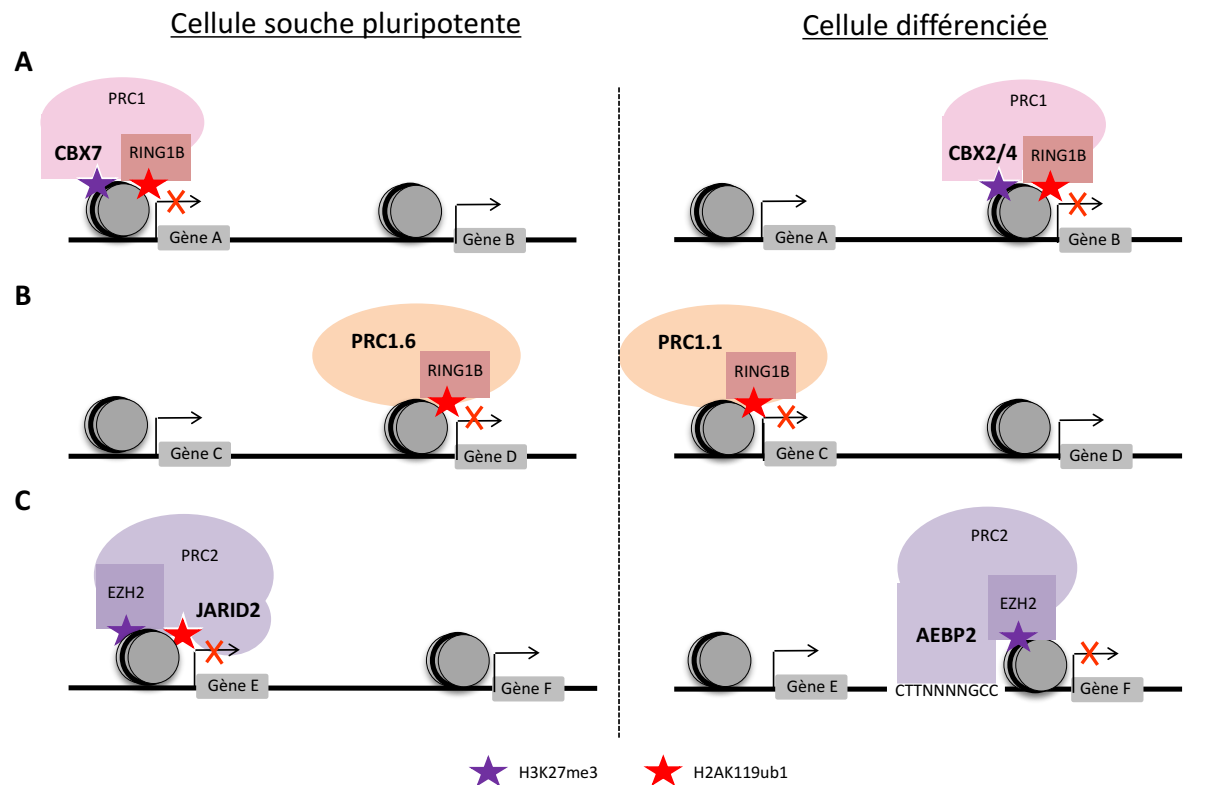


Figure 19 : La variation de la composition des complexes Polycomb induit leur recrutement sur des cibles différentes. A) L'expression de la protéine CBX7 dans les cellules pluripotentes et de CBX2/4 dans les cellules différenciées permettent le recrutement du complexe PRC1 canonique sur des cibles différentes. B) l'expression des membres du complexes PRC1.6 préférentiellement dans les cellules pluripotentes et les membres du complexes PRC1.1 dans les cellules différenciées permettent le recrutement de RING1A/B sur des cibles différentes. C) L'expression de JARID2 dans les cellules pluripotentes et AEBP2 dans les cellules différenciées permet le recrutement de PRC2 sur des cibles différentes.

Ce changement de recrutement du complexe Polycomb est caractérisé par un remodelage de la composition protéique des complexes. Ainsi, le complexe PRC1 canonique est recruté par CBX7, exprimé spécifiquement dans les cellules ES pluripotentes, tandis que lors de la différenciation des cellules ES, CBX2/4 sont préférentiellement exprimés et induisent le recrutement de PRC1 sur des cibles différentes (*Morey et al., 2013*) (**figure 19A**). Le complexe PRC1 non canonique PRC1.6 est exprimé préférentiellement dans les cellules souches permettant le recrutement de RING1A/B sur certains gènes. Au cours de la différenciation des cellules ES, d'autres complexes non canoniques sont exprimés comme PRC1.1. Ce complexe permet alors un recrutement de RING1A/B sur des cibles

différentes (Kloet et al., 2016) (**figure 19B**). De la même manière, le complexe PRC2 peut être recruté par sa protéine accessoire JARID2 qui est plus enrichie dans les cellules non différenciées et AEBP2 dans les cellules différenciées (Kloet et al., 2016) (**figure 19C**).

La variation de l'assemblage des complexes impacte donc leur recrutement et ainsi leur fonction dans un processus biologique donné. Ceci permet à un complexe de se spécialiser dans une fonction particulière. C'est le cas par exemple du complexe PRC1.6 qui est impliqué dans la répression des gènes gamétiques dans les cellules somatiques.

4.4. Mécanismes de répression par les complexes Polycomb

Le recrutement des complexes Polycomb conduit à la répression de la transcription génique. Du fait de la composition différente des deux complexes, les mécanismes conduisant à cette répression sont différents mais résultent tous les deux à une compaction de la chromatine (Margueron et al., 2008; Terranova et al., 2008). Cette compaction a ainsi comme conséquence l'éviction de la machinerie transcriptionnelle et du complexe Trithorax (Kadoch et al., 2017; Stanton et al., 2017).

Mise à part la capacité de recruter le complexe PRC1, la méthylation de la lysine K27 de l'histone H3 peut empêcher son acétylation. En effet la modification d'histone H3K27ac est une marque associée à l'expression des enhancers qui régulent positivement l'expression d'un groupe de gènes (Ferrari et al., 2014).

Le complexe PRC1 est responsable du dépôt de H2AK119ub1 par RING1A/B. Le rôle de cette marque d'histone est toujours débattu : certaines études concluent à un rôle important dans la répression des gènes (Endoh et al., 2008), tandis que d'autres affirment qu'elle y joue un rôle très limité (Illingworth et al., 2015; Pengelly et al., 2015). Récemment, une étude a proposé que la modification H2AK119ub1 a un effet différent selon qu'elle soit réalisée par le complexe PRC1 canonique ou non canonique. En effet, l'ablation simultanée de PCGF1,3, 5 et 6 permettant d'inhiber l'activité de RING1B des complexes non canoniques, conduit à la dérégulation des mêmes gènes que la déplétion de RING1B dans les cellules ES. Tandis que l'ablation de PCGF2 et 4, inhibant l'activité de RING1B des complexes canoniques, n'a pas d'effet sur l'expression des gènes (Fursova et al., 2019). Ceci suggère donc deux mécanismes de répression des gènes par le complexe PRC1 : l'un dépendant de l'activité de RING1A/B (pour les complexes PRC1 non canoniques) et l'autre indépendant de leur activité (pour les complexes PRC1 canoniques). Le rôle de H2AK119ub1 dans la répression des gènes par les complexes PRC1 non canoniques peut alors être soit de recruter le complexe PRC2 (Kalb et al., 2014), soit d'empêcher le recrutement de certaines marques d'histones permissives à la transcription (Aihara et

al., 2016) (**figure 20A**). Le complexe PRC1 canonique a un rôle important dans l'organisation tridimensionnelle du noyau dans les cellules ES. Celui-ci favorise des contacts intra et inter-chromosomiques entre différents promoteurs de gènes régulés par Polycomb comme les gènes *Hox* (*Schoenfelder et al., 2015*). Cette activité est alors indépendante de RING1A/B et de H2AK119ub1 (*Kundu et al., 2017*), et pourrait impliquer la protéine CBX2 qui a la capacité d'interagir avec plusieurs résidus H3K27me3 localisés sur des nucléosomes différents (*Lau et al., 2017*) (**figure 20B**).

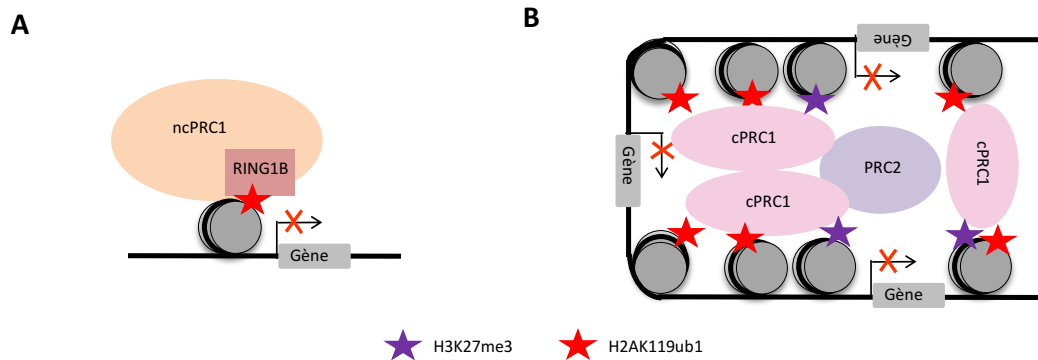


Figure 20 : Le complexe PRC1 canonique et non canonique réprime les gènes par des mécanismes différents. A) Le complexe PRC1 non canonique permet le recrutement de H2AK119ub1, qui joue un rôle important pour la répression de ces gènes cibles. B) Le complexe PRC1 canonique est impliqué dans la compaction locale de la chromatine, et H2AK119ub1 ne semble pas nécessaire dans ce mécanisme.

4.5. Le complexe PRC1.6

4.5.1. Description du complexe PRC1.6

Ce complexe a été initialement décrit suite à l'identification des interactants de la protéine E2F6 qui est connue comme un répresseur de l'activité des gènes (*Ogawa et al., 2002*). Cette étude a ainsi mis en avant qu'E2F6 interagit avec plusieurs protéines du complexe Polycomb. On a alors appelé ce complexe le « E2F6-comp ». Suite à la volonté d'identifier les membres des divers complexes PRC1, plusieurs études d'interactome ont été conduites dans des cellules humaines différenciées (*Gao et al., 2012; Hauri et al., 2016; Trojer et al., 2011*) et des cellules murines (*Endoh et al., 2017; Qin et al., 2012; Sánchez et al., 2007*).

Le complexe PRC1.6 est un complexe PRC1 non canonique, caractérisé par la présence des protéines RYBP/YAF2 et l'absence des protéines CBX2/4/6/7/8 responsables de l'interaction avec H3K27me3. De plus, le complexe PRC1.6 se distingue des autres complexes non canoniques par la présence exclusive de la protéine PCGF6 (*Gao et al., 2012*) (**tableau 2 ; figure 21**).

source	Endoh et al. 2016	Qin et al. 2012	Hauri et al. 2016	Gao et al. 2012	Trojer et al. 2011	Ogawa et al. 2002	Sanchez et al. 2007
type cellulaire	mESC	mESC	HEK 293T	HEK 293T	HEK 293F	HeLa	Murine erythroleukemia (MEL) cells
protéine tagguée	PCGF6	L3MBTL2	RING1A/B, PCGF6, RYBP/YAF2, MAX, E2G6, DP1, L3MBTL2, CBX1/3, WDR5	RING1B, RYBP, YAF2, PCGF6	L3MBTL2	E2F6	RING1B
interactants	E2F6	x	x	x	x	x	x
	DP1	x	x	x	x	x	x
	MAX			x	x	x	x
	MGA	x	x	x	x	x	x
	L3MBTL2	x	x	x	x	x	x
	CBX1	x	x	x			
	CBX3	x	x	x	x	x	x
	G9A		x				x
	GLP						x
	HDAC1	x	x		x		
	HDAC2	x	x		x		
	PCGF6	x	x	x	x	x	x
	YAF2	x		x	x	x	x
	RING1A	x		x		x	x
	RING1B	x	x	x	x	x	x
RYBP	x	x	x	x			
WDR5	x	x	x	x		x	

Tableau 2 : Détermination des membres du complexes PRC1.6. Plusieurs expériences d'interactome ont été conduites dans les cellules murines et humaines. Ces expériences ont permis de déterminer la composition protéique du complexe PRC1.6.

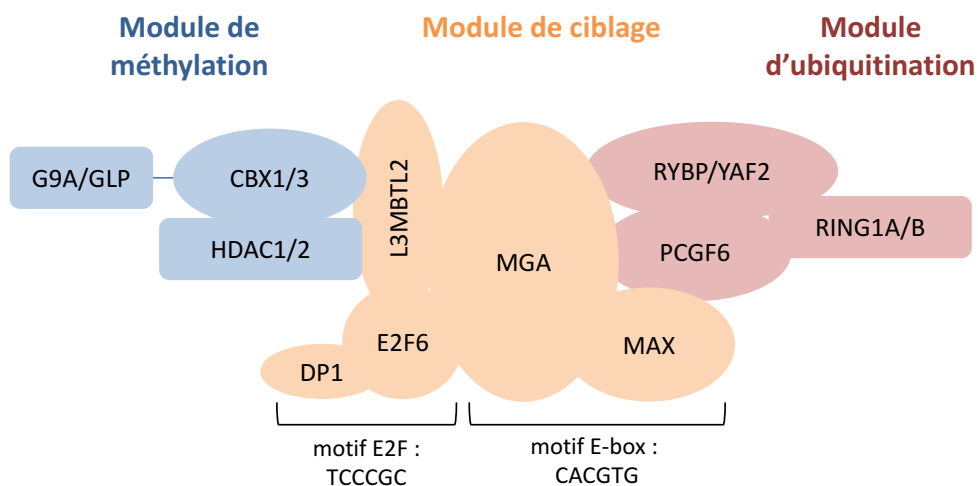


Figure 21 : Le complexe Polycomb PRC1.6. Ce complexe comprend différents modules : un module de ciblage représenté par des facteurs de transcription se liant à l'ADN, un module d'ubiquitination des histones représenté par RING1A/B, et un module de méthylation des histones avec CBX1/3 capable de recruter les histones méthyltransférases G9A/GLP.

Au sein du complexe PRC1.6, on peut distinguer 3 modules différents :

- 1) Un module de ciblage du complexe, composé de plusieurs facteurs de transcription qui se lient à l'ADN génomique comme E2F6, MAX ou MGA.
- 2) Un module d'ubiquitination des histones, composé de RING1A/B et leur protéines régulatrices RYBP/YAF2 et PCGF6.
- 3) Un module de méthylation des histones, composé de HDAC1/2 et de CBX1/3 capables d'interagir à la fois avec la marque H3K9me et les histones méthyltransférases G9A/GLP.

Bien que les membres du complexe PRC1.6 aient été identifiés, ses fonctions biologiques et les mécanismes lui permettant de réguler l'expression des gènes sont encore mal connus. Une étude systématique de ses membres est ainsi nécessaire afin de comprendre son rôle dans la répression des gènes et son implication dans d'éventuels processus biologiques (**tableau 3**).

		léthalité embryonnaire	rôle dans les cellules ES					
			maintien de la pluripotence		contrôle de la prolifération		répression du programme gamétique	
module de ciblage	E2F6	viable	non		non		oui	
	DP1	E12.5	?		non?		?	
	MAX	E5.5-6.5	oui		oui		oui	
	MGA	E3.5	oui		oui		oui	
	L3MBTL2	E9.5	oui		oui		oui	
module de méthylation	CBX1	périnatal	oui		oui		?	non*
	CBX3	périnatal	oui		?		?	
	G9A	E9.5	?		oui		oui	
	GLP	E9.5	?		?		oui	
	HDAC1	E9.5-10.5	non	oui*	non	oui*	?	non*
	HDAC2	viable	non		non		?	non*
module d'ubiquitination	PCGF6	viable	oui		oui		oui	
	YAF2	?	non		non		non	
	RING1A	viable	?		?	oui*	non	oui*
	RING1B	E10.5	oui		non		oui	
	RYBP	E5.5	oui		oui		oui	

Tableau 3 : Rôle du complexe PRC1.6 dans le développement embryonnaire murin. Les membres du complexe PRC1.6 semblent être impliqués dans le maintien de la pluripotence des cellules ES, la prolifération des cellules ES, ainsi que dans la répression du programme gamétique (* : DKO).

Le module de ciblage

MGA : C'est un facteur de transcription qui se lie au motif E-box CACGTG lorsqu'il s'hétérodimérise avec MAX (Hurlin et al., 1999). L'ablation de MGA dans les cellules ES murines et dans des cellules humaines empêche le recrutement du complexe PRC1.6 au niveau de certains gènes cibles,

démontrant l'implication de MGA dans le ciblage du complexe (Endoh et al., 2017; Stielow et al., 2018). Du fait de sa taille très volumineuse (330kDa), MGA joue également un rôle allostérique en stabilisant les autres membres du complexe au niveau protéique (Stielow et al., 2018). Le facteur de transcription MGA joue un rôle dans la répression de gènes gamétiques dans les cellules ES murines (Endoh et al., 2017), ainsi que dans le maintien de leur pluripotence (Hu et al., 2009; Washkowitz et al., 2015). Ceci est probablement aussi le cas dans les cellules embryonnaires *in vivo* puisque les embryons KO pour *Mga* ne se développent pas au-delà du stade blastocyste (Washkowitz et al., 2015).

MAX : C'est un facteur de transcription qui se lie au motif E-box CACGTG. Sa fixation à l'ADN nécessite une hétérodimérisation mutuellement exclusive avec MGA ou MYC (Suzuki et al., 2016). De ce fait, MAX est soit un répresseur de l'expression génique lorsqu'il interagit avec MGA, soit un activateur lorsqu'il interagit avec MYC. Les cellules ES KO pour *Max* ne sont pas viables. En effet elles présentent un défaut de prolifération qui est imputé à une perte de fonction de la protéine MYC (Hishida et al., 2011). L'absence de MAX dans les cellules ES induit également la dérégulation de gènes gamétiques et de gènes 2-cellules (Suzuki et al., 2016). Les embryons KO pour *Max* sont létaux au stade embryonnaire E5.5 (Shen-Li et al., 2000).

E2F6 : C'est un facteur de transcription se liant au motif TCCCGC (voir chapitre 4.6 : le facteur de transcription E2F6)

DP1 : La protéine DP1 est impliquée dans la régulation des facteurs de transcription E2Fs. En effet, il a été démontré que DP1 interagit avec E2F6 et permet sa fixation à l'ADN (Cartwright et al., 1998). Les embryons KO pour *Dp1* ne se développent pas au-delà du stade E12.5 dû à des malformations des tissus extra-embryonnaires (Kohn et al., 2003). Les cellules ES KO pour *Dp1* ne présentent pas de défauts de pluripotence ni de prolifération (Kohn et al., 2004).

L3MBTL2 : Cette protéine n'a pas de propriété de liaison à l'ADN, mais elle est également impliquée dans le recrutement du complexe PRC1.6. En effet, le KO de *L3mbtl2* conduit à une perte de recrutement du complexe sur certains gènes cibles dans les cellules humaines HEK293 (Stielow et al., 2018). Elle pourrait ainsi jouer un rôle dans la cohésion du complexe. Les embryons KO pour *L3mbtl2* ne sont pas viables et ne se développent plus au-delà du stade E9.5 (Qin et al., 2012). L'absence de L3MBTL2 dans les cellules ES conduit à un défaut de prolifération des cellules ainsi qu'une perte de leur état de pluripotence (Huang et al., 2018; Qin et al., 2012). La protéine L3MBTL2 régule l'expression des gènes gamétiques par un processus impliquant H3K9me2 et G9A/GLP (Qin et al., 2012). La protéine L3MBTL2 peut être sumoylée dans sa partie C-terminale. Cette modification post-traductionnelle

n'affecte néanmoins pas son activité de répression des gènes dans les cellules HEK293 (Huang et al., 2018). Cependant, il a récemment été montré que le KD (knock down) d'UBC9 impliqué dans la sumoylation des protéines conduit à la dérégulation des gènes gamétiques dans les cellules ES (Cossec et al., 2018). L'effet de la sumoylation de L3MBTL2 dans la régulation des gènes gamétiques dans les cellules ES reste donc une piste à explorer.

Le module d'ubiquitination

RING1A/B : Ce sont les protéines impliquées dans l'ubiquitination de l'histone H2A (Napoles et al., 2004), grâce à leurs domaines E3 ubiquitin ligase (Wang et al., 2004). RING1A et RING1B sont deux protéines paralogues qui partagent une grande similitude dans leurs séquences en acides aminés. Il a été proposé que RING1A et RING1B ont un effet synergique, puisque l'absence de RING1B conduit à une surexpression de RING1A dans les cellules ES (Endoh et al., 2008). Les souris KO pour *Ring1a* sont viables mais ont un phénotype de transformations homéotiques (Lorente et al., 2000), tandis que les embryons KO pour *Ring1b* ne se développent pas au-delà du stade E10.5 (Voncken et al., 2003). Ceci suggère un rôle préférentiel de RING1B dans le développement embryonnaire, et un rôle de RING1A dans les cellules plus différenciées. Les cellules ES KO pour *Ring1b* perdent leur état de pluripotence (Leeb et Wutz, 2007). Certains gènes gamétiques sont dérégulés dans les cellules *Ring1a/b* DKO (double knock out) (Endoh et al., 2017). Le rôle de la modification H2AK119ub1 dans la répression des gènes est toujours débattu. Récemment une étude a proposé qu'H2AK119ub1 est impliqué dans la répression des gènes lorsque cette modification est mise en place par les complexes PRC1 non canoniques (Fursova et al., 2019) (voir chapitre 4.4 : mécanismes de répression des complexes Polycomb).

PCGF6 : C'est la présence de cette protéine qui définit le complexe PRC1.6 (Gao et al., 2012). Elle ne semble pas être impliquée dans le recrutement du complexe au niveau de gènes cibles (Stielow et al., 2018), mais elle joue un rôle dans la régulation de RING1A/B. En effet la protéine PCGF6 possède un domaine RING qui permet d'interagir et de moduler l'activité de RING1A/B (Akasaka et al., 2002; Zhao et al., 2017). La protéine PCGF6 est impliquée dans la prolifération et l'auto-renouvellement des cellules ES (Endoh et al., 2017; Hu et al., 2009; Yang et al., 2016), mais également dans la répression de gènes gamétiques via son interaction avec RING1A/B (Endoh et al., 2017; Zhao et al., 2017). Les souris KO pour *Pcgf6* sont viables mais elles ont des transformations homéotiques qui sont caractérisées par un défaut d'identité de certaines vertèbres (Endoh et al., 2017).

RYBP/YAF2 : Ces protéines définissent les complexes Polycomb non canoniques (Gao et al., 2012). En effet, on observe une compétition entre RYBP et CBX2 (PRC1 canonique) pour l'interaction avec RING1A/B. L'interaction de RYBP/YAF2 permet alors de réguler l'activité de RING1A/B (Rose et al., 2016; Tavares et al., 2012; Zhao et al., 2018). Les cellules ES KO pour *Rybp* ou *Yaf2* ne présentent pas de défauts de pluripotence ou de prolifération (Gao et al., 2012; Hisada et al., 2012; Zhao et al., 2018). Néanmoins, l'inactivation de *Rybp* conduit à la dérégulation de gènes gamétiques et de gènes 2-cellules (Hisada et al., 2012; Morey et al., 2013). Les embryons KO pour *Rybp* meurent au stade E5.5 (Pirity et al., 2005).

Le module de méthylation

CBX1/3 : Les protéines CBX1 et 3 sont des protéines qui se distinguent des autres membres de la famille des protéines CBX par l'incapacité d'interagir avec H3K27me3. Des études biochimiques ont montré que CBX1/3 interagissent avec H3K9me grâce à leurs chromodomaines (Cheutin et al., 2003). Elles ont ainsi été associées à la compaction de la chromatine, notamment au niveau de l'hétérochromatine péri-centromérique (Hiragami-Hamada et al., 2016; Smallwood et al., 2008). Néanmoins, ces protéines sont également impliquées dans l'activation de la transcription en régulant l'épissage alternatif des gènes (Smallwood et al., 2012). La protéine CBX3 est capable d'interagir avec l'histone méthyltransférase G9A (Smallwood et al., 2007). En effet, CBX3 reconnaît la méthylation de la lysine en position 185 de G9A qui se trouve dans le même contexte de séquence d'acides aminés que l'histone H3 autour de la lysine 9 : HRKS (Chin et al., 2007; Ruan et al., 2012). Les protéines CBX1 et CBX3 sont impliquées dans la prolifération, et le maintien de la pluripotence des cellules ES (Huang et al., 2017; Mattout et al., 2015; Sridharan et al., 2013), mais ne semblent pas intervenir dans la répression de certains gènes gamétiques (Endoh et al., 2017). Les embryons KO pour *Cbx1* ou *Cbx3* se développent correctement mais les souris meurent peu après la naissance dû à des défauts de développement musculaire (Aucott et al., 2008; Takada et al., 2011).

HDAC1/2 : Ce sont des protéines impliquées dans la déacétylation des histones. Les cellules ES doubles mutantes pour *Hdac1/2* montrent une augmentation de l'acétylation de la lysine en position 9 (H3K9ac) ainsi qu'une légère diminution de la méthylation sur cette même lysine (H3K9me) (Jamaladdin et al., 2014). Ceci pourrait suggérer que la déacétylation des histones en position H3K9 est un prérequis pour permettre leur méthylation et induire la répression des gènes médiée par le complexe PRC1.6. Ces cellules doubles mutantes pour *Hdac1/2* montrent également un défaut de prolifération et de maintien de la pluripotence (Jamaladdin et al., 2014), ce qui n'est pas observé dans les cellules déplétées pour *Hdac1* ou *Hdac2* individuellement (Dovey et al., 2010). Ceci suggère ainsi

une activité redondante de ces protéines. Les embryons KO pour *Hdac1* ne se développent pas au-delà du stade E9.5 tandis que les animaux *Hdac2* KO meurent après la naissance à cause de malformations cardiaques (Montgomery et al., 2007). Ceci suggère un rôle prépondérant de HDAC1 au cours du développement embryonnaire chez la souris.

G9A/GLP ? : Ce sont des histones méthyltransférases permettant la mono et la diméthylation de H3K9 (Ogawa et al., 2002; Tachibana et al., 2001). Elles ne sont pas enrichies dans toutes les expériences d'interactome et semblent être des partenaires occasionnels du complexe PRC1.6. En effet, CBX3 est capable d'interagir avec G9A en se liant à la méthylation de la lysine en position 185 (Ruan et al., 2012). Cette modification post traductionnelle est réalisée par G9A elle-même (Chin et al., 2007).

Les protéines G9A et GLP interagissent via leurs domaines SET respectifs et semblent agir de manière coopérative (Tachibana et al., 2005, 2008). Elles sont impliquées dans la répression d'un grand nombre d'éléments génomiques comme les gènes soumis à l'empreinte (Zhang et al., 2016) ou les gènes gamétiques (Maeda et al., 2013). Plusieurs évidences montrent un lien entre G9A et la méthylation de l'ADN *de novo* (Auclair et al., 2016; Dong et al., 2008) ou la maintenance de la méthylation de l'ADN (Ferry et al., 2017).

Les cellules ES KO pour *G9a* présentent un défaut de prolifération (Tachibana et al., 2002).

Les embryons KO pour *G9a* et *Glp* meurent au stade embryonnaire E9.5 (Tachibana et al., 2002, 2005).

Autre protéine

WDR5 : Plusieurs études ont montré que WDR5 fait partie du complexe PRC1.6. Néanmoins, elle fait partie du groupe de protéines Trithorax, et participe à l'activation de l'expression des gènes en induisant la déposition de H3K4me3 par les protéines MLL (Dou et al., 2006; Wysocka et al., 2005).

L'interaction de WDR5 avec PRC1.6 peut s'expliquer par le fait qu'ils colocalisent sur le même locus d'ADN. En effet les gènes ciblés par les complexes Polycomb possèdent généralement un promoteur bivalent avec l'enrichissement à la fois de H3K27me3 et de H3K4me3 au niveau de leur promoteur (voir chapitre 4.4 : rôles des complexes Polycomb au cours du développement embryonnaire). La protéine WDR5 pourrait ainsi être un faux positif des expériences d'interactome.

4.5.2. Rôles du complexe PRC1.6

Le complexe PRC1.6 joue un rôle fondamental au cours du développement embryonnaire murin comme le prouve la létalité des embryons knock-out pour ses membres. Le complexe joue

notamment un rôle dans la prolifération et l'autorenouveauement des cellules ES. On retrouve également une implication du complexe dans la répression du programme gamétique dans les cellules somatiques (**tableau 3**).

Une étude protéomique a montré que le complexe PRC1.6 est le complexe PRC1 le plus abondant dans les cellules ES. Une fois ces cellules différenciées en cellules progénitrices neurales (NPC), ce complexe est moins exprimé au profit des autres complexes PRC1 (*Kloet et al., 2016*). Ceci confirme un rôle de ce complexe dans l'auto-renouveauement des cellules souches. Néanmoins, les mécanismes régulant cet auto-renouveauement ne sont pas connus. Le complexe PRC1.6 régule également la prolifération des cellules ES. Cette régulation n'est pas dépendante d'E2F6, qui est le seul membre connu qui régule le cycle cellulaire (voir chapitre 4.6 : le facteur de transcription E2F6). Les défauts de prolifération pourraient directement découler d'un manque d'auto-renouveauement, puisqu'il est connu que les cellules différenciées prolifèrent moins que les cellules ES pluripotentes.

Un des rôles principaux du complexe PRC1.6 semble être la répression du programme gamétique dans les cellules ES. En effet l'inactivation de plusieurs membres du complexes (*Max, Rybp, Pcgf6, L3mbtl2*) conduit à la surexpression de gènes gamétiques dans les cellules ES (*Suzuki et al., 2016; Hisada et al., 2012; Endoh et al., 2017; Qin et al., 2012*). Cette répression est indispensable pour empêcher les cellules d'entrer en méiose (*Suzuki et al., 2016*). De plus, le complexe PRC1.6 peut également être impliqué dans la régulation fine de l'expression des gènes gamétiques durant la maturation des cellules germinales. Ainsi il a été montré que le KO spécifique de *Ring1b* dans les cellules germinales affecte la spermatogenèse (*Maezawa et al., 2017*) et l'ovogenèse (*Yokobayashi et al., 2013*) suite à des défauts de régulation de gènes gamétiques.

De même que pour la restriction des gènes gamétiques, le complexe PRC1.6 est impliqué dans la répression des gènes 2-cellules. En effet, un criblage à base d'ARN interférants (siARN) a permis d'identifier un grand nombre de membres du complexe PRC1.6 impliqué dans la répression du gène 2-cellules *Zscan4d* (*Rodriguez-Terrones et al., 2018*). Cette étude est confirmée par la dérégulation d'un grand nombre de gènes 2-cellules dans les ES KO pour *Max* (*Suzuki et al., 2016*) ou *Rybp* (*Hisada et al., 2012*).

4.6. Le facteur de transcription E2F6

Introduction aux facteurs de transcription E2Fs

Les facteurs de transcription E2Fs sont impliqués dans de nombreux processus biologiques comme le cycle cellulaire, l'apoptose ou le métabolisme.

La famille des facteurs de transcription E2Fs comprend 8 membres qui sont soit des activateurs (E2F1-3) soit des répresseurs de la transcription (E2F4-8) (*Herlihy and De Bruin, 2017*). Leurs activités peuvent être modulées grâce à l'interaction avec la protéine Retinoblastome (RB) via le domaine de transactivation, ou l'interaction avec d'autres protéines comme des protéines Polycomb qui interagissent avec le domaine marked-box de E2F6 (*Trimarchi et al., 2001*) (**figure 22A**).

Ces protéines partagent une grande similarité dans leur domaine de liaison à l'ADN ce qui leur permet d'interagir avec le motif TCCCGC défini comme le motif E2F. L'interaction avec l'ADN nécessite leur hétérodimérisation avec DP1/2. Seuls les facteurs E2F7 et 8 qui ont deux domaines d'interaction avec l'ADN en tandem, peuvent interagir avec l'ADN indépendamment de l'interaction avec DP1/2 (*Herlihy and De Bruin, 2017*) (**figure 22A**).

Les facteurs de transcriptions E2Fs régulent le cycle cellulaire au niveau de la phase S et la réplication de l'ADN. Les activateurs (E2F1-3) induisent l'expression des gènes cibles E2Fs impliqués dans la réplication cellulaire, tandis que les répresseurs (E2F4-8) répriment ces gènes à la fin de la phase S permettant la poursuite du cycle cellulaire (*Herlihy and De Bruin, 2017*). Les gènes cibles E2Fs peuvent donc être régulés par des facteurs de transcription différents, par exemple : le gène *Art27* est activé par E2F1 et est réprimé par E2F6 (*Oberley et al., 2003*). De plus, la perte de fonction de E2F6 peut être compensée par l'activité de E2F4 dans la répression des gènes cibles E2Fs (*Giangrande et al., 2004*).

Le facteur de transcription E2F6

Ce facteur de transcription est caractérisé par l'absence du domaine de transactivation et d'interaction avec la protéine RB. Le domaine de liaison à l'ADN lui permet de se lier à la séquence TCCCGC en présence de DP1 et de réprimer l'expression des gènes (*Cartwright et al., 1998*). E2F6 interagit avec des protéines Polycomb et fait partie du complexe PRC1.6 (*Ogawa et al., 2002; Trimarchi et al., 2001*) dans lequel il participe à son recrutement sur le génome dans les cellules HEK293 (*Stielow et al., 2018*) (**figure 22B**). Les souris KO pour *E2f6* sont viables et fertiles mais ont des transformations homéotiques : la vertèbre Thoracique 13 est transformée en Lombaire 1 et la vertèbre Lombaire 6 est transformée en Sacrale 1 (*Courel et al., 2008; Storre et al., 2002*) (**figure 22C**).

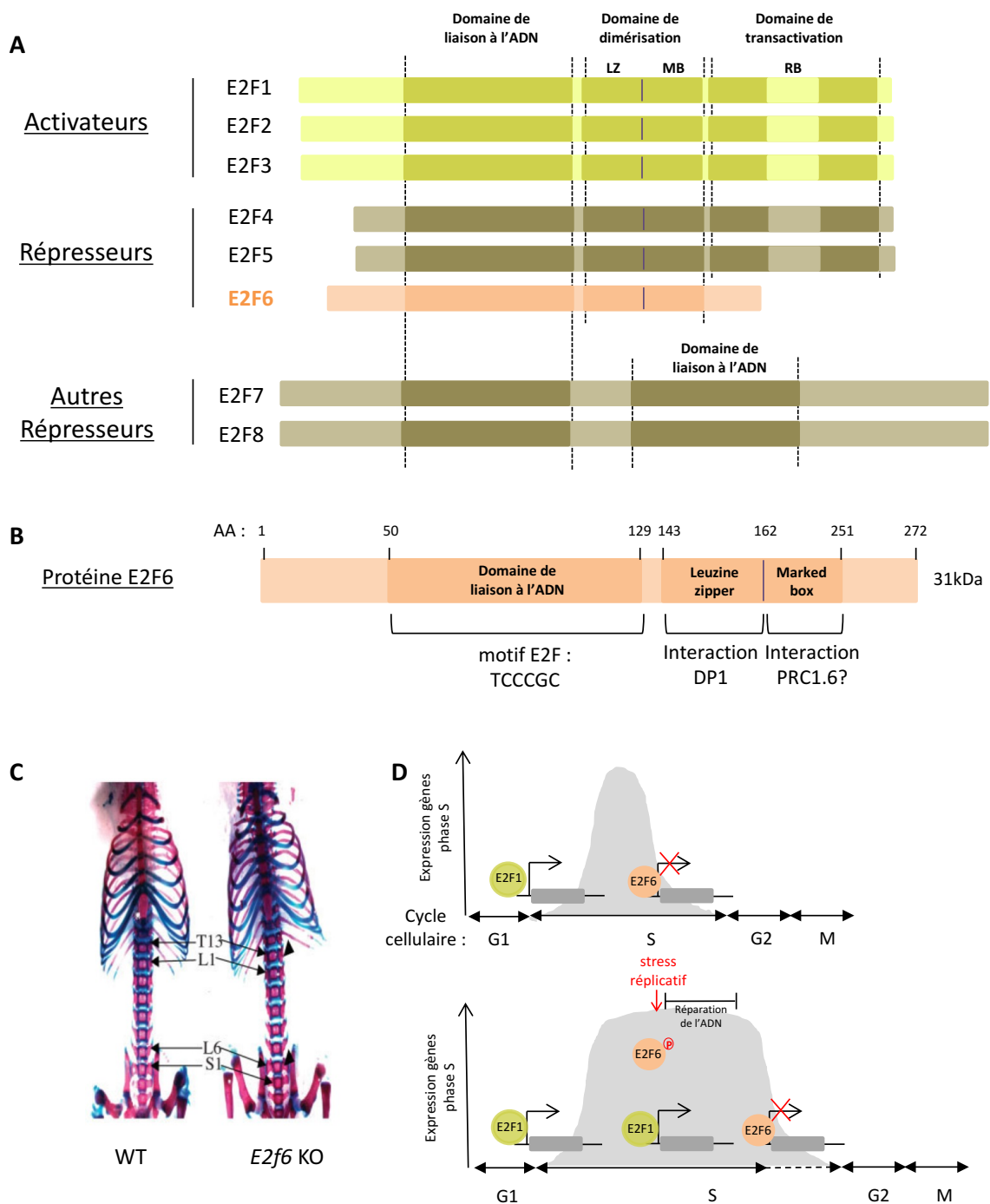


Figure 22 : Le facteur de transcription E2F6. A) E2F6 est un répresseur de la famille des facteurs de transcription E2Fs. B) E2F6 possède un domaine de liaison lui permettant d'interagir avec le motif d'ADN TCCCGC, un domaine de dimérisation comprenant les domaines leucine-zipper (LZ) et marked-box (MB) lui permettant d'interagir avec des co-répresseurs, mais ne possède pas le domaine de transactivation comprenant le domaine d'interaction avec la protéine rétinoblastome (RB). C) Les souris mutantes pour *E2f6* sont viables et fertiles mais ont comme phénotype des transformations homéotiques (de Courrel et al. 2008). D) E2F6 réprime les gènes impliqués dans la phase S permettant la poursuite du cycle cellulaire (haut). Dans le cas d'un stress réplcatif, la phase S est maintenue grâce à l'inhibition d'E2F6 permettant la réparation de l'ADN (bas)

Rôle d'E2F6 dans la régulation du stress réplicatif

La génération de KO de *E2f6* dans des cellules humaines conduit à une diminution de la prolifération et une accumulation de cellules en phase S du cycle cellulaire (Cartwright et al., 1998). En effet, E2F6 réprime les gènes impliqués dans la réplication de l'ADN (Giangrande et al., 2004) ce qui permet la sortie de la phase S et la poursuite du cycle cellulaire.

La réplication est une étape cruciale du cycle cellulaire. En effet, des défauts de copie de l'ADN peuvent conduire à une instabilité génomique des cellules filles et l'initiation de processus oncogéniques. Les défauts de réplication de l'ADN conduisent à une réponse cellulaire appelé stress réplicatif qui stoppe le cycle cellulaire (Herlihy and De Bruin, 2017). Le stress réplicatif active une cascade moléculaire qui conduit à l'activation de *Chk1* qui phosphoryle et inhibe E2F6 (Bertoli et al., 2013). Les gènes cibles E2Fs restent donc exprimés, et la cellule reste en phase S le temps que la machinerie de réparation de l'ADN corrige ces défauts (figure 22D).

Rôle d'E2F6 dans la restriction du programme gamétique

Comme pour l'ensemble du complexe PRC1.6, le facteur de transcription E2F6 a été démontré comme impliqué dans la répression de certains gènes gamétiques (Kehoe et al., 2008; Pohlers et al., 2005; Storre et al., 2002). Les mécanismes impliqués dans cette répression ne sont pas clairs. En effet certaines études suggèrent qu'E2F6 réprime les gènes gamétiques en recrutant des modifications d'histones comme H3K9me3 (Storre et al., 2005) ou H3K27me3 (Leseva et al., 2013). Une autre étude propose qu'E2F6 participe au recrutement de la méthylation de l'ADN au niveau de leurs promoteurs (Velasco et al., 2010).

Lien entre E2F6 et la méthylation de l'ADN

Cette hypothèse découle de l'étude de l'effet d'une mutation hypomorphique de *Dnmt3b* chez la souris. Cette mutation conduit à une hypométhylation et une dérégulation de certains gènes gamétiques, qui sont également ciblés par le facteur de transcription E2F6. Les auteurs ont alors montré une interaction entre E2F6 et DNMT3B dans les MEFs et une perte d'enrichissement de DNMT3B dans les cellules KO pour *E2f6*. Cette étude suggère alors le recrutement de la méthylation de l'ADN par le facteur de transcription E2F6 sur certains gènes gamétiques. Enfin, les souris possédant la mutation hypomorphique de *Dnmt3b* présentent le même phénotype de transformation homéotique que les souris KO pour *E2f6*, renforçant ainsi l'idée d'une action synergique entre ces deux protéines (Velasco et al., 2010).

OBJECTIFS

Le modèle initialement admis de la mise en place de la méthylation de l'ADN implique DNMT3A et DNMT3B dans la méthylation *de novo*, et DNMT1 dans son maintien au cours des divisions cellulaires. Néanmoins, plusieurs études ont proposé un modèle coopératif selon lequel les différentes DNMTs coopèrent dans la méthylation *de novo* et son maintien. Ce modèle n'a néanmoins pas été validé *in vivo* dans les embryons. La première problématique de ma thèse a donc été d'étudier **la contribution des différentes ADN méthyltransférases dans la méthylation du génome lors de l'embryogenèse.**

Le rôle de la méthylation de l'ADN dans la régulation du génome chez la souris a surtout été étudié sur des modèles cellulaires tels que les cellules ES. Les cellules ES Dnmt-TKO qui sont totalement dépourvues de méthylation de l'ADN ne montrent qu'une réactivation limitée des gènes et éléments transposables. De plus les cellules ES hypométhylées sont viables, tandis que les embryons invalidés pour certaines DNMTs ne se développent pas après la phase de reméthylation du génome. Ceci suggère ainsi un rôle plus important de la méthylation de l'ADN dans la régulation de la transcription dans l'embryon comparé aux cellules ES. A ce jour, aucune étude transcriptomique n'a été conduite sur des embryons fortement hypométhylés. La deuxième problématique de ma thèse a donc été de déterminer **le rôle de la méthylation de l'ADN dans la régulation de l'intégrité du transcriptome dans l'embryon.**

Ces problématiques seront traitées dans la partie 1 : rôle des ADN méthyltransférases au cours du développement chez la souris.

La phase de reméthylation du génome dans l'embryogenèse précoce s'accompagne de l'acquisition de la méthylation de l'ADN sur les îlots CpG des promoteurs de gènes gamétiques. Néanmoins, les mécanismes permettant le recrutement spécifique de la méthylation de l'ADN sur les promoteurs de gènes gamétiques ne sont pas élucidés. Le complexe Polycomb non canonique PRC1.6 est associé à la répression des gènes gamétiques dans les cellules ES et pourrait participer à l'adressage de la méthylation de l'ADN sur le génome. Parmi les membres du complexe PRC1.6, le facteur de transcription E2F6 est très peu caractérisé, cependant des études suggèrent un lien fonctionnel entre E2F6 et DNMT3B. Afin de mieux comprendre les mécanismes de ciblage de la méthylation de l'ADN dans le génome, je me suis donc intéressé dans un second temps au **rôle du facteur de transcription E2F6 dans la répression épigénétique des gènes gamétiques durant le développement embryonnaire.**

Cette problématique sera traitée dans la partie 2 : Rôle du facteur de transcription E2F6 dans le ciblage épigénétique des gènes gamétiques dans l'embryon.

RÉSULTATS

1. Rôle des ADN méthyltransférases au cours du développement chez la souris

1.1. Contexte et objectifs

La méthylation de l'ADN est présente sur tout le génome hormis sur certaines régions spécifiques riches en CpG, appelées îlots CpG. Elle est catalysée par la famille des ADN méthyltransférases dont font partie DNMT1, DNMT3A, DNMT3B et DNMT3C (exprimé uniquement dans les gamètes). La protéine DNMT3L est inactive d'un point de vue catalytique, mais interagit et potentialise l'activité des autres enzymes DNMT3. Chez la souris, le développement embryonnaire est caractérisé par une large reprogrammation de la méthylation de l'ADN. Le zygote est déméthylé après la fécondation jusqu'au stade blastocyste qui sera ensuite reméthylé après son implantation.

Cependant, aucune étude systématique sur la contribution des différentes DNMTs à la méthylation de l'ADN dans l'embryon n'a été réalisée à l'échelle génomique. Jusqu'à présent, les études sur les embryons invalidés pour *Dnmt3a/b* ou *Dnmt1* se sont focalisées sur quelques locus spécifiques (Hirasawa et al., 2008; Lei et al., 1996; Oda et al., 2006; Okano et al., 1999; Walsh et al., 1998). Ces études ont montré que la phase de reméthylation du génome nécessite l'activité de DNMT3A et DNMT3B, car leur absence conduit à l'hypométhylation des rétrotransposons IAP et des gènes *Rhox* (Oda et al., 2006; Okano et al., 1999). Néanmoins, l'inactivation de *Dnmt3a* et *Dnmt3b* dans les cellules ES et dans des fibroblastes différenciés conduit à la déméthylation de ces cellules, suggérant un rôle de ces enzymes dans le maintien de la méthylation au cours des divisions cellulaires (Chen et al., 2003; Dodge et al., 2005; Leung et al., 2014b; Li et al., 2015; Thakur et al., 2016). L'enzyme DNMT1 est considérée comme la principale ADN méthyltransférase de maintien, mais certaines études proposent également une activité *de novo* dans les cellules ES (Arand et al., 2012; Vilkaitis et al., 2005). Ces différentes études ont conduit à proposer un modèle de régulation de la méthylation de l'ADN plus complexe, selon lequel DNMT3A, DNMT3B et DNMT1 coopèrent dans la méthylation *de novo* et dans son maintien au cours des divisions cellulaires (Jones and Liang, 2009; Walton et al., 2011). Dans ce contexte, nous avons souhaité clarifier le rôle des 3 enzymes DNMT1, DNMT3A et DNMT3B dans la méthylation de l'ADN *in vivo* au cours du développement chez la souris, en utilisant des approches génétiques et génomiques.

Parallèlement, aucune étude transcriptomique n'a été réalisée sur des embryons déméthylés pour étudier l'impact de la méthylation de l'ADN sur la régulation du génome. A ce jour, les études sur des embryons mutants pour les *Dnmts* se sont concentrés sur des locus spécifiques et montrent que la méthylation de l'ADN est impliquée dans la répression des gènes gamétiques (Borgel et al., 2010), des

gènes soumis à l’empreinte parentale (Li et al., 1993), des gènes *Rhox* (Oda et al., 2006) ou de certains rétrotransposons (Walsh et al., 1998). Une étude transcriptomique a été réalisée précédemment dans le laboratoire sur des embryons invalidés pour le gène *Dnmt3b* (Auclair et al., 2014). Cependant, cette inactivation conduit seulement à une déméthylation partielle du génome et ne permet donc pas d’étudier les conséquences d’une hypométhylation importante (Auclair et al., 2014). Par ailleurs, le rôle de la méthylation de l’ADN dans la régulation du génome dans l’embryon murin a surtout été étudié dans les cellules ES. Les cellules ES triple mutantes pour les *Dnmts* (TKO) n’ont révélé qu’une contribution modeste de la méthylation de l’ADN dans la régulation de l’expression des gènes et des rétrotransposons (Karimi et al., 2011). Ceci peut être causé par des mécanismes de compensation pour l’absence de la méthylation de l’ADN comme il l’a été montré pour les rétrotransposons (Walter et al., 2016). De plus, les cellules ES hypométhylées sont viables, tandis que les embryons déplétés pour certaines DNMTs ne se développent pas après la phase de reméthylation du génome (Li et al., 1993; Okano et al., 1999). Ceci suggère que la méthylation de l’ADN pourrait jouer un rôle plus important dans la régulation de la transcription dans l’embryon par rapport aux cellules ES.

Dans une première partie, nous nous sommes intéressés à la contribution des différentes ADN méthyltransférases dans la méthylation du génome, ainsi qu’au rôle de la méthylation de l’ADN dans la régulation de l’intégrité transcriptomique.

1.2. Stratégie expérimentale

Pour répondre à ces problématiques, nous avons inactivé les gènes codant pour les ADN méthyltransférases *de novo* *Dnmt3a* et *Dnmt3b*, ainsi que l’enzyme de maintien de la méthylation *Dnmt1* dans les embryons.

Les souris *Dnmt1-2lox* ont deux sites LoxP de part et d’autre de l’exon 4 et de l’exon 5. Le croisement de cette lignée avec une lignée exprimant une Cre ubiquitaire sous la dépendance de promoteur de *Actb* a conduit à l’excision de la séquence entre les sites loxP, ce qui crée un décalage du cadre de lecture engendrant une protéine tronquée qui ne contient pas les domaines de régulation ni le domaine catalytique (**Figure 23A**).

Les souris *Dnmt3a-2lox* et *Dnmt3b-2lox* ont des sites loxP qui entourent les exons 19 et 17-20 respectivement. Ces exons codent pour des parties essentielles du domaine catalytique. Ainsi, la recombinaison des sites loxP engendre une perte de fonction de ces gènes par inactivation du domaine catalytique (**Figure 23B, C**).

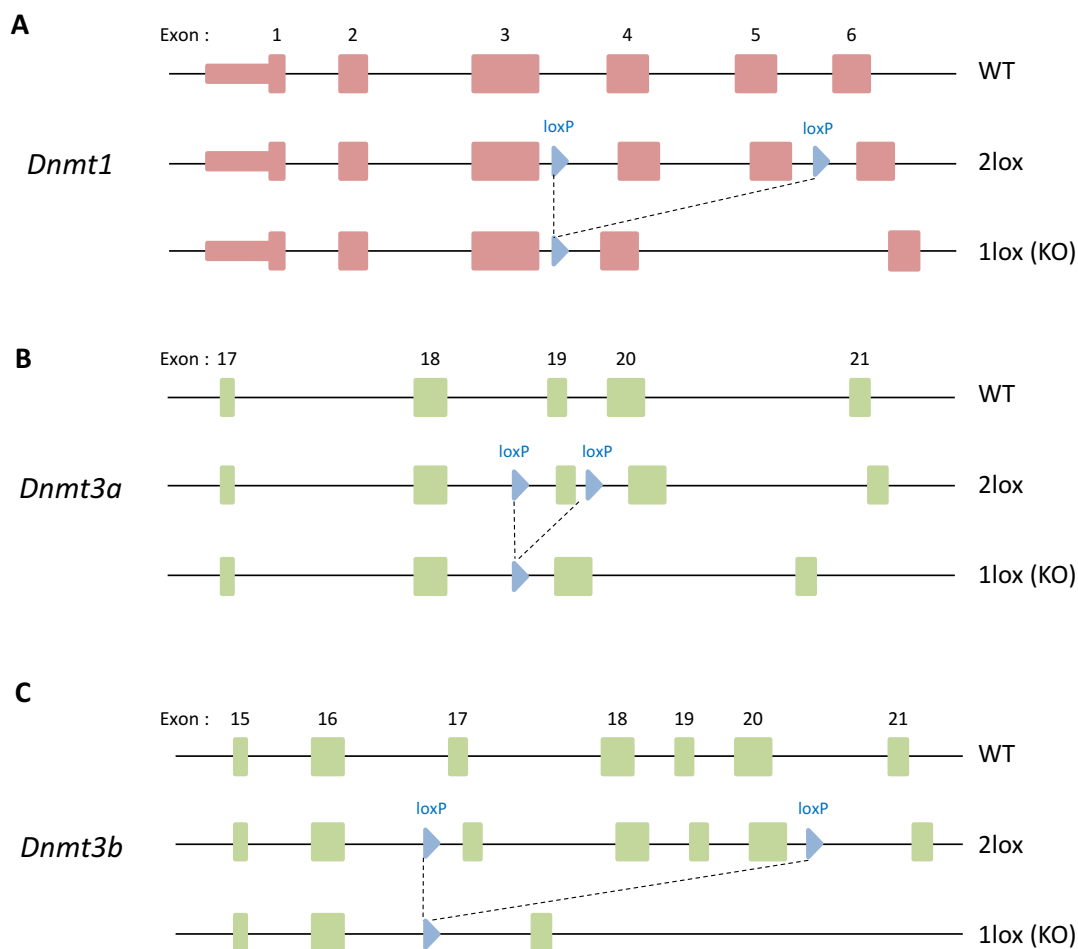


Figure 23 : Génération de souris invalidées pour les ADN méthyltransférases. Les souris *Dnmt1* (A), *Dnmt3a* (B) et *Dnmt3b* (C) floxées ont été croisées avec une souris *Actb-Cre* permettant l'excision d'exons engendrant une perte de fonction des gènes.

Les souris mutées pour *Dnmt1* meurent avant le stade embryonnaire E13.5 (*Li et al., 1992*), tandis que les embryons doubles mutés pour *Dnmt3a* et *Dnmt3b* ne se développent pas au-delà de E9.5 (*Okano et al., 1999*). Pour étudier le rôle de ces enzymes dans le développement, nous avons choisi de collecter des embryons mutants après la phase de reméthylation du génome au stade E8.5 pour des analyses comparatives de méthylome et de transcriptome.

1.3. Travaux réalisés

Mon implication dans ce projet :

- gestion des lignées de souris, génération des embryons KO et dissection des embryons au stade embryonnaire E8.5.
- préparation des échantillons d'ADN et d'ARN.

- préparation des bibliothèques RRBS et WGBS.
- validation de gènes dérégulés par RT-qPCR.
- Inactivation conditionnelle de *Dnmt3a* et *Dnmt3b* dans les MEFs (en partenariat avec Patryk Ngondo).
- participation aux expériences de déméthylation ciblée du promoteur de gènes gamétiques en utilisant une technique d'édition de l'épigénome. Ce travail a majoritairement été réalisé par Andrea Argüeso Lleida et Hala Al Adhami.
- participation à l'analyse bioinformatique des données de séquençage de méthylome et de transcriptome. Cette partie du travail a été majoritairement réalisée par Michael Dumas et Michael Weber.

1.4. Article 1: Genome-wide analyses in the mouse embryo reveals the importance of DNA methylation for transcription integrity

(En révision à Nature Communications)

Genome-wide analyses in the mouse embryo reveals the importance of DNA methylation for transcription integrity

Thomas Dahlet^{1,2}, Andrea Argüeso Lleida^{1,2}, Hala Al Adhami^{1,2}, Michael Dumas^{1,2}, Ambre Bender^{1,2}, Richard P Ngondo^{1,2,3}, Manon Tanguy^{1,2}, Judith Vallet^{1,2}, Ghislain Auclair^{1,2}, Anaïs F Bardet^{1,2}, Michael Weber^{1,2}#

¹ University of Strasbourg

² CNRS UMR7242, Biotechnology and Cell Signaling, 300 Bd Sébastien Brant, 67412 Illkirch Cedex, France

³ IBMP, CNRS UPR 2357, Strasbourg

#Corresponding author: michael.weber@unistra.fr

Abstract

Mammalian embryos acquire global DNA methylation of their genome during implantation. However the exact roles of DNA methyltransferases (DNMTs) in embryos have not been studied comprehensively. Here we systematically analyzed the consequences of genetic inactivation of *Dnmt1*, *Dnmt3a* and *Dnmt3b* on the methylome and transcriptome of mouse embryos. We found a strict division of function between DNMT1, solely responsible for maintenance methylation, and DNMT3A/B, solely responsible for methylation acquisition in development. By analyzing severely hypomethylated embryos, we uncover multiple functions of DNA methylation that is used as a mechanism of repression for a panel of genes including not only imprinted genes and germline genes, but also lineage-committed genes and 2 cell-specific genes. DNA methylation also suppresses multiple transposons and illegitimate transcripts initiating from cryptic promoters in transposons and gene bodies. Our work provides a thorough analysis of the roles of DNA methyltransferases and the importance of DNA methylation for transcriptome integrity in mammalian embryos.

Keywords: DNA methylation, Dnmt1, Dnmt3a, Dnmt3b, embryogenesis

Introduction

Cytosine methylation of DNA in vertebrate genomes has important functions in gene regulation, development and diseases¹. This modification occurs at CpGs and is abundant genome-wide, except at CpG islands (CGIs) that are refractory to DNA methylation. The mouse genome encodes four active DNA methyltransferase (DNMT) enzymes: DNMT1, DNMT3A, DNMT3B and DNMT3C. Another member of the family, DNMT3L, is catalytically inactive but interacts with and stimulates the activity of the other DNMT3 enzymes. DNMT3C and DNMT3L are expressed only in germ cells and their inactivation leads to impaired gametic DNA methylation and infertility^{2,3}. In contrast, the inactivation of DNMT1, DNMT3A or DNMT3B in mice leads to embryonic or postnatal lethality^{4,5}, which illustrates the essential role of these enzymes in somatic development.

Global genome methylation is established after implantation of the embryo in the mouse and subsequently maintained in most cell lineages. DNMT3A and DNMT3B are thought to perform all *de novo* methylation of DNA during development, based on data showing that the double knockout of *Dnmt3a/b* impairs *de novo* methylation of Intracisternal A-particle (IAP) transposons and *Rhox* genes in embryos^{4,6}. However, the consequences of the double inactivation on the methylome have not been investigated genome-wide. Moreover, the single inactivation of *Dnmt3a* or *Dnmt3b* has only a moderate impact on DNA methylation levels in mouse embryos^{4,7}, suggesting either strong redundancy or involvement of other enzymes in *de novo* methylation. In contrast, DNMT1 preferentially binds hemimethylated CpGs and is thought to be the main enzyme responsible for maintenance DNA methylation after replication. However, DNMT1 also shows capabilities for *de novo* DNA methylation *in vitro*, in mouse embryonic stem (ES) cells and oocytes⁸⁻¹². Conversely, late passage *Dnmt3a/b* knockout ES cells show reduced methylation genome-wide¹³⁻¹⁵ and at imprinted germline DMRs (gDMRs)¹⁶, suggesting that DNMT3A/B are also required for the faithful maintenance of CpG methylation in development. Despite these studies suggesting complex functions of DNMTs, the *in vivo* roles of these enzymes in embryonic development remain elusive. Previous investigations of the roles of DNMTs in embryos were limited to locus-specific analysis^{4,6,17-19}, which highlights the necessity for complete methylomes of *Dnmt* mutant embryos to validate models of DNMT functions *in vivo*.

Besides deciphering the role of DNMTs, further work is also needed to illuminate the transcriptional roles of DNA methylation in development. Previous studies in *Dnmt* knockout embryos showed that DNA methylation is required to repress some imprinted genes^{20,21}, *Rhox* genes⁶, germline genes²² and IAP transposons¹⁷. However, no genome-wide transcriptome analysis in strongly hypomethylated embryos has been conducted so far. Moreover, transcriptome profiling in *Dnmt* triple knockout (TKO) mouse ES cells devoid of DNA methylation revealed only a minor impact on the expression of genes and transposable elements (TEs)^{23,24}. One explanation is that ES cells use

other mechanisms to compensate for the loss of DNA methylation, mimicking what is happening during epigenetic reprogramming in preimplantation embryos and primordial germ cells (PGCs)²⁵. Indeed, the repression of ERV retroviruses in mESCs is primarily mediated by KAP1 and SETDB1 responsible for H3K9me3, rather than DNA methylation^{23,26-28}. In contrast, IAP repression becomes dependent on DNA methylation in differentiated cells²⁹, supporting the model that DNA methylation is not important for initial repression in early embryonic cells but for the transition to long term silencing.

Here we performed a comprehensive investigation of the role of DNMTs in methylation and transcription during the phase of global genome remethylation in the mouse embryo. We report genome-wide methylomes in *Dnmt1* knockout and *Dnmt3a/b* double knockout postimplantation embryos (embryonic day 8.5), which elucidates the *in vivo* roles of these enzymes in setting up DNA methylation patterns. We show that severely hypomethylated embryos overexpress a panel of genes, transposons and illegitimate transcripts initiating from cryptic promoters, revealing the multiple roles of DNA methylation for the maintenance of transcriptional integrity in mouse embryos.

Results

Methylome profiling of *Dnmt* mutant embryos

To systematically assess the contribution of DNMTs to DNA methylation *in vivo*, we generated base-resolution DNA methylomes in *Dnmt* mutant embryos. Using a *Dnmt1* 2lox allele³⁰, we created *Dnmt1* mutant embryos lacking the exons 4 and 5, which creates an out-of-frame splice and a functional null allele. As shown previously¹⁹, the *Dnmt1*^{-/-} embryos were distinguishable at embryonic day (E) 8.5 and showed growth retardation (**Figure 1A**). For DNMT3 enzymes, we previously showed that methylation is only partially reduced in single knockouts, suggesting redundancy between DNMT3A and DNMT3B⁷. To address this, we generated double knockout (DKO) embryos for *Dnmt3a* and *Dnmt3b*. Confirming previous observations⁴, DKO embryos resembled *Dnmt1* null embryos and showed growth retardation at E8.5 (**Figure 1A**). We performed Reduced Representation Bisulfite Sequencing (RRBS) in three *Dnmt1*^{-/-} and WT littermate controls at E8.5, as well as three DKO E8.5 embryos and controls from the same litters (**Annexe 1.1**). The RRBS data were highly reproducible between replicates (**Supplementary Figure 1A-C**). To have a more complete view, we also performed whole genome bisulfite sequencing (WGBS) in WT, *Dnmt1*^{-/-} and DKO E8.5 embryos (**Annexe 1.2**). The average sequencing depth after deduplication was 12.5X and close to 90% of the CpGs in the reference genome were sequenced at least 5X in each dataset

(Supplementary Figure 1D). The data obtained independently by RRBS and WGBS correlate well (Supplementary Figure 1E), demonstrating the reliability of the datasets.

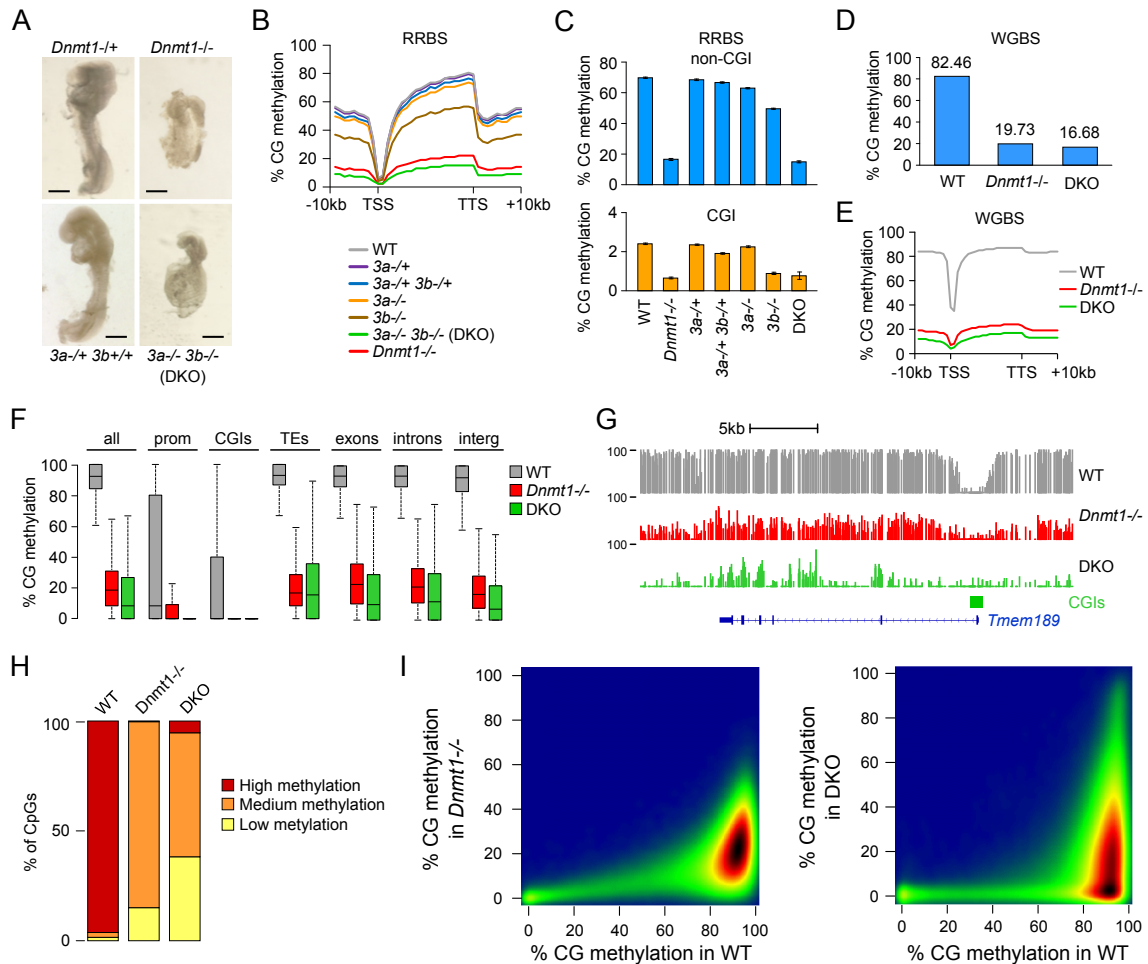


Figure 1. Methylome profiling in *Dnmt* mutant embryos. **A.** Images of *Dnmt1*^{-/-} and DKO embryos dissected at E8.5 compared to littermate heterozygous controls (which are identical to WT). Black bars: 300 μ m. **B.** Average distribution on CG methylation over RefSeq genes and 10 kb flanking sequences calculated from the RRBS data in E8.5 embryos. Our previous data from single *Dnmt3a*^{-/-} and *Dnmt3b*^{-/-} embryos⁷ was included for comparison. TSS= Transcription start site, TTS= Transcription termination site. **C.** Methylation scores of CpGs located outside of CpG islands (non-CGI, top) or in CpG islands (CGI, bottom) measured by RRBS in *Dnmt* mutant and control embryos. The data is represented as the mean \pm SEM ($n=3$ embryos for all genotypes except $n=2$ for *Dnmt3a*^{-/-} and *Dnmt3b*^{-/-}). **D.** Bar plot representing the mean CG methylation level measured by WGBS in WT, *Dnmt1*^{-/-} and DKO E8.5 embryos. **E.** Metaplots of CG methylation levels in RefSeq genes and 10 kb flanking sequences calculated from the WGBS data in WT, *Dnmt1*^{-/-} and DKO E8.5 embryos. **F.** Box plots of the CG methylation levels measured by WGBS in different genomic features in WT, *Dnmt1*^{-/-} and DKO E8.5 embryos. **G.** Example of genome browser view of WGBS methylation profiles in WT, *Dnmt1*^{-/-} and DKO E8.5 embryos. Each track shows the percent methylation of individual CpGs between 0 and 100%. CpG islands (CGIs) depicted by green rectangles and RefSeq gene annotations are shown below the tracks. **H.** Stacked bar graph representing the proportions of CpGs with high (>50%), medium (10-50%) and low (<10%) methylation in WT, *Dnmt1*^{-/-} and DKO embryos. **I.** Density scatter plots comparing WGBS methylation scores in 1kb windows between WT, *Dnmt1*^{-/-} and DKO E8.5 embryos. The density of points increases from blue to dark red

DNMT1 is required to sustain DNA methylation genome-wide

We found a strong reduction of genomic methylation in *Dnmt1*^{-/-} embryos (**Figure 1B**). The mean CpG methylation level measured by RRBS in non-CGI sequences dropped from 69.8% in WT to 16.7% in *Dnmt1*^{-/-} embryos, whereas CpG methylation of CGIs dropped from 2.4% in WT to 0.7% in *Dnmt1*^{-/-} embryos (**Figure 1C**). Confirming the RRBS results, the global CG methylation levels measured by WGBS in WT and *Dnmt1*^{-/-} embryos were 82.46% and 19.73%, respectively (**Figure 1D-E**). The loss of methylation upon inactivation of *Dnmt1* is truly global and occurs uniformly across all genomic sequences including exons, introns, intergenic regions and transposable elements (**Figure 1F-G**). All sequences have low to intermediate methylation and no sequences retain high methylation in *Dnmt1*^{-/-} embryos (**Figure 1H-I**), indicating that DNMT1 is universally required to sustain DNA methylation at all genomic sequences in embryonic development.

De novo DNA methylation is abolished in *Dnmt3a*^{-/-} *Dnmt3b*^{-/-} embryos

Next, we analyzed the methylome of *Dnmt3a/b* double mutant embryos. The data were also compared to our previous data from single *Dnmt3* knockouts⁷. In contrast to the single knockouts, the double inactivation of *Dnmt3a/b* leads to a strong reduction of DNA methylation, demonstrating redundancy between DNMT3A and DNMT3B genome-wide (**Figure 1B**). The mean CpG methylation level measured by RRBS in non-CGI sequences dropped from 69.8% in WT to 63.0% in *Dnmt3a*^{-/-}, 49.6% in *Dnmt3b*^{-/-}, and 15.0% in DKO embryos (**Figure 1C**). Accordingly, the CpG methylation level measured by WGBS dropped from 82.5% in WT to 16.7% in DKO embryos (**Figure 1D-E**). The loss of methylation in DKO embryos affects all genome compartments (**Figure 1F**). However, in contrast to *Dnmt1*^{-/-} embryos, DKO embryos do not display a uniform loss of methylation across the genome (**Figure 1G**). Instead they contain a high proportion of fully demethylated sequences as well as a high proportion of highly methylated sequences (**Figure 1H-I**).

To determine the origin of high methylation in DKO embryos, we compared the WGBS methylation patterns of DKO embryos to those of pre-implantation inner cell mass (ICM)³¹. Strikingly, visual inspection of the methylation patterns in ICM and DKO embryos revealed that they are highly similar (**Figure 2A** and **Supplementary Figure 2A**). To confirm this, we performed a pairwise comparison of WGBS methylation scores and revealed a strong positive correlation between methylation in ICM and DKO embryos (**Figure 2B**). Next, we determined which sequences retain high DNA methylation in DKO embryos and found that the most highly methylated sequences are enriched for transposable elements (TEs) (**Supplementary Figure 2B**). Amongst the most methylated TE families are IAP, RLTR6 and MMERVK10C elements, which carry high methylation in DKO embryos at levels identical to blastocysts (**Supplementary Figure 2C-D**). This strongly suggests that in absence of DNMT3A/B, methylation in E8.5 DKO embryos arises from the maintenance of pre-existing DNA

methylation from the blastocysts. Finally, to determine if *de novo* methylation happens in DKO embryos, we selected all the sequences that are hypomethylated in blastocysts and gain methylation in E8.5 embryos, and found that *de novo* methylation at these sequences is completely abolished in DKO embryos (**Figure 2C**). Altogether this indicates that DNMT3A/B are responsible for all *de novo* methylation in embryos and DNMT1, when alone, is not able to methylate previously unmethylated sequences in embryos.

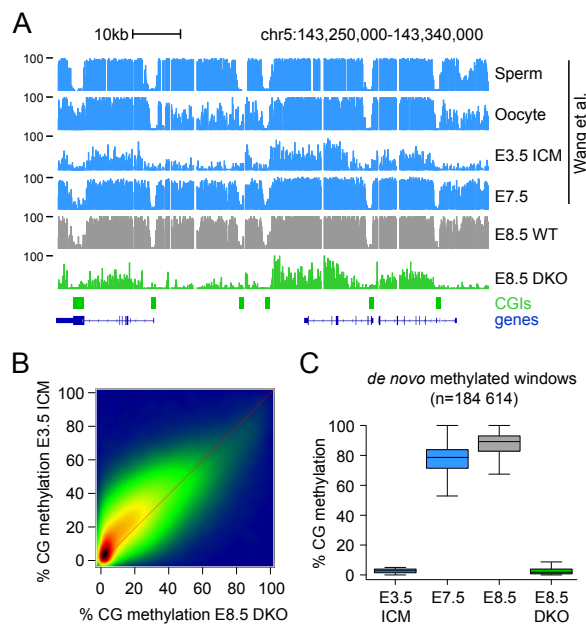


Figure 2. Absence of *de novo* methylation in DKO embryos. **A.** Genome browser tracks of WGBS methylation profiles in gametes, early embryos³¹ and E8.5 DKO embryos, highlighting the strong similarity of methylation patterns between the ICM (Inner Cell Mass) and DKO embryos. CpG islands (CGIs) depicted by green rectangles and RefSeq gene annotations are shown below the tracks. **B.** Density scatter plot correlating WGBS methylation scores in 1kb windows between the E3.5 ICM and E8.5 DKO embryos. The density of points increases from blue to dark red. **C.** Box plots of the CG methylation levels in ICM, postimplantation embryos and E8.5 DKO embryos of all genomic windows *de novo* methylated at implantation, illustrating that *de novo* methylation is abolished in DKO embryos.

Role of DNMT3A/B in methylation maintenance

Given the proposed function of DNMT3A/B in the maintenance of DNA methylation in ES cells, we investigated the role of DNMT3A/B in maintenance DNA methylation *in vivo*. First, we quantified methylation of 20 known gametic DMRs (gDMRs) of imprinted loci, which arise from the post-fertilization maintenance of allelic methylation established in the parental gametes. The methylation of all gDMRs is preserved in DKO embryos (**Figure 3A-B**), indicating that DNMT3A/B have no significant contribution to the maintenance of gametic-derived methylation imprints in embryos.

This agrees with two previous studies showing that DNMT3A/B are dispensable for maintenance methylation of the *Igf2r*, *H19* and *Dlk1/Gtl2* gDMRs *in vivo*^{6,18}. In contrast, all gDMRs are demethylated in *Dnmt1*^{-/-} embryos, confirming that DNMT1 is the main enzyme propagating methylation imprints in embryos (**Figure 3A-B**). Moreover, extending previous findings^{20,21}, RNA-seq in *Dnmt* mutant embryos (**Annexe 1.3**) revealed marked changes in expression of imprinted genes in *Dnmt1*^{-/-} but not DKO embryos, including, as expected, both downregulation (*Grb10*, *Igf2*, *Kcnq1*...) and ~2-fold upregulation (*Snrpn*, *Peg3*, *H19*...) (**Figure 3C**). The exception is *Zdbf2* that shows strongly reduced expression in DKO embryos (**Figure 3C**), validating the model that embryonic *de novo* methylation is required for an unusual switch from a maternal to paternal DMR and activate paternal-specific transcription of *Zdbf2*³². In summary, we show that DNMT1 alone mediates maintenance of methylation imprints and provide *in vivo* validation for the role of DNA methylation at all imprinted loci.

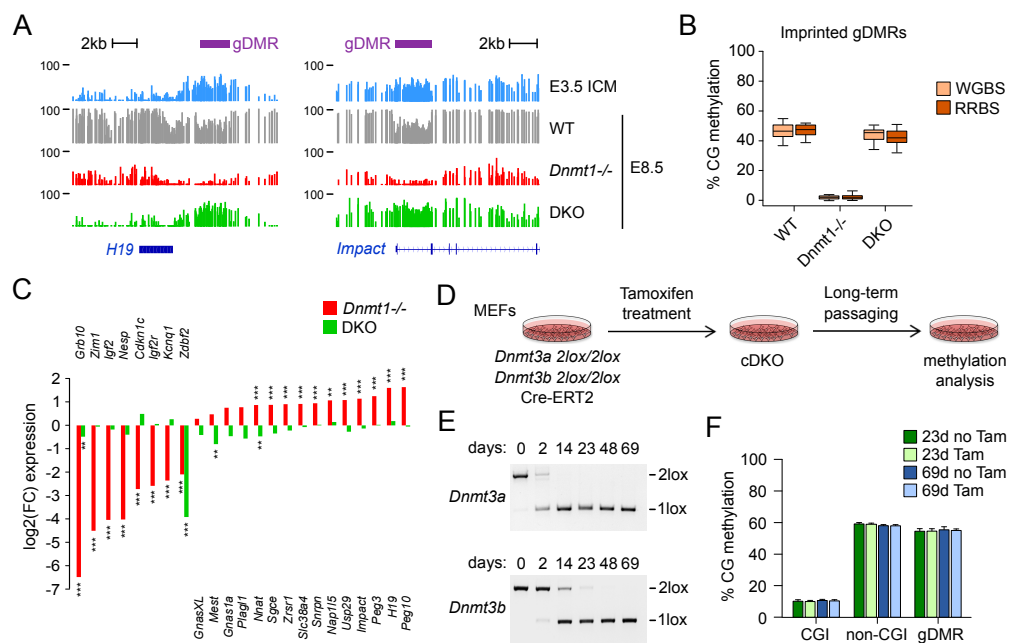


Figure 3. Role of DNMTs in methylation imprints and maintenance methylation. **A.** Genome browser tracks of WGBS methylation profiles at the *H19* and *Impact* imprinted loci in early embryos and *Dnmt* mutant embryos. The positions of the imprinted germline DMRs (gDMRs) are depicted by purple rectangles. **B.** Box plots of the CG methylation levels of all known imprinted germline DMRs measured by WGBS and RRBS in WT, *Dnmt1*^{-/-} and DKO embryos. **C.** Fold change of expression of imprinted genes measured by RNA-seq in *Dnmt1*^{-/-} (red bars) and DKO (green bars) embryos. **: $p < 0.01$, ***: $p < 0.001$ (DESeq adjusted p value). **D.** Experimental outline for investigating the maintenance function of DNMT3A/B in MEFs. *Dnmt3a* 2lox/2lox *Dnmt3b* 2lox/2lox Cre-ERT2 immortalized MEFs expressing Cre-ERT2 were treated with Tamoxifen to generate conditional double knock-out (cDKO) cells, and methylation was quantified by RRBS after long-term passaging. **E.** Evaluation by PCR genotyping of the Cre-mediated recombination of the *Dnmt3a* and *Dnmt3b* 2lox alleles in cDKO fibroblasts. The number of days of Tamoxifen treatment are indicated above the gels. **F.** CG methylation levels quantified by RRBS in cDKO fibroblasts. The graph shows the methylation levels in CpG islands (CGI), non-CGI regions and imprinted gDMRs in cells treated with Tamoxifen (Tam) or not treated with Tamoxifen (no Tam) after 23 and 69 days of culture (n = 3 independent replicates), revealing no global hypomethylation in cDKO MEFs.

To investigate a possible maintenance function of DNMT3A/B in somatic cells, it is necessary to perform conditional inactivation after *de novo* methylation has been completed. To this aim, we derived *Dnmt3a*^{2lox/2lox} *Dnmt3b*^{2lox/2lox} immortalized MEFs and performed conditional inactivation with a tamoxifen-inducible CRE recombinase (Cre-ERT2) (**Figure 3D**). Tamoxifen treatment led to efficient recombination of the *Dnmt3a* and *Dnmt3b* alleles (**Figure 3E** and **Supplementary Figure 3A**) and the cells were cultivated for up to 69 days to allow multiple cell divisions. RRBS performed at 23 days and 69 days of culture revealed no evidence of decreased methylation genome-wide (**Figure 3F** and **Supplementary Figure 3B**). After 69 days of culture, we only identified 11 hypomethylated DMRs in *Dnmt3a/b* deficient cells that most likely reflect *de novo* methylation events that happened during the course of cell culture (**Supplementary Figure 3C-D**). Taken together, our data indicate no major role of DNMT3A/B in maintenance methylation in embryos and differentiated cells.

DNA methylation suppresses the germline program in embryos

Next, we explored the RNA-seq data to investigate the consequences of the absence of DNA methylation on gene expression in *Dnmt*-mutant embryos. RNA-seq confirmed the knockout of critical exons in the *Dnmt* genes (**Supplementary Figure 4A-B**). For DKO embryos, we used WT and *Dnmt3a*^{-/+} littermate controls after having validated that *Dnmt3a*^{-/+} embryos show no significant gene expression changes compared to WT (data not shown). This analysis identified 377 upregulated and 65 downregulated genes (fold change > 3, adjusted *p* value < 0.001) in *Dnmt1*^{-/-} embryos, and 547 upregulated and 41 downregulated genes in DKO embryos (**Supplementary Figure 4C** and **Annexe 1.4**). Principal component analysis showed high similarity between *Dnmt1*^{-/-} and DKO samples, which cluster separately from the controls (**Figure 4A**). Indeed, there is a good correlation between the expression changes in *Dnmt1*^{-/-} and DKO embryos (**Supplementary Figure 4D**), and the genes misregulated in *Dnmt1*^{-/-} and DKO embryos strongly overlap (**Figure 4B**). In contrast, these genes only weakly overlap with the genes upregulated in TKO mESCs (**Supplementary Figure 4E-F**), mostly because many show a strong basal expression in WT ESCs (**Supplementary Figure 4G**).

We then focused on the genes upregulated in DKO embryos and classified them in 3 groups: the group 1 includes genes with Low CpG Promoters (LCP), the group 2 includes genes with unmethylated CpG-rich promoters (Intermediate or High CpG Promoters, ICP or HCP), and the group 3 includes genes with methylated CpG-rich promoters (**Figure 4C**). The genes in the group 2 show only weak derepression compared to the two other groups and could partly reflect indirect effects (**Supplementary Figure 4H**). Whereas no significant ontology terms are associated with the groups 1 and 2, the group 3 is strongly enriched for ontology terms related to germline functions (**Figure 4C**).

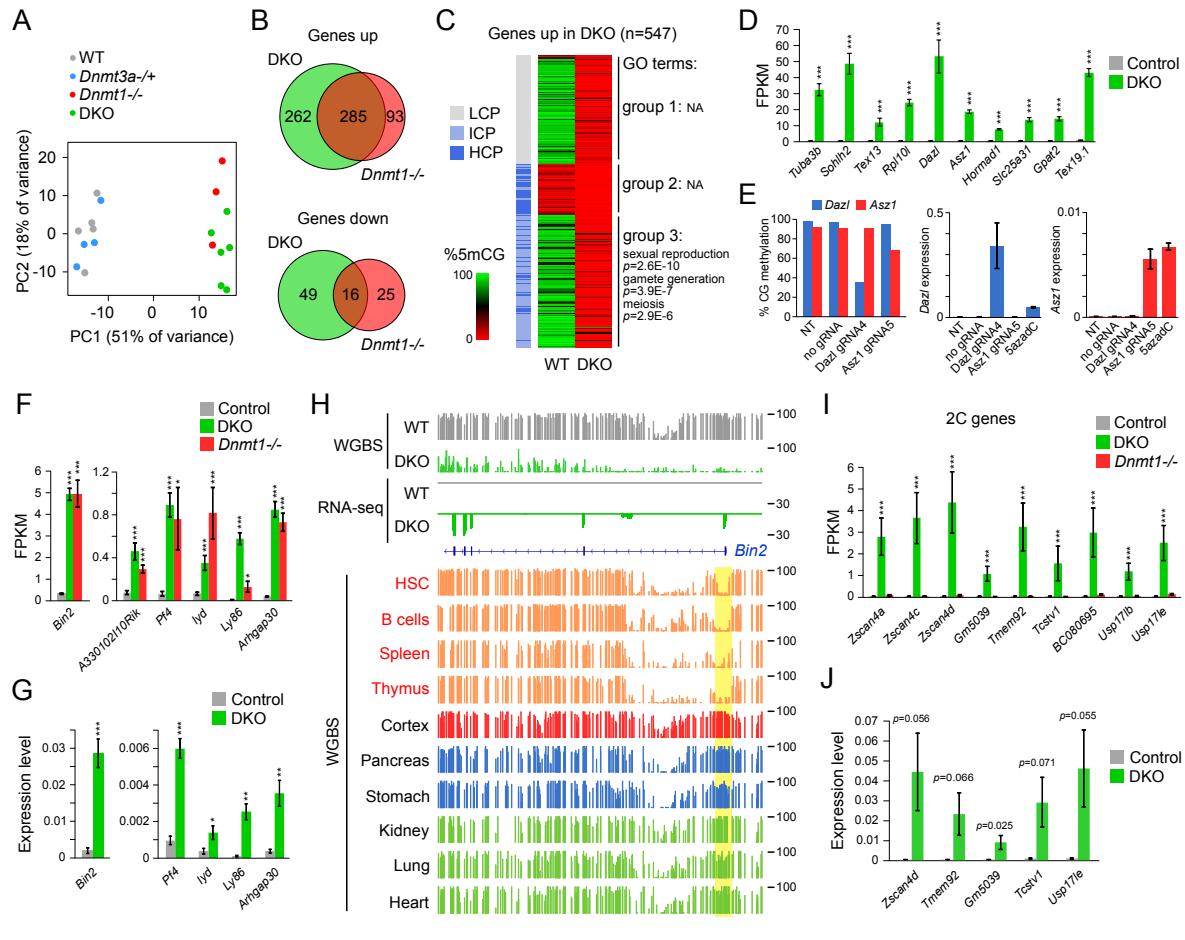


Figure 4. Transcriptome analysis of *Dnmt* mutant embryos. **A.** Principal component analysis (PCA) of RNA-seq data in control, *Dnmt1*^{-/-} and DKO embryos. **B.** Venn diagram comparing the lists of upregulated (top) and downregulated (bottom) genes identified in *Dnmt1*^{-/-} and DKO embryos. **C.** Heatmap representing the three groups of genes upregulated in DKO embryos. The genes are classified based on their promoter class (LCP: low CpG promoter, ICP: intermediate CpG promoter, HCP: high CpG promoter) and their level of promoter CpG methylation in WT embryos (measured in -1000 to +500 bp from the TSS). Group 1: LCP; Group 2: ICP or HCP and promoter methylation < 30%; Group 3: ICP or HCP and promoter methylation \geq 30%. Gene ontology (GO) terms enriched in each group are shown on the right. **D.** RNA-seq expression levels (plotted as FPKM) of selected germline genes in DKO compared to control embryos (mean \pm SEM, n=6). ***: $p < 0.001$ (DESeq adjusted p value). **E.** Targeted promoter demethylation restores expression of *Dazl* and *Asz1* in mouse fibroblasts. The graphs show the percentage of promoter methylation measured by bisulfite sequencing (left) and the expression levels measured by RT-qPCR (right) in non-transfected MEFs (NT) and MEFs expressing dCas9-Suntag-TET1 with no gRNA or a gRNA targeting the *Dazl* or *Asz1* promoter. Cells treated with 0.5 μ M 5-Aza-2'-deoxycytidine (5azadC) for 72h were used as a control for the RT-qPCR. The expression data is presented as the mean \pm SEM (n=3 independent replicates). **F.** RNA-seq expression levels (plotted as FPKM) of somatic lineage-committed genes in DKO and *Dnmt1*^{-/-} compared to control embryos (mean \pm SEM, n=6 for control and DKO, n=3 for *Dnmt1*^{-/-}). *: $p < 0.05$; ***: $p < 0.001$ (DESeq adjusted p value). **G.** RT-qPCR analysis of the expression of five somatic lineage-committed genes in control and DKO embryos (mean \pm SEM, n=4 embryos). The expression was calculated as a ratio relative to the expression of three housekeeping genes (*Mprl32*, *Rpl13a*, *Gusb*). *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$ (two-tailed unpaired t-test). **H.** Genome browser tracks of WGBS methylation profiles at the hematopoietic-specific *Bin2* gene in embryos and adult tissues³⁵⁻³⁷, as well as RNA-seq profiles in one replicate of WT and DKO embryos. The sequence encompassing the *Bin2* promoter (highlighted in yellow) is specifically hypomethylated in hematopoietic cells and tissues (written in red). **I.** RNA-seq expression levels (plotted as FPKM) of 2C-specific genes in DKO and *Dnmt1*^{-/-} compared to control embryos (mean \pm SEM, n=6 for control and DKO, n=3 for *Dnmt1*^{-/-}). ***: $p < 0.001$ (DESeq adjusted p value). **J.** RT-qPCR analysis of the expression of 2C-specific genes in control and DKO embryos (mean \pm SEM, n=5 embryos). The p values are indicated (two-tailed unpaired t-test).

In total, 136 germline genes acquire dense promoter CpG island DNA methylation in WT embryos and are derepressed in DKO embryos (**Annexe 1.4**), with some (e.g. *Tuba3b*, *Sohlh2*, *Tex13*, *Rpl10l*, *Dazl*, *Asz1*, *Hormad1*) reaching up to ~1,000 fold upregulation (**Figure 4D**). This includes numerous germ-cell specific piRNA pathway factors (*PnlDC1*, *Gtsf1*, *Tex19*, *Topaz1*, *Rnf17*, *Piwil2*, *Mov10l1*, *Asz1*, *Ddx4*, *Mael*, *Fkbp6*, *Gpat2*, *Henmt1*). Interestingly, we previously found some germline genes upregulated upon partial hypomethylation in *Dnmt3b*^{-/-} embryos^{7,22}. As expected, all these genes are also upregulated in DKO embryos, nevertheless many additional germline genes are derepressed in DKO embryos (**Supplementary Figure 4I**). Moreover, the degree of reactivation of germline genes is much higher in DKO compared to *Dnmt3b*^{-/-} embryos, which correlates with the degree of methylation loss (**Supplementary Figure 4J**). This supports a direct relationship between CpG island promoter methylation and repression of a large panel of germline genes.

To firmly demonstrate that local CpG island methylation mediates repression of germline genes, we performed dCas9-based targeted demethylation with the TET1 catalytic domain (TET1CD) in embryonic fibroblasts using gRNAs targeting the *Dazl* and *Asz1* promoters (**Supplementary Figure 5A-B**). We first compared the efficiency of dCas9-TET1CD fusion or the dCas9-SunTag-TET1CD system³³ and found that only dCas9-SunTag-TET1CD achieved robust and specific demethylation of *Dazl* and *Asz1* (**Supplementary Figure 5C-E**). Notably, targeted demethylation with the dCas9-SunTag-TET1CD system induced strong derepression of *Dazl* and *Asz1* (**Figure 4E**), demonstrating that dense CpG methylation in the promoters of germline genes plays a causal role in the maintenance of the repressed state. In summary, we reveal an extensive role of DNA methylation in keeping CpG island promoters of the germline program silent in embryos.

Given the strong derepression of germline genes in hypomethylated embryos, we wondered what their expression status is in blastocysts. To address this, we analyzed RNA-seq from E3.5 ICM³⁴ and selected the most derepressed germline genes in DKO embryos (FPKM > 3). Interestingly, while approximately 1/3 of these genes have strongly abundant mRNAs in ICM, the majority has low or undetectable expression in ICM (**Supplementary Figure 6A**). For example, the *Dazl* and *Slc25a31* genes show weak expression in E3.5 ICM despite similar hypomethylation than in E8.5 DKO embryos (**Supplementary Figure 6B**). This suggests that either activators of germline genes are not present in blastocysts or that transient repression mechanisms compensate for erased DNA methylation in preimplantation stages before a switch to strong DNA methylation-dependent repression in postimplantation embryos.

DNA methylation limits early expression of lineage-committed genes

We then investigated what other genes are in the overexpressed groups. Surprisingly, besides expected targets of DNA methylation such as X-linked gene families (*Rhox*, *Xlr*)⁶, we found

that several somatic lineage-committed genes harboring a CpG-dense promoter acquire promoter DNA methylation in WT embryos and are overexpressed in DKO and *Dnmt1*^{-/-} embryos. These genes belong in majority to the group 3 and include genes expressed in the hematopoietic lineage (*Bin2*, *Arhgap30*, *Ly86*, *Pf4*, *Nckap1l*), brain (*A33010210Rik*), eye (*Rbp3*) or digestive tissues (*lyd*, *Gstp2*) (**Figure 4F** and **Annexe 1.4**). The overexpression of these genes was validated by RT-qPCR in DKO embryos (**Figure 4G**). This suggests that DNA methylation of CpG-dense promoters contributes to prevent ectopic expression of some lineage-committed genes. One prediction of this model is that promoter DNA methylation of these genes should be low in the tissues where they are expressed. To test this prediction, we explored public WGBS data from mouse tissues³⁵⁻³⁷ and confirmed that the promoters of these genes are specifically hypomethylated in the tissues where they are expressed (**Figure 4H** and **Supplementary Figure 7**). Taken together, this strongly suggests a role of DNA methylation in suppressing precocious expression of lineage committed genes in embryos.

***De novo* DNA methylation is required to repress 2C-specific genes**

We noticed that many genes specifically expressed in 2-cell embryos and 2C-like ES cells (2C-specific genes) are strongly derepressed (up to ~1,000 fold) in DKO embryos, such as *Zscan4* genes, *Tmem92*, *Tcstv1/3*, *Eif1a*-like genes (*Gm5662*, *Gm2022*, *BB287469*, *Gm2016*, *Gm21319*, *Gm8300*, *Gm5039*) and *Usp17*-like genes (**Figure 4I**), whose differential expression was validated by RT-qPCR (**Figure 4J**). To confirm this finding, we compared the list of genes upregulated in DKO embryos with genes upregulated in 2C-like ES cells³⁸ and found a significant overlap (**Supplementary Figure 8A**). These 2C-specific genes are frequently organized in clusters in heterochromatin regions of chromosomes (**Supplementary Figure 8B**). Intriguingly, most of these 2C-specific genes contain CpG-poor promoters and are much less overexpressed in *Dnmt1*^{-/-} embryos compared to DKO embryos despite similar hypomethylation (**Figure 4I** and **Supplementary Figure 8C**), suggesting an indirect regulation by DNA methylation. Previous studies have established that another feature of 2C embryos is the expression of MERVL retrotransposons and that many 2C-specific genes initiate from promoters in ERVL LTRs³⁹⁻⁴¹. Accordingly, we found a 3 fold activation of MERVL-int transposons in DKO embryos but not in *Dnmt1*^{-/-} embryos (**Supplementary Figure 8D**). In addition, several 2C-specific genes derepressed in DKO embryos (i.e. *Zfp352*, *Tcstv1*, *Tcstv3*, *B020031M17Rik*, *AF067061*, *Gm20767*, *Ubtfl1*, *Gm2022*, *AA792892*) initiate from MERVL LTRs (annotated as MT2_Mm) or other ERVL transposons (MT2B-C, ORR1B) (**Supplementary Figure 8E**). Thus *de novo* methylation is required to extinguish the ERVL-driven and 2C-specific transcriptional network in post-implantation embryos.

DNA methylation represses a high number of TEs and chimeric transcripts

The contribution of DNA methylation to the regulation of TEs in the embryo has not been studied comprehensively, which prompted us to analyze the expression of TEs in *Dnmt* mutant embryos. TE expression was quantified either by counting reads in RepeatMasker annotations or by mapping reads on Repbase sequences (see methods) (**Annexe 1.5**). IAPs showed a dramatic reactivation (50-100 fold) in *Dnmt1*^{-/-} embryos (**Figure 5A**), confirming previous data by northern blot¹⁷. In addition to IAPs, several other retrotransposon families of the LINE-1, ERV1 and ERVK families were significantly upregulated in *Dnmt1*^{-/-} embryos (**Figure 5A** and **Supplementary Figure 9A**). The same set of transposons was upregulated in DKO embryos but with a lower magnitude, which correlates with higher residual methylation of TEs in DKO embryos (**Figure 5A**). The RepBase method yielded similar results (**Supplementary Figure 9B**) and revealed that among the LINE1 elements, only the most recent subfamilies are upregulated (**Supplementary Figure 9C**). Having shown that DNA methylation is required to repress TEs at the family level, we analyzed the expression of individual copies of TEs. This analysis identified 4,593 activated TE copies (fold change > 3, adjusted *p* value < 0.001) in *Dnmt1*^{-/-} embryos (**Supplementary Figure 9D-F** and **Annexe 1.5**). For IAPs, the most active class of retroelements in the mouse, some families showed a massive reactivation of up to 30% annotated copies (**Figure 5B**), in particular IAPeZ-int that represent more than half (2,484 out of 4,593) of all upregulated TEs in *Dnmt1*^{-/-} embryos (**Supplementary Figure 9E**). For other ERVs like ERVB4_1B-I_MM-int, MMERGLN-int, MMEtn-int and MMERVK10C-int, increased expression reflects the activation of a limited set of copies representing no more than 4% of all annotated copies (**Figure 5B**). Interestingly, these activated copies have a higher size and presumably correspond to full length, potentially active copies (**Figure 5C**).

Next, we investigated the impact of the derepression of ERVs on the expression of neighboring genes. We identified 715 genes located close to activated ERVs (<20kb from the TSS) and found that they were significantly more upregulated than control genes in *Dnmt1*^{-/-} embryos (**Figure 5D-E**), indicating that aberrantly activated ERVs alter the expression of proximal genes. Out of the 377 upregulated genes in *Dnmt1*^{-/-} embryos, more than 10% (n= 44 genes) are upregulated in association with derepression of an intragenic or proximal ERV (**Annexe 1.4**). The most frequent case is derepression of an intronic IAP element leading to cryptic intragenic initiation in the host gene, as exemplified by the *Trpm2* and *Capn11* genes (**Figure 5F**). Numerous intergenic ERVs also initiate long RNAs that can extend into adjacent genes and produce chimeric transcripts by splicing from a cryptic splice donor site to an internal genic exon, as exemplified by the *Cyp2b23*, *Serpinb1c* and *Olfir316* genes (**Supplementary Figure 10**).

In summary, we conclude that DNMT1 is the main enzyme involved in TE protection and that DNA methylation is required to repress intact, potentially active copies of many retrotransposon

families. Moreover DNA methylation prevents TEs from initiating long RNAs and disturbing expression of nearby genes in post-implantation embryos.

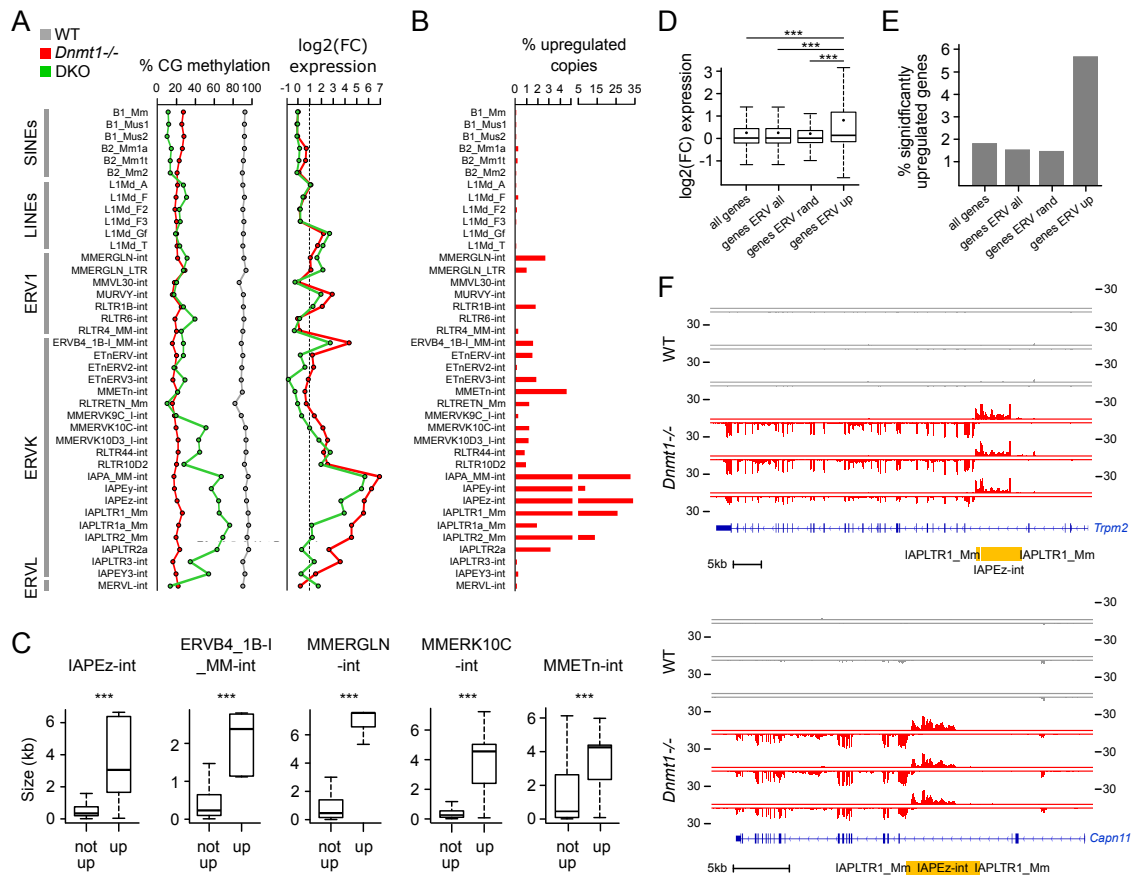


Figure 5. *Dnmt* mutant embryos show widespread transposon upregulation. **A.** DNA methylation measured by WGBS and fold change of expression measured by RNA-seq of SINE, LINE, ERV1, ERVK and EVRL retrotransposon families in *Dnmt1*^{-/-} and DKO embryos. Methylation and expression was calculated on RepeatMasker annotations. **B.** Percentage of significantly upregulated genomic copies (fold change>3, padj<0.001) within each retrotransposon family in *Dnmt1*^{-/-} embryos. **C.** Box plots comparing the size of transposon copies for several ERV families depending on whether they are significantly upregulated in *Dnmt1*^{-/-} embryos (up) or not upregulated (not up). ***: *p*<0.001 (Wilcoxon test). **D.** Box plot representing the log₂ fold change of expression of genes located close to upregulated ERVs (genes ERV up, n=715) compared to all genes, genes located close to all ERVs (genes ERV all) and a random selection of 715 genes located close to ERVs (genes ERV rand). In each boxplot, the horizontal bar represents the median and the dot represents the mean of the distribution. ***: *p*<0.001 (Wilcoxon test). **E.** Barplot representing the percentage of significantly upregulated genes in *Dnmt1*^{-/-} embryos for genes located close to upregulated ERVs and control genes. **F.** RNA-seq tracks in WT and *Dnmt1*^{-/-} embryos illustrating that derepression of intragenic IAPEz elements on the opposite strand leads to internal initiation of the *Trpm2* and *Capn11* genes. For each sample, RNA-seq signals from the top (above the line) and bottom (below the line) strands are shown. IAP elements annotated by RepeatMasker are displayed in yellow below the tracks.

DNA methylation suppresses cryptic initiation sites in gene bodies

Upon further exploring the transcriptional changes in mutant embryos, we noticed that several genes upregulated in DKO embryos initiate from intragenic sequences. For example the *Mgl2*, *Mlana*, *C8b*, *Plekhd1* genes produce truncated mRNAs initiating in the gene body (**Figure 6A** and **Supplementary Figure 11A-C**). To measure if cryptic intragenic transcription initiation is a general phenomenon, we calculated the ratio of expression of downstream exons versus the first exon for all genes. This analysis revealed no significant increase in DKO compared to WT embryos (**Figure 6B**), suggesting that cryptic internal initiation is limited to a subset of genes. Using a bioinformatic pipeline (see Methods), we identified 50 genes in DKO and 29 genes in *Dnmt1*^{-/-} embryos upregulated from intragenic sequences not annotated as transposons or alternative promoters (**Annexe 1.6**). Consistent with a primary role of DNA methylation, there was a large overlap between these genes identified in DKO and *Dnmt1*^{-/-} embryos (**Supplementary Figure 11D**). The sites of cryptic intragenic initiation in these genes tend to be CpG-rich and, in contrast to canonical promoter sequences, are strongly methylated in WT embryos (**Figure 6C**).

To explore the mechanisms of intragenic initiation, we focused on the *Mgl2* gene. Interestingly, the activation of the cryptic *Mgl2* promoter is recapitulated in TKO ES cells (**Figure 6D**), making these cells a good model to investigate *Mgl2* regulation. Analysis of previous datasets generated in TKO ES cells²⁴ revealed that the absence of DNA methylation is associated with the appearance of a new DNase-I hypersensitive site (DHS) and binding of the methylation-sensitive transcription factor NRF1 at the site of intragenic initiation in the intron 6 (**Figure 6D**). The sequence of this intron contains three repetitions of the NRF1 binding motif GAGCATGCGC (**Supplementary Figure 11E**). This suggests that internal binding of NRF1 in absence of DNA methylation creates an intragenic initiation site in the *Mgl2* gene. To validate this hypothesis, we monitored *Mgl2* expression in TKO ES cells knocked down for NRF1 and found that *Mgl2* internal initiation is abolished in these cells (**Figure 6E**). Taken together, this reveals that DNA methylation is critical to prevent methylation-sensitive transcription factors from creating cryptic intragenic initiation sites in embryos.

Discussion

In this manuscript, we interrogated the contribution of DNMTs to the establishment of genome-wide DNA methylation patterns in mouse embryos. Our results provide an important clarification on the function of these enzymes and support a strict division of function between DNMT1 and DNMT3A/B *in vivo*. DNMT3A and DNMT3B are strongly redundant and responsible for all *de novo* methylation in development, confirming what was speculated since the discovery of these enzymes⁴ and excluding a *de novo* function of DNMT1 in embryonic development.

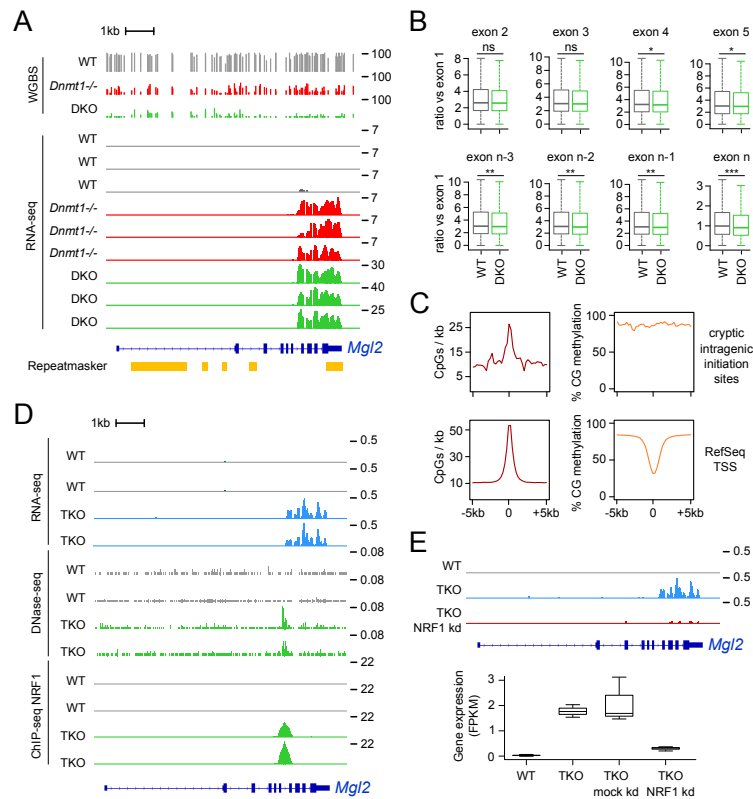


Figure 6. Transcripts initiate from cryptic intragenic promoters in *Dnmt* mutant embryos. **A.** Genome browser tracks of WGBS and RNA-seq profiles at the *Mgl2* locus in WT and *Dnmt* mutant embryos. RepeatMasker annotations are displayed in yellow below the tracks. **B.** Boxplot of the ratio of RNA-seq read counts in downstream exons compared to the first exon in WT and DKO embryos. Exon n represents the last exon of the gene. All genes with at least 5 exons were included in the calculations. ns: not significant; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$ (Wilcoxon test). **C.** Metaplots representing the CpG density and the CpG methylation levels in 5 kb sequences flanking cryptic intragenic initiation sites or canonical RefSeq TSS. **D.** Genome browser tracks of RNA-seq, DNase-seq and NRF1 ChIP-seq profiles at the *Mgl2* locus in WT and *Dnmt* triple knockout (TKO) ES cells. Two replicates are shown. **E.** The genome browser tracks on the top display *Mgl2* RNA-seq profiles upon NRF1 knockdown in TKO ES cells. The boxplot on the bottom shows the quantification of *Mgl2* expression measured by RNA-seq (plotted as FPKM) upon mock and NRF1 knockdown in TKO ES cells ($n = 3$ replicates).

Conversely, DNMT1 alone mediates the faithful maintenance of DNA methylation patterns in developing embryos with no contribution of DNMT3A/B, as supported by several lines of evidence: 1) DNMT1 alone is sufficient to maintain preexisting patterns of DNA methylation from the blastocysts to the E8.5 stage in DKO embryos; 2) All gametic methylation imprints are faithfully maintained in DKO embryos; 3) Global patterns of DNA methylation are unaffected upon conditional inactivation of *Dnmt3a/b* over multiple cell divisions in immortalized embryonic fibroblasts.

The lack of evidence for a role of DNMT3A/B in maintenance methylation is in contradiction with several studies showing that DNMT3A/B are required for the faithful maintenance of CpG methylation in ES cells^{13-16,42}. One possible explanation for this discrepancy is the recent discovery that ES cells continuously cycle in and out of a transient hypomethylated state marked by MERV

expression³⁸. Therefore the reduced methylation of *Dnmt3a/b* knockout ES cells could reflect a requirement for continuous *de novo* methylation to exit the MERVL+ state rather than true maintenance methylation. Our results also contradict a previous study that concluded on a role of DNMT3B in maintenance methylation in MEFs based on rough estimation of DNA methylation with restriction digestion⁴³. In support of our conclusions, combined acute inactivation of DNMT3A and DNMT3B does not lead to global genome hypomethylation in human embryonic carcinoma cells⁴⁴.

Another key finding of our study is the interrogation of the transcriptional roles of DNA methylation by performing for the first time RNA-seq in severely hypomethylated embryos. Indeed, ever since the discovery of DNMTs, the function of DNA methylation in embryogenesis remained unexplored at the genome level. Notably, we establish that DNA methylation of CpG-island promoters is a primary and causal silencing mechanism for restricting ectopic expression of a large panel of germline genes. While we and others previously found a small number of germline genes reactivated in partially hypomethylated embryos and cells^{7,22,45-47}, the profiling of strongly hypomethylated mutants reveals that the set of germline genes repressed by CpG island DNA methylation is much larger than anticipated. In particular this includes many genes involved in the piRNA pathway, reinforcing the model that this might have evolved as a defense mechanism against transposons to couple genome demethylation with immediate activation of piRNA defense genes⁴⁵. Furthermore, our work suggests that compensating mechanisms limit the expression of germline genes when DNA methylation is low in blastocysts. This is reminiscent of IAP transposons, which are repressed by the KAP1/SETDB1 complex in blastocysts before a switch to a permanent DNA methylation dependent repression later in development^{27,48}. Our data support a similar model whereby repressive complexes limit the expression of germline genes in preimplantation stages and subsequently direct *de novo* CpG island DNA methylation at implantation for permanent repression in life. Future studies should be aimed at identifying the factors that repress and methylate germline genes in development, such as the histone methyltransferase G9a⁴⁹. SETDB1 could also be involved because it was shown to repress and stimulate DNA methylation of germline genes in ES cells^{14,23}.

Our work reveals other basic functions of DNA methylation in gene regulation during development. Notably, we found that a small number of somatic lineage-committed genes acquire promoter DNA methylation in WT embryos and are derepressed in methylation-deficient embryos. Furthermore the same genes display tissue-specific hypomethylation over their core promoter in differentiated tissues. This strongly supports an unexpected role for DNA methylation in limiting precocious expression of lineage-committed genes in embryos. Additionally, we demonstrate that intragenic methylation of CpG-rich sequences is essential to mask cryptic promoters in gene bodies and prevent the production of truncated gene transcripts. Using the *Mgl2* gene as a model, we were able to demonstrate that intragenic DNA methylation directly prevents methylation-sensitive

transcription factors such as NRF1²⁴ from initiating cryptic intragenic transcripts. Previously, other epigenetic factors have been shown to limit cryptic intragenic initiation such as KDM5B and SETD2^{50,51}. In addition, a recent study suggested that gene body DNA methylation, stimulated by the binding of DNMT3B to chromatin harboring methylation of lysine 36 on histone H3, suppresses widespread cryptic intragenic initiation in mouse ES cells⁵². In contrast to this report, we found no evidence for widespread increased intragenic transcription in severely hypomethylated embryos by applying the same quantification of RNA-seq signals in downstream versus the first exon of expressed genes. Instead, our results suggest that DNA methylation limits cryptic intragenic initiation from defined sequences in a small number of genes.

Another surprising finding is that 2C genes are derepressed in DKO embryos. This was unexpected because, according to previous work, DNA hypomethylation does not drive expression of 2C genes in ES cells³⁸, which are instead repressed by other epigenetic factors such as CAF1, KDM1a, KAP1, G9a, HP1 and PRC1 in ES cells^{39,53-56}. Adding to the complexity, 2C genes are not strongly activated in *Dnmt1*^{-/-} embryos, suggesting an indirect regulation by DNA methylation. Recently it was found that *Dux*, *Dppa2* and *Dppa4* activate the 2C program^{57,58}. Interestingly *Dux*, *Dppa2* and *Dppa4* are derepressed in DKO embryos (**Annexe 1.4**), which could provide a potential explanation for the coordinated derepression of 2C genes.

The prevailing model for transposon regulation is that they switch from H3K9me3-mediated silencing in preimplantation embryonic cells to a DNA methylation dominant mechanism in postimplantation embryos. However, the latter aspect of this model lacked experimental evidence in the mouse because, besides IAPs^{17,29}, it was unclear if other TE families require DNA methylation for repression in differentiated embryonic cells. Our analysis demonstrates that DNA methylation is universally required to maintain repression of potentially active copies of numerous ERV and LINE transposons in postimplantation embryos, confirming that, in striking contrast to early embryonic cells, DNA methylation becomes a major epigenetic barrier against transposon expression in differentiated cells. Interestingly, SETDB1 is still required to repress some ERV1 transposons (predominantly MMVL30-int, RLTR6_Mm and MULV-int) in mouse differentiated cells⁵⁹, which we find do not depend on DNA methylation.

In summary, our work clarifies the role of DNMTs in embryogenesis and provides a detailed description of the multiple functions of DNA methylation in maintaining the transcription integrity of mouse embryos. These results have important implications for our understanding of why DNA methylation is essential for mammalian development.

Material and methods

Mouse embryos. We created a *Dnmt1* null allele by crossing *Dnmt1*-2lox mice³⁰ with an ACTB-Cre deleter line⁶⁰, which creates a *Dnmt1* allele lacking the exons 4 and 5. *Dnmt1*^{-/-} embryos were obtained by natural mating of heterozygous males and females. *Dnmt3a* and *Dnmt3b* knockout alleles were created by deleting critical catalytic exons as previously described⁷. We obtained *Dnmt3a*^{-/-} *Dnmt3b*^{-/-} (DKO) embryos by natural mating of *Dnmt3a*^{+/-} *Dnmt3b*^{+/-} males and females. As controls, we recovered WT, *Dnmt3a*^{+/-} and *Dnmt3a*^{+/-} *Dnmt3b*^{+/-} embryos from the same litters. All mouse lines were maintained on a C57BL/6J genetic background. The morning of the vaginal plug was designated E0.5 and embryos were manually dissected in M2 medium at E8.5. We simultaneously prepared genomic DNA and total RNA from the same embryos with the AllPrep DNA/RNA Mini Kit (Qiagen).

Isolation and culture of fibroblasts. Mouse embryonic fibroblasts (MEFs) were isolated from E13.5 embryos and immortalized by serial passages. MEFs were grown in DMEM supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin in a humidified atmosphere containing 5% CO₂ at 37°C. All cells were tested negative for mycoplasmas.

Conditional inactivation of *Dnmt3a* and *Dnmt3b* in MEFs. We crossed *Dnmt3a*-2lox and *Dnmt3b*-2lox mouse lines^{43,61} to generate double floxed mice on a C57BL/6J genetic background. MEFs were isolated from a *Dnmt3a*^{2lox/2lox} *Dnmt3b*^{2lox/2lox} embryo and immortalized by serial passages. To conditionally inactivate *Dnmt3a* and *Dnmt3b*, the MEFs were transduced with a retrovirus coding for the Tamoxifen-inducible Cre-ERT2 recombinase and selected with puromycin (2µg/ml). Recombination was induced by treating MEFs with 2µM 4-OH-Tamoxifen (Sigma) for up to 69 days. The medium containing 4-OH-Tamoxifen was renewed every day during the first two weeks and then every 3 days. The efficiency of the recombination was validated by PCR genotyping and RT-qPCR. The conditional inactivation was performed 3 times independently and the cells were harvested after 23 and 69 days of culture for DNA methylation analysis by RRBS. The oligo sequences for PCR are provided in the **Annexe 1.7**.

Epigenetic editing with dCas9-TET1 fusion. The plasmids coding for the dCas9-TET1 fusion were constructed based on the pdCas9-DNMT3A-EGFP plasmid (Addgene #71666). The EGFP sequence was substituted with the puromycin sequence from the PX459-V2 plasmid (Addgene #62988). Subsequently, the plasmid was digested with BamHI and FseI to replace the DNMT3A fragment with the catalytic domain of human TET1 (hTET1-CD). The sequence coding for hTET1-CD was synthesized

by Integrated DNA Technologies (IDT) and amplified by PCR using forward and reverse primers introducing BamHI and FseI restriction sites. Two BbsI restriction sites within the TET1 sequence were removed by introducing silent mutations by site-directed mutagenesis. The gRNAs targeting *Dazl* (gRNA4: ACGCACTCCGTGGGCGACGT) and *Asz1* (gRNA5: GTGAAAGGCCAGCTCGTGGG) were designed using <http://crispr.mit.edu>, synthesized as pairs of oligonucleotides, annealed and cloned into the BbsI site. Immortalized MEFs isolated from a C57BL6/J embryo were transfected with the dCas9-hTET1CD plasmids (wt and mut) using Polyethylenimine (PEI) transfection reagent. In brief, 10 µg of plasmid and 20 µL of PEI were diluted in 250 µL of 150mM NaCl each, combined and incubated for 30 min at RT. The complexes were added to 70-80% confluent MEF cells in 100-mm dishes. 24 hours after transfection, the cells were selected with 3 µg/µL of puromycin (Gibco, Thermo Fisher Scientific) for 48h before harvesting the cells for DNA/RNA extraction using the AllPrep DNA/RNA Mini Kit (Qiagen).

Epigenetic editing with the dCas9-Suntag-TET1 system. The gRNAs were cloned into the pPlatTet-gRNA2 all-in-one vector³³. pPlatTET-gRNA2 was a gift from Izuhiko Hatada (Addgene # 82559). Briefly, two 60mer oligonucleotides containing the gRNA sequence were annealed and extended to make a 100bp double stranded DNA fragment using Q5 high fidelity polymerase (New England Biolabs #M0491S), and the PCR products were incorporated into the linearized pPlatTet-gRNA2 vector by Gibson assembly (New England Biolabs #E2611S). MEFs were transfected with the Neon electroporation system (Thermo Fisher Scientific) and cells expressing GFP were selected 72 hours post-transfection by flow cytometry using a BD FACS Vantage cell Sorter (BD Biosciences) for DNA/RNA extraction using the AllPrep DNA/RNA Mini Kit (Qiagen).

Gene expression analysis by RT-qPCR. RNAs were reverse transcribed into cDNA with the Maxima first strand cDNA synthesis kit (Thermo Fisher Scientific). qPCR was performed with the KAPA SYBR FAST qPCR kit (KAPA Biosystems) on a StepOnePlus realtime PCR system (Applied Biosystems) using the standard curve method. 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. The expression of target genes was normalized with three housekeeping genes (*Gusb*, *Rpl13a*, *B2m* or *Mrp132*). qPCR reactions were performed in triplicates and we systematically amplified no-RT controls to rule out the presence of contaminating DNA. The oligo sequences for qPCR are provided in the **Annexe 1.7**.

Bisulfite sequencing. 100 ng of genomic DNA were bisulfite converted using the EpiTect Bisulfite Kit (Qiagen) according to the manufacturer's instructions. The target regions were amplified by PCR with the Platinum Taq DNA Polymerase (Thermo Fisher Scientific) using the following conditions: 20 cycles

of 30s at 95°C, 30s at 58–48°C (with a 0.5°C decrease per cycle), 50s at 72°C followed by 35 cycles of 30s at 95°C, 30s at 52°C, 50s at 72°C. The PCR products were cloned by TA cloning in the pCR2.1 vector (TA Cloning Kit, Invitrogen) and 15-30 clones were sequenced. Sequences were aligned with the BISMA software and filtered to remove clonal biases. The oligo sequences are provided in the **Annexe 1.7**.

RRBS. We prepared RRBS libraries from single embryos as previously described ⁷. The libraries were sequenced in paired-end (2x75bp) on an Illumina HiSeq4000 at Integragen SA (Evry, France). Reads were trimmed to remove low quality bases with Trim Galore v0.4.2 and aligned to the mm10 genome with BSMAP v2.74 (parameters -v 2 -w 100 -r 1 -x 400 -m 30 -D C-CGG -n 1). We calculated methylation scores using methratio.py in BSMAP v2.74 (parameters -z -u -g). Only CpGs covered by a minimum of 8 reads were retained for analyses.

WGBS. 50 ng of genomic DNA were fragmented to 350 bp using a Covaris E220 sonicator. DNA was bisulfite-converted with the EZ DNA Methylation-Gold kit (Zymo Research) and WGBS libraries were prepared using the Accel-NGS Methyl-Seq DNA Library Kit (Swift Biosciences) according to the manufacturer's instructions with 7 PCR cycles for the final amplification. The libraries were purified using Ampure XP beads (Beckman Coulter) and sequenced in paired-end (2x100 bp) on an Illumina HiSeq4000 at Integragen SA (Evry, France). Low quality bases as well as the first five bases of reads R1 and ten bases of reads R2 were trimmed with Trim Galore v0.4.2 (parameters -q 20 --clip_R1 5 --clip_R2 10). We aligned reads to the mm10 genome, removed read duplicates and extracted methylation calls using Bismark v0.18.2 with default parameters. Only CpGs covered by a minimum of 5 reads were retained for analyses.

Methylation data analysis. The methylation of genomic features was calculated by intersecting CpG positions with genomic annotations using the IRanges package in R, and averaging methylation of individual CpGs in each feature. We used the CpG island annotation, RefSeq gene annotation and RepeatMasker annotation (only elements > 400 bp) downloaded from the UCSC website. Promoters were defined as -1 kb to +1 kb around RefSeq TSS. Intergenic regions were defined as genomic regions that do not overlap with any of the previous annotations. To identify regions with high residual methylation in ICM and DKO embryos, methylation was averaged in 1kb windows containing at least 3 CpGs. To select regions *de novo* methylated during implantation (Figure 2C), we selected 1 kb windows with <5 % CpG methylation in ICM and >50 % CpG methylation in E8.5 WT embryos. Metaplots of CpG methylation in genes were generated by calculating methylation in twenty equal-sized windows within each RefSeq gene (excluding the X and Y chromosomes) and ten 1 kb windows

of flanking sequences. Genomic coordinates of germline imprinted DMRs were retrieved from the WAMIDEX database⁶². Pairwise correlation plots of methylation scores were generated in 1 kb windows for WGBS and 500 bp windows for RRBS. Promoter classification based on CpG density was done as follows. For each promoter (-1 kb to +1 kb around RefSeq TSS), we calculated the CpG ratio and GC content in 500 bp sliding windows with 20 bp increments. LCP were defined as containing no window with a CpG ratio > 0.45, HCP were defined as containing at least one window with a CpG ratio > 0.65 and a GC content > 55%, and the remaining promoters were defined as ICP. To identify DMRs in *Dnmt3a/b* conditional knockout MEFs, we used eDMR from the methylKit R package with the following criteria: at least 3 differentially methylated CpGs (DMCs), difference in methylation > 20%, adjusted *p*-value < 0.001.

RNA-seq and transcriptome analysis. RNA-seq libraries were prepared from single embryos with the TruSeq Stranded Total RNA Sample Prep Kit with Ribo-Zero ribosomal RNA reduction (Illumina). We prepared libraries from three *Dnmt1*^{-/-} and three WT littermate embryos, as well as six DKO, two WT and four *Dnmt3a*^{+/-} littermate embryos. The libraries were sequenced in paired-end (2x100bp) on an Illumina HiSeq 4000. Reads were mapped to the mm10 genome using TopHat v2.0.13 with a RefSeq transcriptome index and default parameters, reporting up to 20 alignments for multi-mapped reads. For gene expression analysis, unique reads were counted in RefSeq genes with HTSeq v0.7.2 (parameters -t exon -s reverse), and differentially expressed genes (fold change > 3, adjusted *p*-value < 0.001) were identified using DESeq2 v1.18.1. For data visualization, bigwig files were generated using bam2wig.py from the RSeQC package v2.6.4 (parameters -u -t 5000000000) and visualized in the Integrative Genomics Viewer (IGV) browser. Splice junctions were visualized in IGV using the splice junctions files produced by TopHat. The FPKM values and PCA analysis were generated using DESeq2. Gene ontology analysis of differentially expressed genes was performed using the DAVID tool (<https://david.ncifcrf.gov>).

Intragenic transcription. To study intragenic transcription initiation, reads were counted in individual exons of all RefSeq gene isoforms using featureCounts from the Rsubread package v1.30.9, and the ratio of FPKM values of downstream exons over the first exon was plotted for all genes with at least 5 exons and a FPKM score > 1. To identify genes with intragenic initiation, differentially expressed exons (fold change > 3, *p* value < 0.001) were identified using DESeq2 v1.20.0 and the percentage of upregulated exon was calculated for each RefSeq isoform. The following criteria were then applied: (i) percentage of upregulated exons < 100 %, (ii) the first exon is not upregulated, (iii) no other isoform of the same gene with 100% upregulated exons, (iv) if < 5 upregulated exons they should be consecutive, (v) if ≥ 5 upregulated exons a gap of one exon is tolerated, (vi) the fold change of

upregulated exons is > 3 times higher than for the other exons, (vii) the FPKM values of upregulated exons is higher than for the other exons. Finally, the list of genes was manually curated to keep only one RefSeq isoform for each gene, eliminate false positives and eliminate genes initiating from transposons. For each gene, the position of intragenic initiation was defined manually from the bigwig files. Metaplots of CpG density and CpG methylation around cryptic intragenic initiation sites and RefSeq TSS were calculated in 250 bp windows.

Transposon analysis. Methylation of TE families was estimated by intersecting CpG positions with the UCSC RepeatMasker annotation using the IRanges package in R, and averaging the CpG methylation scores in each family. Expression of TEs was analyzed in several ways at the family and copy level. First, unique and multiple-mapping reads were counted in TE families using featureCounts from the Rsubread package v1.30.9 with a GTF file built from the UCSC RepeatMasker annotation, with the option to weight multi-mapping reads by the number of mapping sites (parameters countMultiMappingReads= TRUE, fraction= TRUE, useMetaFeatures= TRUE). In parallel, expression of TE families was also analyzed by mapping reads to RepBase consensus sequences using TopHat v2.0.13 allowing 5 mismatches, and counting reads with HTSeq v0.7.2. Differentially expressed TE families (fold change > 2, adjusted *p* value < 0.001) were identified using DESeq2 v1.20.0. To analyze the expression of individual copies of TEs, only unique reads were counted in individual TEs from the RepeatMasker GTF file using featureCounts (parameters countMultiMappingReads=FALSE, useMetaFeatures=FALSE). We then identified differentially expressed copies (fold change > 3, adjusted *p* value < 0.001) using DESeq2 v1.20.0, and calculated the percentage of upregulated copies within each TE family. The estimation of the age of LINE-1 families was taken from Sookdeo et al.⁶³.

Datasets. The following datasets were used: WGBS in gametes and early embryos (GSE56697), WGBS in mouse adult tissues (GSE42836), WGBS in B cells (GSE100262), WGBS in hematopoietic stem cells (GSE52709), RNA-seq in E3.5 ICM (GSE84236), RNA-seq, DNase-seq and NRF1 ChIP-seq in TKO ES cells (GSE67867).

Statistical analyses. All measurements in embryos were biological replicates taken from individual embryos. Details on the statistical tests and samples sizes are provided in the figure legends.

Data availability

The methylome and RNA-seq sequencing data generated in this study are deposited in the NCBI Gene Expression Omnibus (GEO) under the accession number GSE130735.

Acknowledgments

We thank the staff of the IGBMC GenomEast sequencing platform, and Annie Varrault for advices with the dCas9-Suntag-TET1 experiments. This work was funded by the European Research Council (ERC Consolidator grant no. 615371) and the Institut National du Cancer (INCa). Thomas Dahlet was recipient of a Doctoral fellowship from the French Ministry for Higher Education and Research. Andrea Argüeso Lleida was supported by the Fondation pour la Recherche Médicale (FRM).

Author contributions

T.D. performed embryo dissection and RRBS, WGBS, RNA-seq and gene expression analysis. A.A.L. and H.A.A. performed dCas9-based methylation editing experiments. M.D. and A.F.B. conducted bioinformatic analysis of the sequencing data. A.B. performed embryo dissection and RRBS experiments. J.V. and R.P.N. generated MEFs for conditional inactivation experiments and contributed to RRBS analysis in cDKO MEFs. M.T. and G.A. participated in the generation of mouse lines and embryo dissection. M.W. conceived and supervised the project and performed data analysis. M.W. wrote the manuscript with contributions from T.D. and the other authors.

Declaration of Interests

The authors declare no competing interests.

References

- 1 Smith, Z. D. *et al.* DNA methylation: roles in mammalian development. *Nature reviews. Genetics* **14**, 204-220 (2013).
- 2 Bourc'his, D. *et al.* Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**, 96-99 (2004).
- 3 Barau, J. *et al.* The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science* **354**, 909-912 (2016).
- 4 Okano, M. *et al.* DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257 (1999).

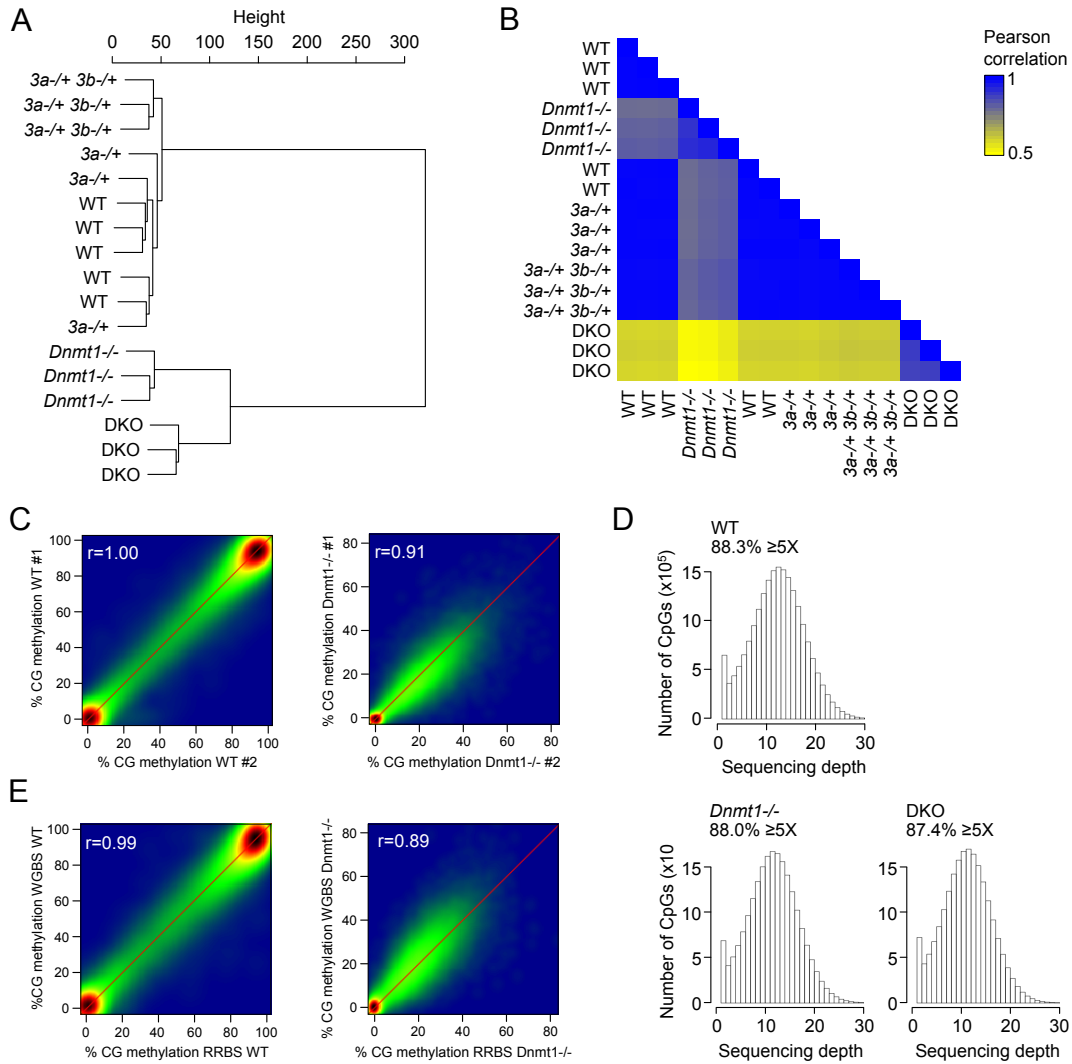
- 5 Li, E. *et al.* Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-926 (1992).
- 6 Oda, M. *et al.* DNA methylation regulates long-range gene silencing of an X-linked homeobox gene cluster in a lineage-specific manner. *Genes & development* **20**, 3382-3394 (2006).
- 7 Auclair, G. *et al.* Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. *Genome biology* **15**, 545 (2014).
- 8 Fatemi, M. *et al.* The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. *Journal of molecular biology* **309**, 1189-1199 (2001).
- 9 Gowher, H. *et al.* De novo methylation of nucleosomal DNA by the mammalian Dnmt1 and Dnmt3A DNA methyltransferases. *Biochemistry* **44**, 9899-9904 (2005).
- 10 Arand, J. *et al.* In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS genetics* **8**, e1002750 (2012).
- 11 Shirane, K. *et al.* Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. *PLoS genetics* **9**, e1003439 (2013).
- 12 Li, Y. *et al.* Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. *Nature* **564**, 136-140 (2018).
- 13 Li, Z. *et al.* Distinct roles of DNMT1-dependent and DNMT1-independent methylation patterns in the genome of mouse embryonic stem cells. *Genome biology* **16**, 115 (2015).
- 14 Leung, D. *et al.* Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 6690-6695 (2014).
- 15 Chen, T. *et al.* Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Molecular and cellular biology* **23**, 5594-5605 (2003).

- 16 Thakur, A. *et al.* Widespread recovery of methylation at gametic imprints in hypomethylated mouse stem cells following rescue with DNMT3A2. *Epigenetics & chromatin* **9**, 53 (2016).
- 17 Walsh, C. P. *et al.* Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature genetics* **20**, 116-117 (1998).
- 18 Hirasawa, R. *et al.* Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes & development* **22**, 1607-1616 (2008).
- 19 Lei, H. *et al.* De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* **122**, 3195-3205 (1996).
- 20 Li, E. *et al.* Role for DNA methylation in genomic imprinting. *Nature* **366**, 362-365 (1993).
- 21 Weaver, J. R. *et al.* Domain-specific response of imprinted genes to reduced DNMT1. *Molecular and cellular biology* **30**, 3916-3928 (2010).
- 22 Borgel, J. *et al.* Targets and dynamics of promoter DNA methylation during early mouse development. *Nature genetics* **42**, 1093-1100 (2010).
- 23 Karimi, M. M. *et al.* DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell stem cell* **8**, 676-687 (2011).
- 24 Domcke, S. *et al.* Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature* **528**, 575-579 (2015).
- 25 Walter, M. *et al.* An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *eLife* **5** (2016).
- 26 Matsui, T. *et al.* Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464**, 927-931 (2010).
- 27 Rowe, H. M. *et al.* KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237-240 (2010).
- 28 Robbez-Masson, L. *et al.* The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. *Genome research* **28**, 836-845 (2018).

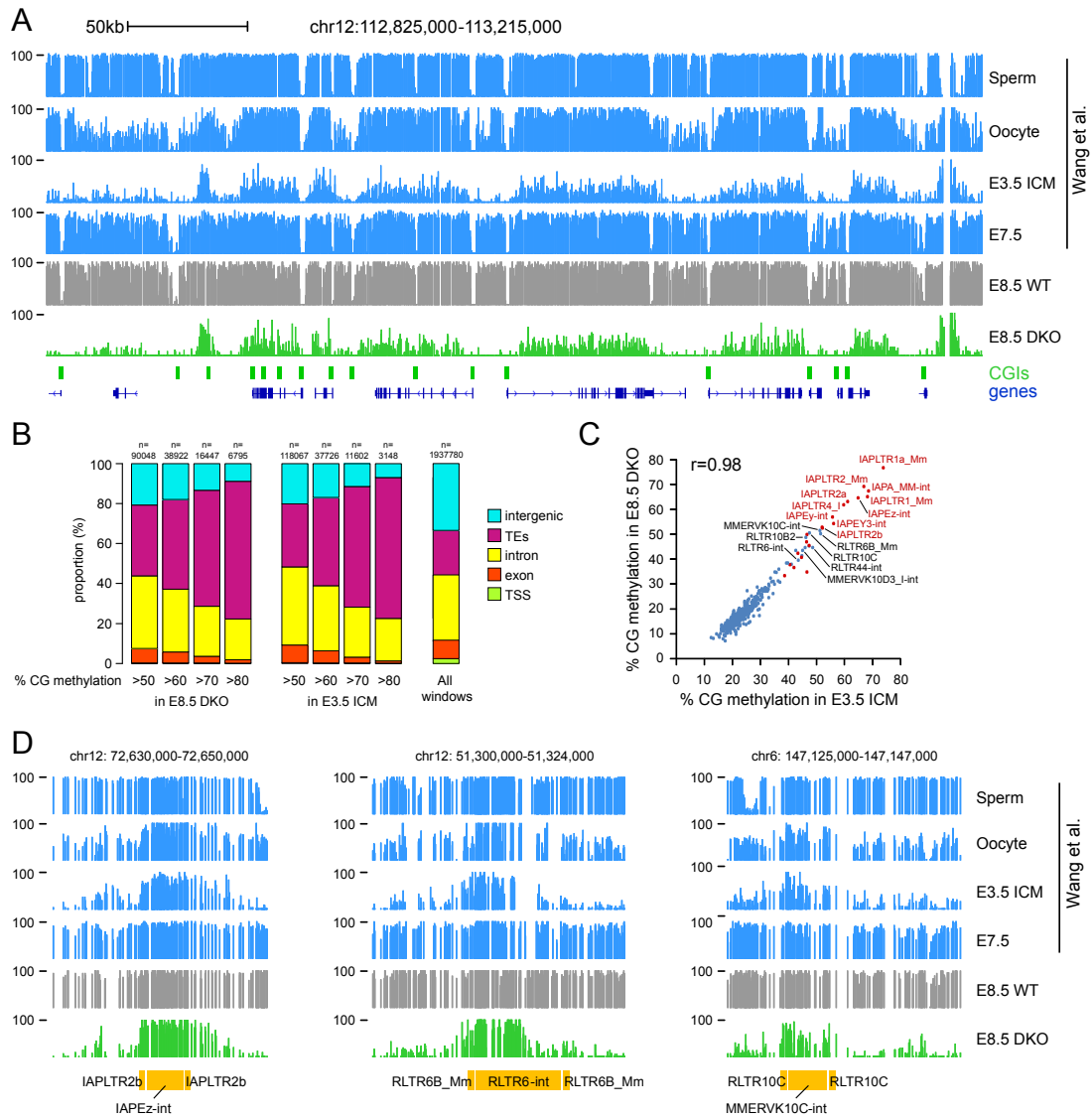
- 29 Hutnick, L. K. *et al.* Repression of retrotransposal elements in mouse embryonic stem cells is primarily mediated by a DNA methylation-independent mechanism. *The Journal of biological chemistry* **285**, 21082-21091 (2010).
- 30 Jackson-Grusby, L. *et al.* Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nature genetics* **27**, 31-39 (2001).
- 31 Wang, L. *et al.* Programming and inheritance of parental DNA methylomes in mammals. *Cell* **157**, 979-991 (2014).
- 32 Duffie, R. *et al.* The Gpr1/Zdbf2 locus provides new paradigms for transient and dynamic genomic imprinting in mammals. *Genes & development* **28**, 463-478 (2014).
- 33 Morita, S. *et al.* Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nature biotechnology* **34**, 1060-1065 (2016).
- 34 Smith, Z. D. *et al.* Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. *Nature* **549**, 543-547 (2017).
- 35 Hon, G. C. *et al.* Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nature genetics* **45**, 1198-1206 (2013).
- 36 Cabezas-Wallscheid, N. *et al.* Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. *Cell stem cell* **15**, 507-522 (2014).
- 37 Duncan, C. G. *et al.* Base-Resolution Analysis of DNA Methylation Patterns Downstream of Dnmt3a in Mouse Naive B Cells. *G3 (Bethesda)* **8**, 805-813 (2018).
- 38 Eckersley-Maslin, M. A. *et al.* MERVL/Zscan4 Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs. *Cell reports* **17**, 179-192 (2016).
- 39 Macfarlan, T. S. *et al.* Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes & development* **25**, 594-607 (2011).
- 40 Macfarlan, T. S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57-63 (2012).

- 41 Peaston, A. E. *et al.* Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Developmental cell* **7**, 597-606 (2004).
- 42 Liang, G. *et al.* Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. *Molecular and cellular biology* **22**, 480-491 (2002).
- 43 Dodge, J. E. *et al.* Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. *The Journal of biological chemistry* **280**, 17986-17991 (2005).
- 44 Tiedemann, R. L. *et al.* Acute depletion redefines the division of labor among DNA methyltransferases in methylating the human genome. *Cell reports* **9**, 1554-1566 (2014).
- 45 Hackett, J. A. *et al.* Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development* **139**, 3623-3632 (2012).
- 46 Velasco, G. *et al.* Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc Natl Acad Sci* **107**, 9281-9286 (2010).
- 47 Maatouk, D. M. *et al.* DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. *Development* **133**, 3411-3418 (2006).
- 48 Rowe, H. M. *et al.* De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. *Development* **140**, 519-529 (2013).
- 49 Auclair, G. *et al.* EHMT2 directs DNA methylation for efficient gene silencing in mouse embryos. *Genome research* **26**, 192-202 (2016).
- 50 Carvalho, S. *et al.* Histone methyltransferase SETD2 coordinates FACT recruitment with nucleosome dynamics during transcription. *Nucleic acids research* **41**, 2881-2893 (2013).
- 51 Xie, L. *et al.* KDM5B regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription. *The EMBO journal* **30**, 1473-1484 (2011).
- 52 Neri, F. *et al.* Intragenic DNA methylation prevents spurious transcription initiation. *Nature* **543**, 72-77 (2017).

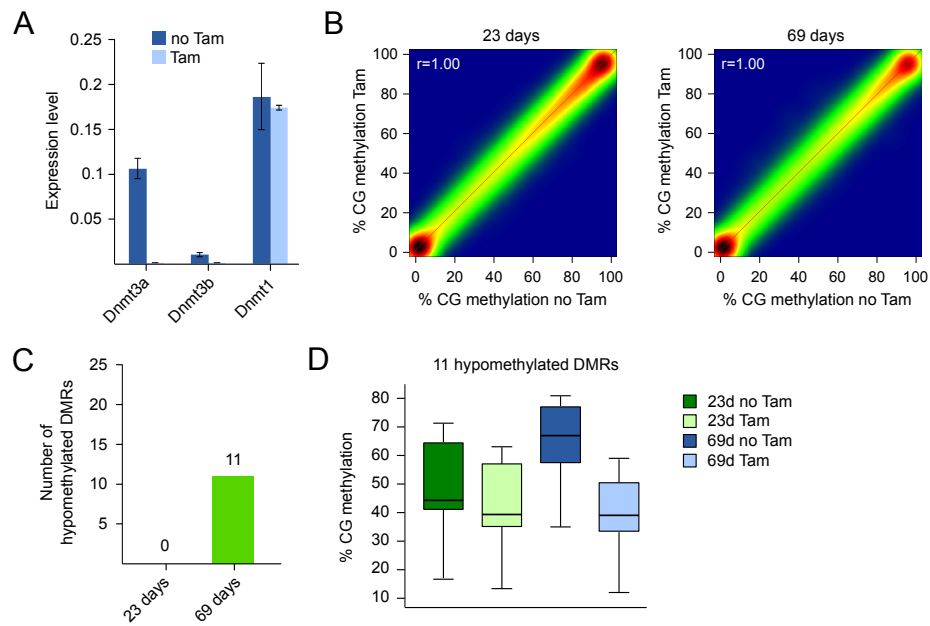
- 53 Ishiuchi, T. *et al.* Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nature structural & molecular biology* **22**, 662-671 (2015).
- 54 Hisada, K. *et al.* RYBP represses endogenous retroviruses and preimplantation- and germ line-specific genes in mouse embryonic stem cells. *Molecular and cellular biology* **32**, 1139-1149 (2012).
- 55 Maksakova, I. A. *et al.* Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. *Epigenetics & chromatin* **6**, 15 (2013).
- 56 Rodriguez-Terrones, D. *et al.* A molecular roadmap for the emergence of early-embryonic-like cells in culture. *Nature genetics* **50**, 106-119 (2018).
- 57 De Iaco, A. *et al.* DPPA2 and DPPA4 are necessary to establish a 2C-like state in mouse embryonic stem cells. *EMBO reports* (2019).
- 58 Eckersley-Maslin, M. *et al.* Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. *Genes & development* **33**, 194-208 (2019).
- 59 Kato, M. *et al.* A somatic role for the histone methyltransferase Setdb1 in endogenous retrovirus silencing. *Nature communications* **9**, 1683 (2018).
- 60 Birling, M. C. *et al.* Highly-efficient, fluorescent, locus directed cre and FlpO deleter mice on a pure C57BL/6N genetic background. *Genesis* **50**, 482-489 (2012).
- 61 Kaneda, M. *et al.* Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429**, 900-903 (2004).
- 62 Schulz, R. *et al.* WAMIDEX: a web atlas of murine genomic imprinting and differential expression. *Epigenetics : official journal of the DNA Methylation Society* **3**, 89-96 (2008).
- 63 Sookdeo, A. *et al.* Revisiting the evolution of mouse LINE-1 in the genomic era. *Mobile DNA* **4**, 3 (2013).



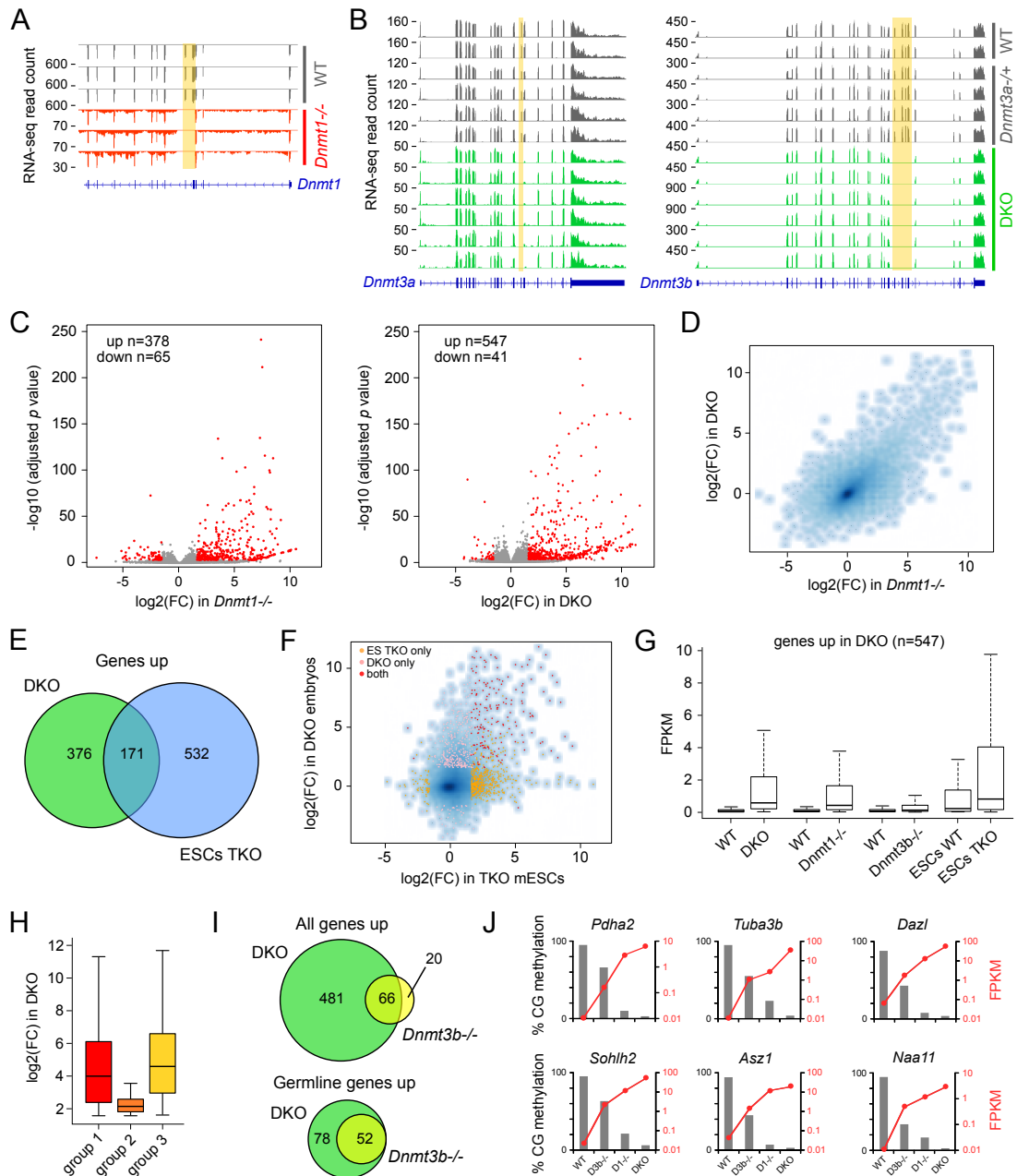
Supplementary Figure 1. Methylation profiling of *Dnmt* mutant embryos by RRBS and WGBS. **A.** Dendrogram of the hierarchical clustering of CpG methylation scores in all RRBS samples. **B.** Matrix of Pearson correlation coefficients of CpG methylation scores between all RRBS samples. **C.** Examples of density scatter plots showing the correlation of methylation scores in 500bp windows between RRBS replicates. The density of points increases from blue to dark red. **D.** Histograms of the sequencing depth at CpGs in each WGBS sample. The percentage of genomic CpGs sequenced at least 5X is indicated for each sample above the graph. **E.** Correlation of methylation scores in 500bp windows between RRBS and WGBS.



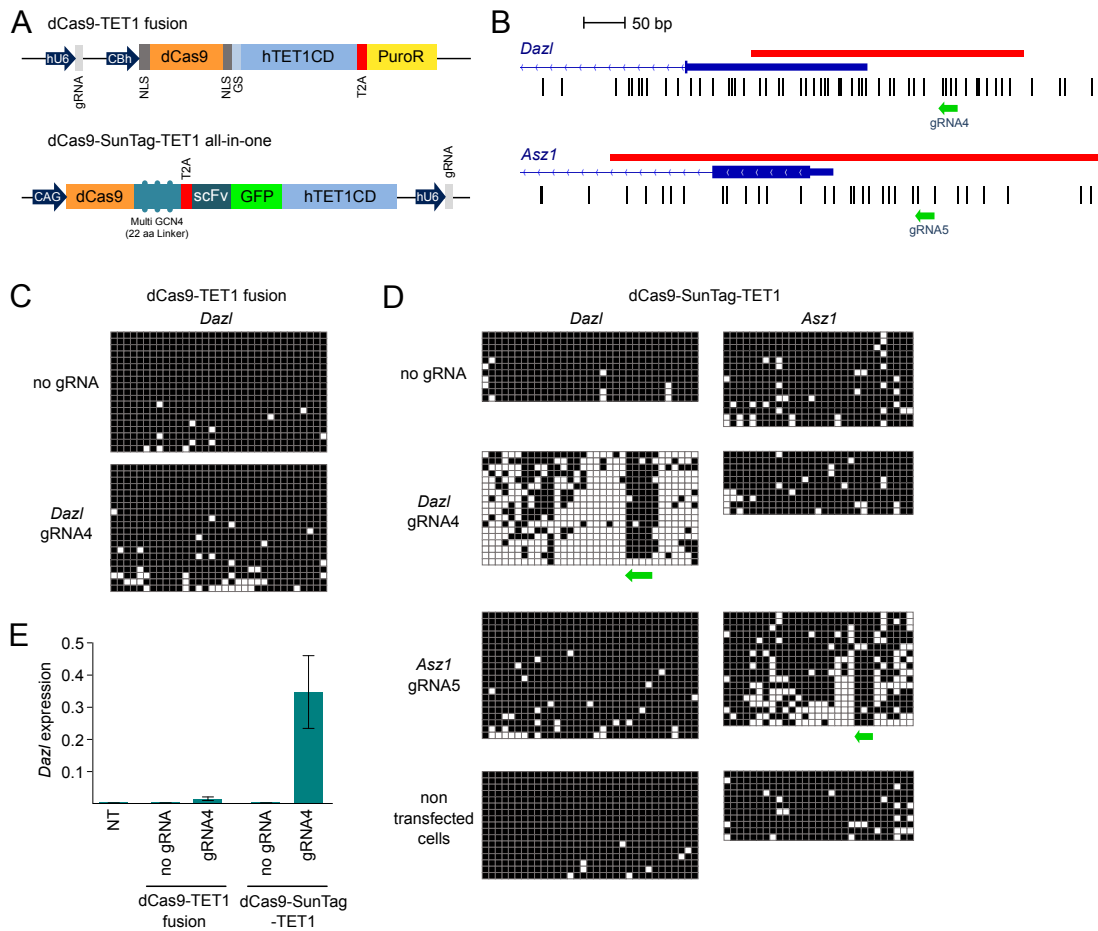
Supplementary Figure 2. Comparison of methylomes in DKO embryos and blastocysts. A. Example of genome browser view of WGBS methylation patterns in gametes, early embryos and E8.5 DKO embryos. CpG islands (CGIs) depicted by green rectangles and RefSeq gene annotations are shown below the tracks. **B.** Genomic features associated with highly methylated sequences (1kb windows) in E8.5 DKO embryos and ICM, which illustrates that the most methylated sequences are enriched for transposable elements (TEs). **C.** Comparison of methylation levels of retrotransposon families in E8.5 DKO embryos and ICM. The red dots represent IAP families. **D.** Examples of WGBS methylation profiles over transposable elements in early embryos and DKO embryos. ERVs annotated by RepeatMasker are displayed in yellow below the tracks.



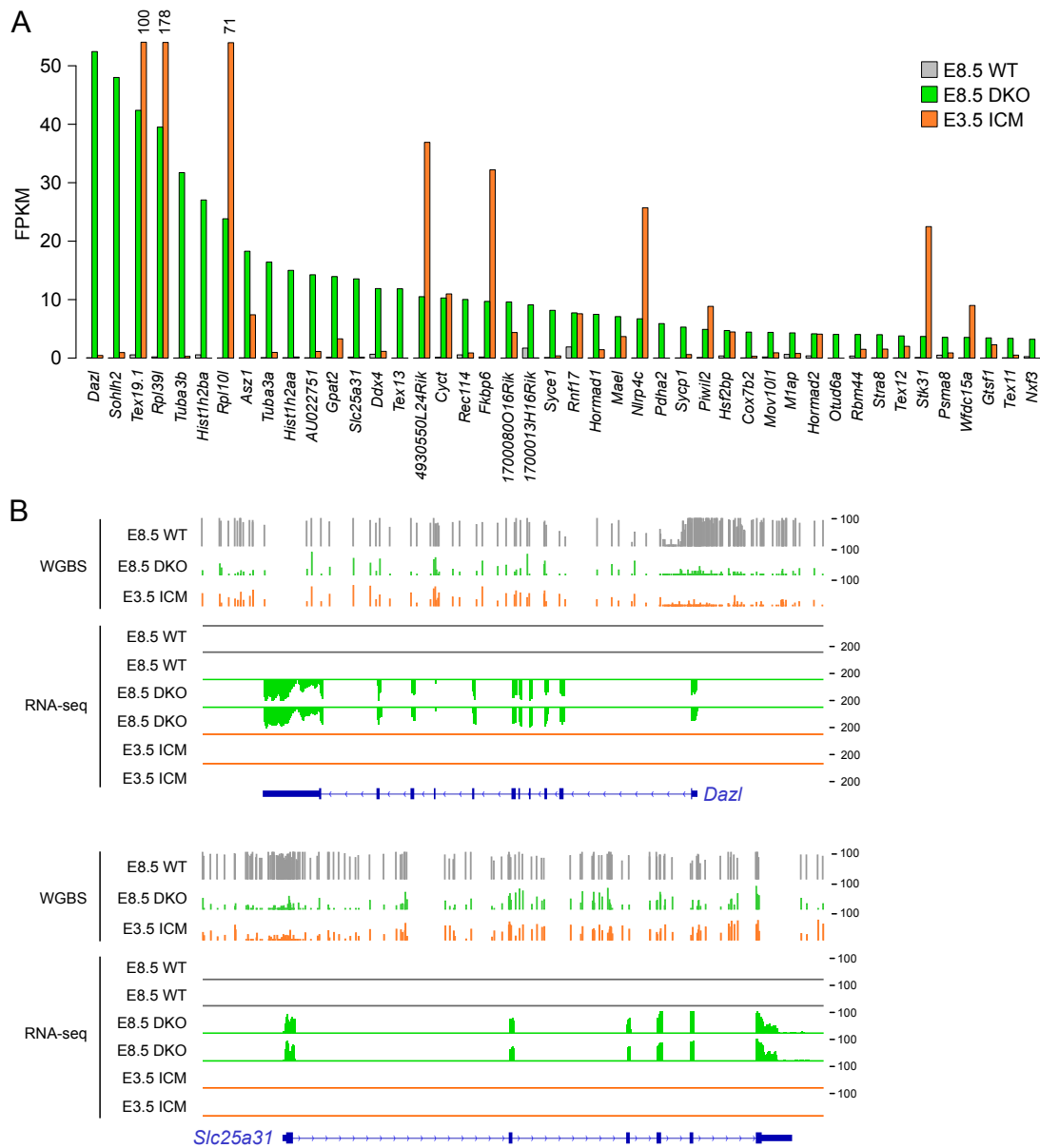
Supplementary Figure 3. Conditional inactivation of DNMT3A/B in MEFs does not affect DNA methylation. A. RT-qPCR analysis of *Dnmt3a*, *Dnmt3b* and *Dnmt1* gene expression levels in cells treated with tamoxifen (Tam) and control cells not treated with tamoxifen (no Tam) at 69 days of culture (mean \pm SEM, $n=2$). The primers used for *Dnmt3a* and *Dnmt3b* were designed in the floxed exons to confirm the efficient deletion. **B.** Density scatter plots showing the correlation of RRBS methylation scores in 500 bp windows between tamoxifen-treated cells (Tam) and control cells not treated with tamoxifen (no Tam) after 23 days and 69 days of culture. The density of points increases from blue to dark red. **C.** Number of hypomethylated DMRs (>20%) identified in tamoxifen-treated cells compared to control cells not treated with tamoxifen after 23 days and 69 days of culture. **D.** Boxplot representing the methylation in the 11 hypomethylated DMRs identified in the tamoxifen-treated cells at 69 days of culture.



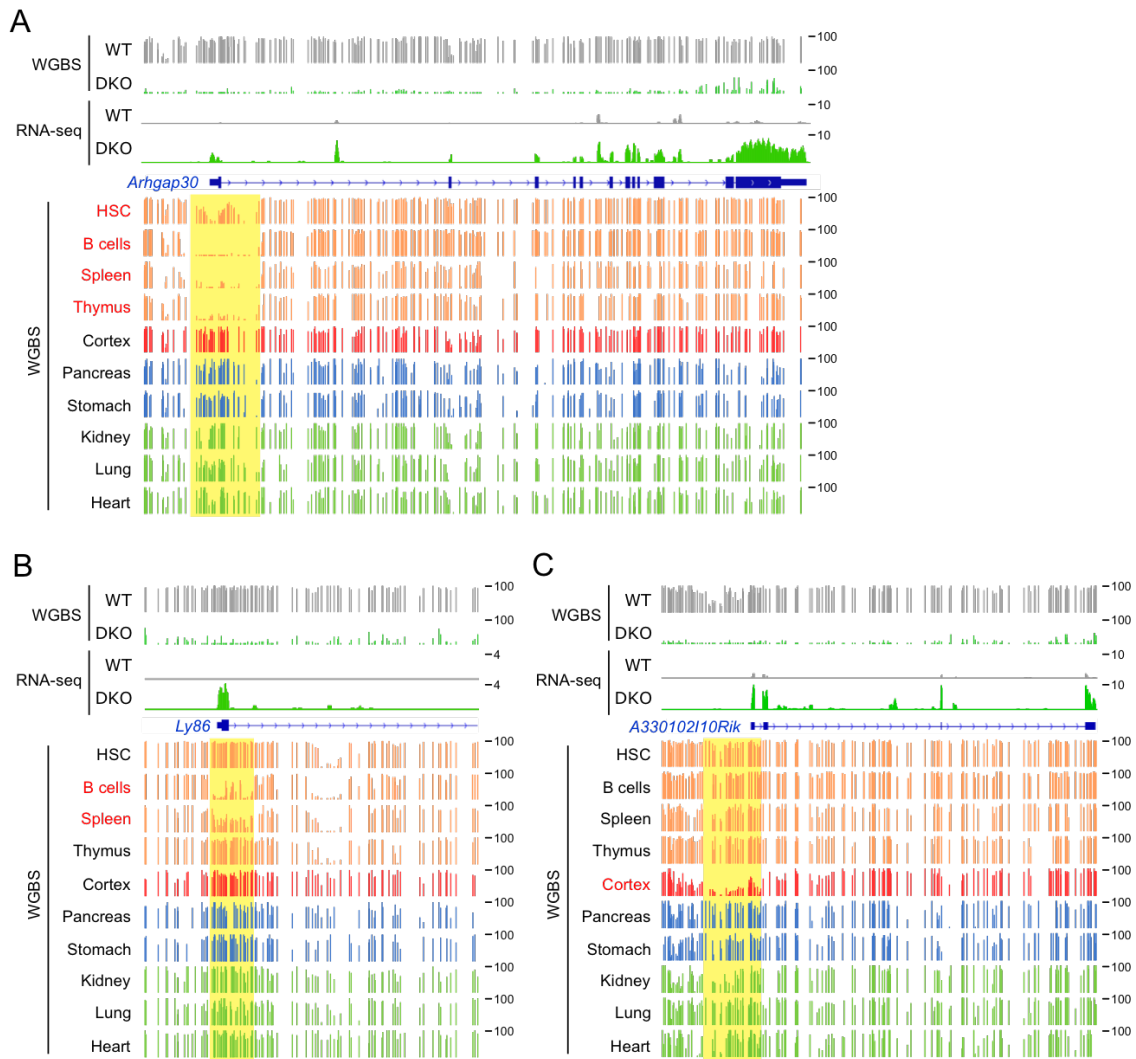
Supplementary Figure 4. Analysis of differentially expressed genes in *Dnmt* mutant embryos. A-B. RNA-seq profiles validate the absence of the exons 4-5 of the *Dnmt1* gene (highlighted in yellow) in *Dnmt1*^{-/-} embryos, and the absence of catalytic exons of the *Dnmt3a* and *Dnmt3b* genes (highlighted in yellow) in DKO embryos. C. Volcano plot representation of gene expression changes in *Dnmt1*^{-/-} and DKO embryos. Red dots indicate significantly differentially expressed genes (fold change >3, adjusted *p* value <0.001). D. Comparison of gene expression changes in *Dnmt1*^{-/-} (horizontal axis) and DKO (vertical axis) embryos. E. Venn diagram comparing the genes upregulated in DKO embryos and *Dnmt* triple knock-out (TKO) ES cells. F. Comparison of gene expression changes in TKO ES cells (horizontal axis) (Domcke et al., 2015) and DKO embryos (vertical axis). Orange, pink and red dots represent, respectively, genes dysregulated in TKO ES cells only, in DKO embryos only or in both. G. Box plots representing the expression level (plotted as FPKM) in *Dnmt* mutant embryos and ES cells of the genes found upregulated in DKO embryos. H. Box plots representing the fold change of expression of the three groups of genes upregulated in DKO embryos (related to Figure 4C). I. Venn diagrams comparing the genes upregulated in DKO and *Dnmt3b*^{-/-} embryos (top), or only germline genes upregulated in DKO and *Dnmt3b*^{-/-} embryos (bottom). J. Correlation between the level of promoter DNA methylation (grey bars) and expression (FPKM, red dots) for six germline genes in *Dnmt3b*^{-/-}, *Dnmt1*^{-/-} and DKO embryos.



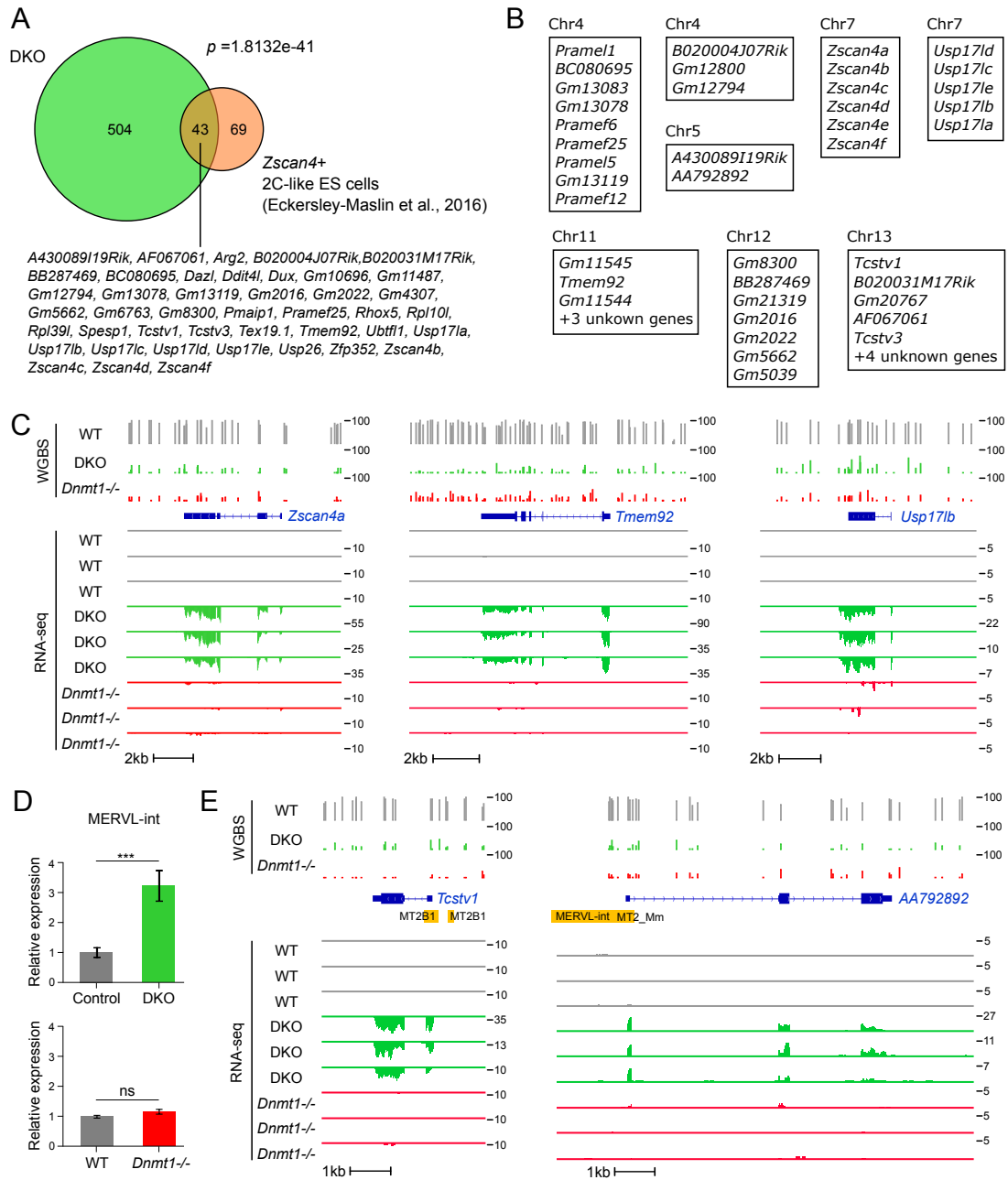
Supplementary Figure 5. Targeted demethylation of germline CpG-rich promoters in mouse fibroblasts using CRISPR- dCas9. **A.** Vectors used in the experiments. We used a vector coding for dCas9 fused to the catalytic domain of the human TET1 protein (hTET1CD, top) and the all-in-one vector coding for the dCas9-SunTag-TET1 system (top) able to recruit multiple copies of hTET1CD (Morita et al., 2016). Arrows indicate promoter sequences. NLS= Nuclear Localisation Signal; GS= GS linker. **B.** Schematic representation of the *Dazl* and *Asz1* promoters. Vertical bars indicate the position of CpGs, green arrows indicate the position of gRNAs, and red bars indicate the amplicons used for bisulfite PCR. **C.** Bisulfite sequencing analysis of the *Dazl* promoter in fibroblasts transfected with the dCas9-TET1 fusion vector containing no gRNA or the *Dazl* gRNA4, indicating low demethylation efficiency. White squares represent unmethylated CpGs and black squares represent methylated CpGs. **D.** Patterns of CpG methylation of the *Dazl* and *Asz1* promoters by bisulfite sequencing analysis in fibroblasts transfected with the dCas9-SunTag-TET1 all-in-one vector containing no gRNA, the *Dazl* gRNA4, the *Asz1* gRNA5 and non-transfected cells. The green arrows indicate to CpGs contained in the gRNA sequences. **E.** Comparison of *Dazl* expression measured by RT-qPCR in fibroblasts transfected with the dCas9-TET1 fusion vector or dCas9-SunTag-TET1 vector containing no gRNA or the *Dazl* gRNA4. Expression is represented as a value normalized to three housekeeping genes (n=3, mean ± SEM). NT= non transfected cells.



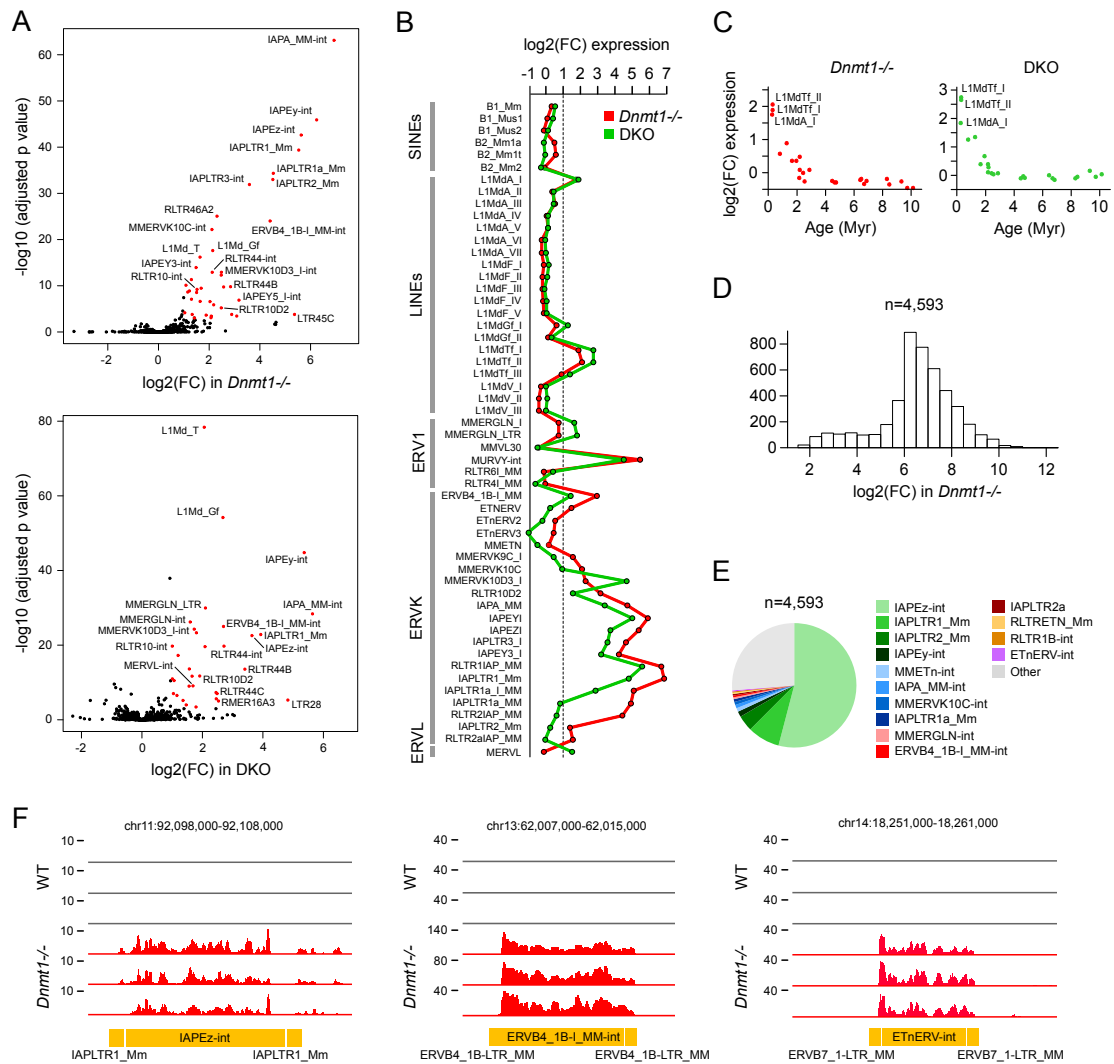
Supplementary Figure 6. Many germline genes are derepressed in E8.5 DKO embryos but not in blastocysts. **A.** Comparison of RNA-seq expression values (plotted as FPKM) in E8.5 embryos and E3.5 ICM for the 44 most derepressed germline genes in DKO embryos. **B.** Genome browser tracks showing WGBS and RNA-seq profiles of the *Dazl* and *Slc25a31* genes in E8.5 embryos and E3.5 ICM. These examples illustrate that several germline genes are not derepressed in E3.5 ICM despite similar hypomethylation than in E8.5 DKO embryos.



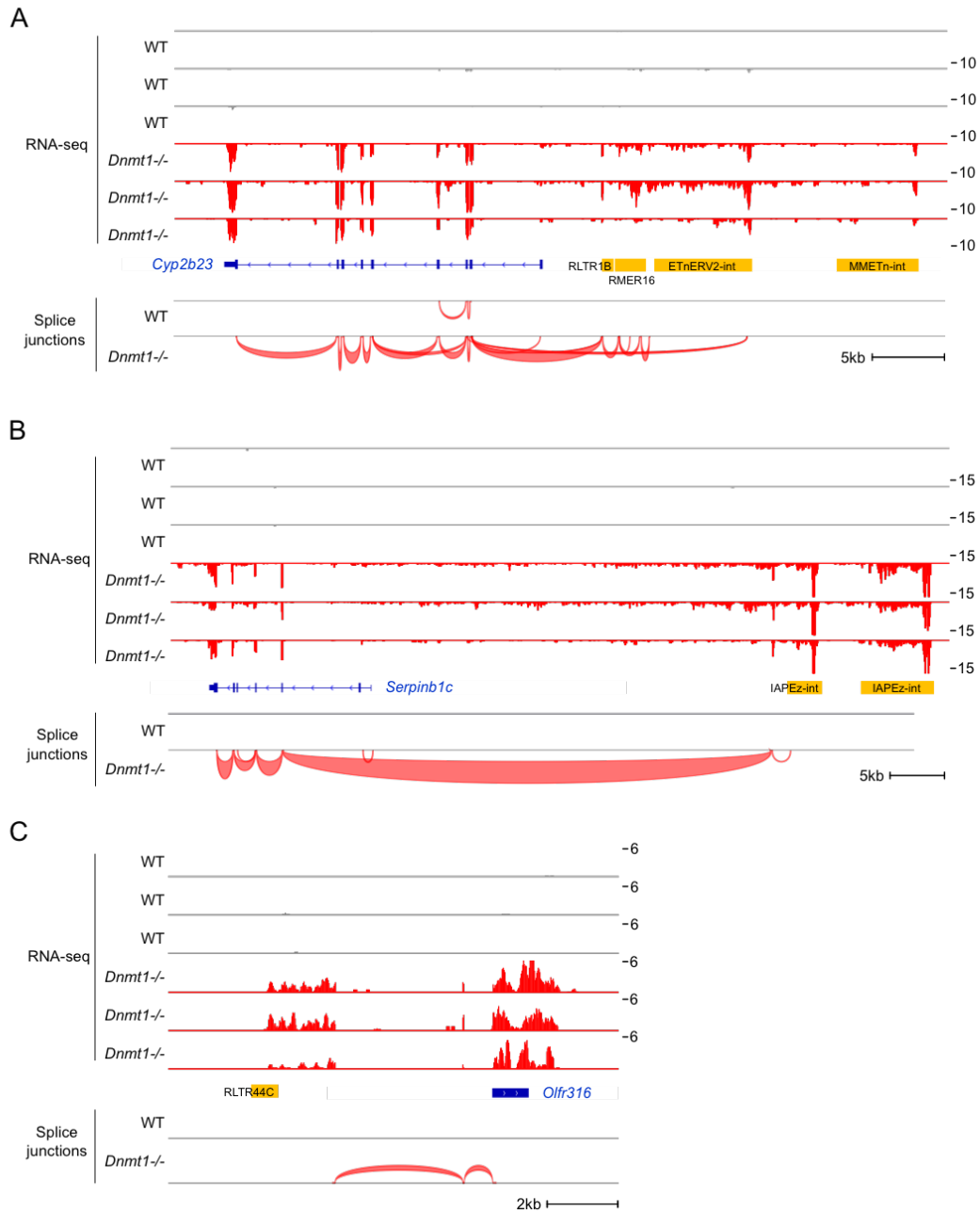
Supplementary Figure 7. DNA methylation profiles of lineage-committed genes upregulated in DKO embryos. Genome browser tracks of WGBS DNA methylation profiles of the *Arhgap30* (A), *Ly86* (B) and *A330102110Rik* (C) genes in embryos and adult tissues (Cabezas-Wallscheid et al., 2014; Duncan et al., 2018; Hon et al., 2013). Also shown are RNA-seq tracks in one replicate of WT and DKO embryos. These genes show tissue-specific DNA methylation profiles in their promoter sequences (highlighted in yellow). The tissues or cells where the genes are expressed are written in red. HSC = Hematopoietic Stem Cells.



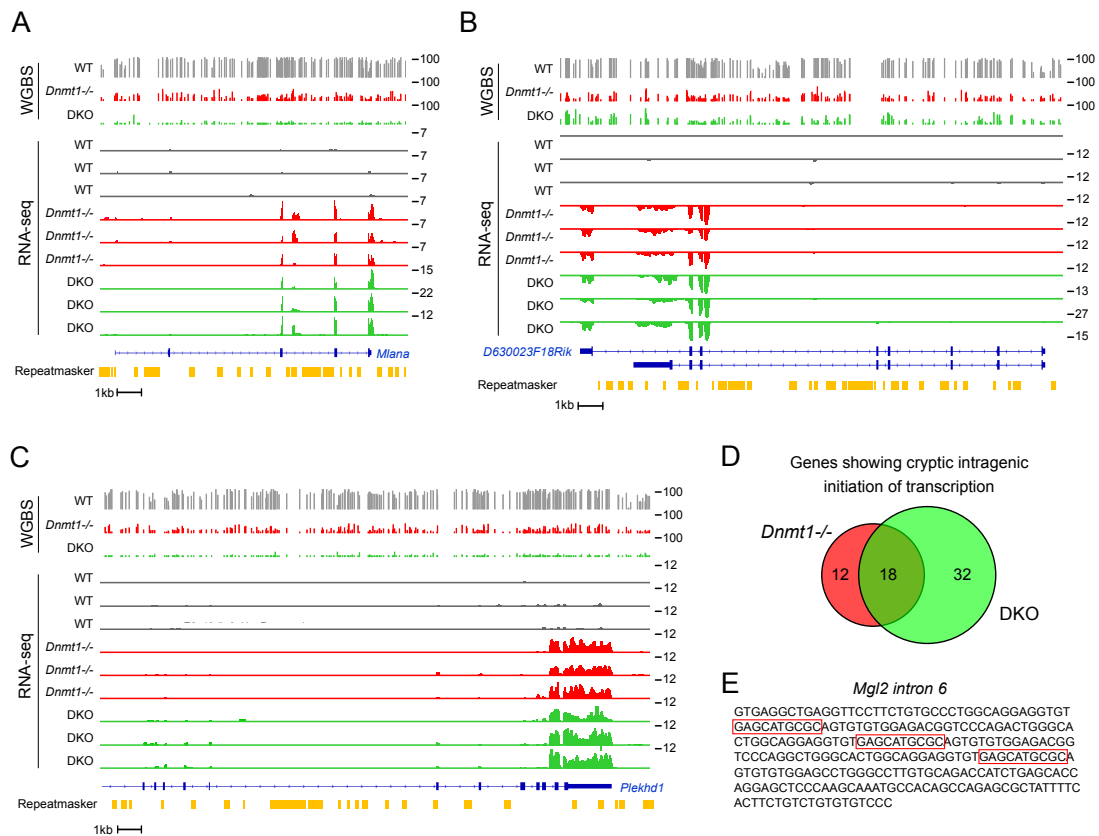
Supplementary Figure 8. 2C-specific genes and MERVL transposons are derepressed in DKO embryos. A. Venn diagram comparing the lists of genes upregulated in DKO embryos and in *Zscan4*⁺ 2C-like ES cells (Eckersley-Maslin et al., 2016). The names of the 43 common genes are shown below the diagram. The p value indicates the significance of the overlap (hypergeometric test). **B.** Examples of gene clusters of 2C-specific genes derepressed in DKO embryos. **C.** Examples of genome browser views of 2C-specific genes derepressed in DKO embryos. The top lanes show WGBS methylation profiles in embryos, and the bottom lanes show RNA-seq profiles in three replicates of WT, DKO and *Dnmt1*^{-/-} embryos. **D.** Expression of MERVL-int transposons (measured as RNA-seq normalized read counts in RepeatMasker MERVL-int sequences) in DKO and *Dnmt1*^{-/-} embryos relative to control embryos (mean \pm SEM, $n=6$ for DKO, $n=3$ for *Dnmt1*^{-/-}). ns: not significant; ***: $p<0.001$ (DESeq adjusted p value). **E.** Examples of genes derepressed in DKO embryos that initiate from ERVL sequences. LTR transposons annotated by RepeatMasker are shown in yellow below the gene.



Supplementary Figure 9. Reactivation of retrotransposons in *Dnmt* mutant embryos. **A.** Volcano plot representing the changes in expression of transposon families in *Dnmt1*^{-/-} and DKO embryos, measured by counting unique and multi-mapping RNA-seq reads in RepeatMasker annotations. Significantly upregulated transposon families (fold change > 2, adjusted *p* value < 0.001) are highlighted in red. **B.** Fold change of expression of SINE, LINE, ERV1, ERVK and ERVL retrotransposon families in *Dnmt1*^{-/-} and DKO embryos, measured by mapping RNA-seq reads to Repbase sequences. **C.** Fold change of expression of Repbase L1Md families in *Dnmt1*^{-/-} and DKO embryos as a function of their evolutionary age in million of years (Myr) (taken from Sookdeo et al., 2013). **D.** Density histogram of the log₂(fold change) expression for the 4,593 individual transposon copies found significantly reactivated (fold change > 3, adjusted *p* value < 0.001) in *Dnmt1*^{-/-} embryos. **E.** Pie chart showing the classification by family of the 4,593 individual transposon copies reactivated in *Dnmt1*^{-/-} embryos. **F.** RNA-seq tracks showing examples of derepressed ERV transposon copies in *Dnmt1*^{-/-} embryos. RepeatMasker annotations are shown below the tracks.



Supplementary Figure 10. Derepressed ERVs produce long RNAs and chimeric transcripts with proximal genes in *Dnmt1*^{-/-} embryos. The figure shows genome browser tracks of RNA-seq profiles in WT and *Dnmt1*^{-/-} embryos, along with splice junctions visualized in the IGV browser in one replicate of WT and *Dnmt1*^{-/-} embryo. ERVs annotated by RepeatMasker are displayed in yellow. **A.** *Cyp2b23* expression is induced by an upstream cluster of ERV elements, which initiates a long RNA that splices into the exon 2 of the gene. **B.** *Serpinb1c* expression is induced by a long RNA, which initiates 40 kb upstream of the gene in a derepressed IAPEz element and splices into the exon 3 of the gene. **C.** *Olf316* expression is induced by an RNA initiating in an upstream RLTR44C element.



Supplementary Figure 11. Cryptic transcription initiation from intragenic sequences in *Dnmt* mutant embryos. **A-C.** Genome E8_13-3_DKO.Reverse.bw browser tracks of WGBS methylation profiles and RNA-seq for the *Mlana*, *Plekhd1* and *D630023F18Rik* genes in WT and *Dnmt* mutant embryos. Three replicates of RNA-seq are shown. The RepeatMasker annotation track is displayed below in yellow. **D.** Venn diagram showing the overlap between the lists of genes with cryptic intragenic initiation identified in *Dnmt1*^{-/-} and DKO embryos. **E.** Nespas DNA sequence of the intron 6 in the *Mgl2* gene. DNA binding motifs of the transcription factor NRF1 are highlighted in red.

2. Rôle du facteur de transcription E2F6 dans le ciblage épigénétique des gènes gamétiques dans l'embryon

2.1. Contexte et objectifs

Le développement embryonnaire fait l'objet d'une large reprogrammation épigénétique au cours de laquelle la méthylation de l'ADN est d'abord effacée après la fécondation, puis est remise en place dans l'embryon après son implantation. Durant la phase de reméthylation du génome, seulement 1% des îlots CpG situés dans les promoteurs de gènes acquièrent la méthylation, suggérant un mécanisme de régulation très spécifique de l'adressage des DNMTs (Auclair et al., 2014). Les îlots CpG qui acquièrent la méthylation sont généralement localisés dans les promoteurs des gènes gamétiques, ce qui engendre leur répression sur le long terme dans les cellules somatiques (Auclair et al., 2014; Borgel et al., 2010).

Plusieurs études suggèrent un adressage de la méthylation par défaut sur le génome. Dans ce modèle, les îlots CpG sont protégés du recrutement des DNMTs par des mécanismes impliquant des protéines ayant un domaine CXXC se liant aux CpG non méthylés comme les enzymes TET ou KDM2B (appelée aussi FBXL10) (Boulard et al., 2015; Dawlaty et al., 2014; Wu et al., 2011). De plus, les modifications des histones pourraient jouer un rôle dans la protection des îlots CpG de la méthylation de l'ADN, notamment la modification d'histone activatrice H3K4me3 qui est antagoniste de la méthylation de l'ADN (Meissner et al., 2008; Weber et al., 2007). En effet, le domaine ADD des DNMTs est capable d'interagir avec H3K4 dans sa version non méthylée uniquement (Ooi et al., 2007; Otani et al., 2009; Zhang et al., 2010). Enfin, un autre modèle suggère que la liaison de facteurs de transcription joue un rôle prépondérant dans la protection des îlots CpG contre la méthylation de l'ADN (Hartl et al., 2019; Krebs et al., 2014).

Dans cette deuxième partie, j'ai tenté de décrypter les voies moléculaires qui permettent de cibler la méthylation de l'ADN vers les promoteurs des gènes gamétiques au cours du développement chez la souris.

Une étude publiée en 2010 suggère que les DNMTs peuvent être adressées spécifiquement au promoteur de gènes gamétiques grâce à leur interaction avec le facteur de transcription E2F6 (Velasco et al., 2010). Le facteur de transcription E2F6 est un membre du complexe Polycomb PRC1.6 (Gao et al., 2012; Ogawa et al., 2002; Qin et al., 2012; Trojer et al., 2011). Ce complexe non canonique ne possède pas les protéines CBX permettant son recrutement par H3K27me3 médiée par le complexe Polycomb PRC2 (Gao et al., 2012). Le ciblage de ce complexe sur le génome nécessite donc la présence

de facteur de transcription se liant à l'ADN, comme E2F6 qui se lie au motif E2F TCCCGC et MAX/MGA qui se lient au motif E-box CACGTG. L'inactivation de plusieurs membres du complexe PRC1.6 conduit à la dérégulation de gènes gamétiques dans les cellules ES (*Endoh et al., 2017; Hisada et al., 2012; Qin et al., 2012; Suzuki et al., 2016*). Néanmoins, le rôle du facteur E2F6 dans la répression et le recrutement de la méthylation de l'ADN au niveau des gènes gamétiques n'a pas été étudié en détail.

Sur la base de ces données de la littérature, nous avons décidé d'étudier le rôle du facteur de transcription E2F6 dans la répression des gènes gamétiques et son implication dans le recrutement de la méthylation de l'ADN au cours du développement chez la souris.

2.2. Stratégie expérimentale

Nous avons étudié le rôle de E2F6 en utilisant comme modèle à la fois des souris et des cellules ES KO pour le gène *E2f6*.

Les souris mutantes pour *E2f6*, initialement développées par le laboratoire de Jacqueline Lees (*Courel et al., 2008*), présentent une cassette PGK-Néomycine insérée entre l'exon 4 et l'exon 8 qui remplace la partie 3' du gène encodant pour une partie du domaine de liaison à l'ADN, le domaine leucine-zipper et le domaine marked-box (**Figure 24A**). Ces souris mutantes sont viables et fertiles, mais présentent des transformations homéotiques de certaines vertèbres (*Courel et al., 2008*).

Le rôle d'E2F6 dans la répression des gènes gamétiques a été déterminé en réalisant des études comparatives à différents stades de l'embryogenèse. La phase précédant la reméthylation du génome a été étudiée dans les blastocystes, et la phase de reméthylation de génome sur des embryons isolés au stade E8.5 ainsi que sur différents tissus adultes. Enfin, nous avons dérivé des cellules embryonnaires comme des cellules ES et des MEFs nous permettant de réaliser des études mécanistiques plus approfondies (**Figure 24B**).

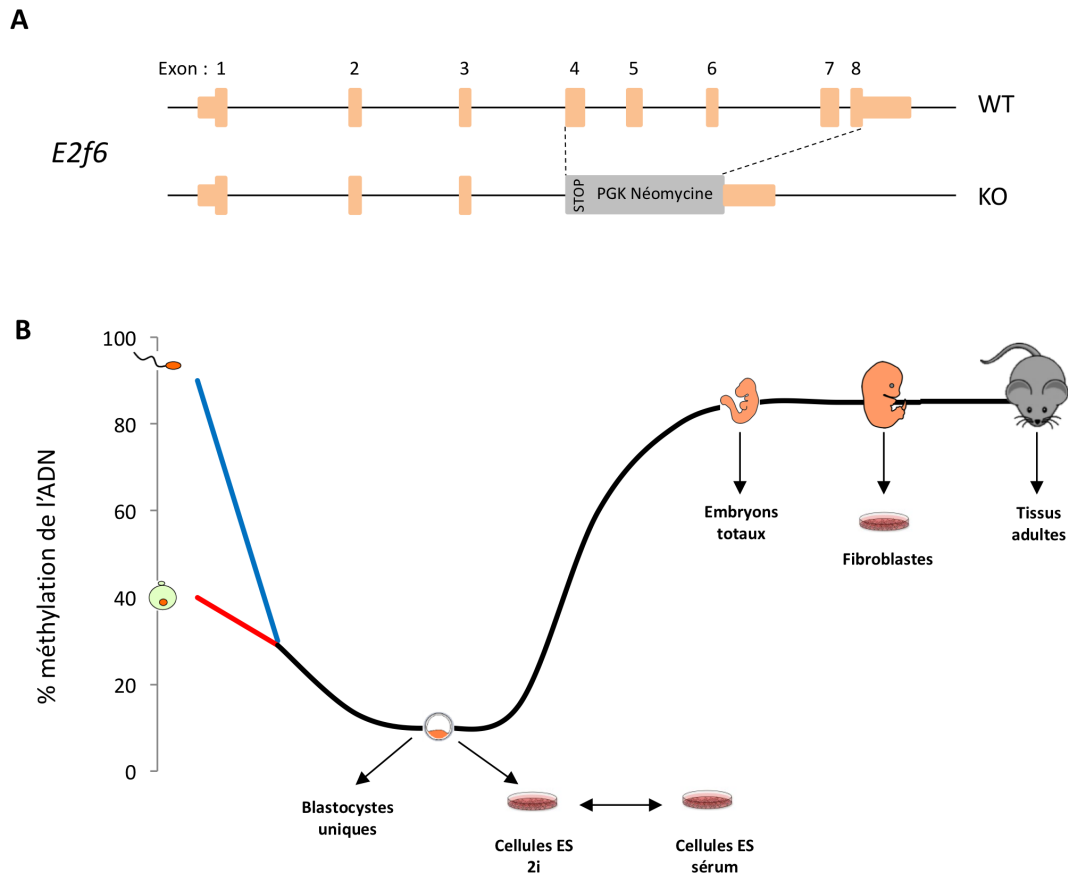


Figure 24 : Stratégie permettant d'étudier le rôle d'E2F6 durant le développement embryonnaire. A) Modèle murin génétique muté du gène *E2f6*. Une cassette PGK-Néomycine a été insérée entre les exons 4 et 8 du gène induisant une mutation perte de fonction. B) Le rôle de E2F6 a été étudié en analysant plusieurs étapes du développement embryonnaire. Des blastocystes uniques ainsi que des cellules ES ont été utilisés par étudier son rôle avant et pendant la phase de reméthylation du génome. Des embryons totaux au stade E8.5, des tissus adultes, ainsi que des MEFs dérivées à partir du stade E13.5 ont été utilisés pour étudier son rôle après la phase de reméthylation du génome.

2.3. Travaux réalisés

Mon implication dans ce projet :

- Dissection d'embryons et de tissus de souris *E2f6* KO pour la préparation d'échantillons d'ADN et ARN.
- Préparation des bibliothèques RRBS et WGBS à partir des embryons et tissus de souris *E2f6* KO.
- Participation à l'analyse de séquençage des méthylomes et transcriptomes.
- Validation par RT-qPCR des gènes cibles d'E2F6 dans les cellules ES et les embryons.
- Mise au point et utilisation d'une stratégie permettant de déterminer l'expression de gènes par RT-qPCR sur des blastocystes uniques.
- Adaptation des cellules ES en 2i et analyse de l'expression des gènes par RT-qPCR.

- Mise au point d'une stratégie d'inactivation du gène *E2f6* par CRISPR-Cas9 dans les MEFs et les cellules ES.
- Création, validation et analyse des clones WT et *E2f6* KO obtenus par CRISPR-Cas9 dans les fibroblastes et les cellules ES
- Expérience de sauvetage de *E2f6* dans les MEFs.

Ce projet a été réalisé en collaboration avec l'équipe de Christian Hagemeier, découvreur du gène *E2f6* en 1997 (*Morkel et al., 1997*), de l'Hôpital Charité à Berlin, Allemagne. Nos collaborateurs ont notamment apporté les cellules ES KO pour *E2f6*, réalisé les expériences de ChIP et ChIP-seq de E2F6 et de membres du complexe PRC1.6 dans ces cellules ES, ainsi que les études mécanistiques de sauvetage avec des versions mutées de *E2f6* dans le domaine de liaison à l'ADN et dans le domaine marked-box dans les cellules ES.

2.4. Article 2: E2F6 initiates stable epigenetic silencing of the germline program during embryonic development

(en cours de soumission)

E2F6 initiates stable epigenetic silencing of the germline program during embryonic development

Thomas Dahlet^{1,2*}, Matthias Truss^{3*#}, Ute Frede³, Anaïs F Bardet^{1,2}, Michael Dumas^{1,2}, Judith Vallet^{1,2}, Sarah Kottnik³, Peter Hansen⁴, Uschi Luz³, Gonzalo Alvarez³, Ghislain Auclair^{1,2}, Jochen Hecht⁴, Peter N. Robinson⁴, Christian Hagemeyer^{3#}, Michael Weber^{1,2#}

1) University of Strasbourg

2) CNRS UMR7242, Biotechnology and Cell Signaling, 300 Bd Sébastien Brant, 67412 Illkirch Cedex, France

3) Labor für Pädiatrische Molekularbiologie, Charité - Universitätsmedizin Berlin, 10117, Berlin, Germany

4) Berlin Brandenburg Center for Regenerative Therapies (BCRT), Charité Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany

* These authors contributed equally

correspondence to: Michael Weber (michael.weber@unistra.fr), Christian Hagemeyer (Christian.Hagemeyer@Charite.de) and Matthias Truss (Matthias.Truss@Charite.de)

Abstract

In the mouse, long-term silencing by CpG island DNA methylation is specifically targeted to germline genes in somatic cells, however the molecular mechanisms of this specificity remain unclear. Here we demonstrate that the transcription factor E2F6, a member of the polycomb repressive complex 1.6 (PRC1.6), is critical to target and initiate epigenetic silencing at germline genes in early embryogenesis. Genome-wide, E2F6 binds preferentially to CpG islands in embryonic cells. E2F6 cooperates with MAX/MGA to silence a subgroup of germline genes in mouse embryonic stem cells and *in vivo*, a function that critically depends on the E2F6 marked box domain. Inactivation of *E2f6* leads to a failure to deposit CpG island DNA methylation at these genes during implantation. Furthermore, E2F6 is required to initiate epigenetic silencing in early embryonic cells but becomes dispensable in differentiated cells. Our findings elucidate the mechanisms of epigenetic targeting of germline genes and provide a paradigm for how transient repression signals by DNA-binding factors in early embryonic cells are translated into long term epigenetic silencing during mammalian development.

Introduction

Methylation of cytosines in CpG dinucleotides by the DNA methyltransferases (DNMTs) represents one of the best-characterized epigenetic control mechanisms for gene transcription. In mice, genome-wide DNA methylation patterns are established during a short wave of *de novo* methylation concomitant to embryo implantation (Borgel et al. 2010; Smith et al. 2012). During this developmental stage, *Dnmt3a* and *Dnmt3b* are up-regulated and DNA methylation progresses uniformly throughout the genome (Auclair et al. 2014).

In contrast to interspersed CpG dinucleotides, the majority of genomic regions of high CpG density, known as CpG islands (CGIs), are protected against DNA methylation (Deaton and Bird 2011). In post-implantation embryos, less than 1% of CGIs at transcription start sites (TSS) have acquired DNA methylation (Auclair et al. 2014). The mechanisms underlying the protection of CGIs from DNA methylation are still poorly understood. CGIs tether proteins containing CXXC zinc finger DNA binding domains that recognize unmethylated CpG containing DNA. Some of these CXXC proteins have been implicated in the protection against DNA methylation (Boulard et al. 2015). CGIs also harbor high levels of H3K4 trimethylation known to inhibit the binding to H3 histone tails and allosteric activation of *de novo* DNA methyltransferases (Ooi et al. 2007; Li et al. 2011). Additionally, several studies suggest that binding of activating transcription factors (TFs) at CpG islands contributes to protection against DNA methylation (Brandeis et al. 1994; Macleod et al. 1994; Boumber et al. 2008; Gebhard et al. 2010; Krebs et al. 2014).

Strikingly, promoter CGI methylation in development is almost exclusively observed at germline genes (Auclair et al. 2014). Suppression of transcription of germline genes in somatic cells depends on DNA methylation, since they are de-repressed in DNA methylation deficient mouse embryos and in response to inhibitors of DNA methylation (Hackett et al. 2012; Auclair et al. 2014). However the molecular mechanisms leading to DNA methylation specifically at CGIs of germ cell genes are unknown. Existing evidence suggests that the transcription factor E2F6 could play a role in this process. E2F6 is a ubiquitously expressed member of the E2F family of transcription factors that control cell proliferation and cell fate (Morkel et al. 1997; Gaubatz et al. 1998; Trimarchi et al. 1998; Giangrande et al. 2004). E2F6 serves as a transcriptional repressor that binds the DNA site TCCCGC, which is highly conserved in promoter regions of meiosis-specific genes (Kehoe et al. 2008). For the *Tuba3a* gene, E2F6-dependent gene repression in somatic cells involves promoter DNA methylation (Pohlers et al. 2005). Interestingly, somatic cells lacking E2F6 or DNMT3B show overlapping sets of de-repressed genes (Borgel et al. 2010; Velasco et al. 2010), further suggesting a link between E2F6 and DNA methylation. Based on experiments in somatic cells, a direct recruitment of DNMT3B by E2F6 has been suggested (Velasco et al. 2010).

In mammalian cells, E2F6 interacts with chromatin modifiers such as G9A and components of the polycomb repressive complex PRC1.6 (Ogawa et al. 2002; Trojer et al. 2011; Qin et al. 2012; Hauri et al. 2016; Kloet et al. 2016). PRC1.6 is one of the non-canonical PRC1 (ncPRC1) complexes that are recruited to chromatin independently of H3K27methylation. Targeting of ncPRC1 complexes can involve recognition of unmethylated CpG islands through the CXXC domain of KDM2B (Farcas et al. 2012). The PRC1.6 complex lacks the KDM2B subunit but contains RYBP, PCGF6, L3MBTL2, CBX3, WDR5 and the DNA binding transcription factors E2F6/DP1 and MGA/MAX (Trojer et al. 2011; Qin et al. 2012). Interestingly MAX, PCGF6, L3MBTL2 and RYBP have been linked to the repression of germline genes in mouse ES cells (Hisada et al. 2012; Qin et al. 2012; Maeda et al. 2013; Endoh et al. 2017; Zhao et al. 2017). In contrast to the other components of the PRC1.6 complex, the biological roles and targets of E2F6 have been much less explored. Here, we investigated the biological functions of E2F6 in the mouse and the hypothesis that E2F6 may play an essential role in initiating long term epigenetic repression of germline genes during embryogenesis.

Results

Identification of E2F6 genomic targets in mouse embryonic stem cells

To start investigating the role of E2F6, we performed ChIP-seq analysis of E2F6 in *E2f6*^{+/+} and *E2f6*^{-/-} mouse ES cells established from *E2f6* knockout and control mice (Pohlers et al. 2005). This analysis generated 25,935,462 unique mapped reads and 2158 high confidence peaks. The established E2F6 target gene *Tuba3a* (Pohlers et al. 2005) revealed a well-defined peak over its TSS (**Figure 1A**). Importantly, most peaks are non-detectable in control ChIP-seq datasets derived from *E2f6*^{-/-} ES cells or with input DNA, thereby confirming the specificity of the immunoprecipitation (**Figure 1A** and **Figure S1A**). The majority of E2F6 peaks were located close to transcription start sites (TSS) in CpG-rich regions (**Figure 1B-C**). Furthermore, sequence motif analysis of the E2F6 ChIP-seq peaks revealed that they are enriched for the E2F6 binding site TCCCGC and the E-box motif CACGTG bound by Max/Mga (**Figure 1D-E** and **Figure S1B**). Gene Ontology (GO) analysis of E2F6 bound gene promoters revealed an enrichment for gene functions associated with metabolic processes, meiosis and cell cycle, which is in line with known E2F6 functions that represses genes activated by E2F transcription factors (E2F-TFs) during G1/S transition of the cell cycle (Giangrande et al. 2004; Bertoli et al. 2013).

Next we compared E2F6 ChIP-seq with publicly available ChIP-seq datasets in mouse ES cells, in particular ChIP-seq of the PRC1.6 subunits. This revealed that E2F6 ChIP-seq profiles cluster with datasets of other components of the PRC1.6 complex including MAX, MGA, L3MBTL2, RYBP and PCGF6 (**Figure S1C**). Indeed, E2F6 peaks in the genome show frequent binding of the other PRC1.6

subunits (**Figure 1F-G** and **Figure S1D**). Furthermore, a comparison of the peaks of E2F6, MGA and L3MBTL2 revealed a high degree of overlap (**Figure S1E**), further indicating that E2F6 and PRC1.6 frequently bind to the same genomic loci. Thus, E2F6 colocalizes with PRC1.6 on chromatin, which is in line with E2F6 being a proposed member of the PRC1.6 complex. To validate that E2F6 is part of the PRC1.6 complex in mouse ES cells, we performed co-immunoprecipitation in ES cells stably expressing an HA-tagged version of E2F6 and confirmed that E2F6 interacts with MGA, PCGF6 and L3MBTL2 (**Figure 1H**). In summary, E2F6 is frequently bound to CpG-rich promoters and colocalizes with the PRC1.6 complex genome-wide in mouse ES cells.

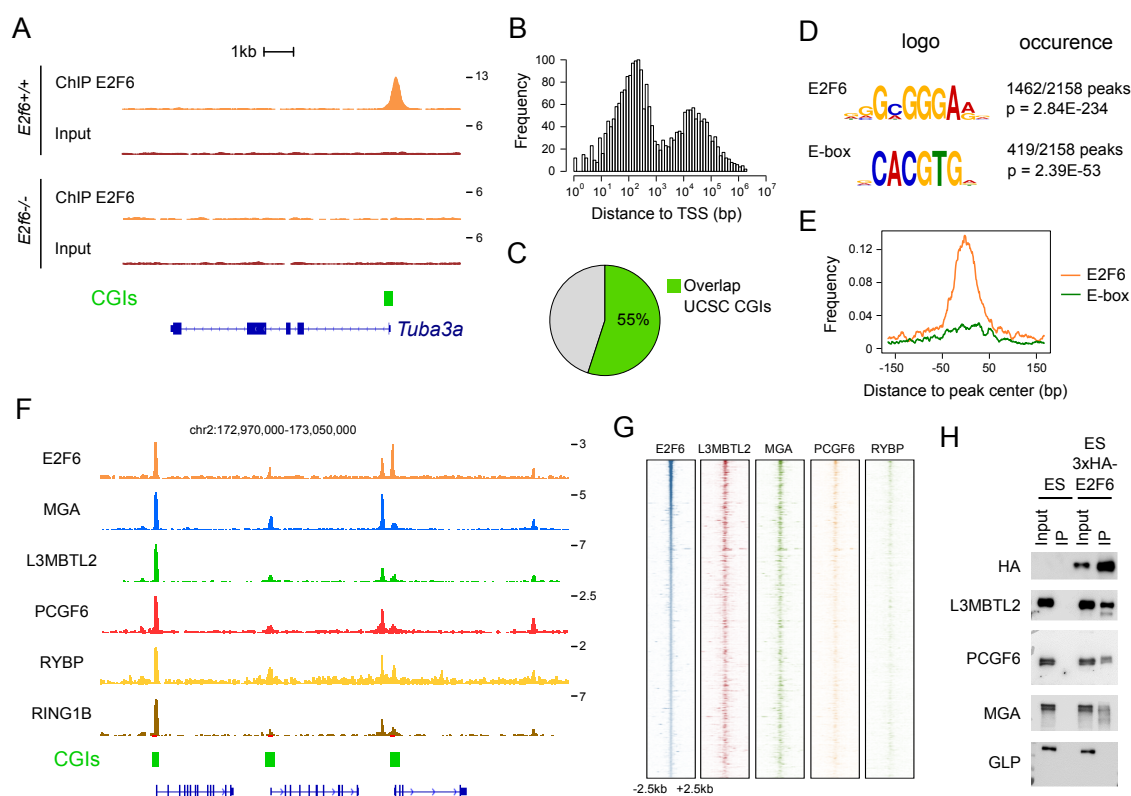


Figure 1. ChIP-seq analysis of E2F6 binding sites in ES cells. **A.** Genome browser tracks showing E2F6 ChIP-seq and input signals in *E2F6*^{+/+} and *E2F6*^{-/-} ES cells over the *Tuba3a* gene. UCSC CGI and RefSeq gene annotations are shown below the tracks. **B.** Density histogram representing the distance of E2F6 peak summits to the closest TSS. **C.** Percentage of E2F6 peaks overlapping with annotated UCSC CpG islands (CGIs). **D.** JASPAR sequence motifs enriched in E2F6 peaks. The E2F6 consensus binding site and the E-box CACGTG are the most enriched sequence motifs identified in E2F6 peaks. **E.** Metaplot representing the position of E2F6 and E-box motifs relative to the center of E2F6 peaks. **F.** Genome browser tracks of ChIP-seq signals in mouse ESCs showing that E2F6 peaks frequently colocalize with peaks of other PRC1.6 subunits. UCSC CGI and RefSeq gene annotations are shown below the tracks. **G.** Heatmap of E2F6, MGA, L3MBTL2, PCGF6 and RYBP ChIP-seq signals \pm 2.5 kb around E2F6 peak summits. **H.** Co-immunoprecipitation shows that E2F6 interacts with the PRC1.6 complex in mouse ES cells. Immunoprecipitation was performed with an anti-HA antibody on nuclear extracts from control ES cells and ES cells expressing 3xHA-E2F6. Western blot was performed on the input and immunoprecipitates using the indicated antibodies. GLP was used as a control nuclear protein that does not interact with E2F6.

E2F6 represses germline genes in mouse embryonic stem cells and *in vivo*

To investigate whether E2F6 binding is causatively involved in transcriptional regulation, we performed gene expression analyses in *E2f6*^{+/+} and *E2f6*^{-/-} ES cells. Microarray experiments in these cells identified 46 genes significantly de-repressed (fold change ≥ 2.5 ; adjusted p-value < 0.01) upon *E2f6* knock out (**Figure 2A** and **Annexe 2.5**). Thus, most genes bound by E2F6 are not transcriptionally deregulated in *E2f6* knockout ES cells. In particular the transcription of G1/S-induced E2F responsive genes known to be occupied by E2F6 was not affected in *E2f6* knock out cells, consistent with previous findings in somatic cells (Giangrande et al. 2004). Rather, gene ontology analysis showed that the upregulated genes are strongly enriched for GO terms related to germ cells (meiotic cell cycle $p = 7.52 \times 10^{-12}$; sexual reproduction $p = 1.12 \times 10^{-4}$; **Annexe 2.4**). These genes include the previously known targets of E2F6 *Tuba3a*, *Slc25a31*, *Smc1b* and *Stag3* (Pohlert et al. 2005; Storre et al. 2005; Kehoe et al. 2008). The derepression of a selected subset of these germline genes in *E2f6*^{-/-} ES cells was validated by RT-qPCR (**Figure 2B**).

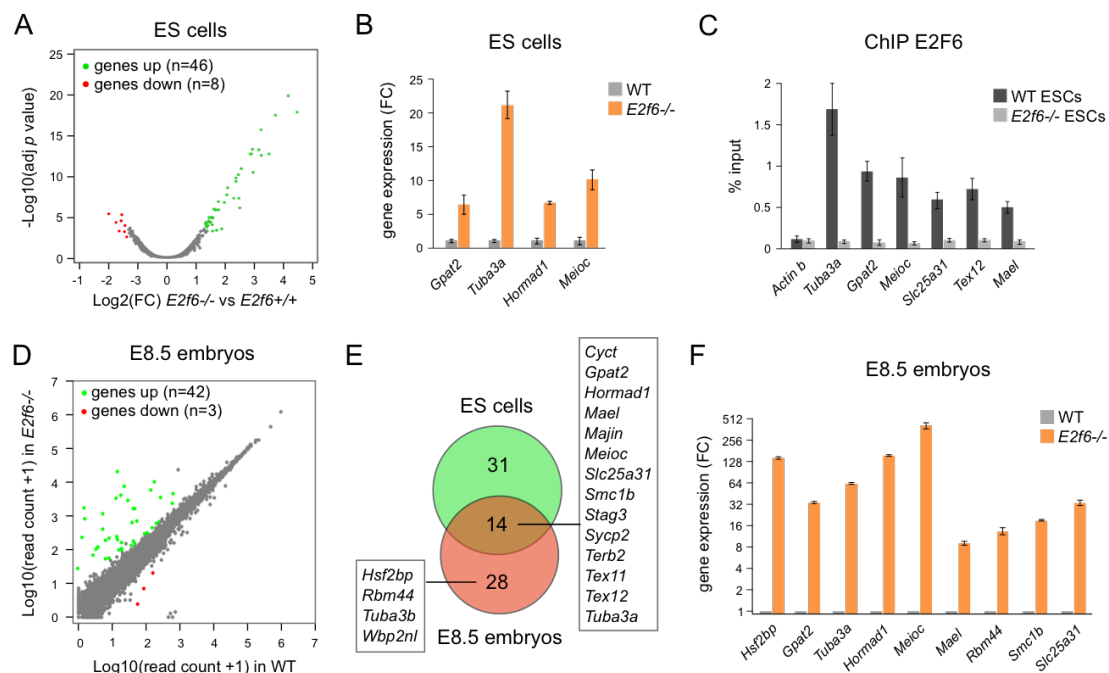


Figure 2. Identification of genes repressed by E2F6 in ES cells and embryos. **A.** Volcano plot representation of differential gene expression levels measured by Affymetrix microarrays in *E2f6*^{-/-} versus *E2f6*^{+/+} ES cells. Genes significantly upregulated or downregulated are highlighted in green and red, respectively. **B.** Validation of the derepression of E2F6 target genes by RT-qPCR in *E2f6*^{-/-} ES cells. Shown is the fold derepression relative to WT cells (mean \pm SEM, n = 3 independent replicates). **C.** ChIP-qPCR analysis of the binding of E2F6 in the promoters of E2F6 target genes in WT and *E2f6*^{-/-} ES cells (mean \pm SEM, n = 3 independent replicates). **D.** Scatter plot comparing the normalized RNA-seq read counts for all RefSeq genes in WT and *E2f6*^{-/-} embryos. Genes significantly upregulated or downregulated are highlighted in green and red, respectively. **E.** Venn diagram comparing the genes upregulated in *E2f6*^{-/-} ES cells and *E2f6*^{-/-} embryos. The names of germline genes found significantly upregulated in both *E2f6*^{-/-} ES cells and *E2f6*^{-/-} embryos or *E2f6*^{-/-} embryos only are shown. **F.** Validation of the derepression of E2F6 target genes by RT-qPCR in *E2f6*^{-/-} embryos (mean \pm SEM, n = 2 embryos).

This demonstrates that E2F6 plays a merely redundant role in the repression of most of its targets but is indispensable for silencing a subgroup of germline genes. These upregulated germline genes have strong E2F6 and PRC1.6 ChIP-seq signals in their promoters, suggesting that they are direct targets of E2F6 (**Figure S1E**). Furthermore, E2F6 binding was validated by ChIP followed by qPCR at all target germline gene promoters tested (**Figure 2C**).

To test if E2F6 plays similar functions *in vivo* in mouse embryos, we analyzed *E2f6* knockout mice (Courel et al. 2008). RNA-seq in *E2f6*^{-/-} and control embryos dissected at E8.5 identified 42 upregulated genes (fold change ≥ 2.5 ; adjusted p-value < 0.001) (**Figure 2D** and **Annexe 2.3 and 2.6**). These upregulated genes were enriched for GO terms related to germ cells (meiotic cell cycle $p = 2.08 \times 10^{-11}$; sexual reproduction $p = 1.51 \times 10^{-4}$; **Annexe 2.4**) and significantly overlapped with the genes upregulated in ES cells, revealing a core set of 14 germline genes repressed by E2F6 in ES cells and embryos whereas a small number of germline genes was found derepressed only in embryos (**Figure 2E**). The derepression of germline genes in *E2f6*^{-/-} embryos was validated by RT-qPCR (**Figure 2F**). Because previous studies showed that E2F6 interacts with G9A and G9A is required to repress a group of germline genes in mouse embryos (Ogawa et al. 2002; Auclair et al. 2016), we further asked if E2F6 and G9A share the same target genes. However, our results indicate that E2F6 and G9A repress distinct sets of genes in mouse embryos, indicating that E2F6 function is independent of G9A (**Figure S2**).

In summary, these results show that E2F6 is indispensable to repress a group of germline genes in ES cells and *in vivo*.

E2F6 cooperates with MAX/MGA for the silencing of germline genes

To investigate if E2F6 participates in PRC1.6 recruitment to germline genes, we performed ChIP-qPCR analysis of the PRC1.6 subunits RYBP and PCGF6 in *E2f6*^{+/+} and *E2f6*^{-/-} ES cells. We investigated six E2F6 target genes that either contain only an E2F6 binding motif in their promoter or both an E2F6 motif and E-box motif (**Figure 3A**). Like E2F6, the non-DNA-binding subunits RYBP and PCGF6 were detected on all E2F6 target gene promoters tested, albeit at low levels on the *Tuba3a* promoter (**Figure 3B**). Interestingly, RYBP and PCGF6 binding was reduced in *E2f6*^{-/-} ES cells specifically in those promoters lacking E-box motifs (**Figure 3B**, genes marked in grey). This suggests that E2F6 participates in the recruitment of PRC1.6 but that in the presence of E-boxes, the MAX/MGA subunits are sufficient to tether PRC1.6 in the absence of E2F6 protein. Against our expectations, the undisturbed presence of PRC1.6 on E-box containing E2F6 target genes in the absence of E2F6 indicates that on these promoters, E2F6-dependent gene repression is in part independent of PRC1.6 recruitment.

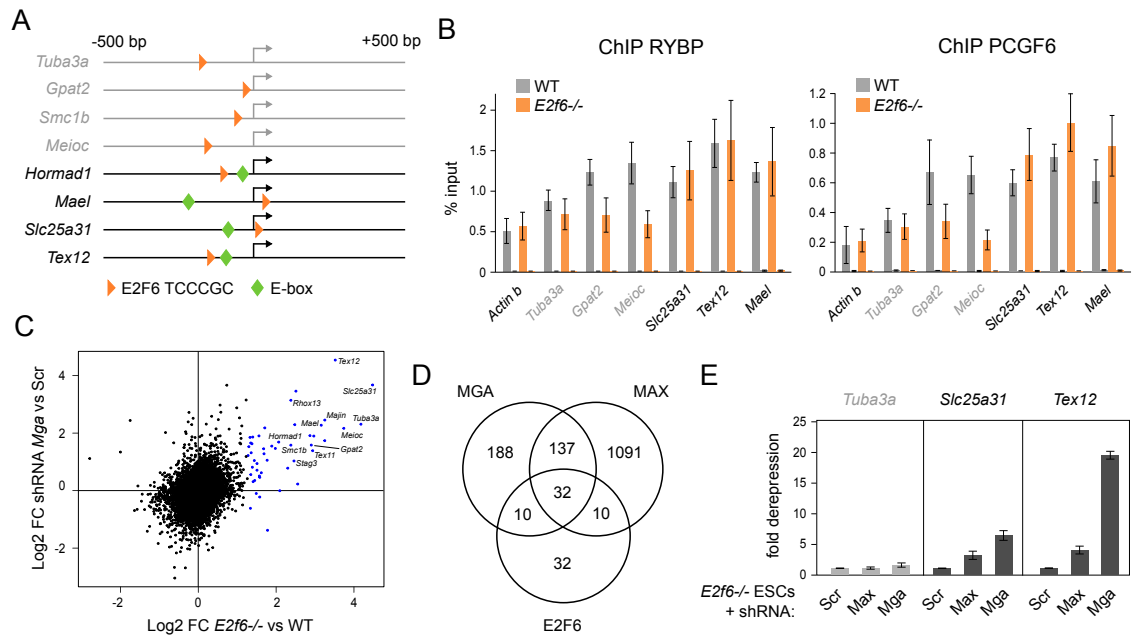


Figure 3. Cooperation between E2F6 and MAX/MGA in ES cells. **A.** Schematic representation of the position of E2F6 binding sites (orange triangles) and E-boxes (green diamonds) in the promoter regions (-500 to +500 bp relative to the start of transcription) of prominent E2F6 target genes. The arrow indicates the position of the TSS. The genes can be classified into two subgroups depending on the presence or absence of a promoter proximal E-box. The genes containing only an E2F6 motif are written in grey. **B.** ChIP-qPCR analysis of the binding of RYBP (left) and PCGF6 (right) to promoters of E2F6 target genes in WT and *E2f6*^{-/-} ES cells. The results are represented as the percentage of signal relative to the input (mean ± SEM, n= 3 independent experiments). For each condition, the values from the ChIP (left) and IgG control (right) are shown. The genes containing only an E2F6 motif are written in grey. **C.** Scatter plot comparing the changes in gene expression in *E2f6*^{-/-} and *Mga* knockdown ES cells, which reveals overlapping target gene specificities of E2F6 and MGA. The datapoints of genes repressed by E2F6 are highlighted in blue and the gene names of prominent E2F6 targets are indicated. **D.** Venn diagram illustrating the overlap of genes repressed by E2F6, MGA (both this study) and MAX (GSE45181) (Maeda et al. 2013) in ES cells. For this comparison genes with a FC > 2 were considered. **E.** MGA mediates E-box dependent repression of E2F6 target genes in the absence of E2F6. The graphs show RT-qPCR quantifications of the expression of E2F6 targets without (*Tuba3a*) or with E-boxes (*Slc25a31*, *Tex12*) in *E2f6*^{-/-} ES cells expressing an shRNA against MGA, MAX or a scramble control (shRNA Scr). The results are represented as a fold increase in expression relative to the scramble control (mean ± SD, n= 3 experiments).

The above results suggest that E2F6 and MAX/MGA, the other DNA binding subunits of PRC1.6, could cooperate for the regulation of germline genes. To address this question, we depleted MGA by shRNA in ES cells. Gene expression analysis of MGA shRNA-transduced ES cells identified 106 upregulated genes (fold change ≥2.5; adjusted p-value <0.01, **Annexe 2.5**). A comparative analysis revealed that most of the E2F6 target genes are concomitantly repressed by MGA, whereas MGA represses additional genes that are not repressed by E2F6 (**Figure 3C**). Additionally, we compared E2F6 and MGA target genes with published transcriptome data generated in MAX depleted ES cells (Maeda et al. 2013). Comparison of the genes repressed by E2F6, MGA and MAX revealed that they share 32 targets, and repression of 189 genes involves collaboration of at least two out of these three transcription factors (**Figure 3D**). GO analysis of the common target genes revealed that they

are enriched for GO terms related to meiotic cell cycle. Thus, E2F6 and MAX/MGA share a set of target genes related to germline functions in ES cells.

However, as previously shown in human somatic cells (Stielow et al. 2018), we observed that knock down of MGA in ES cells resulted in reduced levels of E2F6 protein as measured by western blotting (**Figure S3A**). This is likely due to protein destabilization at the post-transcriptional level because MGA knockdown does not lead to reduced levels of *E2f6* mRNA transcripts (**Figure S3B**), and treatment with the MG132 proteasome inhibitor slightly increases the levels of E2F6 protein in MGA-kd ES cells (**Figure S3A**). To circumvent this confounding effect and test if E2F6 and MAX/MGA synergize for the repression of germline genes, we depleted MAX or MGA by shRNA in *E2f6*^{-/-} cells. In line with our ChIP analysis, knock down of MGA in *E2f6*^{-/-} cells resulted in an additional derepression of the E-box containing genes *Slc25a31* and *Tex12*, but did not further affect the expression of the E-box lacking gene *Tuba3a* (**Figure 3E**). In comparison to MGA, the knock down of MAX had only a weak effect on the derepression of E-box containing genes (**Figure 3E**).

Altogether, these results show that MGA cooperates with E2F6 to silence germline genes with E-box containing promoters and that E2F6-dependent repression is partly decoupled from PRC1.6 recruitment.

E2F6 target gene specificity depends on its marked box domain

We went on to investigate the role of E2F6 domains in the repression of germline genes by performing rescue experiments in *E2f6*^{-/-} ES cells. Stable reexpression of wild type HA-tagged E2F6 in *E2f6*^{-/-} ES cells completely restored the repression of *Tuba3a* (**Figure 4A-B**). In contrast, a mutant version of E2F6 carrying a mutation of the DNA binding domain (E68) failed to bind to the *Tuba3a* promoter and rescue *Tuba3a* silencing (**Figure 4A-B**). Previous studies have shown that the marked box domains of E2F proteins mediate their target gene specificity (Black et al. 2005). To test the importance of the E2F6 marked box domain, we reexpressed a chimeric E2F6 harboring the marked box of E2F4 (E2F6-MB4, **Figure 4A**), an E2F family member that can compensate the loss of E2F6 in cell cycle regulation (Giangrande et al. 2004). This mutant E2F6 was able to bind to the *Tuba3a* promoter but failed to rescue *Tuba3a* repression (**Figure 4B**). Thus, repression of germline genes by E2F6 requires the DNA binding and marked box domains of E2F6.

During cell cycle progression, occupancy of E2F-responsive promoters by repressive and activating E2Fs is sequential and mutually exclusive (Cam and Dynlacht 2003). Our ChIP-seq analysis of E2F6 in non-synchronized ES cells confirmed that a substantial fraction of E2F6 peaks is co-bound by E2F1 (**Figure 4C**). However, a distinctive feature of genes actively repressed by E2F6 is reduced E2F1 binding, suggesting that E2F6 has the ability to specifically evict E2F1 from its target genes (**Figure 4C**). To test this hypothesis, we analyzed E2F1 binding by ChIP at the *Tuba3a* promoter. In

E2f6^{-/-} ES cells, the promoter of the de-repressed *Tuba3a* gene becomes occupied by E2F1 (**Figure 4D**). Reexpression of exogenous WT E2F6 but not the E68 mutant precluded E2F1 binding as efficiently as endogenous E2F6 in WT ES cells (**Figure 4D**), demonstrating that E2F6 efficiently outcompetes E2F1 for promoter binding at germline genes.

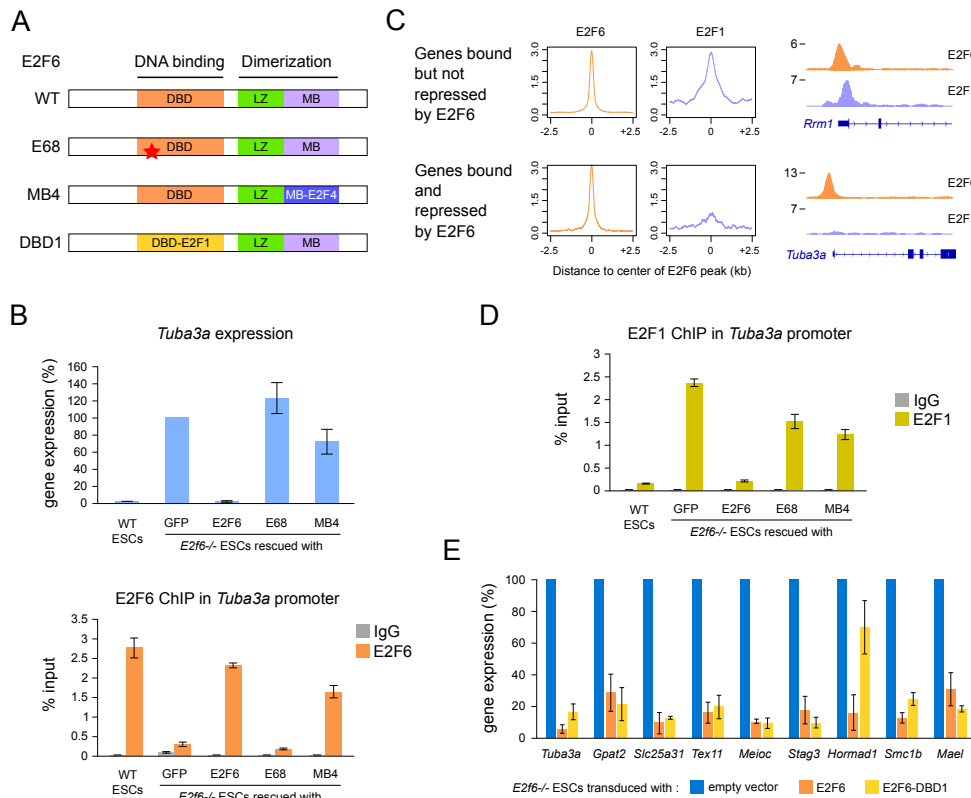


Figure 4. E2F6 function requires the DNA binding and Marked Box domains. **A.** Schematic representation of the E2F6 protein and the mutants generated in the study. The E68 mutant contains a point mutation in the DNA binding domain. DBD: DNA Binding Domain; LZ: Leucine Zipper; MB: Marked Box. **B.** RT-qPCR analysis of *Tuba3a* expression (top) and ChIP-qPCR analysis of E2F6 binding in the *Tuba3a* promoter (bottom) in WT ESCs and *E2f6*^{-/-} ESCs rescued with an empty vector (GFP), WT E2F6 or E68 and MB4 E2F6 mutants. *Tuba3a* expression is represented as a percentage of expression relative to the GFP control. For ChIP experiments, the values for the E2F6 ChIP and IgG control are shown for each condition. All values are mean \pm SEM (n=3). **C.** The metaplots on the left show E2F6 and E2F1 ChIP-seq read densities in E2F6 peaks linked to gene promoters not repressed by E2F6 (top) or repressed by E2F6 (bottom), illustrating that E2F6 repression is linked to reduced E2F1 binding. On the right are shown examples of E2F6 and E2F1 ChIP-seq profiles for the E2F cell-cycle target gene *Rrm1* and the E2F6-repressed gene *Tuba3a*. **D.** ChIP-qPCR analysis of E2F1 binding in the *Tuba3a* promoter in WT ESCs and *E2f6*^{-/-} rescued ESCs (mean \pm SEM, n=3). **E.** DNA binding domains of E2F6 and E2F1 are functionally redundant in mediating repression of E2F6 target genes. RT-qPCR analysis was performed in *E2f6*^{-/-} ES cells transduced with lentiviral vectors expressing wild type E2F6 or E2F6-DBD1. Cells transduced with an empty vector served as negative control. The expression levels are shown as a percentage relative to the negative control (mean \pm SD, n=3).

To investigate if the ability of E2F6 to evict E2F1 results from different DNA binding capabilities, we generated a swap-mutant of E2F6 (E2F6-DBD1) in which the DNA binding domain of E2F6 was replaced by the corresponding E2F1 sequence (**Figure 4A**). All attempts to establish stable transfected *E2f6*^{-/-} ES cell clones expressing E2F6-DBD1 failed and FACS analysis of ES cells pools transduced with lentiviral expression vectors revealed a strong counter selection against E2F6-DBD1 expressing cells (data not shown). However, expression analysis four days after lentiviral transduction into *E2f6*^{-/-} ES cells demonstrates that the DNA binding domains of E2F6 and E2F1 are functionally redundant in mediating repression of E2F6 target genes (**Figure 4E**). Therefore the observed competitive advantage of E2F6 over E2F1 at germline genes is not linked to its DNA binding domain and has to be attributed to other protein domains. In line with that, the chimeric E2F6 harboring the marked box of E2F4 (E2F6-MB4) was unable to outcompete E2F1 from the *Tuba3a* promoter (**Figure 4D**).

Altogether, this shows that E2F6 target gene specificity is linked to its ability to oppose activating transcription factors and critically depends on its marked box domain.

E2F6 recruits DNA methylation to germline genes during development

Given that germline genes undergo long term repression by DNA methylation in differentiated cells, we investigated the interplay between E2F6 and DNA methylation. First, we asked if E2F6 represses germline genes in early embryonic cells when DNA methylation has not yet been established. To this end, we measured the expression of E2F6 target genes in naïve *E2f6*^{+/+} and *E2f6*^{-/-} ES cells grown in 2i conditions, which show global DNA hypomethylation similarly to blastocysts (Ficz et al. 2013). Interestingly E2f6 target genes are strongly repressed by E2F6 in 2i ES cells (**Figure 5A**). To corroborate this finding *in vivo*, we measured the expression of the E2F6 target genes *Tuba3a* and *Hormad1* in *E2f6*^{+/+} and *E2f6*^{-/-} blastocysts, which confirmed that E2F6 represses its target genes in blastocysts (**Figure 5B**). This suggests that E2F6 initially represses germline genes by mechanisms other than DNA methylation in early embryonic cells before DNA methylation acquisition. One such mechanisms could be H3K9me3, which is strongly reduced at E2F6 target genes in 2i *E2f6*^{-/-} ES cells (**Figure S4**).

Next, we investigated if E2F6 is required for the deposition of DNA methylation at germline genes after implantation. To address this question, we performed reduced representation bisulfite sequencing (RRBS) in two *E2f6*^{+/+} and two *E2f6*^{-/-} E8.5 embryos (**Figure S5A**). While global genome methylation is unchanged in *E2f6*^{-/-} embryos (**Figure 5C**), DNA methylation is specifically reduced in the promoter of several E2F6 target germline genes (**Figure 5D** and **Figure S5B**), which was validated using as an alternative method combined bisulfite restriction analysis (COBRA) (**Figure S5C**). In line with our previous results, germline genes containing only the E2F6 binding site (*Tuba3a*, *Majin*,

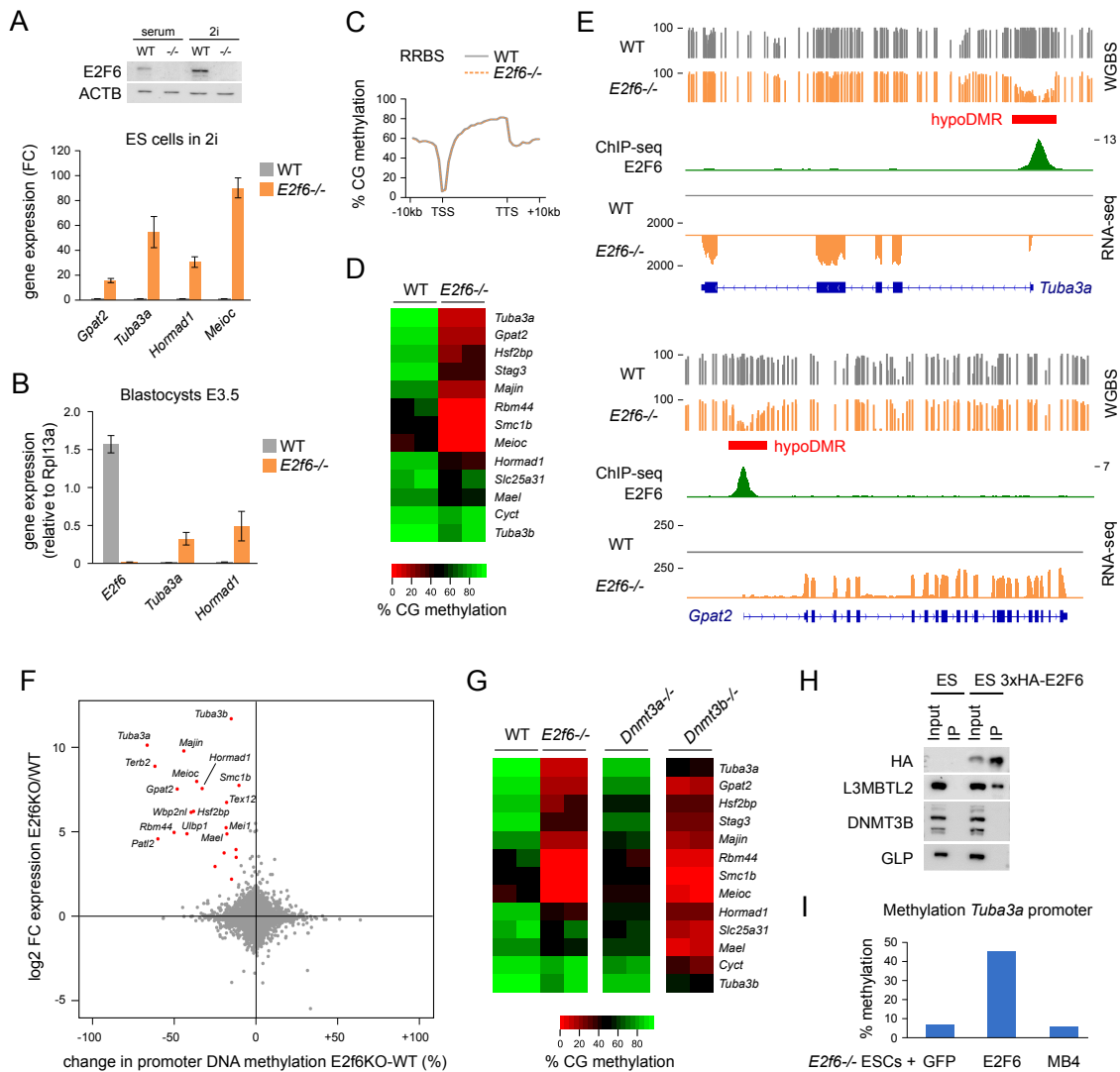


Figure 5. E2F6 recruits DNA methylation to germline genes during development. **A.** qPCR analysis of E2F6 target genes in WT and *E2f6*^{-/-} ES cells cultivated in 2i medium, represented as a fold change relative to the WT (mean ± SEM, n = 3 independent replicates). A Western blot showing E2F6 expression in ES cells cultivated in serum or 2i conditions is shown above the graph. **B.** qPCR analysis of the expression of *E2f6* and two E2F6 target genes in WT and *E2f6*^{-/-} blastocysts, represented as a value normalized to the expression of *Rpl13a* (mean ± SEM, n = 4 blastocysts). **C.** Metaplot showing the average percentage of CpG methylation profiled by RRBS over RefSeq genes and 10 kb of flanking sequences in WT and *E2f6*^{-/-} embryos. **D.** Heatmap representing the percentage of promoter CpG methylation measured by RRBS in E2F6 target germline genes in WT and *E2f6*^{-/-} embryos. Each column represents an independent embryo. **E.** Examples of genome tracks of WGBS, E2F6 Chip-seq and RNA-seq over two E2F6 target genes. Hypomethylated DMRs (hypoDMRs) overlapping the gene promoters are depicted by red rectangles. **F.** Comparison of the changes in gene expression and promoter DNA methylation (measured by WGBS in -500 to +500 bp from TSS) in *E2f6*^{-/-} compared to WT embryos for all genes. The genes that are simultaneously overexpressed (log₂ FC > 2) and hypomethylated (< -10%) are highlighted in red. **G.** Heatmap representing the levels of promoter DNA methylation of E2F6 target genes in E8.5 embryos knockout for *Dnmt3a* or *Dnmt3b*. The values in *E2f6*^{-/-} embryos are shown again as comparison. Each column represents an independent embryo. **H.** Co-immunoprecipitation experiments to investigate the interaction between E2F6 and DNMT3B. Immunoprecipitation was performed with an anti-HA antibody on nuclear extracts from control ES cells and ES cells expressing 3xHA-E2F6. Western blot was performed on the input and immunoprecipitates using anti-HA, anti-L3MBTL2, anti-DNMT3B and anti-GLP antibodies.

Gpat2, *Smc1b*, *Meioc*) show an almost complete reduction of DNA methylation, whereas genes containing an additional E-box (*Slc25a31*, *Hormad1*, *Mael*) show partial hypomethylation, suggesting that MAX/MGA partially compensate the absence of E2F6 for DNA methylation. Reduced DNA methylation at E2F6 target genes is also evident when RRBS was performed in *E2f6*^{+/+} and *E2f6*^{-/-} ES cells cultivated in serum+LIF conditions (**Figure S5B**).

To have a complete view of the impact of E2F6 on DNA methylation, we performed whole genome bisulfite sequencing (WGBS) in *E2f6*^{+/+} and *E2f6*^{-/-} E8.5 embryos (**Figure S5D**), which confirmed that DNA methylation is not affected globally but is locally reduced in the promoters of E2F6 target genes (**Figure 5E** and **Figure S5E-F**). Correlation of gene expression and promoter DNA methylation changes showed that most genes derepressed in *E2f6*^{-/-} embryos undergo concomitant promoter DNA hypomethylation (**Figure 5F**), indicating that E2F6 function is tightly linked to DNA methylation. WGBS identified 64 differentially methylated regions (DMRs) in *E2f6*^{-/-} compared to *E2f6*^{+/+} embryos, including 33 hypomethylated DMRs that contain many promoters of E2F6 repressed germline genes (**Annexe 2.1; Annexe 2.2** and **Annexe 2.7**). Hypomethylated DMRs colocalize with E2F6 ChIP-seq peaks and presumably reflect direct effects of E2F6 in recruiting DNA methylation, in contrast to hypermethylated DMRs that most likely result from indirect effects (**Figure S5G**).

To further investigate if DNA hypomethylation in *E2f6*^{-/-} embryos persists at later developmental stages, RRBS was performed in three tissues (brain, muscle, kidney) of *E2f6*^{-/-} and control adult mice. Promoter DNA hypomethylation of E2F6 target genes was detected in all adult tissues examined at levels similar to the ones in E8.5 embryos (**Figure S5H**), demonstrating that the lack of *de novo* methylation in *E2f6*^{-/-} embryos is not compensated throughout life.

To determine which DNA methyltransferase mediates E2F6-dependent DNA methylation, we compared DNA methylation of E2F6 target genes in *E2f6*^{-/-} embryos with embryos knockout for *Dnmt3a* or *Dnmt3b* (Auclair et al. 2014). E2F6 target genes are strongly hypomethylated in *Dnmt3b*^{-/-} but not *Dnmt3a*^{-/-} mouse embryos (**Figure 5G**), suggesting that DNMT3B is the main enzyme depositing DNA methylation at E2F6 target genes. To test if E2F6 interacts with the DNMT3B protein, we performed co-IP experiments with an HA-tagged version of E2F6 in ES cells. However we found no evidence for an interaction between E2F6 and DNMT3B (**Figure 5H**), suggesting that E2F6 does not stimulate DNA methylation by direct recruitment of DNMT3B. Nevertheless, rescue experiments in *E2f6*^{-/-} ES cells demonstrated that recruitment of DNA methylation by E2F6 depends on the marked box domain (**Figure 5I**).

Altogether, these results demonstrate that E2F6 is indispensable for recruiting DNA methylation-dependent silencing to its target genes during embryonic development.

E2F6 is dispensable for the maintenance of epigenetic silencing in differentiated cells

The general view is that CGI DNA methylation is not an initiating event in gene silencing but acts to lock in the silent state for long term silencing (Deaton and Bird 2011). We thus speculated that E2F6 is required to initiate epigenetic silencing in early development but not to maintain it in differentiated cells. To test this hypothesis, we performed ectopic inactivation of *E2f6* by CRISPR-Cas9 in mouse embryonic fibroblasts (MEFs) and compared these cells to constitutive *E2f6* knockout MEFs derived from *E2f6*^{-/-} embryos (**Figure 6A-B** and **S6A-B**). *E2f6*-deficient MEFs generated by CRISPR-Cas9 did not reactivate E2F6 target genes, in contrast to MEFs derived from *E2f6*^{-/-} embryos that showed a strong derepression of germline genes (**Figure 6C**). Furthermore, RRBS showed that whereas promoter DNA methylation of E2F6 target genes is strongly diminished in MEFs derived from *E2f6*^{-/-} embryos, DNA methylation of E2F6 target genes is unaffected in CRISPR-Cas9 *E2f6*-deficient MEFs (**Figure 6D**). In contrast to MEFs, ectopic inactivation of *E2f6* by CRISPR-Cas9 in ES cells leads to reactivation of germline genes (**Figure S6C-E**). Altogether these results demonstrate that E2F6 is necessary to initiate silencing in early embryonic cells but is replaced by other mechanisms such as DNA methylation at later stages and is not required for the maintenance of epigenetic silencing in differentiated cells. Further supporting this hypothesis, pharmacological inhibition of DNA methylation by 5-azadeoxycytidine in CRISPR-Cas9 *E2f6*-deficient MEFs restores expression of germline genes (**Figure S6F**).

Finally, to examine whether the silencing of germline genes can be rescued by E2F6 in differentiated cells, MEFs derived from *E2f6*^{-/-} embryos were transduced with a viral vector expressing E2F6 or E2F6 mutated in the DNA binding domain and cultivated for 30 days (**Figure 6E**). Reexpression of E2F6 in *E2f6*^{-/-} MEFs was unable to restore silencing of the E2F6 target genes even though it led to a modest decrease in the expression of some of the target genes (**Figure 6F**). Furthermore, restoring expression of E2F6 in *E2f6*^{-/-} MEFs did not induce DNA methylation of the E2F6 target genes even after 30 days of culture as demonstrated by COBRA in the *Tuba3a* and *Gpat2* promoters (**Figure 6G**). These results show that E2F6 is unable to initiate epigenetic silencing of its target germline genes in differentiated cells.

Taken together, these experiments show that E2F6 is required to initiate stable epigenetic silencing at a precise developmental window during early embryogenesis, but E2F6 becomes dispensable for maintaining the repressed state at later developmental stages (**Figure 6H**).

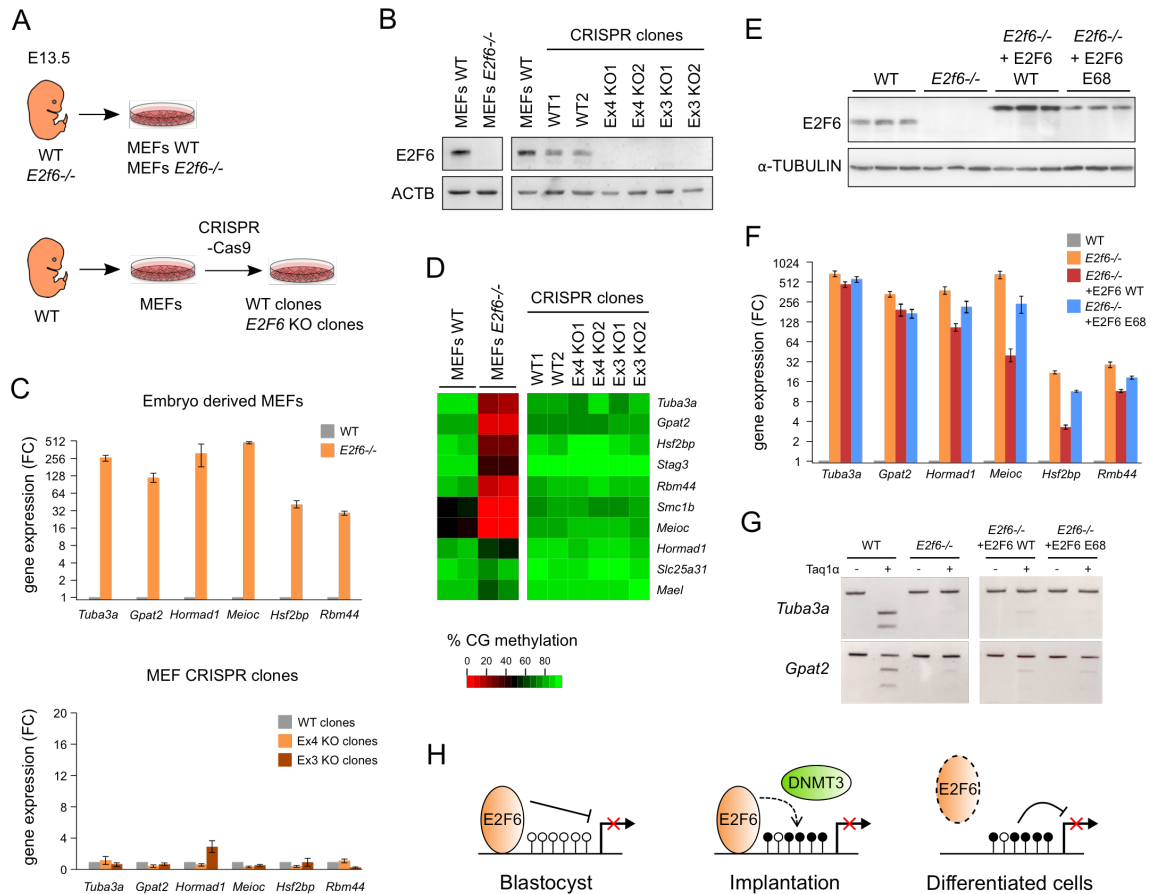


Figure 6. E2F6 is dispensable for the maintenance of epigenetic silencing. **A.** Experimental design. We either derived MEFs from WT and *E2f6*^{-/-} embryos, or generated *E2f6* KO clones and WT control clones by CRISPR-Cas9. **B.** Western blot of E2F6 in MEFs derived from WT and *E2f6*^{-/-} embryos and MEFs CRISPR-Cas9 clones. ACTB was used as a loading control. **C.** Expression of E2F6 target genes measured by qPCR in embryo derived WT and *E2f6*^{-/-} MEFs (top) and MEF CRISPR-Cas9 clones (bottom). The values are represented as the fold change compared to the WT controls (mean ± SEM, n=2 independent cell lines or clones per condition). **D.** Heatmap representing the percentage of CpG methylation profiled by RRBS in the promoters of E2F6 target genes in embryo derived WT and *E2f6*^{-/-} MEFs (left) and MEF CRISPR-Cas9 clones (right). **E.** Western blot of E2F6 in WT and *E2f6*^{-/-} MEFs and *E2f6*^{-/-} MEFs transduced with lentiviruses expressing E2F6 or E2F6 E68 fused to 3x-HA. α-TUBULIN was used as a loading control. **F.** Expression of E2F6 target genes measured by qPCR in WT and *E2f6*^{-/-} MEFs and *E2f6*^{-/-} MEFs rescued with E2F6. The values are represented as the fold change compared to the WT (mean ± SEM, n=3 biological replicates). **G.** Methylation analysis by COBRA of the *Tuba3a* and *Gpat2* promoters in WT and *E2f6*^{-/-} MEFs and *E2f6*^{-/-} MEFs rescued with E2F6. **H.** Model: E2F6 binds to and represses promoters of germline genes in preimplantation stages. At implantation, E2F6 recruits DNA methylation and initiates long term epigenetic silencing of its target germline genes, and subsequently becomes dispensable for the maintenance of the epigenetic repression in differentiated cells.

Discussion

Genome wide studies of DNA methylation during embryonic development revealed that CpG island promoters are constitutively protected from DNA methylation while *de novo* DNA methylation is specifically targeted to a small subset of promoter CGIs of germline genes (Borgel et al. 2010;

Smith et al. 2012; Auclair et al. 2014). However, the molecular pathways explaining this remarkable specificity are unknown. In this study, we extend previous work on the role of the transcription factor E2F6 (Pohlers et al. 2005; Storre et al. 2005; Kehoe et al. 2008; Velasco et al. 2010) and provide evidence that E2F6 is critical to recruit DNA methylation and initiate lifelong epigenetic silencing of a subset of germline genes during mouse development.

Using genetic inactivation in mouse ES cells and *in vivo*, we demonstrate that E2F6 is critical to silence a subgroup of germline genes. E2F6 is a known DNA-binding component of the PRC1.6 complex (Stielow et al. 2018) however, surprisingly, E2F6-dependent repression is in part independent of PRC1.6 recruitment because E2f6 is merely essential for transcriptional repression but dispensable for PRC1.6 recruitment to targets that harbor additional promoter proximal E-boxes. Thus, E2F6 has both PRC1.6 dependent and independent functions. Why E2F6 mediated repression is restricted to a small subset of E2F6 bound promoters associated with germline genes remains unclear. Germline genes contain conserved promoter proximal canonical E2F6 binding TCCGC motifs (Kehoe et al. 2008) but these are not specific for germ cell genes and occur in 1462 out of 2158 E2F6 ChIP-seq peaks. The promoter preference of E2F6 depends on its marked box, a protein interaction domain in E2F transcription factors that acts as key determinant of target gene specificity in E2F1 and E2F3 (Black et al. 2005), and therefore E2F6 function likely involves protein-protein interactions beyond simple DNA recognition. The main distinctive obvious feature that discriminates repressed and non-repressed genomic targets of E2F6 is concurrent high E2F6 occupancy accompanied by lack of E2F1 binding. Genes actively repressed by E2F6 are devoid of E2F1, whereas the vast majority of genomic E2F6 targets are bound by E2F6 and E2F1. This suggests that E2F6 prevents activation of germline genes by its unique ability to preclude binding of activating transcription factors at these genomic sites, presumably by direct binding site competition between E2F6 and E2F1 as has previously been described (Gaubatz et al. 1998; Yang et al. 2008).

Our developmental study showed that E2F6 has a dual function: it represses its target genes during the global hypomethylation phase in preimplantation phases and then is critical to recruit DNA methylation during global genome *de novo* methylation at implantation, thus favoring a transition from DNA methylation-independent to DNA methylation-dependent repression of germline genes in development. The molecular basis for DNA methylation recruitment by E2F6 deserve further investigation. We did not detect a direct interaction between E2F6 and DNMT3B in embryonic cells. Moreover mass spectrometry analysis of E2F6-containing PRC1.6 complexes so far failed to detect the *de novo* methyltransferases DNMT3A and DNMT3B (Ogawa et al. 2002; Trojer et al. 2011; Qin et al. 2012), arguing against a direct recruitment of the DNA methylation machinery by interaction with E2F6 (Velasco et al. 2010). Rather we favor the hypothesis that E2F6 induces a chromatin environment that indirectly enables DNA methylation recruitment. One hypothesis is that

E2F6 dependent H3K9me3 recruits DNA methylation, which is consistent with the H3K9 methyltransferase SETDB1 being required for DNA methylation and silencing of germline genes in mouse ES cells, including some of the most prominent E2F6 targets (*Tuba3a*, *Hormad1*, *Gpat2*) (Karimi et al. 2011; Leung et al. 2014). Alternatively, increasing evidence supports the view that activating transcription factors play a key role in the protection of CpG islands against DNA methylation (Brandeis et al. 1994; Gebhard et al. 2010; Krebs et al. 2014). As E2F6 has the unique property to evict E2F1 from the promoters of germline genes, this suggests that E2F6 could override the intrinsic protection of CGIs against DNA methylation by opposing transcriptional activators such as E2F1. Interestingly, E2F1 was previously linked to the protection against DNA methylation in oogenesis (Saadeh and Schulz 2014). There is also evidence that, in addition to transcriptional inactivity, DNA methylation requires the reconfiguration of nucleosome depleted regions (NDRs). As histone tails regulate DNA methylation by allosterically activating *de novo* methyltransferases (Ooi et al. 2007; Li et al. 2011), backfilling of NDRs could constitute a prerequisite for *de novo* methylation. Interestingly, according to mass spectrometry analysis (Qin et al. 2012), PRC1.6 is associated with the nucleosome remodeling and deacetylation (NuRD) complex and the DNA helicase LSH, both of which facilitate DNA methylation in mammalian cells (Xi et al. 2009; Termanis et al. 2016; Xia et al. 2017). Future work should be aimed at testing these hypotheses.

Some E2F6 target germline genes are only partially hypomethylated in *E2f6*^{-/-} embryos, suggesting that other pathways cooperate with E2F6 for DNA methylation recruitment at germline genes. Because these genes contain promoter-proximal E-boxes, it is likely that E2F6 and MGA, the two main DNA binding modules of PRC1.6, cooperate to induce epigenetic silencing of germline genes. This would predict that DNA methylation of germline genes is also impaired in absence of MGA, however the inability to maintain *Mga*^{-/-} ES cells and the early lethality of *Mga* knock-out embryos at preimplantation stages (Washkowitz et al. 2015) precludes a detailed study of the role of MGA in the epigenetic regulation of germline genes.

Once initiated, the epigenetic silencing of germline genes by E2F6 is stable and does not require the continued presence of E2F6. DNA methylation is a well-defined epigenetic mark mediating long-lasting gene repression and is one prime mediator of E2F6-induced stable silencing. However, our data suggest that E2F6 also induces long term silencing of germline genes by other mechanisms than DNA methylation. Indeed, some E2F6 target genes acquire only low levels of promoter CpG island methylation, which we have previously demonstrated does not have a strong repressive role (Auclair et al. 2014). For example, the *Smc1b* gene acquires 40% CpG methylation and is overexpressed 4-fold upon complete hypomethylation in *Dnmt3b*^{-/-} embryos, compared to a 245-fold derepression in *E2f6*^{-/-} embryos. This suggests that E2F6 induces long term silencing by mechanisms complementary to DNA methylation that remain to be investigated. These mechanisms

could involve polycomb-mediated H3K27 methylation, which is involved in silencing of the *Smc1b* gene in mouse ES cells (Leseva et al. 2013).

Our study contributes to explain how epigenetic repression is targeted with remarkable specificity to germline genes during embryonic development, providing a paradigm for how transient repressive signals by DNA-binding factors are translated into lifelong repressive epigenetic signals. This is reminiscent of KRAB-ZFPs transcription factors establishing stable epigenetic repression of TEs in development (Ecco et al. 2017), suggesting potential evolutionary convergences between the mechanisms that restrict expression of potentially harmful transposons and germline genes in mammalian somatic cells.

Material and Methods

Mouse lines. The *E2f6* knockout mouse line (Courel et al. 2008) was a gift from Jacqueline Lees (Cambridge, MA). Heterozygous mice were maintained in a C57BL/6J background and *E2f6*^{-/-} animals were generated by mating heterozygous males and females. We used WT embryos from the same litters as controls. The morning of the vaginal plug was designated as E0.5 and embryos were dissected at E8.5 in M2 medium. We simultaneously prepared genomic DNA and total RNA from whole single embryos with the AllPrep DNA/RNA Mini Kit (Qiagen).

ES cells. *E2f6*^{-/-} and *E2f6*^{+/+} ES cells were established from blastocysts of *E2f6*^{-/-} and control mice (Pohlers et al. 2005). These cells were cultured on gelatin coated plates in KO DMEM medium supplemented with 10% (v/v) KO serum replacement, 5% (v/v) FCS (Pan Biotech), 1,000 U/mL LIF, 1x non-essential amino acids, 2mM GlutaMAX, 1% penicillin/streptomycin and 0.1 mM 2-mercaptoethanol. ES cells were adapted in 2i conditions in DMEM F12 medium supplemented with 50% (v/v) Neurobasal medium, 1x N2 (Thermo Fisher Scientific), 1x B27 (Thermo Fisher Scientific), 2 mM L-Glutamine, 1% penicillin/streptomycin, 0.1 mM 2-mercaptoethanol, 0.5% BSA (Merck), 1,000 U/mL LIF, 1μM PD0325901 (Merck), 3μM Chir99021 (Merck), and 100 μg/mL Ascorbic Acid (Sigma). ES cell were cultured in 2i medium for 14 days. Mouse ES cell line E14TG2a was cultured on gelatin coated plates in DMEM GlutaMAX medium supplemented with 15% (v/v) FBS, 1,000 U/mL LIF, 0.1 mM non-essential amino acids, 2 mM L-Glutamine, 1% penicillin/streptomycin and 0.1 mM 2-mercaptoethanol. ES cells were passaged with Trypsin-EDTA 0.25% (Thermo Fisher Scientific) every 2 days. Cell pluripotency was assessed by measuring Oct4, Sox2, Nanog expression by RT-qPCR and using an alkaline phosphatase detection kit (Millipore). ES cells were tested negative for mycoplasmas.

Isolation and culture of MEFs. Mouse embryonic fibroblasts were isolated from E13.5 *E2f6*^{-/-} and WT embryos. Primary fibroblasts were immortalized by infecting cells with retroviruses encoding the Antigen T produced in HEK293 cells with the plasmid pBABE-neo largeTcDNA (Addgene #1780). MEFs were grown in DMEM supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin in a humidified atmosphere containing 5% CO₂ at 37°C. MEFs were passaged every 3 days and were tested negative for mycoplasmas.

Inactivation of *E2f6* by CRISPR-Cas9. *E2f6* KO cells were generated by deleting either exon 3 or exon 4 using two gRNAs flanking the exon. The gRNAs around exon 3 (gRNA1: CCCTCATTACTTCAGGATTG; gRNA2: TAGTGCTGCCACTGCGCTAA) and exon 4 (gRNA1: CTGAGAAATGTCGGTTCACA; gRNA2: CATGGTGTGGAAAACCTTAAC) were designed using <http://crispr.mit.edu>. The gRNAs were cloned in the CRISPR-Cas9 plasmids PX458 (pSpCas9(BB)-2A-GFP, Addgene #48138) and PX459 (pSpCas9(BB)-2A-Puro, Addgene #62988). Immortalized MEFs were transfected by the PX458 plasmid using the Neon electroporation system (Thermo Fisher Scientific). Cells expressing GFP were selected 24 hours post-transfection by flow cytometry and plated at a clonal density in 96 well plates. Mouse ES cells E14TG2a were transfected by the PX459 plasmid using Lipofectamine transfection reagent (Invitrogen) and selected for 72h by Puromycin (2µg/mL). Single ES colonies were selected and plated in 96 well-plates. The clones were genotyped by PCR on genomic DNA isolated with the AllPrep DNA/RNA Mini Kit (Qiagen) using primers in the exon 3 or exon 4. We selected WT and *E2f6* KO clones that were further validated by Sanger sequencing, western blotting and RT-qPCR using primers spanning the exons 3 and 4 (see **Annexe 2.8**). For western blotting, whole cell extracts were prepared in RIPA lysis buffer (Clinisciences).

5-AzadC experiments. MEF clones were treated with 5-Aza-2'-Deoxycytidine (A3656, Sigma) at a concentration of 1 µM for 96h with medium renewal every day. After 96h, total RNA was isolated for RT-qPCR analysis.

shRNA knockdown of MGA and MAX in ES cells. Lentiviral shRNA knockdown constructs were generated by cloning synthetic oligos corresponding to MGA shRNA (TRCN0000082083), MAX shRNA (TRCN0000039867) and scramble control of the Genetic Perturbation Platform (Broad Institute) into the pLKO.1 puro shRNA expression vector (Addgene #8453). Knockdown was performed by infecting *E2f6*^{+/+} or *E2f6*^{-/-} ESCs with lentiviral particles produced in HEK293T cells in the presence of 8 µg/ml polybrene. 24h after transduction, ES cells were selected with puromycin (2µg/ml) and RNA was harvested after additional 48h of selection.

Expression vectors for ES cells. For stable transfection experiments in ES cells, the sequences coding for E2F6 WT as well as E68 and MB4 mutants were inserted into the pCAG-EGFP-IB vector (Meilinger et al. 2009). To generate the E68 DNA binding mutant, the amino acids in positions 68 and 69 were changed from Leu-Val to Glu-Ser (Gaubatz et al. 1998). To generate the MB4 mutant, the E2F6 marked box domain (aa 180-240) was replaced by the marked box domain of E2F4 (aa 138-198). These constructs were transfected into ES cells by *electroporation* with the Amaxa Nucleofector II device using the *Amaxa Cell Line Nucleofector kit R*, and stable transfectants were selected with *blastidicin* (7.5 $\mu\text{g/ml}$). For the DNA binding domain swap experiment, the sequences coding for E2F6 WT and DBD1 mutant were inserted in the lentiviral expression vector pCDH-MSCV-MCS-EF1 α -GFP+Puro (System Biosciences). The DBD1 mutant was generated by replacing the DNA binding domain of E2F6 (aa 61-129) with the DNA binding domain of E2F1 (aa 122-187). Viral particles were produced in HEK293T cells and ES cells were infected with 8 $\mu\text{g/ml}$ polybrene. 24h after transduction, puromycin (2 $\mu\text{g/ml}$) was added for 72h before harvesting RNA samples.

Expression vectors for MEFs. The sequences coding for E2F6 WT and E68 fused to 3x-HA were cloned into the pMSCV Puro-IRES-GFP retroviral vector (Addgene #21654) with the In-Fusion HD cloning plus kit (Takara). Viral particles produced in HEK293T cells were used to infect immortalized *E2f6*^{-/-} MEFs with 8 $\mu\text{g/ml}$ polybrene. 24 h after infection, the cells were selected with Puromycin (2 $\mu\text{g/ml}$) for 48h. The cells were harvested 15 days after the infection for protein, DNA and RNA extraction.

Transcriptome analysis by microarrays. Microarray analysis was performed in biological triplicates. RNAs were extracted using RNeasy kits (Qiagen). RNA quality was verified on a 2100 Bioanalyzer (Agilent). 100 to 300ng of total RNA were labelled with the Two-Cycle cDNA synthesis kit (Affymetrix). Subsequently, 5.5 μg of single-strand DNA was fragmented with a combination of uracil DNA glucosylase (UDG) and apurinic/apyrimidinic endonuclease 1 (APE1) and then labeled using the WT terminal labeling kit (Affymetrix). Samples were hybridized to GeneChip Mouse Genome 430 2.0 arrays, washed and scanned at the LFGC Core Facility (Charité) using the Hybridization Wash and Stain Kit (Affymetrix) as described in the Affymetrix GeneChip Expression Analysis Technical Manual. The hybridization cocktail contained 50pM control oligonucleotide B2 and eukaryotic hybridization controls (bioB bioC bioD 1, 5, 25 pM respectively). The arrays were hybridized in a Genechip hybridization oven 640 (Affymetrix) at 45°C and 60 rpm for 16hrs, washed and stained using a GeneChip Fluidic Station 450s (Affymetrix) and then scanned with a GeneChip Scanner 3000 7G System (Affymetrix). MicroArray data were analyzed with the affy and limma Bioconductor packages.

RMA of the affy package was used for processing the CEL files and limma was used for assessing differential expression.

Transcriptome analysis by RNA-seq. RNA-seq libraries were prepared from single embryos using the TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero depletion (Illumina), and sequenced in paired-end 2x100bp on an Illumina HiSeq 4000. Reads were mapped using TopHat v2.0.13 with a RefSeq transcriptome index and counted in RefSeq genes with HTSeq v0.7.2 (parameters `-t exon -s reverse`). Differentially expressed genes (fold change > 2.5, adjusted p-value < 0.001) were identified using DESeq2.

Methylome analysis by RRBS and WGBS. RRBS libraries were prepared from 50-100 ng genomic DNA and sequenced in paired-end 2x75bp on an Illumina HiSeq4000 at Integragen SA (Evry, France). The preparation of RRBS libraries and processing of sequencing reads were performed as described (Auclair et al. 2016). For WGBS, 100 ng genomic DNA was fragmented to 350 bp using a Covaris E220 sonicator, and bisulfite-converted with the EZ DNA Methylation-Gold kit (Zymo Research). WGBS libraries were prepared using the Accel-NGS Methyl-Seq DNA Library Kit (Swift Biosciences) according to the manufacturer's instructions with 6 PCR cycles for the final amplification. The libraries were purified using Ampure XP beads (Beckman Coulter) and sequenced in paired-end 2x100 bp with an Illumina HiSeq4000 at Integragen SA (Evry, France). The sequencing reads were processed as follows: the reads were trimmed to remove low quality bases and the first five bases of reads R1 and ten bases of reads R2 using Trim Galore v0.4.2 (parameters `-q 20 --clip_R1 5 --clip_R2 10`), and aligned to the mouse mm10 genome using Bismark v0.18.2 with default parameters. Methylation calls were extracted using Bismark v0.18.2 and only CpGs covered by more than 5 reads were retained for analyses. Metaplots of methylation over genes were generated by calculating the percentage of CpG methylation in twenty equal-sized windows within each RefSeq gene and ten 1 kb windows of flanking sequences. DMRs were identified using eDMR from the methylKit R package with a minimum of 7 differentially methylated CpGs (DMCs), a difference in methylation > 20% and a q-value < 0.001.

COBRA. 100 ng of genomic DNA was bisulfite-converted with the EpiTect bisulfite kit (Qiagen). The target regions were amplified by touchdown PCR with the Platinum Taq DNA Polymerase (Thermo Fisher Scientific) using the following conditions: 20 cycles of 30s at 95°C, 30s at 60–50°C (with a 0.5°C decrease per cycle), 50s at 72°C followed by 35 cycles of 30s at 95°C, 30s at 50°C, 50s at 72°C. The PCR products were purified using the PCR cleanup kit (Macherey Nagel). 40 ng of PCR product were

digested by Taq1 α (Thermo Fisher Scientific) and loaded on an agarose gel alongside 40 ng of undigested PCR product. The primers are provided in the **Annexe 2.8**.

Pyrosequencing. Genomic DNA was bisulfite-converted with the EpiTect bisulfite kit (Qiagen). The target region was amplified by PCR using the ZymoTaq PreMix (Epigenie) and pyrosequencing was performed by Varionostic (Ulm, Germany). The primers are provided in the **Annexe 2.8**.

ChIP of histone modifications. ES Cells were crosslinked in PBS with 1% formaldehyde for 10 min at RT and quenched with 125 mM glycine for 5 min at RT. Cells were lysed in a buffer containing 0.1% SDS, 1% Triton X-100, 1 mM EDTA, 50 mM HEPES-KOH, 140 mM NaCl, 0.1% sodium deoxycholate, 1x protease inhibitor cocktail (Santa Cruz). The chromatin was sonicated with a Bioruptor (Diagenode) at 4°C. Antibodies were pre-bound to protein A-coated magnetic beads (Diagenode) and incubated overnight at 4°C with chromatin from 200,000 cells per ChIP reaction. We used antibodies against H3K9me3 (ab8898, Abcam), H3 (07-690, Merck), and IgG2A as a control (MAB0031, R&D Systems). Chromatin was successively washed in three buffers containing 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris pH8, 150 mM NaCl, then 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris pH8, 500 mM NaCl and 0.25 M LiCl, 1% NP40, 1% sodium deoxycholate, 10 mM Tris pH8, 1 mM EDTA. Chromatin was decrosslinked for 4h at 65°C and then treated with proteinase K (Euromedex) for 1h at 56°C. DNA was isolated by Phenol-Chloroform extraction and ethanol precipitation for qPCR analysis. The primers are provided in the **annexe 2.8**.

ChIP of E2F6, E2F1 and PRC1.6. ES Cells were crosslinked in PBS with 1% formaldehyde for 15 min on ice and quenched with 125 mM glycine for 5 min. Cells were lysed in 50 mM Tris pH8, 10 mM EDTA, 1 % SDS with 1x cOmplete Mini protease inhibitor cocktail (Sigma Aldrich) and sonicated with a Bioruptor (Diagenode) at 4°C. 80 μ g of chromatin were used per ChIP reaction and incubated with antibodies pre-bound to Dynabeads protein G (Thermo Fischer Scientific) overnight at 4°C. The following antibodies were used: E2F6 (ab53061, Abcam), PCGF6 (24103-1-AP, Proteintech), RYBP (AB3637, Millipore), E2F1 (C-20, Santa Cruz Biotechnology) and as control normal mouse IgG (Millipore). The beads were washed successively in low-salt-buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris pH8, 150 mM NaCl), high-salt-buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris pH8, 500 mM NaCl), LiCl-buffer (0.25 M LiCl, 1% NP40, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris pH8) and TE (10 mM Tris pH8, 1 mM EDTA). Bound chromatin was eluted with elution-buffer (1% SDS, 0.1 M NaHCO₃) and crosslinking was reverted at 62°C overnight in the presence of proteinase K (Qiagen). DNA was purified with a PCR Purification Kit (Qiagen) and qPCR

analysis was performed using a 7500 Fast qPCR system (Applied Biosystems). The primers are provided in the **Annexe 2.8**.

ChIP-seq analysis. E2F6 ChIP-seq libraries were prepared using the DNA sample kit (Illumina) following the manufacturer's instructions. After adapter ligation, the DNA was PCR amplified with Illumina primers for 15 cycles and library fragments of ~250 bp were band isolated from an agarose gel. Libraries were sequenced with an Illumina HiSeq 1500 following the manufacturer's protocols. For all ChIP-seq datasets used in the study, reads were aligned to the mouse reference genome mm10 using bowtie2 and only reads mapping with a quality above 10 were kept using samtools. Read density tracks were generated using genomeCoverageBed from Bedtools from reads extended to 200bp and visualized using the IGV browser. Peaks were called using Peakzilla (Bardet et al. 2013) with default settings. The analysis of motif occurrence and distribution of motifs in peaks was performed using TFmotifView (<http://bardet.u-strasbg.fr/tfmotifview>). Known TF motifs from the JASPAR CORE 2018 database were used to scan the mouse mm10 genome with MAST using each motif information content as p-value threshold. Regions of 151bp around peak summits were overlapped with motif occurrences using intersectBed. Shuffled control regions were generated using shuffledBed and the statistical significance of the motif enrichment in peak regions compared to controls was assessed using the hypergeometric p-value in R. The overlap of E2F6, MGA and L3MBTL2 peaks was generated with the findOverlapsOfPeaks function from the CHIPpeakAnno package in R with a maximum gap of 500 bp.

RT-qPCR. Total RNAs were reverse transcribed into cDNA with the Maxima first strand cDNA synthesis kit (Thermo Fischer Scientific). Quantitative PCR was performed with the KAPA SYBR FAST qPCR kit (KAPA Biosystems) on a StepOnePlus realtime PCR system (Applied Biosystems). The expression of target genes was normalized with one or several housekeeping genes (*Actin b*, *Gusb*, *Rpl13a* and *Mrp132*). No-RT controls were also amplified to exclude the presence of contaminating DNA. The primer sequences for qPCR are provided in the **Annexe 2.8**.

Gene expression analysis in single blastocysts. Blastocysts were isolated from uterine tubes in M2 medium and washed in PBS 1X droplets. Single blastocysts were collected in microtubes in 30 µL of H2O supplemented with RNase inhibitor (Promega) and lysed by performing 5 rounds of freezing in liquid nitrogen and thawing at 37°C for 5 min. RNAs were reverse transcribed into cDNA with the Maxima first strand cDNA synthesis kit (Thermo Fischer Scientific) for qPCR analysis of target genes and *Rpl13a* as a normalization control. The blastocysts were genotyped by expression of the *E2f6* transcript.

Co-immunoprecipitation experiments. ES cell line E14TG2a was transfected with the pCAG-3xHA-E2F6 vector and selected with blasticidine at 0.01 µg/ µL until stable integration of the plasmid. Cell nuclear extracts were prepared as previously described (Gagnon et al. 2014). 1 mg of nuclear extract were incubated with anti-HA magnetic beads (Pierce) overnight at 4°C. The complexes bead-proteins were washed in a buffer containing 15 mM of Tris pH8, 150 mM NaCl, 0.05% Triton. Proteins were eluted in Laemmli buffer (Thermo Fisher Scientific). 25% of eluted proteins and 15 µg of input proteins were loaded on a 10% SDS-polyacrylamide gel and assessed by western blotting.

Western blotting. Proteins were run on SDS-PAGE gels and transferred to 0.2µm nitrocellulose membranes using the Trans-blot Turbo Blotting System (Biorad). The membranes were blocked with 5% milk in TBS-0.05% tween for 1h at RT, incubated with the primary antibodies and then with horseradish peroxidase-conjugated secondary antibodies followed by chemiluminescence detection using the ECL detection reagent (Amersham, GE Healthcare). We used the following antibodies: E2F6 (LLF6-2, Kerafast and ab53061, Abcam), ACTB (A2066, Sigma), α-TUBULIN (T9026, Sigma), β-TUBULIN (Sigma Aldrich), RYBP (AB3637, Millipore), Ubiquitin (Z0458, Dako), L3MBTL2 (39570, Active Motif), PCGF6 (24103-1-AP, Proteintech), MGA (PA5-59934, Thermo Fisher Scientific), GLP (PP-B0422-00, R&D Systems), DNMT3B (NB100-56514, NovusBio), HA (3F10, Roche).

Datasets and data availability. The following datasets were used in the study: ChIP-seq of KDM2B (GSE55698), EZH2 (GSE74330), SUZ12 (GSE74330), RING1B (GSE74330), G9A (GSE46536), RYBP (GSE42466), PCGF6 (PRJEB66757), L3MBTL2 (PRJEB66757), MGA (PRJEB66757), MAX (GSE48666 and GSE48175), ZFX (GSE11431), E2F1 (GSE11431), KLF4 (PRJEB9797), CTCF (GSE30206), TIP60 (GSE48666), DMAP1 (GSE48666), MYC (GSE48666), TET1 (GSE24841), ESRRB (GSE11431), DAX1 (GSE48666), OCT4 (PRJEB9797), NANOG (GSE11431), SOX2 (GSE81168), transcriptome of Max Kd ES cells (GSE45181), RNA-seq of *G9a*^{-/-} embryos (GSE71500). The sequencing data generated in this study are deposited in the NCBI Gene Expression Omnibus (GEO) under the accession number GSEXXXXX.

Acknowledgments

We are grateful to Dr Jacqueline Lees for providing *E2f6* knockout mice. This work was supported by grants from the European Research Council (ERC Consolidator grant no. 615371 to Michael Weber), the Institut National du Cancer (INCa), and the Deutsche Forschungsgemeinschaft (DFG, grant no. 63465200 to Christian Hagemeier). Thomas Dahlet was recipient of a Doctoral fellowship from the French Ministry for Higher Education and Research.

References

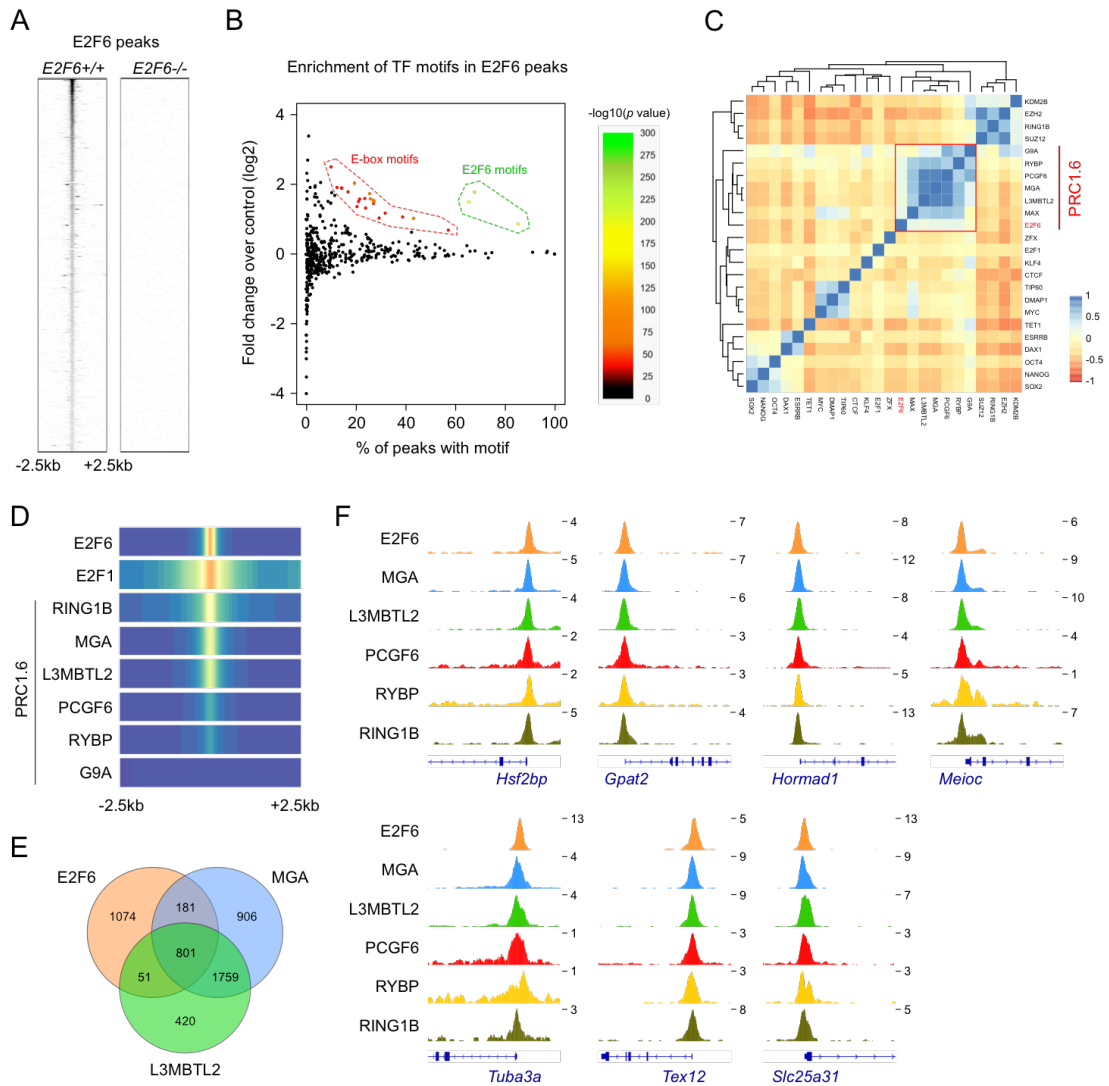
- Auclair G, Borgel J, Sanz LA, Vallet J, Guibert S, Dumas M, Cavelier P, Girardot M, Forne T, Feil R et al. 2016. EHMT2 directs DNA methylation for efficient gene silencing in mouse embryos. *Genome Res* **26**: 192-202.
- Auclair G, Guibert S, Bender A, Weber M. 2014. Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. *Genome Biol* **15**: 545.
- Bardet AF, Steinmann J, Bafna S, Knoblich JA, Zeitlinger J, Stark A. 2013. Identification of transcription factor binding sites from ChIP-seq data at high resolution. *Bioinformatics* **29**: 2705-2713.
- Bertoli C, Klier S, McGowan C, Wittenberg C, de Bruin RA. 2013. Chk1 inhibits E2F6 repressor function in response to replication stress to maintain cell-cycle transcription. *Curr Biol* **23**: 1629-1637.
- Black EP, Hallstrom T, Dressman HK, West M, Nevins JR. 2005. Distinctions in the specificity of E2F function revealed by gene expression signatures. *Proc Natl Acad Sci U S A* **102**: 15948-15953.
- Borgel J, Guibert S, Li Y, Chiba H, Schubeler D, Sasaki H, Forne T, Weber M. 2010. Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet* **42**: 1093-1100.
- Boulard M, Edwards JR, Bestor TH. 2015. FBXL10 protects Polycomb-bound genes from hypermethylation. *Nat Genet* **47**: 479-485.
- Boumber YA, Kondo Y, Chen X, Shen L, Guo Y, Tellez C, Estecio MR, Ahmed S, Issa JP. 2008. An Sp1/Sp3 binding polymorphism confers methylation protection. *PLoS Genet* **4**: e1000162.
- Brandeis M, Frank D, Keshet I, Siegfried Z, Mendelsohn M, Nemes A, Temper V, Razin A, Cedar H. 1994. Sp1 elements protect a CpG island from de novo methylation. *Nature* **371**: 435-438.
- Cam H, Dynlacht BD. 2003. Emerging roles for E2F: beyond the G1/S transition and DNA replication. *Cancer Cell* **3**: 311-316.
- Courel M, Friesenhahn L, Lees JA. 2008. E2f6 and Bmi1 cooperate in axial skeletal development. *Dev Dyn* **237**: 1232-1242.
- Deaton AM, Bird A. 2011. CpG islands and the regulation of transcription. *Genes Dev* **25**: 1010-1022.

- Ecco G, Imbeault M, Trono D. 2017. KRAB zinc finger proteins. *Development* **144**: 2719-2729.
- Endoh M, Endo TA, Shinga J, Hayashi K, Farcas A, Ma KW, Ito S, Sharif J, Endoh T, Onaga N et al. 2017. PCGF6-PRC1 suppresses premature differentiation of mouse embryonic stem cells by regulating germ cell-related genes. *Elife* **6**.
- Farcas AM, Blackledge NP, Sudbery I, Long HK, McGouran JF, Rose NR, Lee S, Sims D, Cerase A, Sheahan TW et al. 2012. KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *Elife* **1**: e00205.
- Ficz G, Hore TA, Santos F, Lee HJ, Dean W, Arand J, Krueger F, Oxley D, Paul YL, Walter J et al. 2013. FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* **13**: 351-359.
- Gagnon KT, Li L, Chu Y, Janowski BA, Corey DR. 2014. RNAi factors are present and active in human cell nuclei. *Cell Rep* **6**: 211-221.
- Gaubatz S, Wood JG, Livingston DM. 1998. Unusual proliferation arrest and transcriptional control properties of a newly discovered E2F family member, E2F-6. *Proc Natl Acad Sci U S A* **95**: 9190-9195.
- Gebhard C, Benner C, Ehrich M, Schwarzfischer L, Schilling E, Klug M, Dietmaier W, Thiede C, Holler E, Andreesen R et al. 2010. General transcription factor binding at CpG islands in normal cells correlates with resistance to de novo DNA methylation in cancer cells. *Cancer Res* **70**: 1398-1407.
- Giangrande PH, Zhu W, Schlisio S, Sun X, Mori S, Gaubatz S, Nevins JR. 2004. A role for E2F6 in distinguishing G1/S- and G2/M-specific transcription. *Genes Dev* **18**: 2941-2951.
- Hackett JA, Reddington JP, Nestor CE, Dunican DS, Branco MR, Reichmann J, Reik W, Surani MA, Adams IR, Meehan RR. 2012. Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development* **139**: 3623-3632.
- Hauri S, Comoglio F, Seimiya M, Gerstung M, Glatter T, Hansen K, Aebersold R, Paro R, Gstaiger M, Beisel C. 2016. A High-Density Map for Navigating the Human Polycomb Complexome. *Cell Rep* **17**: 583-595.

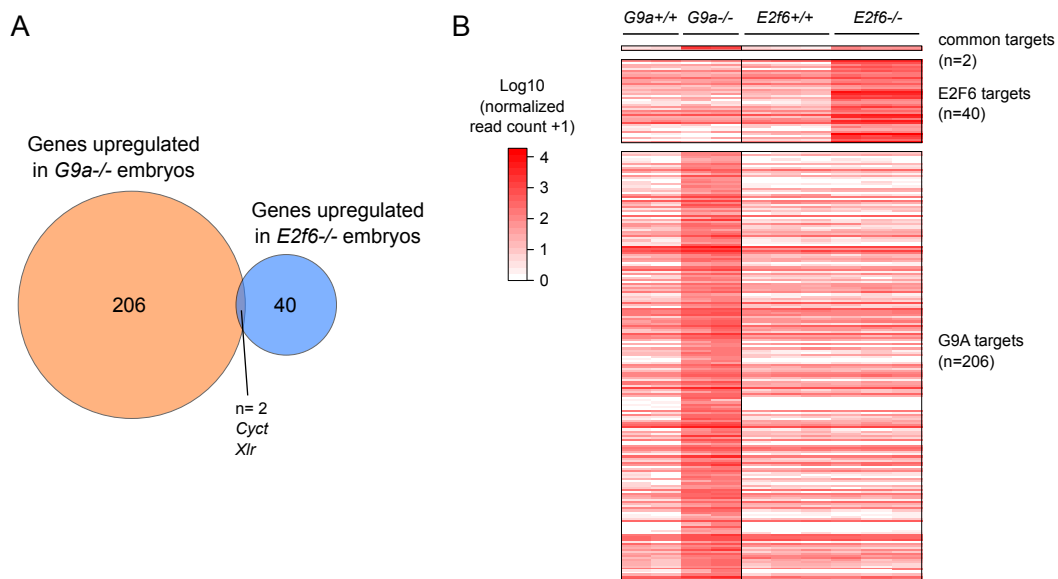
- Hisada K, Sanchez C, Endo TA, Endoh M, Roman-Trufero M, Sharif J, Koseki H, Vidal M. 2012. RYBP represses endogenous retroviruses and preimplantation- and germ line-specific genes in mouse embryonic stem cells. *Mol Cell Biol* **32**: 1139-1149.
- Karimi MM, Goyal P, Maksakova IA, Bilenky M, Leung D, Tang JX, Shinkai Y, Mager DL, Jones S, Hirst M et al. 2011. DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell Stem Cell* **8**: 676-687.
- Kehoe SM, Oka M, Hankowski KE, Reichert N, Garcia S, McCarrey JR, Gaubatz S, Terada N. 2008. A conserved E2F6-binding element in murine meiosis-specific gene promoters. *Biol Reprod* **79**: 921-930.
- Kloet SL, Makowski MM, Baymaz HI, van Voorthuijsen L, Karemaker ID, Santanach A, Jansen P, Di Croce L, Vermeulen M. 2016. The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. *Nat Struct Mol Biol* **23**: 682-690.
- Krebs AR, Dessus-Babus S, Burger L, Schubeler D. 2014. High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions. *Elife* **3**: e04094.
- Leseva M, Santostefano KE, Rosenbluth AL, Hamazaki T, Terada N. 2013. E2f6-mediated repression of the meiotic Stag3 and Smc1beta genes during early embryonic development requires Ezh2 and not the de novo methyltransferase Dnmt3b. *Epigenetics* **8**: 873-884.
- Leung D, Du T, Wagner U, Xie W, Lee AY, Goyal P, Li Y, Szulwach KE, Jin P, Lorincz MC et al. 2014. Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1. *Proc Natl Acad Sci U S A* **111**: 6690-6695.
- Li BZ, Huang Z, Cui QY, Song XH, Du L, Jeltsch A, Chen P, Li G, Li E, Xu GL. 2011. Histone tails regulate DNA methylation by allosterically activating de novo methyltransferase. *Cell Res* **21**: 1172-1181.
- Macleod D, Charlton J, Mullins J, Bird AP. 1994. Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes Dev* **8**: 2282-2292.
- Maeda I, Okamura D, Tokitake Y, Ikeda M, Kawaguchi H, Mise N, Abe K, Noce T, Okuda A, Matsui Y. 2013. Max is a repressor of germ cell-related gene expression in mouse embryonic stem cells. *Nat Commun* **4**: 1754.

- Meilinger D, Fellingner K, Bultmann S, Rothbauer U, Bonapace IM, Klinkert WE, Spada F, Leonhardt H. 2009. Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. *EMBO Rep* **10**: 1259-1264.
- Morkel M, Wenkel J, Bannister AJ, Kouzarides T, Hagemeyer C. 1997. An E2F-like repressor of transcription. *Nature* **390**: 567-568.
- Ogawa H, Ishiguro K, Gaubatz S, Livingston DM, Nakatani Y. 2002. A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science* **296**: 1132-1136.
- Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD et al. 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**: 714-717.
- Pohlers M, Truss M, Frede U, Scholz A, Strehle M, Kuban RJ, Hoffmann B, Morkel M, Birchmeier C, Hagemeyer C. 2005. A role for E2F6 in the restriction of male-germ-cell-specific gene expression. *Curr Biol* **15**: 1051-1057.
- Qin J, Whyte WA, Anderssen E, Apostolou E, Chen HH, Akbarian S, Bronson RT, Hochedlinger K, Ramaswamy S, Young RA et al. 2012. The polycomb group protein L3mbtl2 assembles an atypical PRC1-family complex that is essential in pluripotent stem cells and early development. *Cell Stem Cell* **11**: 319-332.
- Saadeh H, Schulz R. 2014. Protection of CpG islands against de novo DNA methylation during oogenesis is associated with the recognition site of E2f1 and E2f2. *Epigenetics Chromatin* **7**: 26.
- Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, Regev A, Meissner A. 2012. A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**: 339-344.
- Stielow B, Finkernagel F, Stiewe T, Nist A, Suske G. 2018. MGA, L3MBTL2 and E2F6 determine genomic binding of the non-canonical Polycomb repressive complex PRC1.6. *PLoS Genet* **14**: e1007193.
- Storre J, Schafer A, Reichert N, Barbero JL, Hauser S, Eilers M, Gaubatz S. 2005. Silencing of the meiotic genes SMC1beta and STAG3 in somatic cells by E2F6. *J Biol Chem* **280**: 41380-41386.

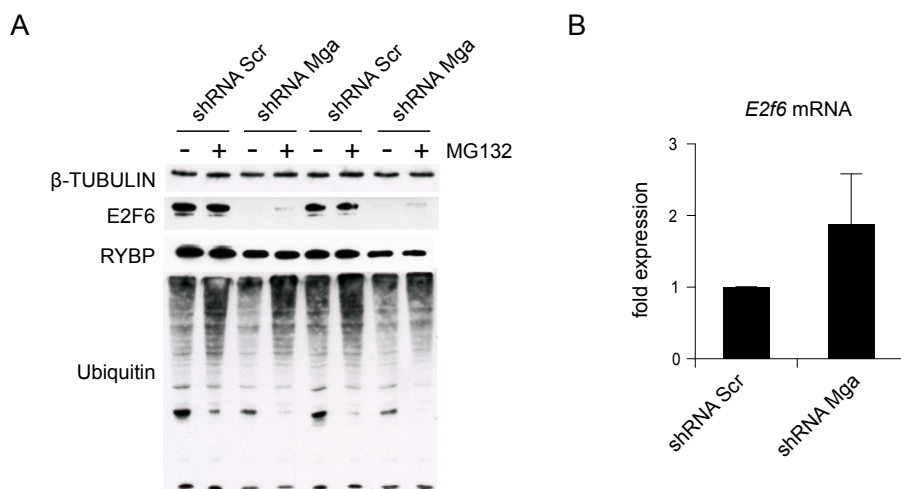
- Termanis A, Torrea N, Culley J, Kerr A, Ramsahoye B, Stancheva I. 2016. The SNF2 family ATPase LSH promotes cell-autonomous de novo DNA methylation in somatic cells. *Nucleic Acids Res* **44**: 7592-7604.
- Trimarchi JM, Fairchild B, Verona R, Moberg K, Andon N, Lees JA. 1998. E2F-6, a member of the E2F family that can behave as a transcriptional repressor. *Proc Natl Acad Sci U S A* **95**: 2850-2855.
- Trojer P, Cao AR, Gao Z, Li Y, Zhang J, Xu X, Li G, Losson R, Erdjument-Bromage H, Tempst P et al. 2011. L3MBTL2 protein acts in concert with PcG protein-mediated monoubiquitination of H2A to establish a repressive chromatin structure. *Mol Cell* **42**: 438-450.
- Velasco G, Hube F, Rollin J, Neuillet D, Philippe C, Bouzinba-Segard H, Galvani A, Viegas-Pequignot E, Francastel C. 2010. Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc Natl Acad Sci* **107**: 9281-9286.
- Washkowitz AJ, Schall C, Zhang K, Wurst W, Floss T, Mager J, Papaioannou VE. 2015. Mga is essential for the survival of pluripotent cells during peri-implantation development. *Development* **142**: 31-40.
- Xi S, Geiman TM, Briones V, Guang Tao Y, Xu H, Muegge K. 2009. Lsh participates in DNA methylation and silencing of stem cell genes. *Stem Cells* **27**: 2691-2702.
- Xia L, Huang W, Bellani M, Seidman MM, Wu K, Fan D, Nie Y, Cai Y, Zhang YW, Yu LR et al. 2017. CHD4 Has Oncogenic Functions in Initiating and Maintaining Epigenetic Suppression of Multiple Tumor Suppressor Genes. *Cancer Cell* **31**: 653-668 e657.
- Yang WW, Shu B, Zhu Y, Yang HT. 2008. E2F6 inhibits cobalt chloride-mimetic hypoxia-induced apoptosis through E2F1. *Mol Biol Cell* **19**: 3691-3700.
- Zhao W, Tong H, Huang Y, Yan Y, Teng H, Xia Y, Jiang Q, Qin J. 2017. Essential Role for Polycomb Group Protein Pcgf6 in Embryonic Stem Cell Maintenance and a Noncanonical Polycomb Repressive Complex 1 (PRC1) Integrity. *J Biol Chem* **292**: 2773-2784.



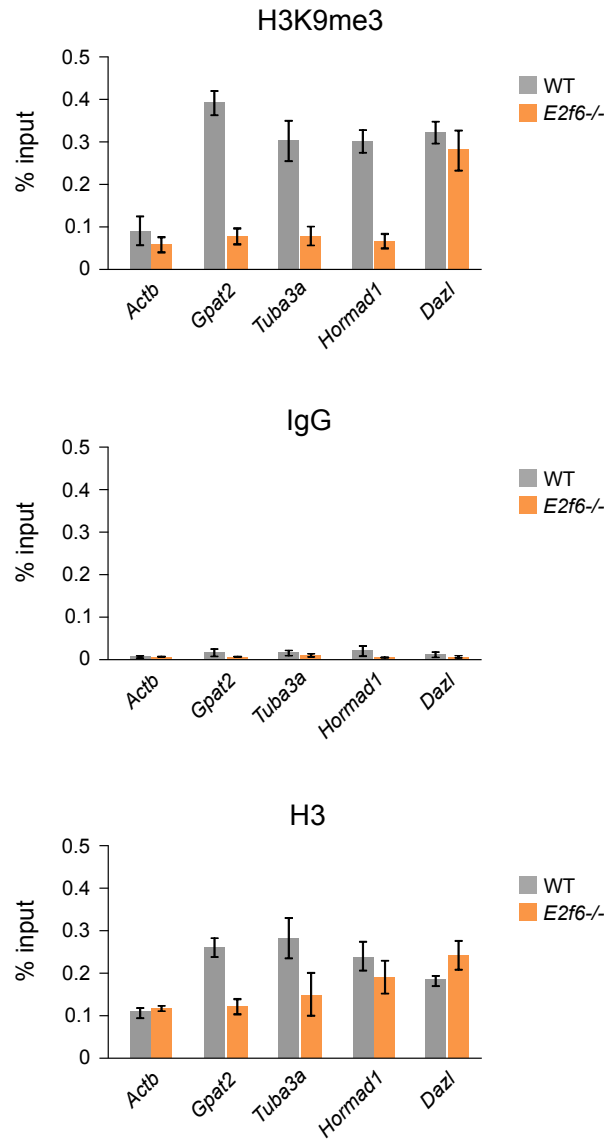
Supplementary Figure 1. Analysis of E2F6 ChIP-seq profiles in mouse ES cells. **A.** Heatmap of E2F6 ChIP-seq signals in *E2f6*^{+/+} and *E2f6*^{-/-} ES cells 5kb around E2F6 peaks, demonstrating the specificity of the E2F6 ChIP-seq dataset. **B.** Analysis of TF motifs in E2F6 peaks. **C.** Clustered correlation matrix between E2F6 ChIP-seq and various ChIP-seq datasets in mouse ES cells. Each point represents the pairwise Spearman correlation of normalized read counts in the unified set of peaks from both datasets. The E2F6 dataset clusters with ChIP-seq from other PRC1.6 components (MAX, MGA, L3MBTL2, PCGF6, RYBP) (highlighted in the red square). **D.** Heatmaps of E2F6, E2F1 and PRC1.6 ChIP-seq signals centered on E2F6 peaks (plus or minus 2.5 kb). **E.** Venn Diagram showing the overlap between E2F6, MGA and L3MBTL2 ChIP-seq peaks in mouse ES cells. **F.** Examples of ChIP-seq profiles of E2F6 and other PRC1.6 components over target germline genes in mouse ES cells. RefSeq gene annotations are shown below the tracks.



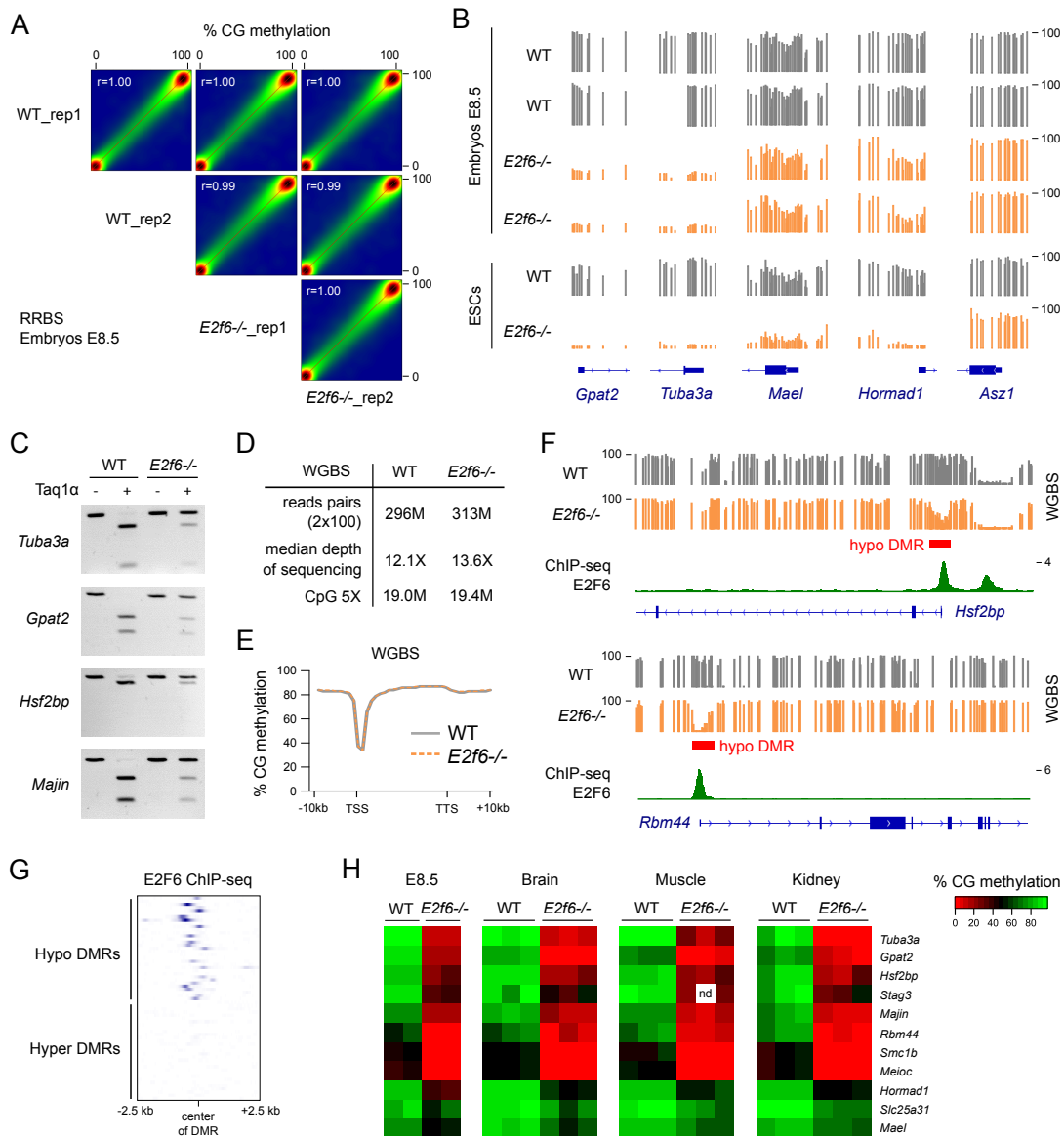
Supplementary Figure 2. E2F6 and G9A repress distinct sets of genes in mouse embryos. **A.** Venn diagram comparing the genes upregulated in *E2f6*^{-/-} and *G9a*^{-/-} E8.5 embryos (Auclair et al., Genome Res 2016). Only two genes (*Xlr* and *Cyct*) are commonly repressed by E2F6 and G9a. **B.** Heatmap of the expression of E2F6 and G9A target genes measured by RNA-seq in *E2f6* and *G9a* knockout embryos, illustrating that *E2f6*^{-/-} and *G9a*^{-/-} embryos overexpress distinct sets of genes. Each column represents a different embryo.



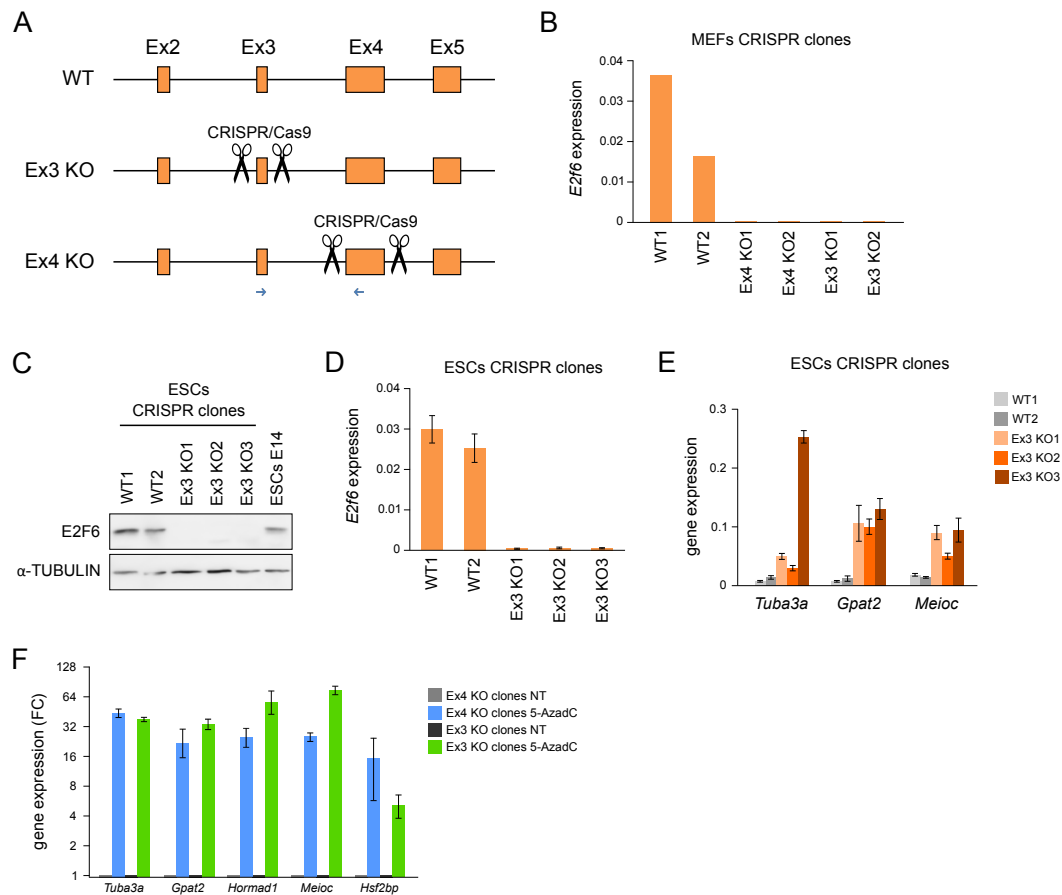
Supplementary Figure 3. MGA depletion causes posttranscriptional downregulation of the E2F6 protein in ES cells. **A.** Western blot showing reduced levels of E2F6 proteins in MGA knockdown ES cells. ES cells were transduced with lentiviral vectors expressing an shRNA against MGA (shRNA Mga) or a scramble control (shRNA scr). After 3 days, half of the cells were treated for 2h with 5M MG132 to block the proteolytic activity of the proteasome complex. Western blot was performed on whole cell extracts with the indicated antibodies. Beta-TUBULIN was used as a loading control. **B.** RT-qPCR analysis of *E2f6* mRNA transcript levels in ES cells transduced with lentiviral vectors expressing an shRNA targeting MGA (shRNA Mga) or a control scramble shRNA (shRNA Scr). The expression levels were normalized to beta-actin and are represented as a fold change relative to shRNA Scr cells (mean \pm SD, n = 3).



Supplementary Figure 4. E2F6 is required for the deposition of H3K9me3 in the promoters of germline genes in naive ES cells. The graphs show ChIP-qPCR quantifications of H3K9me3, H3 and IgG control in the promoters of prominent E2F6 target genes in *E2f6*^{+/+} and *E2f6*^{-/-} ES cultured in 2i conditions (mean ± SEM, n=3 experiments). The *Dazl* gene was used as a control of germline gene that is not strongly repressed by E2F6.



Supplementary Figure 5. DNA methylation analysis in *E2f6*^{-/-} mice and ES cells. A. Density scatter plots showing the correlation of methylation scores in 500 bp windows between RRBS replicates in WT and *E2f6*^{-/-} embryos. The density of points increases from blue to dark red. The Pearson correlation coefficients are indicated in the top left corner. **B.** Examples of RRBS promoter methylation profiles for E2F6 target genes in WT and *E2f6*^{-/-} E8.5 embryos (top) and ESCs (bottom). The *Asz1* germline gene, which is not a target of E2F6, is shown as a control. **C.** DNA methylation analysis by COBRA in 4 promoters of E2F6 target genes in WT and *E2f6*^{-/-} embryos. **D.** Basic sequencing information for the WGBS libraries in WT and *E2f6*^{-/-} embryos. **E.** Metaplot showing WGBS methylation scores over CGI Refseq genes and 10 kb of flanking sequences in WT and *E2f6*^{-/-} embryos. **F.** Examples of WGBS genome tracks in WT and *E2f6*^{-/-} embryos over the *Hsf2bp* and *Rbm44* genes. E2F6 ChIP-seq tracks in ES cells are also shown. Hypomethylated DMRs (hypo DMRs) overlapping the gene promoters are depicted by red rectangles. Refseq gene annotations are shown below the tracks. **G.** Heatmap of E2F6 ChIP-seq signals in 2.5 kb around the center of E2F6_ES-serum_1_peak hypomethylated and hypermethylated DMRs. Only hypo DMRs frequently colocalize with E2F6 peaks. **H.** Heatmaps representing the levels of promoter CpG methylation of E2F6 target genes profiled by RRBS in adult tissues from WT 172979944 and *E2f6*^{-/-} mice. Each column represents an independent animal. The values in E8.5 embryos are shown again as comparison. nd: not determined.



Supplementary Figure 6. Analysis of E2F6 CRISPR-Cas9 clones in MEFs and ES cells. **A.** Schematic representation of the strategy used for inactivation of the *E2f6* gene by CRISPR-Cas9. gRNAs were designed on each side of exons to remove either the exon 3 or the exon 4 and create a frameshift to prevent expression of the E2F6 protein. Blue arrows depict the oligos used for validation of the exon deletion by RT-qPCR. **B.** RT-qPCR analysis of *E2f6* expression in MEFs CRISPR clones using oligos spanning the exon 3-4 region, which confirms the efficient deletion of the exons in KO clones. **C.** Western blot of E2F6 in ESCs CRISPR-Cas9 clones. A protein extract from E14 ESCs was used as a control. Tubulin was used as a loading control. **D.** RT-qPCR analysis of *E2f6* expression in ESCs CRISPR clones using oligos spanning the exon 3-4 region, which confirms the efficient deletion of the exon 3 in KO clones. **E.** RT-qPCR analysis of the expression of three E2F6 target genes in ESCs WT and *E2f6*-KO CRISPR clones (mean \pm SEM, n=3 technical replicates). **F.** RT-qPCR analysis of the expression of E2F6 target genes in CRISPR *E2f6*-KO MEF clones treated with 1 μ M 5-AzadC or not treated (NT). The values are represented as a fold-change CGI (FC) relative to non-treated cells (mean \pm SEM, n=2 clones per condition).

DISCUSSION

1. Rôle de la méthylation de l'ADN au cours du développement embryonnaire

1.1. Rôle des ADN méthyltransférases dans la méthylation du génome

Le rôle des ADN méthyltransférases dans la méthylation du génome a été initialement départagé entre DNMT1 dans le maintien de la méthylation au cours des divisions cellulaires, et DNMT3A et DNMT3B dans la méthylation *de novo*. Néanmoins, un nouveau modèle a émergé ces dernières années proposant un rôle partagé de ces enzymes, notamment dans le maintien des profils de méthylation de l'ADN au cours des divisions cellulaires. Dans notre étude, nous avons validé le modèle initial en montrant un rôle exclusif de DNMT1 dans le maintien de la méthylation sur tout le génome et un rôle strict de DNMT3A et DNMT3B dans la méthylation *de novo* dans l'embryon.

Le modèle révisé proposant une coopération entre DNMT1, DNMT3A et DNMT3B pour le maintien de la méthylation de l'ADN, est notamment influencé par des études sur des cellules ES. Or, ces cellules sont hétérogènes au niveau de la méthylation de l'ADN, puisqu'elles sont méthylées et déméthylées de manière cyclique (Hayashi et al., 2008). De plus, une petite partie d'entre elles adoptent un profil de cellules proches des cellules embryonnaires au stade 2-cellules qui sont hypométhylées (Macfarlan et al., 2012). Pour ces raisons, les cellules ES ne sont pas un bon modèle pour étudier le rôle des différentes ADN méthyltransférases dans la méthylation du génome puisqu'elles subissent à la fois des cycles de méthylation *de novo* et de méthylation de maintien. Ainsi, l'absence de DNMT3A et DNMT3B dans les cellules ES conduirait à une hypométhylation non pas à cause d'un défaut de maintien mais à cause de l'absence de reméthylation *de novo* conduisant une accumulation de cellules hypométhylées.

1.2. Rôle de la méthylation dans la répression du programme gamétique

Le rôle de la méthylation de l'ADN dans la répression du programme gamétique avait déjà été démontré par des travaux précédents de l'équipe (Auclair et al., 2014; Borgel et al., 2010). Notre étude permet cependant d'aller plus loin sur deux niveaux. Premièrement, nous avons identifié la totalité des gènes gamétiques réprimés par la méthylation de l'ADN dans les embryons (~130 gènes). Deuxièmement, nous avons démontré par des expériences de déméthylation ciblée, que la méthylation joue un rôle direct dans la répression de ces gènes au niveau de leur promoteur *en cis* et exclue l'hypothèse que la déméthylation du génome régule un facteur de transcription agissant indirectement *en trans*.

Cependant, plusieurs questions restent ouvertes :

- La répression du programme gamétique par la méthylation de l'ADN est-elle une fonction évolutive conservée chez les vertébrés ?
- Pourquoi les promoteurs de gènes gamétiques sont-ils si sensibles à l'absence de la méthylation ?
- Comment la méthylation de l'ADN réprime-t-elle les promoteurs de gènes gamétiques ?
- Comment la méthylation de l'ADN est-elle ciblée sur les promoteurs de gènes gamétiques ?

Pour répondre en partie à certaines de ces questions, un criblage CRISPR-Cas9 permettant d'identifier des protéines impliquées dans la répression de ces gènes pourrait apporter des éléments de réponse. En effet, cette stratégie pourrait identifier potentiellement des protéines agissant en aval de la mise en place de la méthylation ou bien des protéines interagissant avec l'ADN méthylyé et responsables de la répression par la méthylation.

1.3. Rôle de la méthylation dans la répression des gènes 2-cellules

Nous avons observé un rôle de la méthylation de l'ADN dans la répression des gènes 2-cellules dans l'embryon. Ce résultat était inattendu car un criblage d'ARN interférant visant à déterminer les régulateurs des gènes 2-cellules n'a pas identifié d'effecteurs de la méthylation de l'ADN dans les cellules ES (Rodriguez-Terrones *et al.*, 2018). De plus, une autre étude a montré que la méthylation de l'ADN ne réprime pas ces gènes dans les cellules ES, puisque l'ablation de DNMT1 n'a pas d'effet sur leur régulation (Eckersley-Maslin *et al.*, 2016).

Les promoteurs des gènes 2-cellules sont des promoteurs à faible densité de CpG. La déméthylation de ceux-ci n'est probablement pas suffisante pour permettre leur dérèpression, suggérant un effet indirect. En effet, nous avons observé que les gènes 2-cellules sont dérèprimés dans les embryons DKO, mais pas dans les embryons KO pour *Dnmt1*, malgré un niveau de méthylation équivalent. Ceci suggère que la méthylation ne régule pas ces gènes *en cis* mais *en trans*. En effet, nous observons une dérèpression des rétrotransposons MERVL qui est associé à l'expression des gènes 2-cellules, dans les embryons DKO mais pas dans les embryon *Dnmt1* KO.

La raison pour laquelle les rétrotransposons MERVL sont dérèprimés uniquement dans les embryons DKO est énigmatique. Les rétrotransposons MERVL sont déméthylés et exprimés transitoirement dans l'embryon précoce (Macfarlan *et al.*, 2012). L'absence des enzymes DNMT3A et DNMT3B pourrait ensuite empêcher leur reméthylation après le stade embryonnaire 2-cellules. Par opposition,

l'absence de DNMT1 pourrait insuffisamment diminuer la méthylation de MERVL pour permettre leur dérégulation, à cause de l'expression de DNMT3A et DNMT3B qui méthyle MERVL *de novo* après chaque division cellulaire.

Ces expériences montrent que la méthylation de l'ADN permet de contrôler la sortie de l'état de totipotence des cellules embryonnaires.

1.4. Rôle de la méthylation dans la répression de gènes de cellules différenciées

Mis à part les gènes gamétiques et les gènes *Rhox* (Oda et al., 2006), le rôle de la méthylation de l'ADN dans la régulation directe de l'expression d'autres gènes au cours de l'embryogenèse est débattu et n'a jamais été démontré.

Nous avons observé que des gènes spécifiques de certains lignages cellulaires sont dérégulés dans les embryons en absence de la méthylation de l'ADN. Ces gènes ont des promoteurs denses en CpG et sont déméthylés spécifiquement dans le lignage cellulaire dans lequel ils sont exprimés. Basé sur ces observations, nous proposons le modèle que la méthylation de l'ADN sert à réprimer l'expression précoce de ces gènes dans l'embryon et permet de restreindre leur domaine d'expression. Ce modèle implique que la méthylation est mise en place uniformément dans l'épiblaste embryonnaire et que la différenciation des différents lignages cellulaires s'accompagne de la déméthylation spécifique de ces gènes.

Néanmoins, il faut noter que nous comparons des embryons sauvages avec des embryons mutants présentant d'importants défauts de développement. De plus, notre analyse transcriptomique pourrait être biaisée par le fait que nous étudions des embryons totaux dans lesquels sont présents plusieurs types cellulaires différents (Pijuan-Sala et al., 2019). Ce modèle doit donc être vérifié dans l'embryon en réalisant des expériences plus approfondies. Nous pourrions comparer les embryons sauvages et mutés à un stade embryonnaire plus précoce, par exemple au stade embryonnaire E6.5. En effet la différence morphologique entre les embryons sauvages et mutés à ce stade pourrait être moins marquée, comme pour les embryons mutés pour les 3 gènes *Tet* (Dai et al., 2016). Nous pourrions également réaliser des analyses génomiques sur des cellules embryonnaires uniques.

Les cellules germinales primordiales sont de bons exemples du contrôle de la différenciation cellulaire par la méthylation de l'ADN. Deux modèles sont proposés pour la différenciation de ces cellules. i) Les précurseurs de ces cellules pourraient être spécifiés dès le stade blastocyste avant la

phase de reméthylation du génome qui ne concernerait donc pas les gènes gamétiques ii) Les précurseurs de ces cellules sont spécifiés plus tardivement à partir de cellules de l'épiblaste, et leur différenciation impliquerait donc la déméthylation de leur génome et des gènes gamétiques. Plusieurs évidences tendent à vérifier la seconde hypothèse. Les précurseurs de ces cellules les plus précoces ont été identifiés au niveau de l'épiblaste (*Lesch and Page, 2012*) et l'inactivation de certains effecteurs de la méthylation comme *Dnmt1* ou de *Tet1* dans les PGCs conduit à une dérégulation des gènes gamétiques (*Hargan-Calvopina et al., 2016; Yamaguchi et al., 2012*).

L'exemple des cellules germinales va donc dans le sens de notre modèle selon lequel la méthylation de l'ADN réprime l'expression précoce de marqueurs dans l'embryon, avant leur déméthylation au cours de la différenciation des lignées cellulaires.

1.5. Rôle de la méthylation dans la répression des éléments transposables

La méthylation de l'ADN est considérée comme un acteur majeur de la répression des éléments transposables dans les embryons depuis qu'une étude a montré que les embryons déplétés pour *Dnmt1* présentent une dérégulation des éléments IAPs par northern blot (*Walsh et al., 1998*). Néanmoins, la méthylation ne semble pas avoir un rôle primordial dans la répression des éléments transposables dans les cellules ES, dans lesquelles on observe un mécanisme de compensation pour l'absence de la méthylation de l'ADN (*Walter et al., 2016*).

Notre étude montre que, au-delà des IAPs, la méthylation de l'ADN réprime toutes les familles d'éléments transposables dans l'embryon. Néanmoins, l'absence de méthylation affecte certaines familles préférentiellement, notamment les rétrotransposons qui se sont insérés dans le génome le plus tardivement dans l'évolution. L'activation de ces rétrotransposons permet leur insertion aléatoirement dans le génome, engendrant une potentielle instabilité génomique. Les autres rétrotransposons ont subi des événements des mutations et de troncations les rendant inoffensifs. La méthylation joue donc un rôle essentiel dans la répression des rétrotransposons potentiellement actifs dans l'embryon. De plus, la déméthylation de ceux-ci peut conduire à l'activation de promoteurs cryptiques modulant l'expression de certains gènes.

1.6. Causes de la létalité embryonnaire des embryons hypométhylés

Un des objectifs de ce projet était de comprendre les causes de la létalité des embryons globalement hypométhylés. Nous pouvons proposer plusieurs hypothèses :

- L'expression inadéquate de gènes, comme par exemple les gènes soumis à l'empreinte ou des gènes normalement exprimés dans les cellules différenciées, perturbe la mise en place du plan de formation de l'organisme conduisant ainsi à l'arrêt du développement. C'est le cas par exemple des embryons déplétés pour les protéines TET qui sont létaux au stade de l'épiblaste à cause de défaut de mise en place de l'organisme (*Dai et al., 2016*).
- L'absence de la méthylation de l'ADN pourrait engendrer une instabilité génomique, notamment en raison de la dérégulation des éléments transposables qui sont mutagènes. De plus, la dérégulation de gènes gamétiques pourrait initier des événements méiotiques, comme observé dans les cellules ES (*Suzuki et al., 2016*). Ces événements engendreraient ainsi une instabilité génomique pouvant causer l'apoptose des cellules, et l'arrêt du développement embryonnaire.
- L'activation des rétrotransposons pourrait induire une réponse immunitaire innée conduisant à la mort de l'embryon. En effet, il est observé dans les cellules somatiques que l'activation des rétrotransposons causée par la déméthylation, induit une réponse interférons et une mort cellulaire (*Chiappinelli et al., 2015*). Cependant, en défaveur de cette hypothèse, nous n'avons pas observé de signe d'activation transcriptionnelle de gènes de la réponse interférons dans les embryons mutants.

Les causes de la létalité des embryons hypométhylés restent donc à déterminer et sont probablement multifactorielles. Des études génomiques sur cellules uniques, à des stades embryonnaires plus précoces ou des expériences d'imagerie pour visualiser des aberrations chromosomiques pourraient fournir des éléments de réponse.

2. Rôle du facteur de transcription E2F6 dans la répression des gènes gamétiques

2.1. E2F6 réprime un groupe de gènes gamétiques

Le facteur de transcription E2F6 est lié à de nombreux promoteurs de gènes dans les cellules ES. Néanmoins, cet enrichissement conduit à la répression d'un petit nombre de gènes uniquement. Ceci pourrait s'expliquer par des mécanismes de compensation. En effet, plusieurs membres de la famille E2F ont un domaine de liaison à l'ADN leur permettant de se lier au même motif d'ADN. Ainsi, l'absence d'E2F6 est compensée par l'activité de E2F4 dans la répression du gène *Cdc6* dans les MEFs (Giangrande et al., 2004). De plus, les facteurs de transcription MAX/MGA peuvent également compenser l'absence d'E2F6, dans le cas de gènes gamétiques possédant les deux motifs reconnus par ces facteurs de transcription.

Nous avons observé une quinzaine de gènes gamétiques dérégulés en l'absence de E2F6, sur une centaine que compte le génome murin. La raison pour laquelle plusieurs mécanismes différents sont utilisés par la cellule pour réprimer le programme gamétique est énigmatique. Le facteur de transcription E2F6 ne régule pourtant pas des gènes impliqués dans la même fonction. En effet, TUBA3A est impliqué dans la formation du complexe synaptonémal permettant l'ancrage des microtubules au niveau des régions centromériques, GPAT2 est impliqué dans la biosynthèse des piARN et HORMAD1 régule les recombinaisons intra-chromosomiques.

Les complexes Polycomb sont impliqués dans la régulation de l'expression des gènes gamétiques, durant la différenciation des cellules germinales primordiales (Maezawa et al., 2017; Yokobayashi et al., 2013). Il serait ainsi intéressant de déterminer si les gènes gamétiques réprimés par E2F6 sont exprimés durant la même période au cours de la différenciation des cellules germinales primordiales. Le facteur de transcription E2F6 pourrait ainsi participer à la régulation fine de l'expression temporelle de certains gènes gamétiques durant la différenciation des PGCs.

2.2. Mécanismes de répression des gènes gamétiques par PRC1.6

À l'image des éléments transposables, nous proposons que les gènes gamétiques soient réprimés par des mécanismes indépendants de la méthylation de l'ADN dans les stades embryonnaires précoces. Cette répression favoriserait alors l'acquisition de la méthylation de l'ADN pour une répression durable.

Ce modèle ne semble néanmoins pas s'appliquer à tous les gènes gamétiques car un nombre conséquent de ces gènes est fortement exprimé dans les blastocystes. La raison pour laquelle seuls certains gènes gamétiques sont totalement réprimés dans les blastocystes, comme *Dazl* par exemple, est inconnue. La répression de certains gènes clés pourrait suffire à verrouiller le programme gamétique dans les cellules somatiques afin de prévenir des événements de méioses non désirées dans les cellules embryonnaires. Nous proposons que le complexe PRC1.6 est impliqué dans la répression des gènes gamétiques dans les blastocystes.

Le facteur de transcription E2F6 est un des modules de liaison à l'ADN du complexe Polycomb PRC1.6. Par ailleurs, plusieurs autres membres du complexe PRC1.6 ont été impliqués dans la répression des gènes gamétiques dans les cellules ES (*Suzuki et al., 2016; Hisada et al., 2012; Endoh et al., 2017; Qin et al., 2012*). Ce complexe possède un module d'ubiquitination des histones et un module de méthylation des histones, mais les mécanismes par lesquels il réprime les gènes gamétiques reste à éclaircir.

Le rôle de l'ubiquitination des histones dans la répression des gènes gamétiques a été proposé dans les cellules ES (*Endoh et al., 2017*). De plus, RING1B est impliqué dans la répression des gènes gamétiques dans les cellules germinales primordiales (*Maewawa et al., 2017; Yokobayashi et al., 2013*). Le rôle de l'ubiquitination des histones nécessite donc des investigations plus approfondies. Elle pourrait conduire au recrutement du complexe PRC2 puisque nous avons observé un enrichissement de H3K27me3 au niveau du promoteur de gènes gamétiques (données non montrées). Des études contradictoires montrent un rôle ambigu de H3K27me3 dans la répression des gènes gamétiques. Une étude a montré que le complexe PRC2 réprime les gènes gamétiques ciblés par E2F6 dans les cellules ES (*Leseva et al., 2013*), tandis qu'une autre a montré que ces gènes gamétiques ne sont pas dérèprimés dans les embryons dépourvus de H3K27me3 (*Zylicz et al., 2015*).

La méthylation de H3K9 est également associée à la répression des gènes gamétiques. Le complexe PRC1.6 interagit avec deux histone méthyltransférases responsables de la déposition de la marque H3K9me2 : G9A et GLP. Une précédente étude du laboratoire a montré que l'histone méthyltransférase G9A réprime un groupe de gènes gamétiques (*Auclair et al., 2016*). Néanmoins, ce groupe de gènes est différent de celui réprimé par E2F6. Il n'est pas à exclure que GLP soit impliqué dans la répression des gènes gamétiques médié par E2F6. L'inactivation de *Glp* dans les cellules ES ou dans l'embryon pourrait ainsi fournir des évidences sur les mécanismes de répression utilisés par le complexe PRC1.6. La modification d'histone H3K9me2 pourrait permettre le recrutement de SETDB1

et H3K9me3. En effet l'inactivation de SETDB1 dans les cellules ES conduit à une dérégulation d'un groupe de gènes gamétiques en partie identique à celui réprimé par E2F6 (Karimi et al., 2011).

De manière intéressante, outre l'exemple des gènes gamétiques, la régulation des gènes 2-cellules implique également le complexe PRC1.6 (Rodriguez-Terrones et al., 2018), G9A (Maksakova et al., 2013) et la méthylation de l'ADN *de novo* comme nous l'avons montré dans la première partie. Ceci montre donc une convergence des mécanismes qui répriment les gènes gamétiques et les gènes 2-cellules dans les embryons murins.

Bien que les différents membres du complexe PRC1.6 soient bien étudiés, les mécanismes en jeu permettant la répression des gènes par ce complexe sont mal connus et méritent de futures investigations.

2.3. Mécanismes de recrutement de la méthylation de l'ADN par E2F6

Enfin, notre étude a montré un lien entre E2F6 et la méthylation de l'ADN. Nos résultats montrent que la répression des gènes gamétiques par E2F6 et le complexe PRC1.6 facilite l'acquisition de la méthylation de l'ADN durant la phase de reméthylation du génome.

Ce lien ne semble pas impliquer une interaction directe entre les ADN méthyltransférases et le complexe PRC1.6, puisque nous n'avons pas pu montrer d'interaction entre E2F6 avec DNMT3A ou DNMT3B dans les cellules ES. Ceci confirme de précédentes études protéomiques, dans lesquelles les DNMTs n'ont pas été enrichies dans des expériences d'interaction avec d'autres membres du complexe PRC1.6 (Endoh et al., 2017; Qin et al., 2012).

Deux modèles différents de recrutement de la méthylation de l'ADN sont proposés dans la communauté scientifique. La méthylation de l'ADN peut être recrutée activement au niveau de certains locus, ou inversement elle peut être adressée par défaut sur tout le génome à l'exception de certaines régions, comme des îlots CpG, qui sont protégés.

Le complexe PRC1.6 pourrait permettre le recrutement de la méthylation de l'ADN via l'activité de SETDB1 qui dépose la marque d'histone H3K9me3. En effet, l'inactivation de SETDB1 conduit à une déméthylation de certains gènes gamétiques dans les cellules ES (Leung et al., 2014a). Ce mécanisme moléculaire a déjà un rôle supposé dans la méthylation *de novo* et son maintien au niveau des rétrotransposons au cours de la reprogrammation épigénétique de l'embryon (Leung et al., 2014b; Liu

et al., 2014; Rowe et al., 2013). Néanmoins, aucun lien fonctionnel n'a été formellement établi entre H3K9me3 et la méthylation de l'ADN *de novo*. Une étude biochimique a néanmoins proposé que le domaine ADD des enzymes DNMT3 est capable d'interagir avec la modification d'histone H3K9me3 (*Dhayalan et al., 2011*). Des expériences de complémentation de cellules ES TKO avec des versions mutées de *Dnmt3b* dans ces domaines pourraient mettre en lumière d'éventuels lien entre H3K9me3 et la méthylation de l'ADN.

Le complexe PRC1.6 pourrait également empêcher le recrutement de la machinerie de la transcription et de H3K4me3. En effet, cette modification d'histone est antagoniste de la méthylation de l'ADN. Elle empêche le recrutement des enzymes DNMT3 puisque leur domaines ADD reconnaît uniquement H3K4 non méthylé, tandis que la méthylation de H3K4 inhibe cette interaction (*Ooi et al., 2007*). Ainsi, nos collaborateurs ont observé qu'en l'absence de E2F6, l'activateur E2F1 se fixe sur les promoteurs de gènes gamétiques. Cette activation s'accompagne de la mise en place de H3K4me3 qui pourrait ainsi empêcher le recrutement de la méthylation de l'ADN. Néanmoins, des résultats préliminaires suggèrent que le recrutement de E2F1 dans les cellules ES KO *E2f6* n'est pas nécessaire pour empêcher le recrutement de la méthylation de l'ADN (données non montrées).

Une étude protéomique visant à définir les interactants de E2F6 pourrait également fournir des éléments de réponse sur les mécanismes de recrutement de la méthylation de l'ADN.

2.4. E2F6 induit une répression durable par des mécanismes dépendants ou indépendants de la méthylation de l'ADN

Nos résultats démontrent que E2F6 initie une répression épigénétique durable des gènes gamétiques. L'expérience de sauvetage de *E2f6* dans les MEFs montre que E2F6 n'est pas capable d'induire la méthylation des promoteurs des gènes gamétique dans les cellules différenciées. Ceci suggère ainsi que E2F6 initie la répression épigénétique des gènes gamétiques durant une période définie du développement embryonnaire, et que passé cette période il n'est plus capable de réprimer ces gènes. Cette fenêtre de développement permettant l'initiation de la répression épigénétique des gènes gamétiques corrèle avec la forte expression des enzymes de méthylation *de novo* DNMT3A et DNMT3B qui sont ensuite exprimées à des niveaux beaucoup plus faibles dans les cellules différenciées.

La méthylation de l'ADN est suffisante pour maintenir sur le long terme la répression des gènes gamétiques. En effet, nous avons montré que le facteur de transcription E2F6 n'est plus nécessaire

pour le maintien de la répression de ces gènes dans les cellules différenciées une fois que la méthylation de l'ADN est mise en place. La méthylation de l'ADN peut ainsi servir à verrouiller le programme gamétique dans les cellules somatiques. Cette stratégie est également utilisée pour tous les éléments génomiques nécessitant une répression sur le long terme, comme les éléments transposables par exemple. Elle pourrait ainsi permettre une économie d'énergie à la cellule en rendant dispensables les complexes moléculaires qui initient la répression de ces éléments génomiques tel que le complexe PRC1.6.

Néanmoins, de manière intéressante, on observe que certains gènes gamétiques réprimés par E2F6 acquièrent des niveaux de méthylation d'ADN peu élevés dans les embryons après l'implantation. En conséquence, nos analyses dans les embryons hypométhylés montrent que ces gènes sont faiblement réprimés par la méthylation de l'ADN. L'exemple du gène *Smc1b* est le plus représentatif : il acquiert un niveau de méthylation faible dans l'embryon (40%) et est surexprimé seulement d'un facteur 4 dans les embryons *Dnmt* DKO contre une surexpression d'un facteur 250 dans les embryons *E2f6*^{-/-}. Nous n'avons pas encore vérifié si la présence de E2F6 n'est plus nécessaire pour maintenir la répression de ce gène dans les MEFs, mais si tel est le cas cela démontrera qu'il existe des mécanismes de répression épigénétique durables induits par E2F6 qui sont indépendant de la méthylation de l'ADN. Pour identifier ces mécanismes, nous pourrions envisager de tagger le gène *Smc1b* avec GFP dans les MEFs et de réaliser une expérience de criblage CRISPR-Cas9 en utilisant comme sélection positive la réactivation de *Smc1b* (cellules GFP+).

BIBLIOGRAPHIE

Aihara, H., Nakagawa, T., Mizusaki, H., Yoneda, M., Kato, M., Doiguchi, M., Imamura, Y., Higashi, M., Ikura, T., Hayashi, T., et al. (2016). Histone H2A T120 Phosphorylation Promotes Oncogenic Transformation via Upregulation of Cyclin D1. *Mol. Cell* *64*, 176–188.

Akasaka, T., Takahashi, N., Suzuki, M., Koseki, H., Bodmer, R., and Koga, H. (2002). MBLR, a new RING finger protein resembling mammalian Polycomb gene products, is regulated by cell cycle-dependent phosphorylation. *Genes Cells* *7*, 835–850.

Aloia, L., Stefano, B.D., and Croce, L.D. (2013). Polycomb complexes in stem cells and embryonic development. *Development* *140*, 2525–2534.

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.

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.

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

Auclair, G., Borgel, J., Sanz, L.A., Vallet, J., Guibert, S., Dumas, M., Cavelier, P., Girardot, M., Forné, T., Feil, R., et al. (2016). EHMT2 directs DNA methylation for efficient gene silencing in mouse embryos. *Genome Res.* *26*, 192–202.

Aucott, R., Bullwinkel, J., Yu, Y., Shi, W., Billur, M., Brown, J.P., Menzel, U., Kioussis, D., Wang, G., Reisert, I., et al. (2008). HP1- β is required for development of the cerebral neocortex and neuromuscular junctions. *J. Cell Biol.* *183*, 597–606.

Ballaré, C., Lange, M., Lapinaite, A., Martin, G.M., Morey, L., Pascual, G., Liefke, R., Simon, B., Shi, Y., Gozani, O., et al. (2012). Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nat. Struct. Mol. Biol.* *19*, 1257–1265.

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

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

Bashtrykov, P., Jankevicius, G., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2014). The UHRF1 Protein Stimulates the Activity and Specificity of the Maintenance DNA Methyltransferase DNMT1 by an Allosteric Mechanism. *J. Biol. Chem.* *289*, 4106–4115.

- Baubec, T., 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., Shiekhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. *Genes Dev.* 23, 781–783.
- Berman, B.P., Weisenberger, D.J., Aman, J.F., Hinoue, T., Ramjan, Z., Liu, Y., Noushmehr, H., Lange, C.P.E., van Dijk, C.M., Tollenaar, R.A.E.M., et al. (2011). Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat. Genet.* 44, 40–46.
- Bertoli, C., Klier, S., McGowan, C., Wittenberg, C., and de Bruin, R.A.M. (2013). Chk1 Inhibits E2F6 Repressor Function in Response to Replication Stress to Maintain Cell-Cycle Transcription. *Curr. Biol.* 23, 1629–1637.
- Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells: The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J. Mol. Biol.* 203, 971–983.
- Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. *Nature* 321, 209.
- 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.
- Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L.P., Ito, S., Cooper, S., Kondo, K., Koseki, Y., et al. (2014). Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. *Cell* 157, 1445–1459.
- Blaschke, K., Ebata, K.T., Karimi, M.M., Zepeda-Martínez, J.A., Goyal, P., Mahapatra, S., Tam, A., Laird, D.J., Hirst, M., Rao, A., et al. (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* 500, 222–226.
- Bogdanović, O., Smits, A.H., de la Calle Mustienes, E., Tena, J.J., Ford, E., Williams, R., Senanayake, U., Schultz, M.D., Hontelez, S., van Kruijsbergen, I., et al. (2016). Active DNA demethylation at enhancers during the vertebrate phylotypic period. *Nat. Genet.* 48, 417–426.
- 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* 317, 1760–1764.

- Boulard, M., Edwards, J.R., and Bestor, T.H. (2015). FBXL10 protects Polycomb-bound genes from hypermethylation. *Nat. Genet.* *47*, 479–485.
- Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* *431*, 96.
- Bourc'his, D., Xu, G.-L., Lin, C.-S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the Establishment of Maternal Genomic Imprints. *Science* *294*, 2536–2539.
- Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* *441*, 349.
- Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Names, A., Temper, V., Razin, A., and Cedar, H. (1994). Spl elements protect a CpG island from de novo methylation. *Nature* *371*, 435.
- Brenner, C., Deplus, R., Didelot, C., Loriot, A., Viré, E., Smet, C.D., Gutierrez, A., Danovi, D., Bernard, D., Boon, T., et al. (2005). Myc represses transcription through recruitment of DNA methyltransferase corepressor. *EMBO J.* *24*, 336–346.
- Brinkman, A.B., Gu, H., Bartels, S.J.J., Zhang, Y., Matarese, F., Simmer, F., Marks, H., Bock, C., Gnirke, A., Meissner, A., et al. (2012). Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res.* *22*, 1128–1138.
- Bulut-Karslioglu, A., De La Rosa-Velázquez, I.A., Ramirez, F., Barenboim, M., Onishi-Seebacher, M., Arand, J., Galán, C., Winter, G.E., Engist, B., Gerle, B., et al. (2014). Suv39h-Dependent H3K9me3 Marks Intact Retrotransposons and Silences LINE Elements in Mouse Embryonic Stem Cells. *Mol. Cell* *55*, 277–290.
- Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science* *298*, 1039–1043.
- Cardoso, M.C., and Leonhardt, H. (1999). DNA Methyltransferase Is Actively Retained in the Cytoplasm during Early Development. *J. Cell Biol.* *147*, 25–32.
- Cartwright, P., Müller, H., Wagener, C., Holm, K., and Helin, K. (1998). E2F-6: a novel member of the E2F family is an inhibitor of E2F-dependent transcription. *Oncogene* *17*, 611–623.
- Challen, G.A., Sun, D., Jeong, M., Luo, M., Jelinek, J., Berg, J.S., Bock, C., Vasanthakumar, A., Gu, H., Xi, Y., et al. (2012). Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat. Genet.* *44*, 23–31.
- 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.* *99*, 16916–16921.

- Chen, T., Ueda, Y., Xie, S., and Li, E. (2002). A Novel Dnmt3a Isoform Produced from an Alternative Promoter Localizes to Euchromatin and Its Expression Correlates with Active de Novo Methylation. *J. Biol. Chem.* *277*, 38746–38754.
- Chen, T., Ueda, Y., Dodge, J.E., Wang, Z., and Li, E. (2003). Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* *23*, 5594–5605.
- Cheng, J., Yang, H., Fang, J., Ma, L., Gong, R., Wang, P., Li, Z., and Xu, Y. (2015). Molecular mechanism for USP7-mediated DNMT1 stabilization by acetylation. *Nat. Commun.* *6*, 7023.
- Cheutin, T., McNairn, A.J., Jenuwein, T., Gilbert, D.M., Singh, P.B., and Misteli, T. (2003). Maintenance of Stable Heterochromatin Domains by Dynamic HP1 Binding. *Science* *299*, 721–725.
- Chiappinelli, K.B., Strissel, P.L., Desrichard, A., Li, H., Henke, C., Akman, B., Hein, A., Rote, N.S., Cope, L.M., Snyder, A., et al. (2015). Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell* *162*, 974–986.
- Chin, H.G., Estève, P.-O., Pradhan, M., Benner, J., Patnaik, D., Carey, M.F., and Pradhan, S. (2007). Automethylation of G9a and its implication in wider substrate specificity and HP1 binding. *Nucleic Acids Res.* *35*, 7313–7323.
- Cirio, M.C., Martel, J., Mann, M., Toppings, M., Bartolomei, M., Trasler, J., and Chaillet, J.R. (2008). DNA methyltransferase 1o functions during preimplantation development to preclude a profound level of epigenetic variation. *Dev. Biol.* *324*, 139–150.
- Coluccio, A., Ecco, G., Duc, J., Offner, S., Turelli, P., and Trono, D. (2018). Individual retrotransposon integrants are differentially controlled by KZFP/KAP1-dependent histone methylation, DNA methylation and TET-mediated hydroxymethylation in naïve embryonic stem cells. *Epigenetics Chromatin* *11*, 7.
- Comet, I., Riising, E.M., Leblanc, B., and Helin, K. (2016). Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. *Nat. Rev. Cancer* *16*, 803–810.
- Cooper, S., Grijzenhout, A., Underwood, E., Ancelin, K., Zhang, T., Nesterova, T.B., Anil-Kirmizitas, B., Bassett, A., Kooistra, S.M., Agger, K., et al. (2016). Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. *Nat. Commun.* *7*, 13661.
- Cossec, J.-C., Theurillat, I., Chica, C., Búa Aguín, S., Gaume, X., Andrieux, A., Iturbide, A., Jouvion, G., Li, H., Bossis, G., et al. (2018). SUMO Safeguards Somatic and Pluripotent Cell Identities by Enforcing Distinct Chromatin States. *Cell Stem Cell* *23*, 742-757.e8.
- Courel, M., Friesenhahn, L., and Lees, J.A. (2008). E2f6 and Bmi1 cooperate in axial skeletal development. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* *237*, 1232–1242.

- Dai, H.-Q., Wang, B.-A., Yang, L., Chen, J.-J., Zhu, G.-C., Sun, M.-L., Ge, H., Wang, R., Chapman, D.L., Tang, F., et al. (2016). TET-mediated DNA demethylation controls gastrulation by regulating Lefty–Nodal signalling. *Nature* 538, 528–532.
- Dawlaty, M.M., Breiling, A., Le, T., Barrasa, M.I., Raddatz, G., Gao, Q., Powell, B.E., Cheng, A.W., Faull, K.F., Lyko, F., et al. (2014). Loss of Tet Enzymes Compromises Proper Differentiation of Embryonic Stem Cells. *Dev. Cell* 29, 102–111.
- De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* 49, 941–945.
- Delhommeau, F., Dupont, S., Della Valle, V., 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.
- Deniz, Ö., de la Rica, L., Cheng, K.C.L., Spensberger, D., and Branco, M.R. (2018). SETDB1 prevents TET2-dependent activation of IAP retroelements in naïve embryonic stem cells. *Genome Biol.* 19, 6.
- Desai, P., Mencia-Trinchant, N., Savenkov, O., Simon, M.S., Cheang, G., Lee, S., Samuel, M., Ritchie, E.K., Guzman, M.L., Ballman, K.V., et al. (2018). Somatic mutations precede acute myeloid leukemia years before diagnosis. *Nat. Med.* 24, 1015.
- 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.
- Dhayalan, A., Tamas, R., Bock, I., Tattermusch, A., Dimitrova, E., Kudithipudi, S., Ragozin, S., and Jeltsch, A. (2011). The ATRX-ADD domain binds to H3 tail peptides and reads the combined methylation state of K4 and K9. *Hum. Mol. Genet.* 20, 2195–2203.
- Dodge, J.E., Okano, M., Dick, F., Tsujimoto, N., Chen, T., Wang, S., Ueda, Y., Dyson, N., and Li, E. (2005). Inactivation of Dnmt3b in Mouse Embryonic Fibroblasts Results in DNA Hypomethylation, Chromosomal Instability, and Spontaneous Immortalization. *J. Biol. Chem.* 280, 17986–17991.
- Domcke, S., Bardet, A.F., Adrian Ginno, P., Hartl, D., Burger, L., and Schübeler, D. (2015). Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature* 528, 575–579.
- 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.
- Dou, Y., Milne, T.A., Ruthenburg, A.J., Lee, S., Lee, J.W., Verdine, G.L., Allis, C.D., and Roeder, R.G. (2006). Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat. Struct. Mol. Biol.* 13, 713.

- Dovey, O.M., Foster, C.T., and Cowley, S.M. (2010). Histone deacetylase 1 (HDAC1), but not HDAC2, controls embryonic stem cell differentiation. *Proc. Natl. Acad. Sci.* *107*, 8242–8247.
- Duymich, C.E., Charlet, J., Yang, X., Jones, P.A., and Liang, G. (2016). DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells. *Nat. Commun.* *7*, 11453.
- Eckersley-Maslin, M.A., Svensson, V., Krueger, C., Stubbs, T.M., Giehr, P., Krueger, F., Miragaia, R.J., Kyriakopoulos, C., Berrens, R.V., Milagre, I., et al. (2016). MERVL/Zscan4 Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs. *Cell Rep.* *17*, 179–192.
- Ehrhart, F., Coort, S.L.M., Cirillo, E., Smeets, E., Evelo, C.T., and Curfs, L.M.G. (2016). Rett syndrome – biological pathways leading from MECP2 to disorder phenotypes. *Orphanet J. Rare Dis.* *11*.
- Ehrlich, M., Sanchez, C., Shao, C., Nishiyama, R., Kehrl, J., Kuick, R., Kubota, T., and Hanash, S.M. (2008). ICF, an immunodeficiency syndrome: DNA methyltransferase 3B involvement, chromosome anomalies, and gene dysregulation. *Autoimmunity* *41*, 253–271.
- Endoh, M., Endo, T.A., Endoh, T., Fujimura, Y., Ohara, O., Toyoda, T., Otte, A.P., Okano, M., Brockdorff, N., Vidal, M., et al. (2008). Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development* *135*, 1513–1524.
- Endoh, M., Endo, T.A., Shinga, J., Hayashi, K., Farcas, A., Ma, K.-W., Ito, S., Sharif, J., Endoh, T., Onaga, N., et al. (2017). PCGF6-PRC1 suppresses premature differentiation of mouse embryonic stem cells by regulating germ cell-related genes. *ELife* *6*, e21064.
- Entrevan, M., Schuettengruber, B., and Cavalli, G. (2016). Regulation of Genome Architecture and Function by Polycomb Proteins. *Trends Cell Biol.* *26*, 511–525.
- Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* *292*, 154.
- Farcas, A.M., Blackledge, N.P., Sudbery, I., Long, H.K., McGouran, J.F., Rose, N.R., Lee, S., Sims, D., Cerase, A., Sheahan, T.W., et al. (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *ELife* *1*, e00205.
- Farlik, M., Halbritter, F., Müller, F., Choudry, F.A., Ebert, P., Klughammer, J., Farrow, S., Santoro, A., Ciaurro, V., Mathur, A., et al. (2016). DNA Methylation Dynamics of Human Hematopoietic Stem Cell Differentiation. *Cell Stem Cell* *19*, 808–822.
- Farthing, C.R., Ficiz, G., Ng, R.K., Chan, C.-F., Andrews, S., Dean, W., Hemberger, M., and Reik, W. (2008). Global Mapping of DNA Methylation in Mouse Promoters Reveals Epigenetic Reprogramming of Pluripotency Genes. *PLOS Genet.* *4*, e1000116.
- Feinberg, A.P., and Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* *301*, 89.

- Ferrari, K.J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., Fischle, W., Bonaldi, T., and Pasini, D. (2014). Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity. *Mol. Cell* *53*, 49–62.
- Ferry, L., Fournier, A., Tsusaka, T., Adelmant, G., Shimazu, T., Matano, S., Kirsh, O., Amouroux, R., Dohmae, N., Suzuki, T., et al. (2017). Methylation of DNA Ligase 1 by G9a/GLP Recruits UHRF1 to Replicating DNA and Regulates DNA Methylation. *Mol. Cell* *67*, 550-565.e5.
- Ficz, G., Hore, T.A., Santos, F., Lee, H.J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.-L., Walter, J., et al. (2013). FGF Signaling Inhibition in ESCs Drives Rapid Genome-wide Demethylation to the Epigenetic Ground State of Pluripotency. *Cell Stem Cell* *13*, 351–359.
- Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev.* *17*, 1870–1881.
- Fukuda, K., Okuda, A., Yusa, K., and Shinkai, Y. (2018). A CRISPR knockout screen identifies SETDB1-target retroelement silencing factors in embryonic stem cells. *Genome Res.*
- Fursova, N.A., Blackledge, N.P., Nakayama, M., Ito, S., Koseki, Y., Farcas, A.M., King, H.W., Koseki, H., and Klose, R.J. (2019). Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression. *Mol. Cell.*
- Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. *Mol. Cell* *45*, 344–356.
- Garvilles, R.G., Hasegawa, T., Kimura, H., Sharif, J., Muto, M., Koseki, H., Takahashi, S., Suetake, I., and Tajima, S. (2015). Dual Functions of the RFTS Domain of Dnmt1 in Replication-Coupled DNA Methylation and in Protection of the Genome from Aberrant Methylation. *PLOS ONE* *10*, e0137509.
- Giangrande, P.H., Zhu, W., Schlisio, S., Sun, X., Mori, S., Gaubatz, S., and Nevins, J.R. (2004). A role for E2F6 in distinguishing G1/S- and G2/M-specific transcription. *Genes Dev.* *18*, 2941–2951.
- Gilsbach, R., Preissl, S., Grüning, B.A., Schnick, T., Burger, L., Benes, V., Würch, A., Bönisch, U., Günther, S., Backofen, R., et al. (2014). Dynamic DNA methylation orchestrates cardiomyocyte development, maturation and disease. *Nat. Commun.* *5*, 5288.
- Gjerstorff, M.F., Andersen, M.H., and Ditzel, H.J. (2015). Oncogenic cancer/testis antigens: prime candidates for immunotherapy. *Oncotarget* *6*, 15772–15787.
- 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.

Goodier, J.L., and Kazazian, H.H. (2008). Retrotransposons Revisited: The Restraint and Rehabilitation of Parasites. *Cell* 135, 23–35.

Gordon, C.A., Hartono, S.R., and Chédin, F. (2013). Inactive DNMT3B Splice Variants Modulate De Novo DNA Methylation. *PLOS ONE* 8, e69486.

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.

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, C., Liu, S., Wu, Q., Zhang, L., and Guo, F. (2019). Integrative single-cell analysis of transcriptome, DNA methylome and chromatin accessibility in mouse oocytes. *Cell Res.* 29, 110–123.

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.

Guibert, S., Forné, T., and Weber, M. (2012). Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Res.* 22, 633–641.

Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., Wu, X., Wen, L., Gu, T.-P., Hu, B., et al. (2014). Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote. *Cell Stem Cell* 15, 447–459.

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.

Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H.D., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., et al. (2013). Whole-Genome Bisulfite Sequencing of Two Distinct Interconvertible DNA Methylomes of Mouse Embryonic Stem Cells. *Cell Stem Cell* 13, 360–369.

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

Hagarman, J.A., Motley, M.P., Kristjansdottir, K., and Soloway, P.D. (2013). Coordinate regulation of DNA methylation and H3K27me3 in mouse embryonic stem cells. *PloS One* 8, e53880.

- Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of Cancer: The Next Generation. *Cell* 144, 646–674.
- Hargan-Calvopina, J., Taylor, S., Cook, H., Hu, Z., Lee, S.A., Yen, M.-R., Chiang, Y.-S., Chen, P.-Y., and Clark, A.T. (2016). Stage-Specific Demethylation in Primordial Germ Cells Safeguards against Precocious Differentiation. *Dev. Cell* 39, 75–86.
- Hartl, D., Krebs, A.R., Grand, R.S., Baubec, T., Isbel, L., Wirbelauer, C., Burger, L., and Schübeler, D. (2019). CG dinucleotides enhance promoter activity independent of DNA methylation. *Genome Res.*
- Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* 40, 4841–4849.
- Hauri, S., Comoglio, F., Seimiya, M., Gerstung, M., Glatter, T., Hansen, K., Aebbersold, R., Paro, R., Gstaiger, M., and Beisel, C. (2016). A High-Density Map for Navigating the Human Polycomb Complexome. *Cell Rep.* 17, 583–595.
- Hayashi, K., Lopes, S.M.C. de S., Tang, F., and Surani, M.A. (2008). Dynamic Equilibrium and Heterogeneity of Mouse Pluripotent Stem Cells with Distinct Functional and Epigenetic States. *Cell Stem Cell* 3, 391–401.
- 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.
- Hendrickson, P.G., Doráis, J.A., Grow, E.J., Whiddon, J.L., Lim, J.-W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., et al. (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* 49, 925–934.
- Herlihy, A.E., and De Bruin, R.A.M. (2017). The Role of the Transcriptional Response to DNA Replication Stress. *Genes* 8, 92.
- Hill, P.W.S., Leitch, H.G., Requena, C.E., Sun, Z., Amouroux, R., Roman-Trufero, M., Borkowska, M., Terragni, J., Vaisvila, R., Linnett, S., et al. (2018). Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. *Nature* 555, 392–396.
- Hiragami-Hamada, K., Soeroes, S., Nikolov, M., Wilkins, B., Kreuz, S., Chen, C., Rosa-Velázquez, I.A.D.L., Zenn, H.M., Kost, N., Pohl, W., et al. (2016). Dynamic and flexible H3K9me3 bridging via HP1 β dimerization establishes a plastic state of condensed chromatin. *Nat. Commun.* 7, 11310.
- 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.

Hisada, K., Sánchez, C., Endo, T.A., Endoh, M., Román-Trufero, M., Sharif, J., Koseki, H., and Vidal, M. (2012). RYBP Represses Endogenous Retroviruses and Preimplantation- and Germ Line-Specific Genes in Mouse Embryonic Stem Cells. *Mol. Cell. Biol.* 32, 1139–1149.

Hishida, T., Nozaki, Y., Nakachi, Y., Mizuno, Y., Okazaki, Y., Ema, M., Takahashi, S., Nishimoto, M., and Okuda, A. (2011). Indefinite Self-Renewal of ESCs through Myc/Max Transcriptional Complex-Independent Mechanisms. *Cell Stem Cell* 9, 37–49.

Hitt, M.M., Wu, T.L., Cohen, G., and Linn, S. (1988). De novo and maintenance DNA methylation by a mouse plasmacytoma cell DNA methyltransferase. *J. Biol. Chem.* 263, 4392–4399.

Hon, G.C., Hawkins, R.D., Caballero, O.L., Lo, C., Lister, R., Pelizzola, M., Valsesia, A., Ye, Z., Kuan, S., Edsall, L.E., et al. (2012). Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.* 22, 246–258.

Hon, G.C., Rajagopal, N., Shen, Y., McCleary, D.F., Yue, F., Dang, M.D., and Ren, B. (2013). Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat. Genet.* 45, 1198–1206.

Howell, C.Y., Bestor, T.H., Ding, F., Latham, K.E., Mertineit, C., Trasler, J.M., and Chaillet, J.R. (2001). Genomic Imprinting Disrupted by a Maternal Effect Mutation in the Dnmt1 Gene. *Cell* 104, 829–838.

Hu, G., Kim, J., Xu, Q., Leng, Y., Orkin, S.H., and Elledge, S.J. (2009). A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev.* 23, 837–848.

Huang, C., Su, T., Xue, Y., Cheng, C., Lay, F.D., McKee, R.A., Li, M., Vashisht, A., Wohlschlegel, J., Novitch, B.G., et al. (2017). Cbx3 maintains lineage specificity during neural differentiation. *Genes Dev.* 31, 241–246.

Huang, Y., Zhao, W., Wang, C., Zhu, Y., Liu, M., Tong, H., Xia, Y., Jiang, Q., and Qin, J. (2018). Combinatorial Control of Recruitment of a Variant PRC1.6 Complex in Embryonic Stem Cells. *Cell Rep.* 22, 3032–3043.

Hunkapiller, J., Shen, Y., Diaz, A., Cagney, G., McCleary, D., Ramalho-Santos, M., Krogan, N., Ren, B., Song, J.S., and Reiter, J.F. (2012). Polycomb-Like 3 Promotes Polycomb Repressive Complex 2 Binding to CpG Islands and Embryonic Stem Cell Self-Renewal. *PLOS Genet.* 8, e1002576.

Hurlin, P.J., Steingrimsson, E., Copeland, N.G., Jenkins, N.A., and Eisenman, R.N. (1999). Mga, a dual-specificity transcription factor that interacts with Max and contains a T-domain DNA-binding motif. *EMBO J.* 18, 7019–7028.

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.

Illingworth, R.S., Moffat, M., Mann, A.R., Read, D., Hunter, C.J., Pradeepa, M.M., Adams, I.R., and Bickmore, W.A. (2015). The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development. *Genes Dev.* *29*, 1897–1902.

Inoue, A., Shen, L., Dai, Q., He, C., and Zhang, Y. (2011). Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Res.* *21*, 1670–1676.

Inoue, K., Ichiyanagi, K., Fukuda, K., Glinka, M., and Sasaki, H. (2017). Switching of dominant retrotransposon silencing strategies from posttranscriptional to transcriptional mechanisms during male germ-cell development in mice. *PLOS Genet.* *13*, e1006926.

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.

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

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

Jachowicz, J.W., Bing, X., Pontabry, J., Bošković, A., Rando, O.J., and Torres-Padilla, M.-E. (2017). LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.* *49*, 1502–1510.

Jackson, M., Krassowska, A., Gilbert, N., Chevassut, T., Forrester, L., Ansell, J., and Ramsahoye, B. (2004). Severe Global DNA Hypomethylation Blocks Differentiation and Induces Histone Hyperacetylation in Embryonic Stem Cells. *Mol. Cell. Biol.* *24*, 8862–8871.

Jamaladdin, S., Kelly, R.D.W., O'Regan, L., Dovey, O.M., Hodson, G.E., Millard, C.J., Portolano, N., Fry, A.M., Schwabe, J.W.R., and Cowley, S.M. (2014). Histone deacetylase (HDAC) 1 and 2 are essential for accurate cell division and the pluripotency of embryonic stem cells. *Proc. Natl. Acad. Sci.* *111*, 9840–9845.

Jeltsch, A., and Jurkowska, R.Z. (2014). New concepts in DNA methylation. *Trends Biochem. Sci.* *39*, 310–318.

Jeong, M., Sun, D., Luo, M., Huang, Y., Challen, G.A., Rodriguez, B., Zhang, X., Chavez, L., Wang, H., Hannah, R., et al. (2014). Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat. Genet.* *46*, 17–23.

Jermann, P., Hoerner, L., Burger, L., and Schübeler, D. (2014). Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation. *Proc. Natl. Acad. Sci.* *111*, E3415–E3421.

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.

Jones, P.A., and Liang, G. (2009). Rethinking how DNA methylation patterns are maintained. *Nat. Rev. Genet.* *10*, 805–811.

Kadoch, C., Williams, R.T., Calarco, J.P., Miller, E.L., Weber, C.M., Braun, S.M.G., Pulice, J.L., Chory, E.J., and Crabtree, G.R. (2017). Dynamics of BAF-Polycomb complex opposition on heterochromatin in normal and oncogenic states. *Nat. Genet.* *49*, 213–222.

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.

Kahn, T.G., Dorafshan, E., Schultheis, D., Zare, A., Stenberg, P., Reim, I., Pirrotta, V., and Schwartz, Y.B. (2016). Interdependence of PRC1 and PRC2 for recruitment to Polycomb Response Elements. *Nucleic Acids Res.* *44*, 10132–10149.

Kalb, R., Latwiel, S., Baymaz, H.I., Jansen, P.W.T.C., Müller, C.W., Vermeulen, M., and Müller, J. (2014). Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat. Struct. Mol. Biol.* *21*, 569–571.

Kalmar, T., Lim, C., Hayward, P., Muñoz-Descalzo, S., Nichols, J., Garcia-Ojalvo, J., and Arias, A.M. (2009). Regulated Fluctuations in Nanog Expression Mediate Cell Fate Decisions in Embryonic Stem Cells. *PLOS Biol.* *7*, e1000149.

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.

Karemaker, I.D., and Vermeulen, M. (2018). Single-Cell DNA Methylation Profiling: Technologies and Biological Applications. *Trends Biotechnol.* *36*, 952–965.

Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L., Jones, S., Hirst, M., et al. (2011). DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs. *Cell Stem Cell* *8*, 676–687.

Kazachenka, A., Bertozzi, T.M., Sjoberg-Herrera, M.K., Walker, N., Gardner, J., Gunning, R., Pahita, E., Adams, S., Adams, D., and Ferguson-Smith, A.C. (2018). Identification, Characterization, and Heritability of Murine Metastable Epialleles: Implications for Non-genetic Inheritance. *Cell* *175*, 1259-1271.e13.

Kehoe, S.M., Oka, M., Hankowski, K.E., Reichert, N., Garcia, S., McCarrey, J.R., Gaubatz, S., and Terada, N. (2008). A conserved E2F6-binding element in murine meiosis-specific gene promoters. *Biol. Reprod.* *79*, 921–930.

Kelsey Gavin, and Feil Robert (2013). New insights into establishment and maintenance of DNA methylation imprints in mammals. *Philos. Trans. R. Soc. B Biol. Sci.* *368*, 20110336.

Kim, H., Kang, K., and Kim, J. (2009). AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. *Nucleic Acids Res.* *37*, 2940–2950.

- King, A.D., Huang, K., Rubbi, L., Liu, S., Wang, C.-Y., Wang, Y., Pellegrini, M., and Fan, G. (2016). Reversible Regulation of Promoter and Enhancer Histone Landscape by DNA Methylation in Mouse Embryonic Stem Cells. *Cell Rep.* *17*, 289–302.
- Kloet, S.L., Makowski, M.M., Baymaz, H.I., van Voorthuijsen, L., Karemaker, I.D., Santanach, A., Jansen, P.W.T.C., Di Croce, L., and Vermeulen, M. (2016). The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. *Nat. Struct. Mol. Biol.* *advance online publication*.
- Kobayashi, H., Sakurai, T., Miura, F., Imai, M., Mochiduki, K., Yanagisawa, E., Sakashita, A., Wakai, T., Suzuki, Y., Ito, T., et al. (2013). High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Res.* *23*, 616–627.
- Kohn, M.J., Bronson, R.T., Harlow, E., Dyson, N.J., and Yamasaki, L. (2003). Dp1 is required for extra-embryonic development. *Development* *130*, 1295–1305.
- Kohn, M.J., Leung, S.W., Criniti, V., Agromayor, M., and Yamasaki, L. (2004). Dp1 Is Largely Dispensable for Embryonic Development. *Mol. Cell. Biol.* *24*, 7197–7205.
- Kosmider, O., Gelsi-Boyer, V., Cheok, M., Grabar, S., Della-Valle, V., Picard, F., Viguíé, F., Quesnel, B., Beyne-Rauzy, O., Solary, E., et al. (2009). TET2 mutation is an independent favorable prognostic factor in myelodysplastic syndromes (MDSs). *Blood* *114*, 3285–3291.
- Kouzarides, T. (2007). Chromatin Modifications and Their Function. *Cell* *128*, 693–705.
- 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.
- 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*, e1000242.
- Kulis, M., Merkel, A., Heath, S., Queirós, A.C., Schuyler, R.P., Castellano, G., Beekman, R., Raineri, E., Esteve, A., Clot, G., et al. (2015). Whole-genome fingerprint of the DNA methylome during human B cell differentiation. *Nat. Genet.* *47*, 746–756.
- Kumar, S., Cheng, X., Klimasauskas, S., Sha, M., Posfai, J., Roberts, R.J., and Wilson, G.G. (1994). The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res.* *22*, 1–10.
- Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J.T., Sadreyev, R.I., Dekker, J., and Kingston, R.E. (2017). Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. *Mol. Cell* *65*, 432–446.e5.

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.

Lau, M.S., Schwartz, M.G., Kundu, S., Savol, A.J., Wang, P.I., Marr, S.K., Grau, D.J., Schorderet, P., Sadreyev, R.I., Tabin, C.J., et al. (2017). Mutation of a nucleosome compaction region disrupts Polycomb-mediated axial patterning. *Science* 355, 1081–1084.

Leeb, M., and Wutz, A. (2007). Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. *J. Cell Biol.* 178, 219–229.

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.

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

Lesch, B.J., and Page, D.C. (2012). Genetics of germ cell development. *Nat. Rev. Genet.* 13, 781–794.

Leseva, M., Santostefano, K.E., Rosenbluth, A.L., Hamazaki, T., and Terada, N. (2013). E2f6-mediated repression of the meiotic *Stag3* and *Smc1 β* genes during early embryonic development requires *Ezh2* and not the de novo methyltransferase *Dnmt3b*. *Epigenetics Off. J. DNA Methylation Soc.* 8, 873–884.

Leung, D., Du, T., Wagner, U., Xie, W., Lee, A.Y., Goyal, P., Li, Y., Szulwach, K.E., Jin, P., Lorincz, M.C., et al. (2014a). Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase *Setdb1*. *Proc. Natl. Acad. Sci.* 111, 6690–6695.

Leung, D., Du, T., Wagner, U., Xie, W., Lee, A.Y., Goyal, P., Li, Y., Szulwach, K.E., Jin, P., Lorincz, M.C., et al. (2014b). Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase *Setdb1*. *Proc. Natl. Acad. Sci.* 111, 6690–6695.

Ley, T.J., Ding, L., Walter, M.J., McLellan, M.D., Lamprecht, T., Larson, D.E., Kandoth, C., Payton, J.E., Baty, J., Welch, J., et al. (2010). DNMT3A Mutations in Acute Myeloid Leukemia. *N. Engl. J. Med.* 363, 2424–2433.

Leygo, C., Williams, M., Jin, H.C., Chan, M.W.Y., Chu, W.K., Grusch, M., and Cheng, Y.Y. (2017). DNA Methylation as a Noninvasive Epigenetic Biomarker for the Detection of Cancer. *Dis. Markers* 2017.

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

- Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. *Nature* 366, 362.
- Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., and Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. *Genes Dev.* 24, 368–380.
- Li, G., Liu, Y., Zhang, Y., Kubo, N., Yu, M., Fang, R., Kellis, M., and Ren, B. (2019). Joint profiling of DNA methylation and chromatin architecture in single cells. *Nat. Methods* 1–3.
- Li, H., Liefke, R., Jiang, J., Kurland, J.V., Tian, W., Deng, P., Zhang, W., He, Q., Patel, D.J., Bulyk, M.L., et al. (2017). Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature* 549, 287–291.
- Li, T., Wang, L., Du, Y., Xie, S., Yang, X., Lian, F., Zhou, Z., and Qian, C. (2018a). Structural and mechanistic insights into UHRF1-mediated DNMT1 activation in the maintenance DNA methylation. *Nucleic Acids Res.* 46, 3218–3231.
- 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.
- Li, Y., Zheng, H., Wang, Q., Zhou, C., Wei, L., Liu, X., Zhang, W., Zhang, Y., Du, Z., Wang, X., et al. (2018b). Genome-wide analyses reveal a role of Polycomb in promoting hypomethylation of DNA methylation valleys. *Genome Biol.* 19, 18.
- Li, Z., Dai, H., Martos, S.N., Xu, B., Gao, Y., Li, T., Zhu, G., Schones, D.E., and Wang, Z. (2015). Distinct roles of DNMT1-dependent and DNMT1-independent methylation patterns in the genome of mouse embryonic stem cells. *Genome Biol.* 16, 115.
- 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.
- Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., et al. (2013). Global Epigenomic Reconfiguration During Mammalian Brain Development. *Science* 341, 1237905.
- Liu, N., Lee, C.H., Swigut, T., Grow, E., Gu, B., Bassik, M.C., and Wysocka, J. (2018a). Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature* 553, 228–232.
- Liu, S., Brind'Amour, J., Karimi, M.M., Shirane, K., Bogutz, A., Lefebvre, L., Sasaki, H., Shinkai, Y., and Lorincz, M.C. (2014). *Setdb1* is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes Dev.* 28, 2041–2055.
- Liu, X.S., Wu, H., Krzisch, M., Wu, X., Graef, J., Muffat, J., Hnisz, D., Li, C.H., Yuan, B., Xu, C., et al. (2018b). Rescue of Fragile X Syndrome Neurons by DNA Methylation Editing of the *FMR1* Gene. *Cell* 172, 979-992.e6.

- Liu, Y., Siejka-Zielińska, P., Velikova, G., Bi, Y., Yuan, F., Tomkova, M., Bai, C., Chen, L., Schuster-Böckler, B., and Song, C.-X. (2019). Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nat. Biotechnol.* *37*, 424–429.
- Long, H.K., King, H.W., Patient, R.K., Odom, D.T., and Klose, R.J. (2016). Protection of CpG islands from DNA methylation is DNA-encoded and evolutionarily conserved. *Nucleic Acids Res.* *44*, 6693–6706.
- Lorente, M. del M., Marcos-Gutierrez, C., Perez, C., Schoorlemmer, J., Ramirez, A., Magin, T., and Vidal, M. (2000). Loss- and gain-of-function mutations show a polycomb group function for Ring1A in mice. *Development* *127*, 5093–5100.
- Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* *487*, 57–63.
- Maeda, I., Okamura, D., Tokitake, Y., Ikeda, M., Kawaguchi, H., Mise, N., Abe, K., Noce, T., Okuda, A., and Matsui, Y. (2013). Max is a repressor of germ cell-related gene expression in mouse embryonic stem cells. *Nat. Commun.* *4*, 1754.
- Maenohara, S., Unoki, M., Toh, H., Ohishi, H., Sharif, J., Koseki, H., and Sasaki, H. (2017). Role of UHRF1 in de novo DNA methylation in oocytes and maintenance methylation in preimplantation embryos. *PLOS Genet.* *13*, e1007042.
- Maezawa, S., Hasegawa, K., Yukawa, M., Sakashita, A., Alavattam, K.G., Andreassen, P.R., Vidal, M., Koseki, H., Barski, A., and Namekawa, S.H. (2017). Polycomb directs timely activation of germline genes in spermatogenesis. *Genes Dev.* *31*, 1693–1703.
- Mahé, E.A., Madigou, T., Sérandour, A.A., Bizot, M., Avner, S., Chalmel, F., Paliérne, G., Métivier, R., and Salbert, G. (2017). Cytosine modifications modulate the chromatin architecture of transcriptional enhancers. *Genome Res.* *27*, 947–958.
- Maiti, A., and Drohat, A.C. (2011). Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine POTENTIAL IMPLICATIONS FOR ACTIVE DEMETHYLATION OF CpG SITES. *J. Biol. Chem.* *286*, 35334–35338.
- Maksakova, I.A., Romanish, M.T., Gagnier, L., Dunn, C.A., Lagemaat, L.N. van de, and Mager, D.L. (2006). Retroviral Elements and Their Hosts: Insertional Mutagenesis in the Mouse Germ Line. *PLOS Genet.* *2*, e2.
- Maksakova, I.A., Thompson, P.J., Goyal, P., Jones, S.J., Singh, P.B., Karimi, M.M., and Lorincz, M.C. (2013). Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. *Epigenetics Chromatin* *6*, 15.
- Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C.L., Dynlacht, B.D., and Reinberg, D. (2008). Ezh1 and Ezh2 Maintain Repressive Chromatin through Different Mechanisms. *Mol. Cell* *32*, 503–518.

- Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury III, W.J., Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* *461*, 762–767.
- Marks, H., Kalkan, T., Menafrá, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Francis Stewart, A., Smith, A., et al. (2012). The Transcriptional and Epigenomic Foundations of Ground State Pluripotency. *Cell* *149*, 590–604.
- 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.
- Mattout, A., Aaronson, Y., Sailaja, B.S., Raghu Ram, E.V., Harikumar, A., Mallm, J.-P., Sim, K.H., Nissim-Rafinia, M., Supper, E., Singh, P.B., et al. (2015). Heterochromatin Protein 1 β (HP1 β) has distinct functions and distinct nuclear distribution in pluripotent versus differentiated cells. *Genome Biol.* *16*, 213.
- Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. *Nature* *403*, 501.
- 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.
- Messerschmidt, D.M., Vries, W. de, Ito, M., Solter, D., Ferguson-Smith, A., and Knowles, B.B. (2012). Trim28 Is Required for Epigenetic Stability During Mouse Oocyte to Embryo Transition. *Science* *335*, 1499–1502.
- Messerschmidt, D.M., Knowles, B.B., and Solter, D. (2014). DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes Dev.* *28*, 812–828.
- von Meyenn, F., Iurlaro, M., Habibi, E., Liu, N.Q., Salehzadeh-Yazdi, A., Santos, F., Petrini, E., Milagre, I., Yu, M., Xie, Z., et al. (2016). Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells. *Mol. Cell* *62*, 848–861.
- 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.
- Min, J., Zhang, Y., and Xu, R.-M. (2003). Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev.* *17*, 1823–1828.
- 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.
- Montgomery, R.L., Davis, C.A., Potthoff, M.J., Haberland, M., Fielitz, J., Qi, X., Hill, J.A., Richardson, J.A., and Olson, E.N. (2007). Histone deacetylases 1 and 2 redundantly regulate cardiac morphogenesis, growth, and contractility. *Genes Dev.* 21, 1790–1802.
- Morey, L., Aloia, L., Cozzuto, L., Benitah, S.A., and Di Croce, L. (2013). RYBP and Cbx7 Define Specific Biological Functions of Polycomb Complexes in Mouse Embryonic Stem Cells. *Cell Rep.* 3, 60–69.
- Morkel, M., Wenkel, J., Bannister, A.J., Kouzarides, T., and Hagemeyer, C. (1997). An E2F-like repressor of transcription. *Nature* 390, 567.
- 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.
- Mujtaba, S., Manzur, K.L., Gurnon, J.R., Kang, M., Van Etten, J.L., and Zhou, M.-M. (2008). Epigenetic transcriptional repression of cellular genes by a viral SET protein. *Nat. Cell Biol.* 10, 1114–1122.
- Nakamura, T., Arai, Y., Umehara, H., Masuhara, M., Kimura, T., Taniguchi, H., Sekimoto, T., Ikawa, M., Yoneda, Y., Okabe, M., et al. (2007). PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nat. Cell Biol.* 9, 64–71.
- Nan, X., Ng, H.-H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393, 386.
- Napoles, M. de, Mermoud, J.E., Wakao, R., Tang, Y.A., Endoh, M., Appanah, R., Nesterova, T.B., Silva, J., Otte, A.P., Vidal, M., et al. (2004). Polycomb Group Proteins Ring1A/B Link Ubiquitylation of Histone H2A to Heritable Gene Silencing and X Inactivation. *Dev. Cell* 7, 663–676.
- Narlikar, G.J., Sundaramoorthy, R., and Owen-Hughes, T. (2013). Mechanisms and Functions of ATP-Dependent Chromatin-Remodeling Enzymes. *Cell* 154, 490–503.
- Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., and Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. *Nature* 543, 72–77.
- 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.
- Oberley, M.J., Inman, D.R., and Farnham, P.J. (2003). E2F6 Negatively Regulates BRCA1 in Human Cancer Cells without Methylation of Histone H3 on Lysine 9. *J. Biol. Chem.* 278, 42466–42476.

- 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., 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.
- Okano, M., Xie, S., and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat. Genet.* *19*, 219.
- 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.
- Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* *448*, 714–717.
- 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.
- Panning, B., and Jaenisch, R. (1996). DNA hypomethylation can activate Xist expression and silence X-linked genes. *Genes Dev.* *10*, 1991–2002.
- Pasini, D., Bracken, A.P., Jensen, M.R., Denchi, E.L., and Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J.* *23*, 4061–4071.
- Peat, J.R., Dean, W., Clark, S.J., Krueger, F., Smallwood, S.A., Ficz, G., Kim, J.K., Marioni, J.C., Hore, T.A., and Reik, W. (2014). Genome-wide Bisulfite Sequencing in Zygotes Identifies Demethylation Targets and Maps the Contribution of TET3 Oxidation. *Cell Rep.* *9*, 1990–2000.
- Pengelly, A.R., Kalb, R., Finkl, K., and Müller, J. (2015). Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. *Genes Dev.* *29*, 1487–1492.
- Perino, M., Mierlo, G. van, Karemaker, I.D., Genesen, S. van, Vermeulen, M., Marks, H., Heeringen, S.J. van, and Veenstra, G.J.C. (2018). MTF2 recruits Polycomb Repressive Complex 2 by helical-shape-selective DNA binding. *Nat. Genet.* *50*, 1002.
- Pfeifer, G.P. (2018). Defining Driver DNA Methylation Changes in Human Cancer. *Int. J. Mol. Sci.* *19*.
- Pijuan-Sala, B., Griffiths, J.A., Guibentif, C., Hiscock, T.W., Jawaid, W., Calero-Nieto, F.J., Mulas, C., Ibarra-Soria, X., Tyser, R.C.V., Ho, D.L.L., et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* *566*, 490–495.

Pirity, M.K., Locker, J., and Schreiber-Agus, N. (2005). Rybp/DEDAF Is Required for Early Postimplantation and for Central Nervous System Development. *Mol. Cell. Biol.* *25*, 7193–7202.

Pohlers, M., Truss, M., Frede, U., Scholz, A., Strehle, M., Kuban, R.-J., Hoffmann, B., Morkel, M., Birchmeier, C., and Hagemeyer, C. (2005). A Role for E2F6 in the Restriction of Male-Germ-Cell-Specific Gene Expression. *Curr. Biol.* *15*, 1051–1057.

Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* *463*, 1101–1105.

Pradhan, S., Bacolla, A., Wells, R.D., and Roberts, R.J. (1999). Recombinant Human DNA (Cytosine-5) Methyltransferase I. EXPRESSION, PURIFICATION, AND COMPARISON OF DE NOVO AND MAINTENANCE METHYLATION. *J. Biol. Chem.* *274*, 33002–33010.

Qin, J., Whyte, W.A., Anderssen, E., Apostolou, E., Chen, H.-H., Akbarian, S., Bronson, R.T., Hochedlinger, K., Ramaswamy, S., Young, R.A., et al. (2012). The Polycomb Group Protein L3mbtl2 Assembles an Atypical PRC1-Family Complex that Is Essential in Pluripotent Stem Cells and Early Development. *Cell Stem Cell* *11*, 319–332.

Qin, W., Leonhardt, H., and Spada, F. (2011). Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1. *J. Cell. Biochem.* *112*, 439–444.

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.

Rasmussen, K.D., and Helin, K. (2016). Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev.* *30*, 733–750.

Robbez-Masson, L., Tie, C.H.C., Conde, L., Tunbak, H., Husovsky, C., Tchasovnikarova, I.A., Timms, R.T., Herrero, J., Lehner, P.J., and Rowe, H.M. (2018). The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. *Genome Res.*

Rodriguez-Terrones, D., Gaume, X., Ishiuchi, T., Weiss, A., Kopp, A., Kruse, K., Penning, A., Vaquerizas, J.M., Brino, L., and Torres-Padilla, M.-E. (2018). A molecular roadmap for the emergence of early-embryonic-like cells in culture. *Nat. Genet.* *50*, 106.

Rose, N.R., King, H.W., Blackledge, N.P., Fursova, N.A., Ember, K.J., Fischer, R., Kessler, B.M., and Klose, R.J. (2016). RYBP stimulates PRC1 to shape chromatin-based communication between Polycomb repressive complexes. *ELife* *5*, e18591.

Rountree, M.R., Bachman, K.E., and Baylin, S.B. (2000). DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nat. Genet.* *25*, 269.

- Rousseaux, S., Debernardi, A., Jacquiau, B., Vitte, A.-L., Vesin, A., Nagy-Mignotte, H., Moro-Sibilot, D., Brichon, P.-Y., Lantuejoul, S., Hainaut, P., et al. (2013). Ectopic Activation of Germline and Placental Genes Identifies Aggressive Metastasis-Prone Lung Cancers. *Sci. Transl. Med.* *5*, 186ra66-186ra66.
- Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P.V., Layard-Liesching, H., Verp, S., Marquis, J., et al. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* *463*, 237–240.
- Rowe, H.M., Friedli, M., Offner, S., Verp, S., Mesnard, D., Marquis, J., Aktas, T., and Trono, D. (2013). De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. *Development* *140*, 519–529.
- Rowley, M.J., and Corces, V.G. (2018). Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* *19*, 789.
- Ruan, J., Ouyang, H., Amaya, M.F., Ravichandran, M., Loppnau, P., Min, J., and Zang, J. (2012). Structural Basis of the Chromodomain of Cbx3 Bound to Methylated Peptides from Histone H1 and G9a. *PLOS ONE* *7*, e35376.
- Sánchez, C., Sánchez, I., Demmers, J.A.A., Rodriguez, P., Strouboulis, J., and Vidal, M. (2007). Proteomics Analysis of Ring1B/Rnf2 Interactors Identifies a Novel Complex with the Fbxl10/Jhdm1B Histone Demethylase and the Bcl6 Interacting Corepressor. *Mol. Cell. Proteomics* *6*, 820–834.
- 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.
- Schermelleh, L., Haemmer, A., Spada, F., Rösing, N., Meilinger, D., Rothbauer, U., Cardoso, M.C., and Leonhardt, H. (2007). Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res.* *35*, 4301–4312.
- Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben-Shushan, E., Reubinoff, B.E., et al. (2007). Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat. Genet.* *39*, 232–236.
- Schoenfelder, S., Sugar, R., Dimond, A., Javierre, B.-M., Armstrong, H., Mifsud, B., Dimitrova, E., Matheson, L., Tavares-Cadete, F., Furlan-Magaril, M., et al. (2015). Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat. Genet.* *47*, 1179–1186.
- Schuettengruber, B., Bourbon, H.-M., Croce, L.D., and Cavalli, G. (2017). Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* *171*, 34–57.
- 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.

- Sendžikaitė, G., Hanna, C.W., Stewart-Morgan, K.R., Ivanova, E., and Kelsey, G. (2019). A DNMT3A PWWP mutation leads to methylation of bivalent chromatin and growth retardation in mice. *Nat. Commun.* *10*, 1884.
- Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., et al. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* *450*, 908–912.
- Shen, L., Wu, H., Diep, D., Yamaguchi, S., D’Alessio, A.C., Fung, H.-L., Zhang, K., and Zhang, Y. (2013). Genome-wide Analysis Reveals TET- and TDG-Dependent 5-Methylcytosine Oxidation Dynamics. *Cell* *153*, 692–706.
- Shen, L., Inoue, A., He, J., Liu, Y., Lu, F., and Zhang, Y. (2014). Tet3 and DNA Replication Mediate Demethylation of Both the Maternal and Paternal Genomes in Mouse Zygotes. *Cell Stem Cell* *15*, 459–471.
- Shen-Li, H., O’Hagan, R.C., Hou, H., Horner, J.W., Lee, H.-W., and DePinho, R.A. (2000). Essential role for Max in early embryonic growth and development. *Genes Dev.* *14*, 17–22.
- Shi, D.-Q., Ali, I., Tang, J., and Yang, W.-C. (2017). New Insights into 5hmC DNA Modification: Generation, Distribution and Function. *Front. Genet.* *8*.
- Shirane, K., Toh, H., Kobayashi, H., Miura, F., Chiba, H., Ito, T., Kono, T., and Sasaki, H. (2013). Mouse Oocyte Methylomes at Base Resolution Reveal Genome-Wide Accumulation of Non-CpG Methylation and Role of DNA Methyltransferases. *PLOS Genet.* *9*, e1003439.
- Shlush, L.I., Zandi, S., Mitchell, A., Chen, W.C., Brandwein, J.M., Gupta, V., Kennedy, J.A., Schimmer, A.D., Schuh, A.C., Yee, K.W., et al. (2014). Identification of pre-leukemic hematopoietic stem cells in acute leukemia. *Nature* *506*, 328–333.
- 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.
- 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, A., Black, J.C., Tanese, N., Pradhan, S., and Carey, M. (2008). HP1-mediated silencing targets Pol II coactivator complexes. *Nat. Struct. Mol. Biol.* *15*, 318–320.
- Smallwood, A., Hon, G.C., Jin, F., Henry, R.E., Espinosa, J.M., and Ren, B. (2012). CBX3 regulates efficient RNA processing genome-wide. *Genome Res.* *22*, 1426–1436.
- Smallwood, S.A., Tomizawa, S., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S.R., and Kelsey, G. (2011). Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat. Genet.* *43*, 811–814.

- Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* *11*, 817–820.
- Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* *14*, 204–220.
- Smith, A.G., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M., and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* *336*, 688.
- 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.
- Sridharan, R., Gonzales-Cope, M., Chronis, C., Bonora, G., McKee, R., Huang, C., Patel, S., Lopez, D., Mishra, N., Pellegrini, M., et al. (2013). Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1 γ in reprogramming to pluripotency. *Nat. Cell Biol.* *15*, 872–882.
- 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.
- Stanton, B.Z., Hodges, C., Calarco, J.P., Braun, S.M.G., Ku, W.L., Kadoch, C., Zhao, K., and Crabtree, G.R. (2017). *Smarca4* ATPase mutations disrupt direct eviction of PRC1 from chromatin. *Nat. Genet.* *49*, 282–288.
- Stewart, K.R., Veselovska, L., Kim, J., Huang, J., Saadeh, H., Tomizawa, S., Smallwood, S.A., Chen, T., and Kelsey, G. (2015). Dynamic changes in histone modifications precede de novo DNA methylation in oocytes. *Genes Dev.* *29*, 2449–2462.
- Stielow, B., Finkernagel, F., Stiewe, T., Nist, A., and Suske, G. (2018). MGA, L3MBTL2 and E2F6 determine genomic binding of the non-canonical Polycomb repressive complex PRC1.6. *PLoS Genet.* *14*, e1007193.
- Storre, J., Elsässer, H.-P., Fuchs, M., Ullmann, D., Livingston, D.M., and Gaubatz, S. (2002). Homeotic transformations of the axial skeleton that accompany a targeted deletion of E2f6. *EMBO Rep.* *3*, 695–700.
- 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.

- Stuart, T., and Satija, R. (2019). Integrative single-cell analysis. *Nat. Rev. Genet.* *20*, 257.
- Suzuki, A., Hirasaki, M., Hishida, T., Wu, J., Okamura, D., Ueda, A., Nishimoto, M., Nakachi, Y., Mizuno, Y., Okazaki, Y., et al. (2016). Loss of MAX results in meiotic entry in mouse embryonic and germline stem cells. *Nat. Commun.* *7*, 11056.
- Suzuki, M., Yamada, T., Kihara-Negishi, F., Sakurai, T., Hara, E., Tenen, D.G., Hozumi, N., and Oikawa, T. (2006). Site-specific DNA methylation by a complex of PU.1 and Dnmt3a/b. *Oncogene* *25*, 2477.
- 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.
- Takada, Y., Naruse, C., Costa, Y., Shirakawa, T., Tachibana, M., Sharif, J., Kezuka-Shiotani, F., Kakiuchi, D., Masumoto, H., Shinkai, Y., et al. (2011). HP1 γ links histone methylation marks to meiotic synapsis in mice. *Development* *138*, 4207–4217.
- Takahashi, N., Coluccio, A., Thorball, C.W., Planet, E., Shi, H., Offner, S., Turelli, P., Imbeault, M., Ferguson-Smith, A.C., and Trono, D. (2019). ZNF445 is a primary regulator of genomic imprinting. *Genes Dev.* *33*, 49–54.
- Talbert, P.B., and Henikoff, S. (2017). Histone variants on the move: substrates for chromatin dynamics. *Nat. Rev. Mol. Cell Biol.* *18*, 115–126.
- Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H., et al. (2012). RYBP-PRC1 Complexes Mediate H2A Ubiquitylation at Polycomb Target Sites Independently of PRC2 and H3K27me3. *Cell* *148*, 664–678.

- Terranova, R., Yokobayashi, S., Stadler, M.B., Otte, A.P., Lohuizen, M. van, Orkin, S.H., and Peters, A.H.F.M. (2008). Polycomb Group Proteins Ezh2 and Rnf2 Direct Genomic Contraction and Imprinted Repression in Early Mouse Embryos. *Dev. Cell* *15*, 668–679.
- Thakur, A., Mackin, S.-J., Irwin, R.E., O’Neill, K.M., Pollin, G., and Walsh, C. (2016). Widespread recovery of methylation at gametic imprints in hypomethylated mouse stem cells following rescue with DNMT3A2. *Epigenetics Chromatin* *9*, 53.
- 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*, 7870.
- Thompson, J.J., Kaur, R., Sosa, C.P., Lee, J.-H., Kashiwagi, K., Zhou, D., and Robertson, K.D. (2018). ZBTB24 is a transcriptional regulator that coordinates with DNMT3B to control DNA methylation. *Nucleic Acids Res.* *46*, 10034–10051.
- Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R.W., Deaton, A., Andrews, R., James, K.D., et al. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* *464*, 1082–1086.
- Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K., and Niwa, H. (2008). Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* *135*, 909–918.
- Trimarchi, J.M., Fairchild, B., Wen, J., and Lees, J.A. (2001). The E2F6 transcription factor is a component of the mammalian Bmi1-containing polycomb complex. *Proc. Natl. Acad. Sci.* *98*, 1519–1524.
- Trojer, P., Cao, A.R., Gao, Z., Li, Y., Zhang, J., Xu, X., Li, G., Losson, R., Erdjument-Bromage, H., Tempst, P., et al. (2011). L3MBTL2 Protein Acts in Concert with PcG Protein-Mediated Monoubiquitination of H2A to Establish a Repressive Chromatin Structure. *Mol. Cell* *42*, 438–450.
- 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.
- Velasco, G., Grillo, G., Touleimat, N., Ferry, L., Ivkovic, I., Ribierre, F., Deleuze, J.-F., Chantalat, S., Picard, C., and Francastel, C. (2018). Comparative methylome analysis of ICF patients identifies heterochromatin loci that require ZBTB24, CDCA7 and HELLS for their methylated state. *Hum. Mol. Genet.* *27*, 2409–2424.
- Verkerk, A.J.M.H., Pieretti, M., Sutcliffe, J.S., Fu, Y.-H., Kuhl, D.P.A., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F., et al. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* *65*, 905–914.

- Verma, N., Pan, H., Doré, L.C., Shukla, A., Li, Q.V., Pelham-Webb, B., Teijeiro, V., González, F., Krivtsov, A., Chang, C.-J., et al. (2018). TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells. *Nat. Genet.* *50*, 83–95.
- Veselovska, L., Smallwood, S.A., Saadeh, H., Stewart, K.R., Krueger, F., Maupetit-Méhouas, S., Arnaud, P., Tomizawa, S.-I., Andrews, S., and Kelsey, G. (2015). Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. *Genome Biol.* *16*, 209.
- Vidal, E., Sayols, S., Moran, S., Guillaumet-Adkins, A., Schroeder, M.P., Royo, R., Orozco, M., Gut, M., Gut, I., Lopez-Bigas, N., et al. (2017). A DNA methylation map of human cancer at single base-pair resolution. *Oncogene* *36*, 5648–5657.
- 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.
- Vincent, J.J., Huang, Y., Chen, P.-Y., Feng, S., Calvopiña, J.H., Nee, K., Lee, S.A., Le, T., Yoon, A.J., Faull, K., et al. (2013). Stage-Specific Roles for Tet1 and Tet2 in DNA Demethylation in Primordial Germ Cells. *Cell Stem Cell* *12*, 470–478.
- Voncken, J.W., Roelen, B.A.J., Roefs, M., Vries, S. de, Verhoeven, E., Marino, S., Deschamps, J., and Lohuizen, M. van (2003). Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc. Natl. Acad. Sci.* *100*, 2468–2473.
- Waddington, C.H. (2012). The Epigenotype. *Int. J. Epidemiol.* *41*, 10–13.
- Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* *20*, 116–117.
- Walter, M., Teissandier, A., Pérez-Palacios, R., and Bourc’his, D. (2016). An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *ELife* *5*, e11418.
- Walton, E.L., Francastel, C., and Velasco, G. (2011). Maintenance of DNA methylation: Dnmt3b joins the dance. *Epigenetics* *6*, 1373–1377.
- Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature* *431*, 873.
- Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W., et al. (2014). Programming and Inheritance of Parental DNA Methylomes in Mammals. *Cell* *157*, 979–991.
- Washkowitz, A.J., Schall, C., Zhang, K., Wurst, W., Floss, T., Mager, J., and Papaioannou, V.E. (2015). Mga is essential for the survival of pluripotent cells during peri-implantation development. *Development* *142*, 31–40.
- Watanabe, D., Suetake, I., Tada, T., and Tajima, S. (2002). Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. *Mech. Dev.* *118*, 187–190.

- Weber, A.R., Krawczyk, C., Robertson, A.B., Kuśnierczyk, A., Vågbø, C.B., Schuermann, D., Klungland, A., and Schär, P. (2016). Biochemical reconstitution of TET1–TDG–BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nat. Commun.* *7*, 10806.
- 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.
- Whiddon, J.L., Langford, A.T., Wong, C.-J., Zhong, J.W., and Tapscott, S.J. (2017). Conservation and innovation in the DUX4-family gene network. *Nat. Genet.* *49*, 935–940.
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* *8*, 973–982.
- Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., and Walter, J. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat. Commun.* *2*, 241.
- Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* *18*, 517–534.
- Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Eve Sun, Y., and Zhang, Y. (2011). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* *473*, 389–393.
- Wu, T.P., Wang, T., Seetin, M.G., Lai, Y., Zhu, S., Lin, K., Liu, Y., Byrum, S.D., Mackintosh, S.G., Zhong, M., et al. (2016). DNA methylation on N^6 -adenine in mammalian embryonic stem cells. *Nature* *532*, 329–333.
- Wu, X., Johansen, J.V., and Helin, K. (2013). Fbxl10/Kdm2b Recruits Polycomb Repressive Complex 1 to CpG Islands and Regulates H2A Ubiquitylation. *Mol. Cell* *49*, 1134–1146.
- Wysocka, J., Swigut, T., Milne, T.A., Dou, Y., Zhang, X., Burlingame, A.L., Roeder, R.G., Brivanlou, A.H., and Allis, C.D. (2005). WDR5 Associates with Histone H3 Methylated at K4 and Is Essential for H3 K4 Methylation and Vertebrate Development. *Cell* *121*, 859–872.
- Xiao, C.-L., Zhu, S., He, M., Chen, D., Zhang, Q., Chen, Y., Yu, G., Liu, J., Xie, S.-Q., Luo, F., et al. (2018). N^6 -Methyladenine DNA Modification in the Human Genome. *Mol. Cell* *71*, 306-318.e7.
- Xie, W., Schultz, M.D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J.W., Tian, S., Hawkins, R.D., Leung, D., et al. (2013). Epigenomic Analysis of Multilineage Differentiation of Human Embryonic Stem Cells. *Cell* *153*, 1134–1148.

- 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.
- Yamaguchi, S., Shen, L., Liu, Y., Sandler, D., and Zhang, Y. (2013). Role of Tet1 in erasure of genomic imprinting. *Nature* 504, 460–464.
- Yamaji, M., Ueda, J., Hayashi, K., Ohta, H., Yabuta, Y., Kurimoto, K., Nakato, R., Yamada, Y., Shirahige, K., and Saitou, M. (2013). PRDM14 Ensures Naive Pluripotency through Dual Regulation of Signaling and Epigenetic Pathways in Mouse Embryonic Stem Cells. *Cell Stem Cell* 12, 368–382.
- Yang, C.-S., Chang, K.-Y., Dang, J., and Rana, T.M. (2016). Polycomb Group Protein Pcgf6 Acts as a Master Regulator to Maintain Embryonic Stem Cell Identity. *Sci. Rep.* 6, 26899.
- Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* 453, 519–523.
- Yokobayashi, S., Liang, C.-Y., Kohler, H., Nestorov, P., Liu, Z., Vidal, M., van Lohuizen, M., Roloff, T.C., and Peters, A.H.F.M. (2013). PRC1 coordinates timing of sexual differentiation of female primordial germ cells. *Nature* 495, 236–240.
- Yu, W., McIntosh, C., Lister, R., Zhu, I., Han, Y., Ren, J., Landsman, D., Lee, E., Briones, V., Terashima, M., et al. (2014). Genome-wide DNA methylation patterns in LSH mutant reveals de-repression of repeat elements and redundant epigenetic silencing pathways. *Genome Res.* 24, 1613–1623.
- Yuan, T., Jiao, Y., de Jong, S., Ophoff, R.A., Beck, S., and Teschendorff, A.E. (2015). An integrative multi-scale analysis of the dynamic DNA methylation landscape in aging. *PLoS Genet.* 11, e1004996.
- 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, T., Termanis, A., Özkan, B., Bao, X.X., Culley, J., Alves, F. de L., Rappsilber, J., Ramsahoye, B., and Stancheva, I. (2016). G9a/GLP Complex Maintains Imprinted DNA Methylation in Embryonic Stem Cells. *Cell Rep.* 15, 77–85.
- 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.

Zhang, Y., Xiang, Y., Yin, Q., Du, Z., Peng, X., Wang, Q., Fidalgo, M., Xia, W., Li, Y., Zhao, Z., et al. (2018). Dynamic epigenomic landscapes during early lineage specification in mouse embryos. *Nat. Genet.* *50*, 96.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.-J., and Lee, J.T. (2008). Polycomb Proteins Targeted by a Short Repeat RNA to the Mouse X Chromosome. *Science* *322*, 750–756.

Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq. *Mol. Cell* *40*, 939–953.

Zhao, W., Tong, H., Huang, Y., Yan, Y., Teng, H., Xia, Y., Jiang, Q., and Qin, J. (2017). Essential Role for Polycomb Group Protein Pcgf6 in Embryonic Stem Cell Maintenance and a Noncanonical Polycomb Repressive Complex 1 (PRC1) Integrity. *J. Biol. Chem.* *292*, 2773–2784.

Zhao, W., Liu, M., Ji, H., Zhu, Y., Wang, C., Huang, Y., Ma, X., Xing, G., Xia, Y., Jiang, Q., et al. (2018). The polycomb group protein Yaf2 regulates the pluripotency of embryonic stem cells in a phosphorylation-dependent manner. *J. Biol. Chem.* *293*, 12793–12804.

Ziller, M.J., Gu, H., Müller, F., Donaghey, J., Tsai, L.T.-Y., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature* *500*, 477–481.

Zylicz, J.J., Dietmann, S., Günesdogan, U., Hackett, J.A., Cougot, D., Lee, C., and Surani, M.A. (2015). Chromatin dynamics and the role of G9a in gene regulation and enhancer silencing during early mouse development. *ELife* *4*, e09571.

(2001). Initial sequencing and analysis of the human genome. *Nature* *409*, 860.

(2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* *420*, 520.

ANNEXES

1. Rôle de la méthylation de l'ADN dans la régulation de l'intégrité transcriptomique

Annexe 1.1 : Statistiques de séquençage des bibliothèques RRBS

Sample	sex	read pairs (2x75 bp)	% bisulfite conversion	median depth	CpG 8X
WT	M	38 871 210	99.68	77X	1 300 606
WT	F	43 050 321	99.69	50X	1 265 422
WT	F	45 345 745	99.64	92X	1 378 414
Dnmt1-/-	M	29 941 359	99.72	35X	1 236 187
Dnmt1-/-	F	30 217 435	99.66	51X	1 330 663
Dnmt1-/-	M	34 954 873	99.65	68X	1 257 331
WT	F	38 276 481	98.53	75X	1 045 177
WT	M	52 382 980	98.89	110X	1 045 177
Dnmt3a-/+	F	46 831 250	99.75	90X	1 347 352
Dnmt3a-/+	M	18 981 740	99.81	52X	1 092 934
Dnmt3a-/+	F	41 117 131	99.68	94X	1 017 974
Dnmt3a-/+ Dnmt3b-/+	M	29 391 912	99.88	65X	1 344 204
Dnmt3a-/+ Dnmt3b-/+	F	28 919 109	99.88	57X	1 304 744
Dnmt3a-/+ Dnmt3b-/+	M	37 617 612	99.75	102X	1 003 420
DKO	M	43 684 564	99.23	89X	1 062 550
DKO	F	39 043 128	99.24	119X	947 512
DKO	M	35 233 905	99.79	56X	1 261 249
MEF_day69_TAM	F	43 685 660	99.68	68X	1 266 329
MEF_day69_TAM	F	23 358 364	99.77	71X	1 243 865
MEF_day69_TAM	F	41 697 544	99.81	93X	1 029 797
MEF_day69_no TAM	F	34 326 404	99.83	69X	1 374 412
MEF_day69_no TAM	F	41 142 254	99.83	87X	1 257 422
MEF_day69_no TAM	F	42 890 920	99.82	88X	1 340 905
MEF_day23_TAM	F	27 915 809	99.66	62x	1 318 796
MEF_day23_TAM	F	38 811 224	99.69	86x	1 293 752
MEF_day23_TAM	F	38 696 177	99.64	80x	1 346 863
MEF_day23_no TAM	F	40 523 255	99.18	99x	1 197 481
MEF_day23_no TAM	F	45 557 733	99.51	95x	1 398 281
MEF_day23_no TAM	F	32 890 630	99.42	71x	1 301 220

Annexe 1.2 : Statistiques de séquençage des bibliothèques WGBS

sample	sex	read pairs (2x100bp)	% bisulfite conversion	mean depth	CpG 5X
WT	M	301 591 204	98.8	13.06X	19 297 647
Dnmt1-/-	M	295 405 005	99.1	12.31X	19 233 021
DKO	M	317 554 047	99.2	11.89X	19 106 834

Annexe 1.3 : Statistiques de séquençage des bibliothèques RNA-seq

genotype	sex	read pairs (2x100bp)	% uniquely mapped reads
WT	M	122 739 365	83.33
WT	F	104 213 248	83.04
WT	F	121 124 866	80.21
Dnmt1-/-	M	114 210 283	72.77
Dnmt1-/-	F	109 543 827	71.88
Dnmt1-/-	M	102 199 179	71.64
WT	M	87 649 101	82.89
WT	M	91 463 316	83.05
Dnmt3a-/+	F	76 419 000	82.02
Dnmt3a-/+	F	80 239 982	76.14
Dnmt3a-/+	F	56 665 660	76.88
Dnmt3a-/+	M	63 281 010	64.84
DKO	M	77 483 594	79.22
DKO	F	88 310 497	81.43
DKO	F	81 180 800	81.14
DKO	M	80 244 675	78.73
DKO	M	53 957 539	59.30
DKO	M	68 230 903	56.05

Annexe 1.4 : liste des gènes dérprimés dans les embryons DKO (top 40)

gene.id	log2FoldChange	padj
Tuba3b	11,68113912	2,51647E-62
Zscan4d	11,3013744	2,31311E-19
1700080016Rik	11,01331085	3,41817E-36
Tmem92	11,00200982	4,77193E-20
BC080695	10,83006986	1,77136E-17
Zscan4a	10,82778601	3,32377E-20
Sohlh2	10,793829	4,5437E-157
Zscan4c	10,75392884	5,36983E-19
Mageb4	10,72112613	2,10686E-33
Pet2	10,60304123	7,41343E-33
Gm13119	10,54531109	7,56669E-17
Tex13	10,40770242	1,17918E-50
Tktl2	10,18980433	1,34653E-42
4930550L24Rik	10,17624299	4,19E-16
Pramef25	10,08770463	2,97926E-16
Zscan4b	10,02874288	4,50253E-13
Rhox2a	10,01688659	3,87187E-29
Otud6a	9,992555949	5,52519E-13
Rpl10l	9,969364613	2,61306E-74
1700018B24Rik	9,961272324	1,69465E-28
Dazl	9,940761188	1,5695E-163
Zscan4f	9,856106737	1,25011E-16
Dux	9,65227501	1,24498E-12
Gm12794	9,514033915	2,35174E-14
1700029P11Rik	9,489918829	2,73472E-26
Pramel3	9,27611009	1,73324E-21
Pdha2	9,161834411	2,06726E-66
Rhox2f	9,093474607	5,37079E-19
Rhox2h	8,991487998	3,06105E-23
Rhox2b	8,965505215	4,20275E-23
Gm5635	8,916022873	2,37313E-21
Tktl1	8,857969067	1,17877E-22
AV320801	8,808510813	4,73766E-20
Magea10	8,805046052	3,90254E-19
Rhox2c	8,719701111	1,88208E-19
Asz1	8,712290893	4,4588E-162
AF067061	8,685326662	1,01598E-13
Gm2022	8,682418684	2,65448E-15
Usp17le	8,641879459	2,80324E-15
Tcstv1	8,640415261	9,75638E-13
A430089I19Rik	8,520239645	1,09702E-13

Annexe 1.5 : Liste des éléments transposables déréprimés dans les embryons DKO et *Dnmt1* KO

Liste des éléments transposables déréprimés dans les embryons DKO après une analyse Rebase et RepeatMasker (RMSK)

rebase			RMSK		
transposon.id	log2FoldChange	padj	transposon.id	log2FoldChange	padj
RLTR1IAP_MM	5,59	4,07E-39	IAPA_MM-int	5,64	4,12E-29
IAPEYI	5,04	5,04E-29	IAPEy-int	5,36	1,61E-45
IAPLTR1_Mm	4,83	1,49E-36	LTR28	4,82	5,17E-06
MMERVK10D3_I	4,68	2,33E-23	IAPLTR1_Mm	3,93	1,42E-23
IAPEZI	3,75	1,28E-23	IAPEz-int	3,64	2,75E-23
L1Md_Gf_5end	3,46	1,20E-44	RLTR44B	3,40	2,77E-14
IAPA_MM	3,43	2,94E-14	RLTR44-int	2,71	1,86E-20
ERV4_2-LTR_MM	3,30	4,99E-11	ERV4_1B-I_MM-int	2,70	9,91E-26
IAPEY3_I	3,19	1,12E-30	L1Md_Gf	2,68	6,36E-55
IAPEY_LTR	2,90	3,71E-22	RMER16A3	2,53	9,35E-06
IAPLTR1a_I_MM	2,88	4,17E-16	ERV4_2-LTR_MM	2,48	9,45E-08
MERV1_LTR	2,86	2,89E-05	RLTR44A	2,46	2,37E-06
L1MdTf_I	2,74	3,45E-46	RLTR44C	2,45	4,97E-08
L1MdTf_II	2,72	1,38E-61	MMERGLN_LTR	2,10	1,10E-30
GSAT_MM	2,11	3,06E-04	IAPEY_LTR	2,08	2,50E-20
L1MdA_I	1,85	2,93E-11	L1Md_T	2,06	4,22E-79
MMERGLN_LTR	1,79	1,79E-21	RLTR10D2	1,91	1,85E-12
IAP1-MM_I	1,74	2,00E-09	RLTR34B_MM	1,80	4,72E-24
MMERGLN_I	1,63	1,62E-22	MMERVK10D3_LTR	1,79	3,68E-04
MERVL	1,49	1,06E-07	MMERVK10D3_I-int	1,73	5,29E-25
ERV4_1B-I_MM	1,43	7,34E-08	MERVL-int	1,69	8,19E-10
L1_MM	1,41	7,66E-05	RLTR27	1,66	2,07E-12
L1MdTf_III	1,36	1,30E-14	MMERGLN-int	1,60	6,23E-27
L1MdGf_I	1,26	9,62E-08	RLTR16C_MM	1,57	2,39E-14
ERV4_2-I_MM	1,13	2,79E-05	MER34	1,56	9,70E-10
			Ricksha_0	1,46	1,03E-04
			IAPLTR3-int	1,36	6,56E-06
			RLTR21	1,19	5,69E-18
			IAPLTR2_Mm	1,15	2,99E-07
			RLTR9E	1,06	3,05E-11
			RLTR16B_MM	1,04	9,38E-08
			L1Md_A	1,01	1,08E-11
			RLTR10-int	1,00	1,86E-20

Liste des éléments transposables dérprimés dans les embryons *Dnmt1* KO après une analyse Rebase et RepeatMasker (RMSK)

rebase			RMSK		
transposon.id	log2 FoldChange	padj	transposon.id	log2 FoldChange	padj
IAPLTR1_Mm	6,90	9,75E-92	IAPA_MM-int	6,92	7,54E-64
RLTR1IAP_MM	6,73	4,25E-73	IAPEy-int	6,24	1,24E-46
IAPEYI	5,93	1,00E-53	IAPEz-int	5,63	2,38E-43
ERV4_1-LTR_MM	5,56	5,08E-05	IAPLTR1_Mm	5,53	4,28E-40
IAPEZI	5,40	2,92E-51	LTR45C	5,32	1,60E-04
IAPLTR1a_I_MM	5,10	6,86E-50	IAPLTR1a_Mm	4,53	4,64E-35
IAPLTR1a_MM	4,93	4,66E-60	IAPLTR2_Mm	4,51	1,01E-33
IAPA_MM	4,72	4,50E-34	ERV4_1B-I_MM-int	4,40	9,53E-25
IAPLTR3_I	4,63	6,82E-10	IAPLTR3-int	3,59	1,21E-32
RLTR2IAP_MM	4,44	2,95E-44	IAPEY5_I-int	3,17	1,30E-07
IAPEY3_I	4,23	3,90E-58	IAPLTR4_I	3,09	3,58E-04
MERV1_LTR	3,62	3,75E-06	MURVY-int	2,89	1,60E-04
IAPEY_LTR	3,46	3,96E-32	RLTR44B	2,84	1,57E-10
RLTR10D2	3,16	9,18E-04	IAPLTR2a	2,57	1,81E-10
RLTR1F_Mm	3,08	5,91E-04	MMERVK10D3_I-int	2,48	1,22E-13
ERV4_1B-I_MM	2,95	3,45E-14	RLTR10D2	2,47	6,01E-06
IAPLTR4_I	2,80	2,07E-19	IAPEY_LTR	2,47	5,01E-13
IAP5-Mm_I	2,74	1,59E-05	RLTR46A2	2,31	8,69E-26
L1Md_Gf_5end	2,48	9,05E-10	RLTR1F_Mm	2,18	1,31E-06
L1_MM	2,36	2,22E-14	L1Md_Gf	2,14	2,39E-18
IAP1-MM_I	2,32	3,50E-07	RLTR44-int	2,12	1,22E-13
MMERVK10D3_I	2,27	8,46E-06	MMERVK10C-int	2,11	6,68E-23
MMERVK10C	2,07	3,61E-16	IAPLTR4	2,08	3,72E-04
L1MdTf_II	2,06	2,85E-23	RLTR1D2_MM	2,07	8,12E-04
L1MdTf_I	1,89	1,28E-17	RLTR9A2	2,03	2,64E-07
L1MdA_I	1,75	5,12E-15	RMER16A3	1,95	5,67E-06
RLTR2aIAP_MM	1,54	4,46E-04	RLTR9A3A	1,88	2,29E-04
MMERVK9C_I	1,53	6,35E-06	RLTR44C	1,77	1,12E-05
ETNERV	1,48	8,33E-06	RLTR27	1,68	3,41E-10
IAPLTR2b	1,41	3,33E-04	RLTR10F	1,64	2,42E-07
IAPLTR2_Mm	1,39	3,28E-08	L1Md_T	1,64	6,34E-17
IAPEY5_I	1,20	6,17E-05	RLTR10-int	1,52	6,11E-10
			IAPEY2_LTR	1,50	2,83E-09
			IAPEY3-int	1,48	1,13E-14
			RLTR13D3A	1,42	9,33E-04
			IAPLTR2b	1,41	8,91E-04
			MMERVK9C_I-int	1,33	1,60E-04
			MuRRS4-int	1,30	8,33E-08
			ETnERV2-int	1,30	4,52E-12
			RLTR10C	1,22	1,31E-09
			RLTR17	1,15	2,29E-09
			L1Md_A	1,08	7,74E-11
			RLTR10D	1,04	6,87E-05

Annexe 1.6 : Liste des gènes avec une initiation interne de la transcription

in DKO		in Dnmt1KO
4930558C23Rik_NR_015490	Gm973_NM_001013771	4933433C11Rik_NM_028961
4933433C11Rik_NM_028961	Gm8579_NR_036696	9530053A07Rik_NM_001164655
9530026F06Rik_NR_040483	Hpse2_NM_001081257	Ackr4_NM_145700
9530053A07Rik_NM_001164655	Kcng4_NM_025734	BC051019_NM_001040700
A430089119Rik_NM_177913	Krt12_NM_010661	C8b_NM_133882
Abca8a_NM_153145	Mgl2_NM_145137	Cd28_NM_007642
Ackr4_NM_145700	Mlana_NM_029993	Cd6_NM_009852
Acpp_NM_019807	Myh3_NM_001099635	Cerkl_NM_001048176
Aifm2_NM_178058	Myt1l_NM_008666	Ckm_NM_007710
BC021614_NM_144869	Nckap5_NM_001081756	Crtam_NM_019465
C8b_NM_133882	Nlrp9b_NM_194058	Ctsk_NM_007802
Cdc20b_NM_001281487	Nos1_NM_008712	D630023F18Rik_NM_175293
Cerkl_NM_001048176	Phkg1_NM_011079	Gabrr1_NM_008075
Clec10a_NM_010796	Plekha4_NM_148927	Gm15107_NM_001081648
Crtam_NM_019465	Plekhd1_NM_001177503	Gm973_NM_001013771
Crybb3_NM_021352	Pnoc_NM_010932	Itgal_NM_008400
D630023F18Rik_NM_175293	Ptprq_NM_001081432	Lor_NM_008508
Dclk3_NM_172928	Pzp_NM_007376	Mgl2_NM_145137
Gabrr1_NM_008075	Rhox7a_NM_001025086	Mlana_NM_029993
Gal3st2_NM_199366	Rpgrip1_NM_001168515	Myh3_NM_001099635
Gh_NM_008117	Scd3_NM_024450	Nos1_NM_008712
Gm15107_NM_001081648	Slc11a1_NM_013612	Plekhd1_NM_001177503
Gm15127_NM_001114400	Slc12a3_NM_019415	Ptprq_NM_001081432
Gm20765_NM_001270644	Sult6b1_NM_001163625	Rhox7a_NM_001025086
Gm8267_NM_001162954	Tph2_NM_173391	Rpl3l_NM_001163945
		Tph2_NM_173391
		Trp73_NM_011642
		Vmn2r5_NM_001104618
		Wdfy4_NM_001146022

Annexe 1.7 : Liste des oligos utilisés dans cette étude

RT-qPCR		Bisulfite sequencing	
Gusb	GATTCAGATATCCGAGGGAAAGG GCCAACGGAGCAGGTTGA	Dazl	GTGAGGTGGTTGTTTAGT ATCCATCTATCTATCAATAAACTC
Rpl13a	GAGTCCGTTGGTCTTGAGGA GGTTACGGAAACAGGCAGAA	Asz1	TAAGTTAAAGTTTTGGGAAAAAGTTAG TTATTTCTAATAAAAAATCCATCTCACC
B2m	TATGCTATCCAGAAAACCCCTCAA GTATGTTCCGGCTTCCCATTCTC	PCR genotyping	
Mrpl32	AGGTGCTGGGAGCTGCTACA AAAGCGACTCCAGCTCTGCT	Dnmt3a	GCACCAGGGAAAGATCATGT TAGCTGAGGCTGTCTGCATCGGACA
Dazl	AACAACCCCTGAGATGAGTT TTCAGAAGAAGTCGGGCAGT	Dnmt3b	CGCAGGAAAGATTGGAACAT GGATGTTCCGAGAATGTTGTGGCC CAGGTCAGACCTCTCTGGTGACAAG
Asz1	CCCGAATCCACACACTGTAAC CGGGAAAATGACCCAACTC		
Bin2	ATGGATGCTGCTCTTTTGCT GCAGAGGTGACAGGGAAG		
Pf4	ACCCTGAAGAATGGGAGGAA TTTCTTATATAGGGGTGCTTGC		
lyd	GGTCACTACCACTCCCCTCA CAGGCACTGTGGCATCTCT		
Ly86	GAGAATATCAGCTCTTGCTGGAA GAGGAGGTGACAGTGGCATT		
Arhgap30	TGTCCTGGCTCCTAACTGCT CCCTACACCTTCCCAGTTGA		
Zscan4d	TGCAGCTACTCAAGTATATAGAA CAGAATGATATATTCTCTTGTTAATGG		
Tmem92	GGTTGGTGGTGGATCCAAG GTGGATCATCAGACTCCTAT		
Gm5039	GCTACTACAGTCTTAGGATGT TGTGGATCTTCTACACTTATATCA		
Tcstv1	GCTTGCTCAGGGAGAAAGGT ACTGTCAGAAACAGGATCCC		
Usp17le	GCACGGATGCTCCAGACAAA CCCTGACTACAGAGCTGCCC		
Dnmt3a (exon 18)	GCCCATTGACCTGGTGATT GGCAGGGTTGACAATGGAGA		
Dnmt3b (exon 18)	CCTCCGAAGTCTGTGCAGAG TGGTGATTTTCCGGACGTCA		
Dnmt1	CAAAAGAGGAGGCTGCTACCA AAAAACATCAGTGCACGTGGG		

2. Rôle du facteur de transcription E2F6 dans le ciblage épigénétique des gènes gamétiques dans l'embryon

Annexe 2.1 : Statistiques de séquençage des bibliothèques RRBS

Sample	sex	read pairs (2x75 bp)	% bisulfite conversion	median depth	CpG 8X
Brain WT	F	28 834 363	99.57	62x	1 382 465
Brain WT	F	53 340 010	99.38	120x	1 375 539
Brain WT	M	41 011 679	99.33	82x	1 289 185
Brain KO	M	32 586 536	99.37	71x	1 367 777
Brain KO	F	49 331 068	99.36	92x	1 483 780
Brain KO	F	35 825 738	99.27	66x	1 396 428
Kidney WT	M	28 396 054	99.88	64x	1 400 866
Kidney WT	F	34 294 774	99.79	82x	1 323 792
Kidney WT	M	45 348 662	99.81	72x	1 263 911
Kidney KO	M	30 478 333	99.81	69x	1 438 098
Kidney KO	M	27 535 319	99.84	65x	1 321 848
Kidney KO	F	53 034 247	99.81	95x	1 339 053
Muscle WT	F	34 187 745	99.84	84x	1 258 396
Muscle WT	F	34 765 432	99.82	79x	1 344 217
Muscle WT	M	31 396 025	99.78	63x	1 346 507
Muscle KO	M	38 970 624	99.77	90x	1 324 741
Muscle KO	M	36 383 012	99.79	75x	1 224 801
Muscle KO	F	40 051 830	99.72	75x	1 287 312
MEFp WT	F	41 971 141	99.86	103x	1 348 415
MEFp WT	M	42 267 577	99.89	97x	1 394 623
MEFp KO	M	40 977 683	99.87	91x	1 365 800
MEFp KO	F	33 900 305	99.81	78x	1 350 091
MEF CRISPR exon 4 WT	F	44 230 681	98.79	101x	1 244 315
MEF CRISPR exon 4 WT	F	39 250 576	99.66	93x	1 259 729
MEF CRISPR exon 4 KO	F	28 356 995	99.69	68x	1 160 466
MEF CRISPR exon 4 KO	F	50 646 477	99.8	113x	1 208 589
MEF CRISPR exon 3 KO	F	40 858 850	99.86	91x	1 303 844
MEF CRISPR exon 3 KO	F	37 189 418	99.89	79x	1 403 708
E8.5 WT	M	38 029 577	99.83	67x	1 353 443
E8.5 WT	F	28 528 851	99.81	55x	1 277 888
E8.5 KO	M	19 188 994	99.86	39x	1 237 516
E8.5 KO	F	18 196 336	99.84	34x	1 228 972
ES WT	F	42 358 303	99.42	87x	1 397 365
ES KO	M	38 721 388	99.49	84x	1 354 448

Annexe 2.3 : Statistiques de séquençage des librairies WBGS

Sample	sex	read pairs (2x100 bp)	% bisulfite conversion	median depth	CpG 5X
WT	M	591 719 992	99.4	12.08x	19 016 288
KO	M	626 044 202	99.52	13.62x	19 380 561

Annexe 2.2 : Statistiques de séquençage des librairies RNA-seq

genotype	sex	read pairs (2x100bp)	% uniquely mapped reads
E2F6 WT	M	139 723 508	63,23
E2F6 WT	M	130 261 502	63,39
E2F6 WT	M	111 925 612	52,94
E2F6 KO	M	135 183 728	71,24
E2F6 KO	M	115 873 206	59,73
E2F6 KO	M	129 695 230	75,61

Annexe 2.4 : Analyse ontologique des gènes régulés par E2F6

Category	Term	microarray ES cells		RNA-seq E8.5	
		PValue	Benjamini	PValue	Benjamini
GOTERM_BP_FAT	GO:0051327~M phase of meiotic cell cycle	6,25E-12	1,95E-09	1,77E-11	4,63E-09
GOTERM_BP_FAT	GO:0007126~meiosis	6,25E-12	1,95E-09	1,77E-11	4,63E-09
GOTERM_BP_FAT	GO:0051321~meiotic cell cycle	7,52E-12	1,17E-09	2,08E-11	2,72E-09

Annexe 2.5 : Liste des gènes dérprimés en l'absence d'E2F6 dans les cellules ES

Gene Symbol	log2FC_E2f6	adj_p_value_E2f6	log2FC_Mga	adj_p_Value_Mga
Slc25a31	4,48	1,17E-18	3,67	3,27E-06
Tuba3a	4,18	1,15E-20	2,31	1,64E-05
Gm1564	3,74	2,79E-18	2,17	2,42E-04
Tex12	3,52	1,68E-13	4,54	8,82E-06
Scml2	3,25	2,54E-13	1,74	8,03E-04
1700123I01Rik	3,24	1,83E-16	2,45	4,86E-06
Mael	3,16	5,28E-14	2,27	6,24E-06
Dazl	2,97	3,19E-11	1,90	3,37E-05
Tex11	2,94	4,65E-14	1,39	3,72E-04
Gpat2	2,91	1,68E-13	1,56	5,84E-05
Cox7b2	2,87	1,68E-13	1,91	4,13E-05
Uty	2,56	1,13E-11	0,22	3,81E-02
1700013H16Rik	2,51	7,70E-07	3,46	8,97E-06
1700018B24Rik	2,48	4,91E-08	2,29	2,45E-04
Stag3	2,46	6,31E-11	1,03	5,42E-04
Rhox13	2,38	4,17E-10	3,14	3,04E-05
Smc1b	2,38	1,56E-10	1,58	1,43E-03
Tex13	2,3	4,87E-08	0,78	1,17E-03
Kdm5d	2,09	2,55E-09	-0,01	8,98E-01
Hormad1	2,06	1,88E-08	1,69	2,85E-04
Sycp2	1,98	1,68E-07	1,47	7,15E-05
Sycp1	1,89	3,02E-04	1,55	7,21E-04
Lef1	1,78	1,33E-06	-1,38	1,38E-04
Rnf130	1,77	1,30E-07	1,06	3,09E-04
Taf7l	1,71	4,44E-04	2,18	1,66E-04
AU022751	1,69	1,40E-06	1,28	8,65E-06
Syce1	1,68	1,40E-06	1,45	8,93E-04
Sohlh2	1,62	1,09E-06	0,62	5,00E-03
Col1a2	1,58	5,62E-04	0,46	1,95E-03
Gm9	1,58	1,44E-05	-0,22	1,98E-01
Hdgfrp3	1,57	1,10E-05	0,51	1,98E-03
Cpeb1	1,56	9,25E-05	1,90	3,56E-05
Mlf1	1,51	4,36E-05	0,94	1,03E-02
Ms4a4d	1,51	1,19E-05	0,71	6,07E-03
Efhc2	1,49	4,81E-05	-0,10	6,65E-01
Rnf219	1,45	1,15E-05	0,35	6,95E-03
4933406J08Rik	1,44	3,03E-05	1,05	7,96E-04
Cyct	1,41	1,13E-04	0,31	2,16E-02
Cxcl10	1,4	4,92E-05	1,86	2,67E-05
Tchh	1,39	5,58E-04	0,81	7,81E-05
4933427D06Rik	1,38	1,89E-04	1,63	4,73E-06
Brdt	1,36	6,67E-04	1,32	1,33E-03
Gtsf1	1,35	1,32E-04	1,66	4,92E-05
E2f6	1,35	4,65E-05	0,56	1,70E-03
Ivns1abp	1,34	8,54E-04	-0,62	2,55E-02
Stra8	1,33	7,01E-05	1,85	2,79E-05

Annexe 2.6 : liste des gènes dérprimés en l'absence de E2F6 dans l'embryon (E8.5)

gene symbol	log2FC	pvalue
Tuba3b	11,96	1,36E-22
Tuba3a	10,37	0,00E+00
1700123I01Rik	10,02	1,43E-22
4933406J08Rik	9,10	3,82E-14
Gm1564	8,19	0,00E+00
Smc1b	7,95	2,30E-195
Hormad1	7,76	3,38E-80
Gpat2	7,73	6,06E-236
Tex12	6,92	1,93E-07
Hsf2bp	6,39	1,93E-238
Wbp2nl	6,33	9,26E-48
Zcwpw1	5,66	0,00E+00
Mei1	5,40	2,24E-11
Cyct	5,23	3,32E-16
Tex11	5,21	4,77E-06
Stag3	5,16	1,82E-189
Rbm44	5,11	3,79E-104
Ulbp1	5,03	2,14E-177
Mael	5,03	1,79E-19
Patl2	4,73	9,28E-34
Slc25a31	4,08	2,72E-07
Dmc1	3,88	3,36E-40
4732471J01Rik	3,67	2,31E-17
1700001L05Rik	3,62	2,76E-23
4930432K21Rik	3,54	8,55E-71
Ddx43	3,11	6,05E-09
Cntd1	3,03	2,80E-20
Slfn9	2,75	6,36E-84
Sycp2	2,72	3,52E-08
Slfn8	2,58	3,06E-11
Xlr	2,50	6,25E-12
Cyp2r1	2,30	6,85E-09
Ccdc36	1,91	1,33E-11
Syce2	1,82	7,37E-56
Agbl3	1,71	4,71E-09
Adcy7	1,48	5,97E-19
Tex16	1,46	2,14E-07
Snord49a	1,42	2,05E-06
Slc4a5	1,39	1,11E-15
C77370	1,35	1,31E-09
Cntln	1,35	9,52E-16
Casp7	1,33	9,79E-26

Annexe 2.7 : Liste des DMRs identifiées dans l'embryon E8.5 en l'absence de E2F6 (WGBS)

Liste des DMR hyperméthylées

Dist to TSS	GeneName_TSS	Diff meth KO-WT	DMR.qvalue
23552	Rnf157	0,802	2,00E-16
73	Gstm6	0,725	9,00E-14
-23182	Dhrs7	0,59	2,00E-12
194333	Magi1	0,565	1,00E-10
114139	1700030N03Rik	0,56	3,00E-08
-22118	4930413E15Rik	0,529	9,00E-07
-10120	Cxcl17	0,522	7,00E-13
26104	Dnmt3a	0,509	4,00E-12
-504	Hs3st5	0,5	8,00E-13
138367	9530068E07Rik	0,489	1,00E-07
190	Wee2	0,489	1,00E-10
50901	Ykt6	0,483	1,00E-08
-39938	AF067061	0,444	1,00E-08
922	Tnfrsf1a	0,442	1,00E-10
3107	Ulbp1	0,426	5,00E-08
-20053	Mki67	0,425	1,00E-07
-8173	Aim1l	0,407	6,00E-08
17004	2210039B01Rik	0,389	5,00E-09
-3933	Igsf5	0,389	5,00E-08
-19527	Mki67	0,38	8,00E-07
343	Rec8	0,38	1,00E-06
379	Fam161b	0,367	2,00E-07
-107752	1700028D13Rik	0,36	8,00E-10
18012	Marc1	0,34	2,00E-08
37993	Gpam	0,329	3,00E-08
-3385	Cox14	0,322	4,00E-07
-123104	Celrr	0,317	2,00E-07
23322	Zfp777	0,308	7,00E-08
-32310	Fbrsl1	0,296	2,00E-06
31926	Gde1	0,29	1,00E-07
1829	Impact	0,245	2,00E-07

Liste des DMR hypométhylées

Dist to TSS	GeneName_TSS	Diff meth KO-WT	DMR.qvalue
-70	Rbm44	-0,753	1,00E-19
-64	Tuba3a	-0,701	3,00E-14
15017	Slc15a2	-0,699	5,00E-08
311	Patl2	-0,697	1,00E-13
-14248	Xpnpep1	-0,669	1,00E-16
-78	Ulbp1	-0,657	1,00E-13
230	Gpat2	-0,61	2,00E-17
265	4933406J08Rik	-0,59	2,00E-17
62114	Akr1d1	-0,58	2,00E-09
121	1700123I01Rik	-0,575	5,00E-13
2847	Ttc12	-0,511	7,00E-12
-12990	Xpnpep1	-0,502	3,00E-10
182	Hsf2bp	-0,499	7,00E-10
523	Zcwpw1	-0,471	6,00E-13
532	Stag3	-0,467	7,00E-08
184	Hormad1	-0,462	2,00E-11
298	Wbp2nl	-0,46	5,00E-11
22763	Tmem30c	-0,46	7,00E-10
-71333	Gatsl2	-0,448	1,00E-07
91513	Xrcc6bp1	-0,423	2,00E-09
18663	Cdk5rap1	-0,411	5,00E-08
122280	BC039966	-0,392	6,00E-09
402	4930432K21Rik	-0,382	5,00E-12
18	4930481A15Rik	-0,374	2,00E-06
12498	Dmrt3	-0,369	4,00E-06
146	Gm1564	-0,367	5,00E-12
173	Smc1b	-0,364	4,00E-10
11169	1700047L14Rik	-0,352	3,00E-06
-7397	Galp	-0,331	4,00E-06
46772	Pea15b	-0,306	3,00E-07
-378297	5730507C01Rik	-0,299	6,00E-07
-3352	Ccdc36	-0,266	1,00E-05
447	1700001L05Rik	-0,265	2,00E-09

Annexe 2.8 : Liste des oligos utilisés dans cette étude

RT-qPCR

Gus	GATTCAGATATCCGAGGGAAAGG GCCAACGGAGCAGGTTGA	Gpat2 (2)	GCTGCCAGACCTGTACTCCT AGCCCAGGTCCATTATGCTT
Rpl13	GAGTCCGTTGGTCTTGAGGA GGTTACGGAAACAGGCAGAA	Tuba3a (2)	GATTGGTGTCCGACTGGATT GCACTTTGCCAGGTCTC
Mrpl32	AGGTGCTGGGAGCTGCTACA AAAGCGACTCCAGCTCTGCT	Hormad1 (2)	GACAAAGATGATGTGGAAGATGA AATCTAAATTTGGTTCTTTGATTTGAG
Tuba3a	GTTGAGGACCAGTGGTGAGG CATGCTGAACTCCGGAGAC	Meioc (2)	ACAGGATAAGTCGCAGCCATA GTTTCATTGGCCAACTGACTT
Hormad1	CAATAAGATATCAACTGAGCATCAATC TAGGTGATGCAAGATACTGAAACAG	Smc1b (2)	TCTACAAGAAGCTCTGCAGGAA TTCAGGGTTCTCTGGGCTAA
Gpat2	AGCATCTGAGTGCAAAGCTG CCTGGCTCCTCAAAGTCATC	Slc25a31	TTTCATGTCTGGTGTTAATAAAGAAAA GTGGGTATACTACACACAAGGATGTT
Rbm44	ACACACCACCACCTCTTTCC GAGGTGGCATCTTTGTTGGT	Tex11	AAGAGAGTAGAACTGTCTGTGAGGAA CCATGCATATATACCAGCTTGC
Meioc	TGTGCACTGCAGATGACCTT GGAAAGCCTTTTCCATGTCA	Stag3	AATGGCAACGTGGA ATAGGTCCAACCCA
Hsf2bp	TGCAAGGGTAAGGAGCTCTG GGCTCATCCAGTCTGTGGTT	RT-qPCR (blastocyst experiment)	
Terb2	CGCAGGAAGACCAAGTCTCT ACAAGTCCCTCCAGCTCTGA	E2f6	TGGAATAGGAGCCTCTCCATCTA GCTACTAGCCGCTCACTCTGAGT
Smc1b	TCCACAGGGTCAGGGTAAAG GAAGCCCAGAGTACCACGAG	Tuba3a	GTTGAGGACCAGTGGTGAGG CCCCACGTGGATAGAGATACAC
E2f6	AAGCTGCAGGCAGAACTCTC CAACAGTTGCTGAGCACAATC	Rpl13	GAGTCCGTTGGTCTTGAGGA TTCGGCTGAAGCCTACCAG
E2f6 (CRISPR)	AAGAAAATGTACAGTATGTGTCCATGAG CTTCCGAGTTAAGTATACCAGTGACA	Hormad1	TCCAGAACGTGCTTATGGGA GCTGTGAAGAACCTGGACAATTT
Actin b	AAGGCCAACCGTGAAAAGAT GTGGTACGACCAGAGGCATAC		

ChIP-qPCR

Actb	CGCTGTGGCGTCCTATAAAA AAGGAGCTGCAAAGAAGCTG
Gpat2	TAGAGCGCGTTCTGATTGG CGAGTCTCAGGAAGCTCACC
Hormad1	CTTTTGGCGGGAATAGTGG GCGCCTTAACTGTCACTGG
Tuba3a	ATGTGGCTCAAGTTGCAGTG GCGCAGATAACATACGCAGA
Dazl	ATTTACAAAATGCCCGCAGA AACCTTCTCAATGTGGCTGAA
Actin b	TGCAGGCCTAGTAACCGAGA AGGCCCAGATGTACAGGAATAG
Tuba3a (2)	ATGCTCCTCAATGCTTGCC TCTGATTGGATGTGGCTTTG
Gpat2 (2)	AGAGCGCGTTCTGATTGG AACAGCCGTGTGACTGAGC
Slc25a31	CAGTTAATCACCGGGTTGGT AATCAGTGCACAGCAGCTTTT
Tex12	TGCTCATCTGCTGAGATTCC TAAGCTTCAAAGCGGGTCTT
Mael	AGGCTGTTTTACCCCCACAGCAGG TGGCCCTGCGGTTGGGCATG

COBRA

Gpat2	GGATAAGTAGGTTTTTAAGTTAGAATGTTT CCACCAACTTACCTCTAACTATTTATCTAT
Tuba3a	GTTTTTTATTATTGGTTTTTAAT ACTAAACAACCTTTTCTACCCCAAATA
Hsf2bp	GAGTTTTAAGGTTTAAGGTTTGGGT CAATAAAATCCAAATCCATCTCATC
Majin	GAGGTGTTGGAGGAATTTTAGTTAGT CAATCTACACCAATACCCAACCTAC

Pyrosequencing

Tuba3a	TAGATTAGAGGCGGTTGAGGATTAGTGGT AAAAAACCCCTAACCTCAAAAAAAC GATTAGTGGTGAGGAA (seq)
--------	--

Méthylation de l'ADN : Fonctions et ciblage au cours du développement chez la souris

Résumé :

La méthylation des cytosines est une modification épigénétique catalysée par la famille des ADN méthyltransférases (DNMTs). C'est une marque répressive lorsqu'elle est adressée sur les îlots CpG des promoteurs de gènes. Le développement embryonnaire murin est caractérisé par une reprogrammation de la méthylation de l'ADN qui est critique pour le développement de l'embryon. Cependant, la contribution des différentes DNMTs dans la méthylation du génome ainsi que les mécanismes qui ciblent la méthylation de l'ADN vers certains gènes durant le développement embryonnaires sont mal connus.

En combinant des approches de cartographie génomique avec des lignées génétiquement modifiées de souris, mes travaux de Thèse ont permis de clarifier la contribution des différentes DNMTs dans la méthylation du génome dans l'embryon : DNMT3A et DNMT3B sont strictement impliqués dans la méthylation *de novo*, et DNMT1 est strictement impliqué dans son maintien au cours des divisions cellulaires. De plus, l'analyse d'embryons globalement déméthylés a révélé de nombreuses fonctions de la méthylation de l'ADN dans le maintien de l'intégrité transcriptomique de l'embryon en réprimant des gènes gamétiques, des gènes du développement, des promoteurs cryptiques ainsi qu'un large panel de transposons.

Dans un deuxième temps, j'ai étudié le rôle du facteur de transcription E2F6 dans le ciblage de la méthylation de l'ADN *in vivo* chez la souris. Mes résultats démontrent que E2F6 facilite l'acquisition de la méthylation de l'ADN au niveau du promoteurs des gènes gamétiques et est nécessaire pour initier leur répression à long terme au cours de l'embryogenèse.

Dans leur ensemble, ces travaux contribuent à mieux comprendre les fonctions et mécanismes de ciblage de la méthylation de l'ADN au cours de l'embryogenèse des mammifères.

Mots clés : épigénétique, méthylation de l'ADN, DNMT, développement

Résumé en anglais :

Cytosine methylation is an epigenetic modification catalyzed by the family of DNA methyltransferases (DNMTs). This modification is involved in gene repression when it is addressed to CpG islands in gene promoters. Global DNA methylation reprogramming occurs in mice during the early phases of embryogenesis, which is critical for proper embryo development. However, the contribution of different DNMTs in genome methylation and the mechanisms that target DNA methylation to specific genes during embryonic development are poorly understood.

By combining genomic mapping with genetically modified mouse lines, my Thesis work clarified the contribution of the different DNMTs in genome methylation in the embryo: DNMT3A and DNMT3B are strictly involved in *de novo* methylation, and DNMT1 is strictly involved in the maintenance of DNA methylation during cellular divisions. In addition, the analysis of globally demethylated embryos revealed numerous functions of DNA methylation in maintaining the transcriptomic integrity of the embryo by repressing gametic genes, developmental genes, cryptic promoters as well as a large panel of transposons.

In the second part of my Thesis, I studied the role of the E2F6 transcription factor in the targeting of DNA methylation *in vivo* in mice. My results demonstrate that E2F6 facilitates the acquisition of DNA methylation in the promoters of gametic genes and is required to initiate their long-term epigenetic silencing during embryogenesis.

Collectively, this work contributes to a better understanding of the functions and targeting mechanisms of DNA methylation during mammalian embryogenesis.

Keywords: epigenetics, DNA methylation, DNMT, development