



Formes atypiques d’empreinte génomique : transitoire, tissu-spécifique et lignée-spécifique

Sophie Ajjan

► To cite this version:

Sophie Ajjan. Formes atypiques d’empreinte génomique : transitoire, tissu-spécifique et lignée-spécifique. Génétique. Université Pierre et Marie Curie - Paris VI, 2015. Français. NNT : 2015PA066251 . tel-01236091

HAL Id: tel-01236091

<https://theses.hal.science/tel-01236091>

Submitted on 1 Dec 2015

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

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

Université Pierre et Marie Curie

Ecole doctorale Complexité du Vivant, ED515

Laboratoire de Génétique et Biologie du Développement

Equipe Décisions Epigénétiques et Reproduction chez les Mammifères

Thèse de doctorat de Génétique

Présentée par

Sophie Ajjan

**Formes atypiques d'empreinte génomique :
transitoire, tissu-spécifique et lignée-spécifique**

Dirigée par Déborah Bourc'his

Présentée et soutenue publiquement le 10 juillet 2015

Devant un jury composé de :

Dr. Peronnet Frédérique

Présidente

Dr. Defossez Pierre-Antoine

Rapporteur

Dr. Feil Robert

Rapporteur

Pr. Bloyer Sébastien

Examineur

Dr. Bourc'his Déborah

Directrice de Thèse

Université Pierre et Marie Curie

Ecole doctorale Complexité du Vivant, ED515

Laboratoire de Génétique et Biologie du Développement

Equipe Décisions Epigénétiques et Reproduction chez les Mammifères

Thèse de doctorat de Génétique

Présentée par

Sophie Ajjan

**Formes atypiques d'empreinte génomique :
transitoire, tissu-spécifique et lignée-spécifique**

Dirigée par Déborah Bourc'his

Présentée et soutenue publiquement le 10 juillet 2015

Devant un jury composé de :

Dr. Peronnet Frédérique

Présidente

Dr. Defossez Pierre-Antoine

Rapporteur

Dr. Feil Robert

Rapporteur

Pr. Bloyer Sébastien

Examineur

Dr. Bourc'his Déborah

Directrice de Thèse

A ma grand-mère...

Remerciements

Alors voilà, c'est fini...

Je tenais à remercier tous ceux qui ont partagé avec moi, de près ou de loin, cette expérience : une thèse en science. Quatre ans et demi déjà que j'ai commencé à chercher ces fameux gènes soumis à empreinte.

Je remercie les membres du jury d'avoir accepté d'évaluer mon travail, Pierre-Antoine Defossez et Robert Feil, mes rapporteurs, Sébastien Bloyer, mon examinateur et Frédérique Perronet qui préside ce jury.

Je tenais à remercier tout particulièrement Déborah qui m'a permis d'intégrer le laboratoire et de mener à bien mes recherches. Elle a été tout au long de ces quatre années beaucoup plus que ma directrice de thèse et a toujours eu un regard maternel à mon égard. Je n'ai pas été l'étudiante en thèse la plus facile à suivre que tu aies eu je pense. Tu auras eu un bon entraînement pour le futur... ;-)

Une spéciale dédicace aux membres de mon équipe qui ont su me supporter, et c'est pas une mince affaire au quotidien : Rachel, avec qui j'ai commencé ma thèse et qui m'a initiée à toutes les techniques du laboratoire ; Julian, qui a toujours la réponse à mes questions ; Max, que j'adore embêter pour me détendre et qui supporte mes folies et ma maniaquerie ; Marius, qui me déteste aujourd'hui car j'ai décalé sa date de soutenance, il va donc être obligé de rester au labo ; Juliane, dernière arrivée et déjà maitre CRISPR malgré sa jeunesse ; Joan, sans qui on mourait de faim à 18h ; Raquel, toujours de bonne humeur même tôt le matin (car c'est une lève-tôt) ; Natasha, pour les histoires de filles ; Tomek, qui à peine arrivé a su intégrer mon caractère.

Je voulais aussi remercier mes acolytes de master Antoine et Mahéva qui ont partagé tous les moments forts, angoissants, joyeux et toute ma vie au labo et en dehors depuis 5 ans. Vous avez été d'un grand soutien et une force pour moi. Cette expérience n'aurait pas eu la même saveur sans vous les loulous.

Ma vie au labo n'aurait pas été la même sans tous les membres de la BDD avec qui les échanges scientifiques ou gazeux lors des tea time ont été toujours bénéfiques et enrichissants. Travailler dans cet environnement est vraiment une chance.

Un grand merci aux Meufs pour tous les à côté. Ma vie tumultueuse n'aurait pas été la même sans vous et vous m'avez toujours poussée et épaulée. Gaëlle, pour les un an et demi en colloc avec nos horaires décalés et ton oreille attentive ; la cousine, toi même tu sais c'est la millefa j'ai pas besoin d'en dire plus ; Léa, même à distance skype est là ; Diu, toujours au top avec tes surnoms et ta bonne humeur communicative ; Ouillie, toujours le mot pour me redonner l'envie quand je suis vidée.

Ma Camille, pour ton art de parler de tout et de rien, pour me connaître mieux que quiconque et savoir quoi dire surtout quand je ne vais pas bien. Et puis pour toutes les photos hilarantes qui me distraient lors de mes pauses dans mes journées.

Ma famille, ah la famille... Je vous adore mais vous m'en faites voir quand même. N'est ce pas Papa ? Si je vous avais dit il y a 10 ans que je passerais une thèse de génétique, vous en auriez douté. Heureusement que les gens grandissent entre 20 et 30 ans.

Un petit mot pour mon Papi, qui a aussi fait beaucoup de génétique lors de sa carrière et qui m'a inspiré lors de mon choix d'orientation. Tu as réussi à publier plus de 600 articles avec 5 enfants dans les pattes, je ne sais pas si j'en ferai autant, mais quel exemple.

Et enfin la plus méritante dans tout ça, celle qui me supporte au quotidien et qui a toujours été la pour moi, ma merveilleuse Bichette.

Résumé :

Les gènes soumis à empreinte (GSE) se distinguent du reste du génome par une expression mono-allélique et parent-spécifique. Cette forme de régulation génique dépend de marques de méthylation différentielles héritées des gamètes parentaux au niveau de régions *cis*-régulatrices appelées ICR (« Imprinting Control Region »). Une centaine de GSE contrôlés par 20 ICR ont été répertoriés chez la souris et ils sont en général conservés chez l'Homme. Mon projet de thèse a consisté à caractériser de nouvelles ICR maternelles et à analyser leur impact sur la régulation génique, à partir d'un criblage génomique de méthylation réalisé chez la souris. J'ai ainsi participé à la révélation de l'existence de trois formes d'empreinte, qui résultent de sensibilité différente des ICR face aux changements développementaux des profils de méthylation génomique: 1) une empreinte persistante tout au long de la vie et ubiquitaire, qui caractérise les ICR classiques déjà connues, 2) transitoire, avec une existence limitée au développement pré-implantatoire, et 3) persistante tout au long de la vie mais tissu-spécifique. Plus précisément, j'ai déterminé les profils d'histones associées aux ICR des loci *Cdh15* et *Gpr1/Zdbf2*, et mis en évidence la conservation de l'empreinte transitoire au locus *GPR1/ZDBF2* chez l'humain. Je me suis ensuite focalisée sur l'ICR candidate associée au gène *Socs5*, dont l'empreinte s'est avérée être tissu-spécifique mais également, de façon inédite, polymorphique en fonction des lignées de souris. Cette ICR en position intragénique présente les caractéristiques d'une séquence « enhancer », hypothèse que je teste actuellement par invalidation fonctionnelle (système CRISPR/Cas9) chez la souris. La découverte de ces formes atypiques d'empreinte génomique permet de mieux cerner l'étendue du phénomène d'empreinte parentale et d'évaluer son impact sur les phénotypes.

Mots clés : méthylation de l'ADN, empreinte parentale, développement, épigénétique, modifications d'histones, modèle animal souris

Abstract :

Genomic imprinting refers to the functional non-equivalence of the two parental genomes in mammals. Imprinted genes are expressed only from the paternal or maternal allele: this mono-allelic expression is regulated by parent-inherited DNA methylation of specific *cis*-regulatory regions called ICRs (Imprinting Control Regions). There are currently around 120 imprinted genes known in the mouse genome, which are under the control of 20 characterized ICRs, and are generally conserved in Human. My thesis project aimed at characterizing new maternal ICRs and at analyzing their impact on gene regulation, based on a genome-wide methylation screen conducted in the mouse. I participated to revealing the existence of three forms of genomic imprinting, which reflects variable susceptibility to developmentally-regulated DNA methylation changes: 1) ubiquitous and life-long imprinting, which refers to the 20 canonical ICRs, 2) transient, whose existence is limited to preimplantation development, and 3) tissue-specific. More specifically, I deciphered the histone modification profiles of two new maternal ICR associated with the *Cdh15* and the *Gpr1/Zdbf2* loci and confirmed that the *GPR1/ZDBF2* locus is also subject to transient imprinting in Human. My main achievement concerns the characterization of a candidate ICR associated with the *Socs5* gene, which I found to be tissue-specific but also strain-specific, pointing towards a new form of imprinting polymorphism. This ICR has an intragenic position and has the characteristics of an enhancer, hypothesis that I am functionally testing *in vivo* by a CRISPR/Cas9-mediated deletion. The discovery of these new forms of genomic imprinting provides a better understanding of this phenomenon and its impact on phenotypes.

Key words : DNA methylation, genomic imprinting, development, histone modifications, epigenetics, mouse model

TABLE DES MATIERES

INTRODUCTION	16
1- La méthylation de l'ADN chez les mammifères	18
1.1- Les ADN méthyltransférases : Dnmt	18
1.1.1- L'ADN méthyltransférase de maintenance, Dnmt1	20
1.1.2- Les ADN méthyltransférases de novo, Dnmt3A et Dnmt3B	22
1.1.3- Le cofacteur des ADN méthyltransférases de novo, Dnmt3L	24
1.2- Distribution de la méthylation dans le génome des mammifères	26
1.2.1- Contexte CpG versus non-CpG	28
1.2.2- Les îlots CpG (CGI)	28
1.2.3- Les régions intra- et inter-géniques	32
1.2.4- Les régions régulatrices : enhancers et insulateurs	34
1.3- La méthylation de l'ADN et les modifications d'histones	36
1.3.1- Les modifications post-traductionnelles des histones	36
1.3.2- Marques positivement corrélées à la méthylation de l'ADN : H3K9me3 et H3K36me3	38
1.3.3- Marques antagonistes à la méthylation de l'ADN : H3K4me3 et H3K27me3	40
1.4- La dynamique de méthylation du génome au cours du développement	44
1.4.1- Déméthylation du génome dans les PGC	44
1.4.2- Mise en place de la méthylation dans les gamètes	46
1.4.3- Déméthylation de la méthylation dans l'embryon après la fécondation	48
1.4.4- Etablissement de la méthylation dans l'embryon après l'implantation	52
2- L'empreinte génomique parentale	54
2.1- Découverte, définition et rôles de l'empreinte génomique parentale	54
2.1.1- Découverte de l'empreinte génomique parentale	54
2.1.2- Le point sur l'empreinte génomique parentale aujourd'hui : définitions et faits	56
2.1.3- Rôles des gènes soumis à empreinte génomique	58

2.2- Les régions de contrôle de l’empreinte génomique	62
2.2.1- Caractéristiques génétiques et épigénétiques générales des ICR	62
2.2.2- Les ICR maternelles	64
2.2.3- Les ICR paternelles	66
2.2.4- Les DMR secondaires	70
2.3- Protection des profils de méthylation différentiels des ICR	72
2.3.1- Protection des allèles méthylés contre la déméthylation après la fécondation	72
2.3.2- Protection des allèles non-méthylés contre la méthylation après l’implantation	76
2.4- Evolution de l’empreinte génomique parentale	78
3- Les différents types de cribles développés pour l’identification de nouveaux GSE	84
3.1- Les cribles basés sur les caractéristiques de séquence des ICR	84
3.2- Les cribles basés sur la méthylation de l’ADN des ICR	86
3.3- Les cribles basés sur l’expression mono-allélique des GSE	90
 RESULTATS	 94
1- Contexte et questions abordées	96
2- Identification de nouvelles DMR maternelles germinales non-ubiquitaires	98
2.1- Notion d’empreinte génomique tissu-spécifique : le cas de Cdh15	100
2.2- Notion d’empreinte génomique transitoire : le cas de Gpr1/Zdbf2	102
Article 1: Protection against de novo methylation is instrumental in maintaining parent-of-origin methylation inherited from the gametes.	
Article 2: The Gpr1/Zdbf2 locus provides new paradigms for transient and dynamic genomic imprinting in mammals.	
3- Caractérisation d’une nouvelle DMR maternelle, tissu-spécifique et lignée-spécifique	184
3.1- Classification des DMR candidates en régions d’empreinte transitoire ou persistante	184

3.2- Résultats	188
3.2.1- Identification de la DMR de Socs5 en tant que DMR maternelle non-transitoire	188
3.2.2- Variation de la méthylation de la DMR de Socs5 en fonction des tissus et des lignées murines	188
3.2.3- Complexité du caractère lignée-spécifique de l'empreinte de la DMR de Socs5	194
3.2.4- Dynamique développementale de la méthylation de la DMR de Socs5 dans des croisements hybrides homogènes	196
3.2.5- Marques d'histones associées à la DMR	198
3.2.6- Activité transcriptionnelle au niveau du locus Socs5	202
3.2.7- Conservation et évolution du locus Socs5 dans d'autres espèces	206
3.2.8- Délétion génétique de la DMR de Socs5 in vivo par le système CRISPR/Cas9	208
3.3- Figures et Légendes	212

DISCUSSION 246

1- Importance de la protection des allèles non-méthylés contre la méthylation.	248
2- La gDMR de Socs5 : le dernier locus soumis à empreinte persistante ?	252
2.1- Pourquoi le locus Socs5 est-il soumis à empreinte ?	252
2.2- Comment expliquer le caractère lignée-spécifique de la gDMR de Socs5 ?	254
2.3- Quelle est la fonction du gène Socs5 ?	258
2.4- Comment fonctionne la gDMR de Socs5 ?	262
3- Combien de gènes soumis à empreinte reste-il à identifier ?	268
4- Conclusion	270

REFERENCES BIBLIOGRAPHIQUES 274

INTRODUCTION

1- La méthylation de l'ADN chez les mammifères

La méthylation de l'ADN chez les mammifères fait référence à la 5-méthylcytosine (5mC) qui résulte de l'ajout d'un groupement méthyle (-CH₃) sur le carbone 5 d'une des 4 bases azotées de l'ADN, la cytosine (Figure1). Cette base modifiée a été pour la première fois mise en évidence en 1948 par séparation chromatographique d'ADN de thymus de veau (Hotchkiss, 1948), avant même la découverte de la structure double-hélice de l'ADN par Watson et Crick. De part ses fonctions biologiques, elle est considérée comme la cinquième base de l'ADN. En effet, cette modification épigénétique joue un rôle crucial sur l'expression des gènes en association avec les modifications d'histones. Dans les cellules somatiques, le génome est globalement méthylé et seules certaines régions échappent à cette modification, notamment grâce à des mécanismes de protection (Lister et al., 2009 ; Meissner et al., 2008 ; Duffié and Bourc'his, 2013).

La méthylation de l'ADN est impliquée dans un large nombre de processus tel que la répression des éléments répétés afin d'assurer la stabilité et l'intégrité du génome (Bird, 1997 ; Howard et al., 2008 ; Yoder et al., 1997), l'inactivation du chromosome X (Yen et al., 2007) et l'empreinte génomique parentale (Hore et al., 2007 ; Schulz et al., 2010), sujet qui a fait l'objet de ma thèse. D'une manière générale, elle est associée à la répression génique lorsqu'elle se situe au niveau de régions promotrices. La méthylation de l'ADN est essentiellement présente dans un contexte de dinucléotides 5'-CpG-3' (CpG) : l'ajout d'un groupement méthyle sur la cytosine nécessite l'intervention de protéines capables de reconnaître l'ADN et de par leur activité enzymatique, de lier de façon covalente le groupement méthyle. Il s'agit des ADN méthyltransférases.

1.1- Les ADN méthyltransférases : Dnmt

Les ADN méthyltransférases, ou Dnmts, catalysent la réaction de transfert d'un groupement méthyle à partir du donneur S-Adénosyl-1-méthionine sur les cytosines (Figure1) (Chen and Li, 2004). La partie carboxy-terminale, tout d'abord définie par homologie à celui des méthyltransférases bactériennes, est responsable de l'activité catalytique de méthylation de l'ADN (Bestor et al., 1988). Il existe deux mécanismes de méthylation : 1) la méthylation *de novo*, qui correspond à la mise en place des profils de

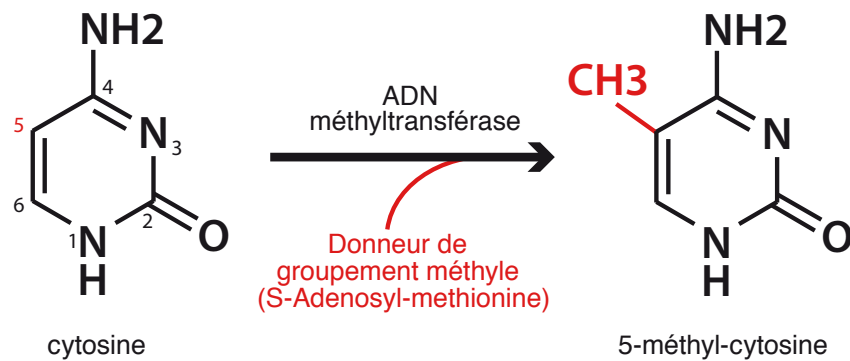


Figure 1 : Réaction de méthylation de l'ADN et structures chimiques de la cytosine (C) et de la 5méthylcytosine (5mC).

La méthylation de l'ADN a lieu via la liaison covalente d'un groupement méthyle (-CH₃) sur le carbone 5 de la cytosine. La réaction nécessite l'activité enzymatique d'une ADN méthyltransférase ainsi qu'un cofacteur donneur de groupement méthyle, la S-Adénosyl-méthionine.

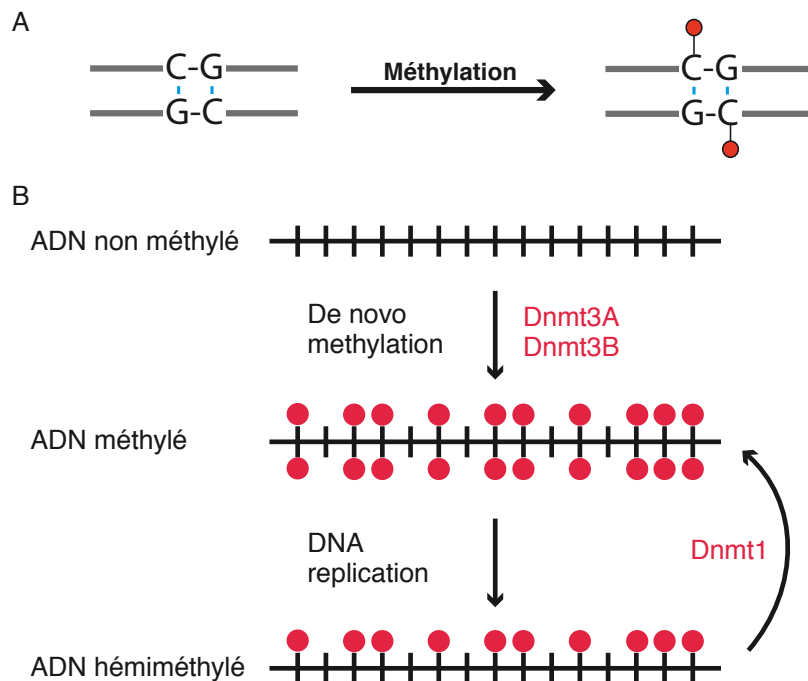


Figure 2 : Mécanismes de méthylation in vivo : méthylation *de novo* et de maintenance.

A. Méthylation symétrique en contexte CpG.

B. Représentation schématique des mécanismes de méthylation *de novo* et de maintenance. Chaque barre verticale représente une paire de CpG symétrique. L'ADN est méthylé *de novo* par Dnmt3A et Dnmt3B. La réplication de l'ADN qui est semi-conservative, génère des brins hémiméthylés. La symétrie est restaurée par l'enzyme de maintenance Dnmt1.

méthylation à partir de substrats naïfs, et qui a lieu dans les gamètes et l'embryon précoce principalement, et 2) la méthylation de maintenance, qui recopie sur le brin néo-synthétisé les profils de méthylation du brin parental lors de la réplication, assurant la fidélité des profils de méthylation lors des divisions cellulaires (Figure2) (Reik, 2007). La famille des ADN méthyltransférases est ainsi composée de quatre protéines divisées en deux groupes : une ADN méthyltransférase de maintenance, Dnmt1 et deux ADN méthyltransférases *de novo* associées à un cofacteur, et qui forment le groupe des Dnmt3 (Figure3) (Goll and Bestor, 2005).

1.1.1- L'ADN méthyltransférase de maintenance, Dnmt1

En 1975, Riggs et Holliday & Pugh sont les premiers à proposer un mécanisme de maintenance des marques de méthylation de l'ADN. Ils prédisent l'existence d'une méthyltransférase qui méthylerait l'ADN non-méthylé avec beaucoup de difficulté mais qui méthylerait efficacement l'ADN hémiméthylé (Riggs, 1975 ; Holliday and Pugh, 1975). L'hypothèse d'une activité de maintenance repose notamment sur le mode de réplication semi-conservatif de l'ADN, qui génère des substrats hémiméthylés, et la symétrie des motifs CpGs (Figure2).

L'existence d'une telle activité de maintenance a ensuite été démontrée dans des cellules de mammifères par des expériences montrant la propagation de la méthylation à travers les divisions cellulaires (Wigler et al., 1981 ; Stein et al., 1982) et par la caractérisation de l'enzyme responsable de cette propagation, Dnmt1 (Bestor and Ingram, 1983). Dnmt1 a une préférence pour l'ADN hémiméthylé et son activité est 30 à 40 fois plus élevée que sur de l'ADN nu (Bestor et al., 1988 ; Yoder et al., 1997). Elle a été localisée au niveau des fourches de réplication de l'ADN, son recrutement se faisant via un domaine spécifique de sa partie N-terminale (Leonhardt et al., 1992) et via l'interaction avec le facteur NP95/UHRF1 qui possède un domaine de liaison spécifique aux CpGs hémiméthylés (Bostick et al., 2007 ; Sharif et al., 2007).

Dnmt1 existe sous trois isoformes, provenant de transcrits différents initiés par le recrutement de promoteurs alternatifs. La forme somatique, Dnmt1s, est exprimée de façon ubiquitaire ; la mutation du gène est létale à 8dpc (day post-coïtum, ou jour de développement embryonnaire) et entraîne une perte globale de la méthylation de l'ADN, ce qui démontre bien l'importance du maintien des profils de méthylation très tôt au

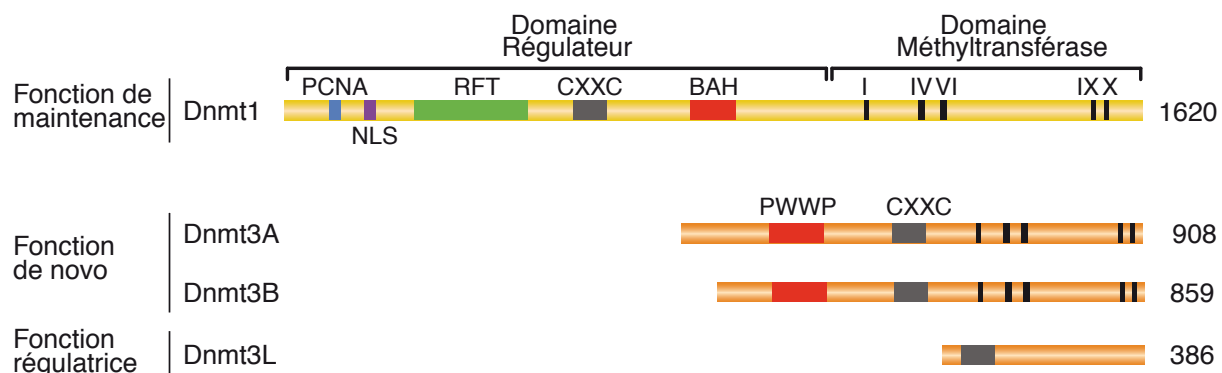


Figure 3 : Famille des ADN méthyltransférases mammifères.

Les DNMTs présentent une partie N-terminale d'interaction avec l'ADN et des protéines régulatrices, et une partie C-terminale portant le domaine d'activité catalytique. Les domaines catalytiques de Dnmt1, Dnmt3A et Dnmt3B sont conservés en particulier au niveau des motifs I, IV, VI, IX et X, retrouvés chez toutes les ADN méthyltransférases de cytosines. Les régions N-terminales sont plus spécifiques à chaque fonction, dépendant de modes de régulation différents. Le ciblage de Dnmt1 au noyau se fait par la séquence NLS (nuclear localisation signal), son recrutement à la fourche de réplication par interaction avec la protéine PCNA (proliferating cellular nuclear antigen) et le domaine RFT (replication foci targeting domain). Les domaines CXXC, riches en cystéines, permettent la liaison à des séquences d'ADN comportant des CpGs. Le domaine BAH (bromodomain) participe à des interactions protéine-protéine, le domaine PWWP contenant un motif riche en proline-tryptophane-tryptophane-proline hautement conservé joue sur l'association avec l'hétérochromatine.

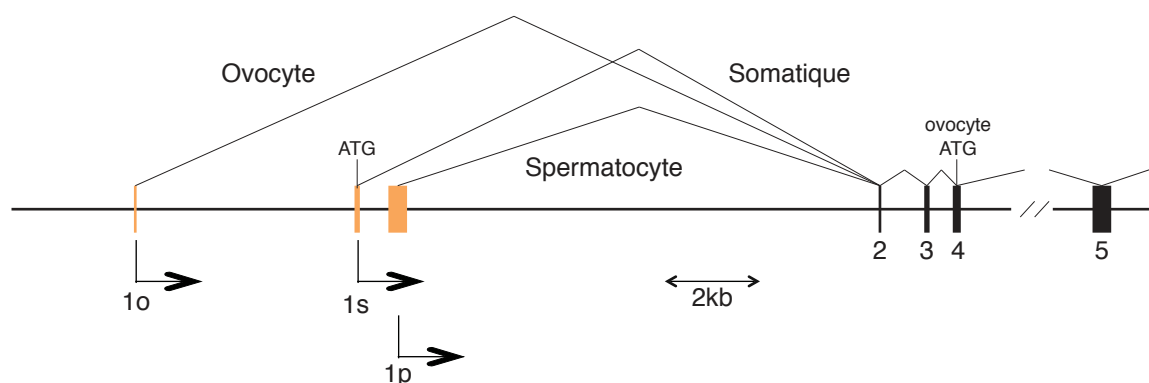


Figure 4 : Transcrit alternatif sexe spécifique du gène *Dnmt1*.

Le gène *Dnmt1* est situé sur le chromosome 9 chez la souris. L'exon 1o est spécifique de l'ovocyte, l'exon 1s est spécifique des cellules somatiques chez les deux sexes et l'exon 1p est restreint aux spermatocytes au stade pachytène. Le codon ATG de l'exon 1s est utilisé pour l'initiation de la traduction dans les cellules somatiques et une forme tronquée est traduite à partir du codon ATG de l'exon 4 dans les ovocytes (adapté de Bestor, 2000).

cours du développement (Ruchirawat et al., 1987 ; Li et al., 1992 ; Arand et al., 2012). Une forme spécifique de l'ovocyte, Dnmt1o, est stockée dans le cytoplasme de l'ovocyte, transmise à l'embryon pré-implantatoire lors de la fécondation et joue un rôle notamment au stade 8 cellules (Carlson et al., 1992 ; Mertineit et al., 1998). Dnmt1o est remplacée par la forme somatique Dnmt1s dans le noyau après l'implantation (Figure4) (Bestor, 2000). Il existe aussi une forme spécifique des spermatocytes au stade pachytene, Dnmt1p, qui inhiberait la production de la forme somatique au moment de la méiose (Mertineit et al., 1998).

1.1.2- Les ADN méthyltransférases *de novo*, Dnmt3A et Dnmt3B

La famille des ADN méthyltransférases *de novo* est composée de deux gènes *Dnmt3a* et *Dnmt3b* qui sont très fortement exprimés dans les gamètes, l'embryon péri-implantatoire et les cellules ES mais dont l'expression est très faible dans les tissus somatiques (Okano et al., 1998 ; Okano et al., 1999 ; Watanabe et al., 2002 ; La Salle et al., 2004 ; Sakai et al., 2004 ; Hirasawa et al., 2008). Les deux enzymes sont capables de méthyliser de l'ADN natif et n'ont pas de préférence particulière pour l'ADN hémiméthylé (Okano et al., 1998 ; Chédin et al., 2002).

Ces deux protéines sont responsables de la méthylation *de novo* du génome : Dnmt3A est préférentiellement requise pour la mise en place de la méthylation dans les gamètes et les tissus adultes, et Dnmt3B pour la mise en place dans l'embryon (Okano et al., 1999 ; Kaneda et al., 2004 ; Borgel et al., 2010 ; Kaneda et al., 2010). Lors de la gamétogenèse, Dnmt3A est notamment requise au niveau des régions soumises à empreinte génomique parentale (Chen et al., 2002 ; Kaneda et al., 2004 ; Kato et al., 2007). Dnmt3B est quant à elle plus spécialisée dans la méthylation des îlots CpG présents sur les autosomes ainsi que les régions du chromosome X inactif (Okano et al., 1999 ; Auclair et al., 2014).

Une délétion constitutive de *Dnmt3A* a pour conséquence la mort des individus juste après la naissance, vers 28dpp (days post-partum), tandis que la délétion de *Dnmt3B* cause une létalité très précoce vers 9,5dpc (Okano et al., 1999). Les souris double mutante *Dnmt3A* et *Dnmt3B* présentent une létalité à 8,5dpc et leur génome est complètement déméthylé, illustrant la coopération des deux enzymes (Okano et al., 1999). En plus des motifs catalytiques responsables de la réaction de transméthylation,

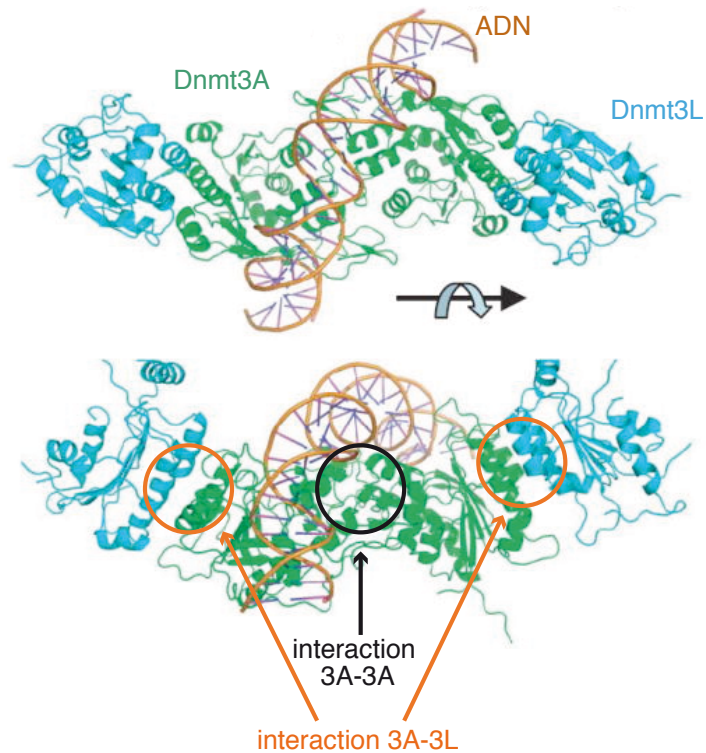


Figure 5 : Modèle d'interaction du complexe Dnmt3A-Dnmt3L avec l'ADN. Vue sous deux angle du complexe Dnmt3A-Dnmt3L organisé sous forme d'hétérotétramère Dnmt3L-Dnmt3A-Dnmt3A-Dnmt3L, en interaction avec une molécule d'ADN de 25 nucléotides présentant des cibles pour les 2 sites actifs du complexe. (d'après Jia et al., 2007).

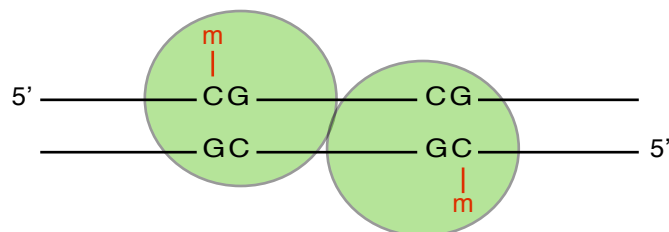


Figure 6 : Modèle pour la méthylation des CpGs par le dimère Dnmt3A. (d'après Cheng and Blumenthal, 2008).

Dnmt3A et Dnmt3B ont toutes les deux un domaine CXXC, riche en cystéines, permettant la liaison à des séquences d'ADN contenant des CpGs (Figure3). Les deux enzymes possèdent aussi un domaine PWWP fréquemment présent dans les protéines associées à la chromatine (Stec et al., 2000) et participerait au ciblage de ces enzymes sur la chromatine, particulièrement en présence de certaines modifications d'histones, telle que la marque H3K36me3 (Chen et al., 2004 ; Dhayalan et al., 2010). Enfin, le domaine ADD (ATRX-DNMT3-DNMT3L) des Dnmt3 leur confère une répulsion vis-à-vis de la méthylation en H3K4 (Ooi et al., 2007 ; Otani et al., 2009 ; Zhang et al., 2010 ; Morselli et al., 2015).

1.1.3- Le cofacteur des ADN méthyltransférases de novo, Dnmt3L

Dnmt3L (*Dnmt3-Like*) est une protéine homologue aux protéines Dnmt3A et Dnmt3B, avec les mêmes domaines en partie N-terminale, mais elle ne possède pas de domaines catalytiques en C-terminale nécessaires à la réaction de méthylation de l'ADN (Figure3) (Aapola et al., 2000 ; Bourc'his et al., 2001 ; Chedin et al., 2002 ; Hata et al., 2002). Cette protéine inactive est apparue spécifiquement chez les mammifères placentaires, probablement par duplication du gène *Dnmt3A* (Yokomine et al., 2006).

Dnmt3L interagit avec Dnmt3A et Dnmt3B, *in vivo* et *in vitro*, pour stimuler leur activité de méthylation *de novo* (Chédin et al., 2002 ; Suetake et al., 2004 ; Moarefi and Chédin, 2011). Des études de cristallographie ont permis de démontrer que Dnmt3A et Dnmt3L formaient un complexe physique sous forme d'hétéro-tétramère Dnmt3L-Dnmt3A-Dnmt3A-Dnmt3L (Figure5) (Jia et al., 2007). La présence de deux sites catalytiques Dnmt3A suggère la possibilité de méthyler deux CpGs en même temps s'ils sont séparés de 8 à 10 paires de bases lorsque le complexe est lié à une même molécule d'ADN (Figure6). Dnmt3L est très fortement exprimée pendant la gamétogenèse sous le contrôle de deux promoteurs sexe-spécifiques (Bourc'his et al., 2001 ; Shovlin et al., 2007). Dans les pro-spermatogonies, le transcrit canonique est exprimé alors que dans les ovocytes, un transcrit plus long prend son origine au niveau d'un promoteur en amont du promoteur canonique. Les deux transcrits produisent cependant une protéine de même taille (O'Doherty et al., 2011). Enfin, comme pour *Dnmt1*, un transcrit plus court spécifique du stade pachytène est également produit.

La stimulation de la machinerie de méthylation *de novo* par Dnmt3L est indispensable pour la méthylation des gamètes, mais elle n'est en revanche pas nécessaire à la méthylation *de novo* dans l'embryon (Chedin et al., 2002 ; Hata et al., 2002 ; Bourc'his and Bestor, 2004 ; Guenatri et al., 2013). Les souris homozygotes mutantes pour *Dnmt3L* sont ainsi viables, mais l'absence de méthylation dans la lignée germinale est la cause de stérilité dans les deux sexes (Bourc'his et al., 2001 ; Hata et al., 2002 ; Bourc'his and Bestor, 2004). Chez les mâles, les animaux déficients pour *Dnmt3L* présentent un hypogonadisme et une azoospermie complète dus à la réactivation des rétrotransposons pendant la spermatogenèse (Bourc'his et al., 2001 ; Bourc'his and Bestor, 2004 ; Webster et al., 2005). Chez les femelles, la méthylation des régions soumises à empreinte génomique maternelle n'est pas établie en absence de *Dnmt3L*, ce qui n'empêche pas la production d'ovocytes matures mais résulte en une létalité embryonnaire systématique des embryons *Dnmt3L*^{-/+}, par effet maternel (Bourc'his et al., 2001 ; Hata et al., 2002 ; Smallwood et al., 2011 ; Kobayashi et al., 2012).

1.2- Distribution de la méthylation dans le génome des mammifères

L'existence de cytosines modifiées chez les mammifères est connue depuis plus de 60 ans, mais leur distribution dans le génome n'a été établie de façon précise que récemment. Grâce aux nouvelles techniques d'analyse pangénomique, telles que MeDIP-on-chip (Methylated DNA ImmunoPrecipitation), MeDIP-seq, WGBS (Whole Genome Bisulfite Sequencing) et RRBS (Reduced Representation Bisulfite Sequencing), nous disposons aujourd'hui de cartes détaillées de la méthylation de l'ADN, appelées méthylomes, dans divers types cellulaires et à différents stades du développement (Smallwood et al., 2011 ; Smith et al., 2012 ; Kobayashi et al., 2012 ; Xie et al., 2012 ; Hon et al., 2013). On observe chez les mammifères une distribution sur tout le génome avec un enrichissement au niveau des séquences répétées et des régions intergéniques (Weber and Schubeler, 2007 ; Lister et al., 2009) mais la méthylation n'est pas répartie de façon aléatoire. Comment la méthylation est ciblée dans le génome est l'objet de recherches actives. Dnmt1 et la famille des Dnmt3 n'ayant pas de spécificité de séquence à part les cytosines, la chromatine est suspectée de jouer un rôle dans la régulation des profils de méthylation génomique. Ce point sera discuté plus en détail dans la partie 1.3.

1.2.1- Contexte CpG versus non-CpG

Les dinucléotides CpG ont longtemps été considérés comme étant les seules cibles des Dnmt chez les mammifères. La publication du premier méthylome humain complet par WGBS en 2009 a mis en évidence que la méthylation en contexte nonCG comptait pour 25% de la méthylation totale des cellules ES humaines analysées (Lister et al., 2009). La densité de méthylation nonCG suit celle de la méthylation en contexte CG, avec une préférence pour les dinucléotides CpA et CpT. Cette méthylation est deux fois plus présente dans le corps des gènes au niveau des exons que dans les régions régulatrices. Cette méthylation nonCG avait déjà été observée précédemment dans des cellules ES murines et Dnmt3A avait été impliquée dans la mise en place de cette méthylation (Ramsahoye et al., 2000). Chez les mammifères, le rôle de cette méthylation nonCG est pour le moment inconnu. Les ovocytes possèdent aussi une fraction de méthylation importante en contexte nonCG (Tomizawa et al., 2011 ; Shirane et al., 2013). En absence de Dnmt3A et Dnmt3L, cette méthylation n'est pas présente, confirmant que la machinerie de méthylation *de novo* est impliquée dans l'établissement de la méthylation nonCG (Shirane et al., 2013).

La méthylation nonCG est donc spécifique des contextes cellulaires de forte activité de méthylation *de novo*, à savoir dans les cellules ES et la lignée germinale. Cependant, la méthylation nonCG n'ayant pas lieu dans un contexte symétrique (par exemple un motif CA sur un brin sera un motif TG sur le brin en face), la méthyltransférase de maintenance Dnmt1 ne peut pas propager la méthylation des sites nonCG. De ce fait, la méthylation nonCG est fatalement perdue au cours des divisions cellulaires et n'est plus présente dans les cellules somatiques ou dans le sperme mature à la fin des divisions de la spermatogenèse (Chedin, 2011 ; Ichiyanagi et al., 2013). Par contre, les ovocytes matures ne se divisent pas après que leur méthylation ait été établie et contiennent de ce fait une forte densité de cytosines nonCG méthylées, environ deux tiers de toutes les cytosines méthylées (Shirane et al., 2013).

1.2.2- Les îlots CpG (CGI)

Les îlots CpG sont des régions du génome riches en CpG dont la taille varie entre 0,4 et 2kb et qui possèdent un taux de C+G supérieur à 55% et un ratio de CpGs observés/CpG attendus supérieur à 60% (Bird, 1986). Ils sont en général associés aux

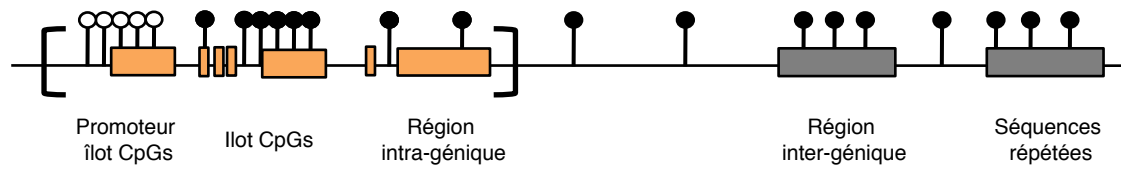
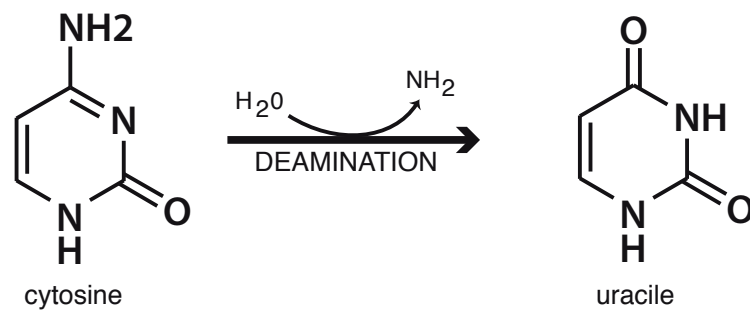


Figure 7 : Répartition de la méthylation de l'ADN en contexte CpGs dans le génome.

Les CpGs sont représentés par des cercles blancs quand ils sont déméthylés et noirs quand ils sont méthylés. Les cadres oranges représentent les exons au sein d'un gène qui est délimité par des crochets. Le génome des mammifères est globalement méthylé à l'exception des îlots CpGs situés au niveau des régions promotrices.

A



B

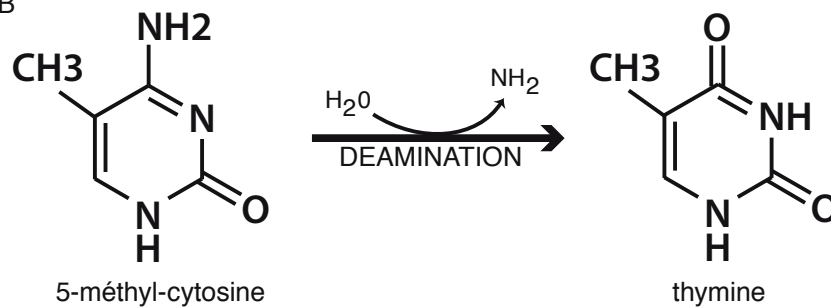


Figure 8 : Réaction de déamination.

A. La déamination de la cytosine donne un uracile.

B. La déamination de la cytosine méthylée produit une thymine.

régions promotrices en 5' des gènes (environ 50% de tous les CGI) où ils sont déméthylés dans 90% des cas. On les retrouve aussi en régions intra- et inter-géniques, où ils sont au contraire globalement méthylés (Figure7) (Weber et al., 2007).

Les CGI existent justement parce qu'ils sont déméthylés. La 5mC peut être spontanément convertie en thymine par une réaction de déamination, ce qui entraîne par conséquent la perte des CpG méthylés au cours de l'évolution, en particulier s'ils sont méthylés dans la lignée germinale (Figure8) (Weber and Schubeler, 2007 ; Smallwood et al., 2011). Bien que la majorité des CGI en régions promotrices soient déméthylés, il existe des exceptions comme les promoteurs des gènes de pluripotence et les gènes spécifiques de la lignée germinale qui sont réprimés dans les cellules somatiques différenciées (Deaton and Bird, 2011 ; Farthing et al., 2008 ; Weber et al., 2007). Les CGI associés aux gènes exprimés de façon monoallélique, comme ceux du chromosome X chez la femelle ou ceux soumis à empreinte génomique parentale, présentent une méthylation différentielle avec seulement un des deux allèles méthylés (Razin and Cedar, 1994 ; Jaenisch and Bird, 2003). Certains CGI sont aussi méthylés de façon tissu-spécifique, en association avec une expression du gène associé seulement dans un certain type de cellules ou certain tissus (Maunakea et al., 2010 ; Lienert et al., 2011 ; Wan et al., 2013).

L'absence de méthylation au niveau des promoteurs qui contiennent des CGI peut être associée à un état actif ou inactif des gènes à proximité, ce qui implique que d'autres mécanismes de régulation entrent en jeu afin de réguler l'expression génique. La présence de modifications d'histones ou la disponibilité de certains facteurs de transcription sont notamment impliqués dans cette régulation. Par contre, la méthylation des promoteurs contenant des CGI est en général associée à un état inactif, agissant comme un verrou. Il existe cependant différentes classes de CGI en fonction de leur densité en CpG : les promoteurs les plus riches en CpG ont un taux de méthylation en corrélation avec leur taux de répression alors que pour les promoteurs les moins enrichis en CpG, la méthylation peut être associée à un état réprimé ou exprimé (Weber et al., 2007).

1.2.3- Les régions intra- et inter-géniques

De façon générale, on appelle régions intra-géniques, ou corps des gènes, les régions transcrites d'un gène sans leurs éléments régulateurs tels que le promoteur. Ces régions sont généralement méthylées et associées à des modifications d'histones spécifiques telle que la marque H3K36me3, déposée lors de la phase d'élongation de la transcription (Ball et al., 2009 ; Hodges et al., 2009). La méthylation de ces régions n'induit pas de répression transcriptionnelle et au contraire on observe un taux de méthylation plus important dans le corps des gènes fortement transcrits (Jones, 1999 ; Lister et al., 2009). Il a été proposé que l'élongation de la transcription en elle-même stimule la méthylation de l'ADN, la marque H3K36me3 associée à l'élongation étant impliquée dans le recrutement des Dnmts via leur domaine PWWP (Dhayalan et al., 2010 ; Hahn et al., 2011 ; Baubec et al., 2015) (voir Introduction section 1.3.2). Mais inversement, il existe aussi des preuves que la transcription pourrait être stimulée par la méthylation intra-génique (Yang et al., 2014).

Plusieurs rôles ont été proposés quant à la méthylation des régions intra-géniques : la répression de promoteurs alternatifs présents dans le corps des gènes et la régulation de l'épissage alternatif des gènes (Suzuki and Bird, 2008 ; Shukla et al., 2011 ; Yearim et al., 2015). Tout d'abord, la méthylation de l'ADN dans le corps des gènes en association avec la marque répressive H3K9me3 permettrait de réprimer l'initiation d'évènements de transcription à partir de promoteurs cryptiques qui viendrait interférer avec la transcription initiée au promoteur canonique (Suzuki and Bird, 2008). D'autre part, il a été montré que la protéine CTCF, qui se lie sur l'ADN non-méthylé, permet l'inclusion de certains exons (à proximité du site CTCF) en obligeant l'ARN Pol II à faire une pause sur ces exons. Lorsque l'ADN est méthylé, CTCF ne se lie plus et par conséquent l'exon à proximité n'est plus incorporé dans le transcrit en cours d'élongation (Shukla et al., 2011). Enfin, la protéine HP1 se lie à la chromatine via les résidus H3K9me3 présents dans le corps des gènes méthylés, et recrute le facteur d'épissage SRSF3, qui va permettre de spécifiquement exclure l'exon sur lequel il se lie (Yearim et al., 2015).

Les régions inter-géniques sont situées entre les gènes et sont généralement pauvres en CpG et largement méthylées (Weber and Schubeler, 2007). Le rôle de la méthylation au niveau de ces régions n'est pas encore bien déterminé, mais comme les régions inter-géniques contiennent en grande partie les éléments transposables du

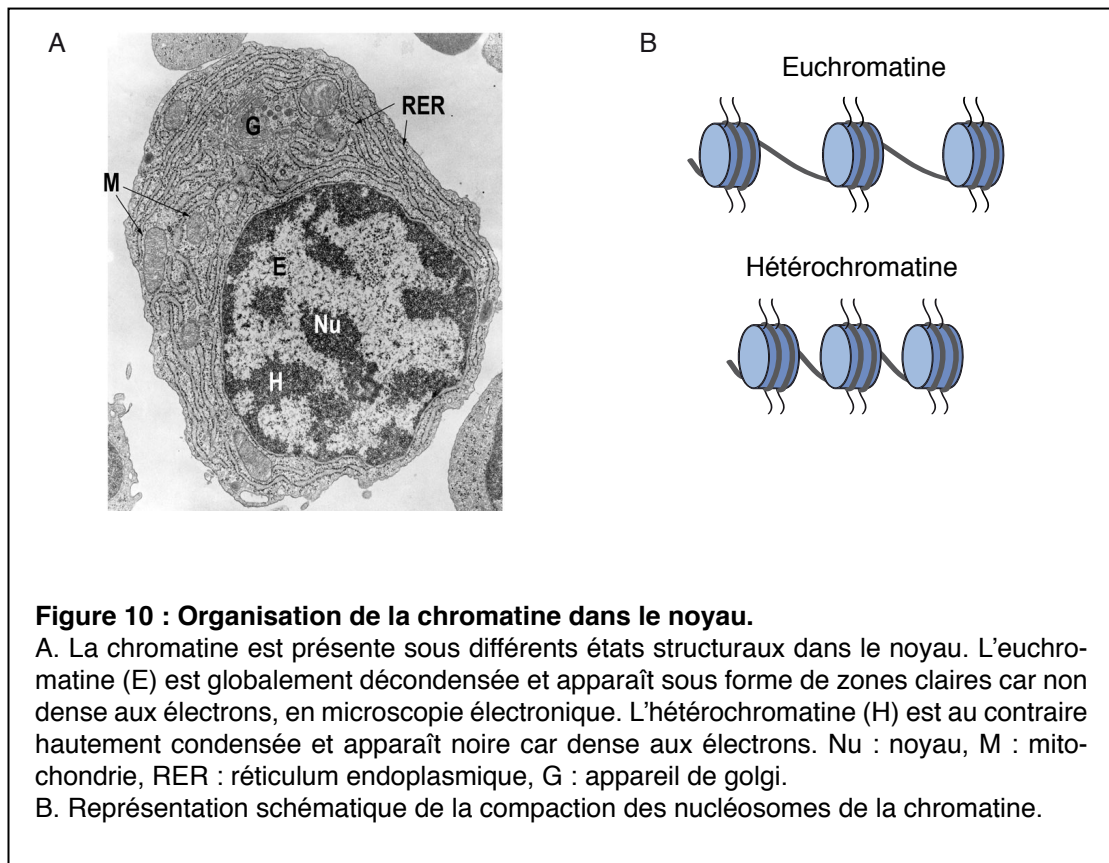
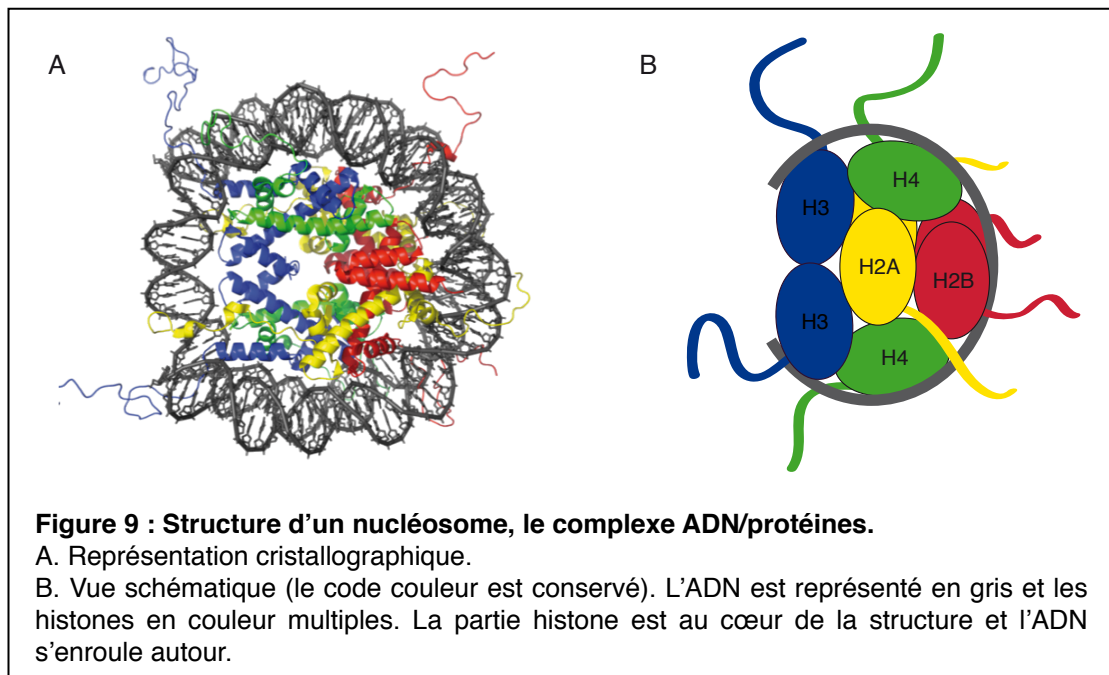
génomique, elles pourraient participer à la stabilité du génome en les maintenant inactifs (Rollins et al., 2006). De plus, les enhancers sont largement retrouvés au niveau de ces régions inter-géniques.

1.2.4- Les régions régulatrices : enhancers et insulateurs

Les enhancers sont situés à des distances variables des promoteurs et sont des éléments clés pour le contrôle de l'expression des gènes par la liaison de facteurs de transcription pendant le développement et la différenciation des cellules. Ils sont en général pauvres en CpG et ont des niveaux de méthylation très variables (Cokus et al., 2008 ; Lister et al., 2009). Dans le génome murin, des enhancers ont été identifiés sur la base qu'ils n'étaient ni 100% méthylés ni 100% déméthylés et ces régions ont été nommées LMRs pour « low methylated regions » (Stadler et al., 2011). Un autre groupe a aussi identifié un large nombre de régions différentiellement méthylées dans des cellules T au niveau d'enhancers de gènes impliqués dans la différenciation (Schmidl et al., 2009).

Le statut de méthylation d'un enhancer et son activité sont liés, comme le montre l'exemple du récepteur aux glucocorticoïdes qui, en se fixant sur sa cible, va entraîner la déméthylation du CpG et l'activation de l'enhancer (Wiench et al., 2011). D'une manière générale, l'hypométhylation d'un enhancer est corrélée avec l'expression du gène sous contrôle de cet enhancer (Carone et al., 2010 ; Sandovici et al., 2011). Les relations entre la méthylation des CpG et la fixation de facteurs de transcription restent tout de même complexes, et nous ne connaissons pas encore les mécanismes qui sont impliqués dans la régulation des enhancers par la méthylation.

Les insulateurs sont des éléments de séquences qui bloquent l'interaction entre un enhancer et un promoteur. Les insulateurs les plus étudiés sont les séquences d'ADN liées par la protéine CTCF. Un des exemples les plus connus est celui du locus soumis à empreinte génomique *Igf2/H19* : la méthylation du site de fixation pour CTCF à ce locus empêche la liaison de la protéine CTCF et promeut l'interaction entre l'enhancer à distance et le promoteur d'*Igf2*. A l'inverse lorsque le site de fixation est déméthylé, CTCF peut se lier et empêche alors l'interaction de l'enhancer avec le promoteur d'*Igf2* (Bell and Felsenfeld, 2000 ; Smits et al., 2008). De manière intéressante, des études sur des cellules embryonnaires souches (ES) murines ainsi que des cellules différenciées ont



suggéré que la méthylation du site de fixation de CTCF ne jouait pas de rôle sur la liaison de CTCF au niveau de régions pauvres en CpG, mais que ce serait sa liaison qui entraînerait la déméthylation (Stadler et al., 2011).

1.3- La méthylation de l'ADN et les modifications d'histones

La chromatine est un complexe ADN/protéines : la double hélice d'ADN (146pb) s'enroule autour d'une unité appelée nucléosome qui est constitué d'un octamère de protéines composé de deux molécules d'histones H2A, H2B, H3 et H4 (Figure9). Les nucléosomes sont l'unité de base de la chromatine et ils jouent un rôle dans la compaction de l'ADN qui elle-même joue un rôle sur l'état transcriptionnel des gènes. Deux états de la chromatine ont été observés : l'euchromatine, décondensée, qui contient aussi bien des gènes actifs qu'inactifs, et l'hétérochromatine, hautement condensée, qui reflète un état silencieux (Figure10). La régulation de ces états peut se faire via le positionnement des nucléosomes, les différentes modifications post-traductionnelles des histones, l'incorporation de variants d'histones et le recrutement de protéines associées aux histones. Il semble évident que l'organisation et les modifications des protéines de la chromatine influencent les profils de méthylation de l'ADN, notamment par modulation de l'accessibilité des Dnmt à la molécule d'ADN.

1.3.1- Les modifications post-traductionnelles des histones

Les histones sont constituées de deux parties, une partie globulaire en C-terminale qui est insérée dans le nucléosome et une partie N-terminale qui sort du nucléosome et qu'on appelle la queue des histones. Ces queues d'histones sont accessibles à différentes protéines et sont le siège de modifications post-traductionnelles sur plusieurs résidus. On connaît aujourd'hui pas moins de 60 résidus pouvant être ciblés par 8 types de modifications ayant une action sur l'état de la chromatine et la régulation des gènes (Figure11) (Kouzarides, 2007). Certaines modifications sont ainsi corrélées à un état ouvert et transcriptionnellement actif de la chromatine comme l'acétylation des histones et la méthylation des lysines 4, 36 ou 79 de H3 ; d'autres sont liés à un état condensé et transcriptionnellement inactif de la chromatine comme la méthylation des lysines 9 et 27 de H3 ou 20 de H4 (Tableau1)



Figure 11 : Sites de modification de la queue des histones.

La queue amino-terminale des histones constitue un quart de la masse d'un nucléosome. Elle abrite la majorité des modifications covalentes connues, comme illustré dans cette figure. Certaines modifications ont lieu aussi dans la partie globulaire (sous forme de rectangle) dont certaines sont indiquées ici. En générale les marques dites actives incluent l'acétylation (drapeaux bleus Ac), la méthylation des arginines (hexagones jaunes Me) et certaines méthylation de lysines comme H3K4 et H3K36 (hexagones verts Me). La méthylation d'H3K79 dans la partie globulaire exerce une fonction anti-répressive. Les marques répressives sont composées par la méthylation de H3K9, H3K27 et H4K20 (Hexagones rouges Me) (d'après Epigenetics © 2007 CSHL press).

	Rôle sur la transcription	Sites modifiés
Groupe 1 - petites modifications		
Acétylation	Activation	H3 (K9, K14, K18, K56) H4 (K5, K8K12, K16) H2A H2B (K6, K7, K16, K17)
Phosphorylation	Activation	H3(S10)
Méthylation	Activation Répression	H3 (K4, K36, K79) H3(K9, K27) H4 (K20)
Groupe 2 - gros peptides		
Ubiquitylation	Activation Répression	H2B (K123) H2A (K119)
Sumoylation	Répression	H3 (?) H4 (K5, K8, K12, K16,) H2A (K126) H2B (K6, K7, K16, K17)

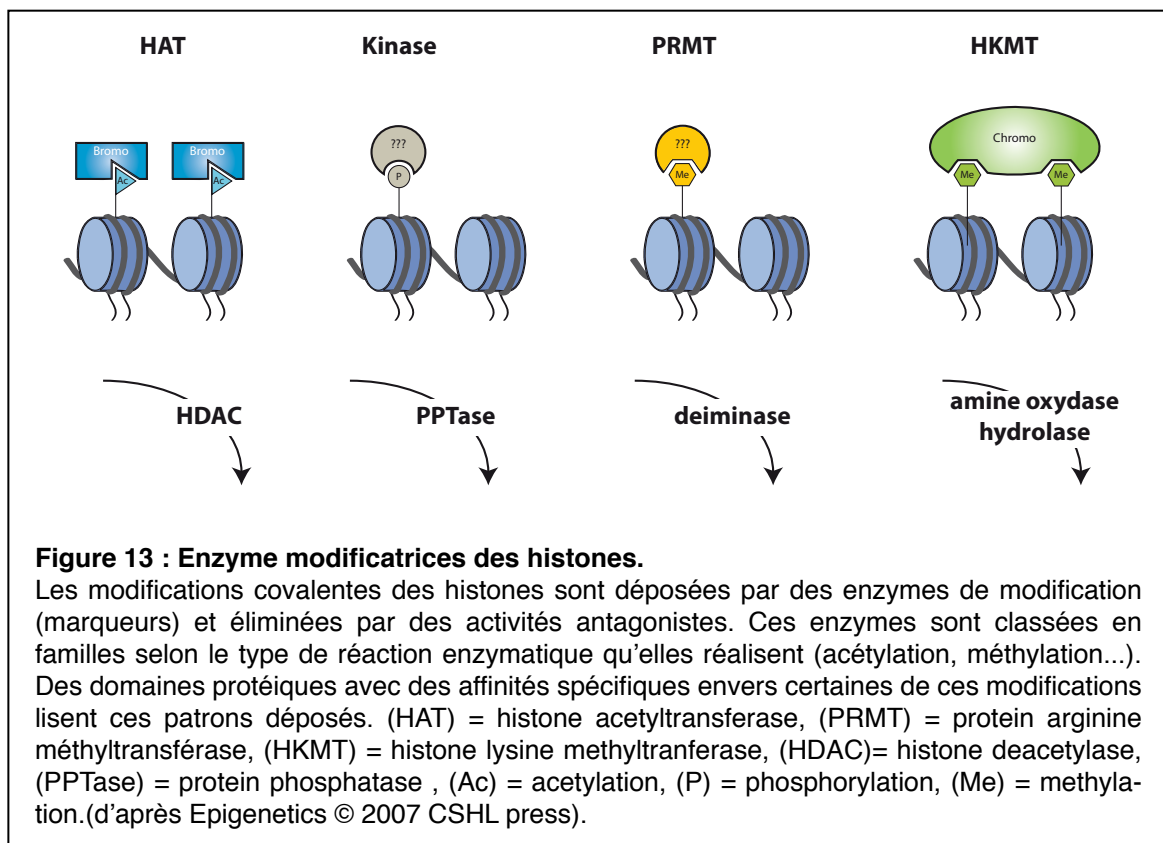
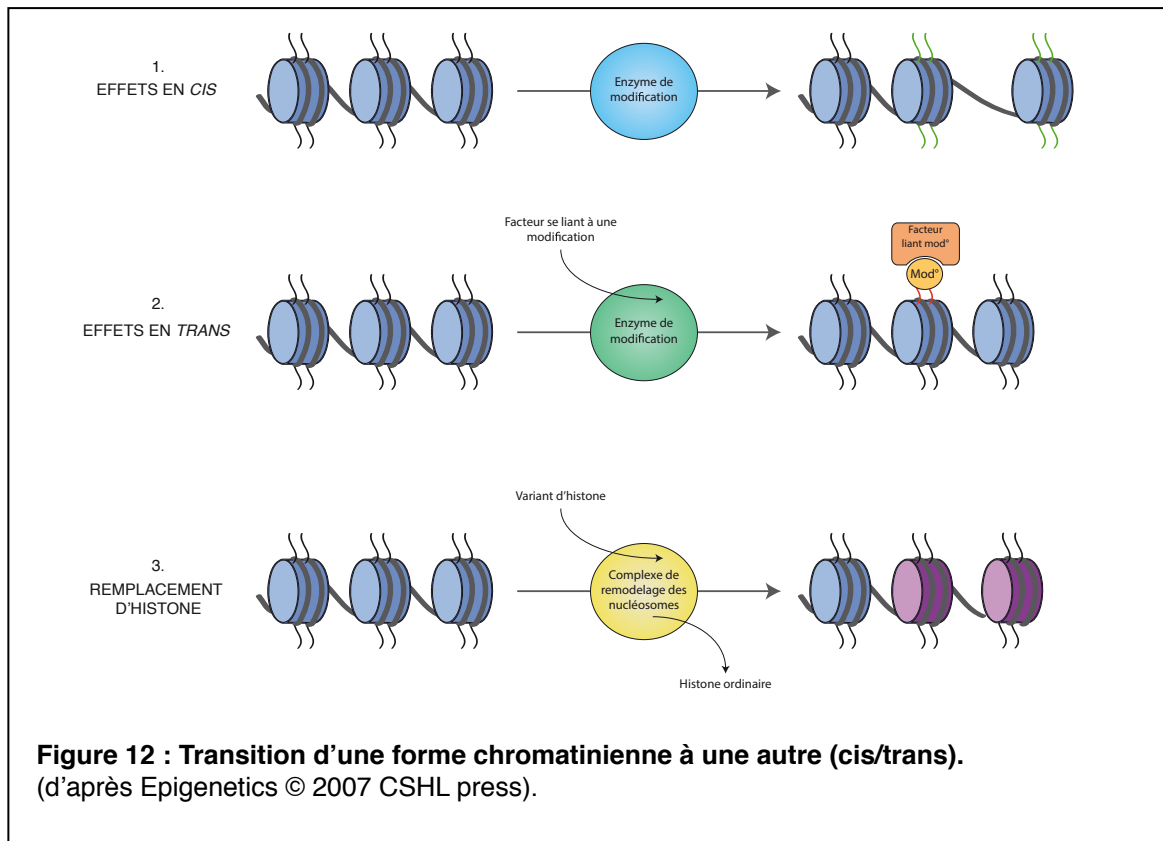
Tableau 1 : Modifications post-traductionnelles des histones.
(d'après Epigenetics © 2007 CSHL press).

(Jenuwein and Allis, 2001 ; Kouzarides, 2007). La partie globulaire des histones peut elle aussi être la cible de modifications post-traductionnelles mais leur rôle est aujourd'hui incertain (Kebede et al., 2015).

Trois modes d'action de ces modifications des queues d'histones sur la régulation des gènes sont connus (Figure12) (Allis et al., 2007). Tout d'abord une action en *cis* peut modifier la structure de la chromatine. C'est le cas de l'acétylation qui permet une décompaction des nucléosomes en neutralisant les charges positives des queues d'histones et donne ainsi accès à la machinerie de la transcription. Ensuite, une action en *trans* permet l'activation ou la répression des gènes par des facteurs ou des complexes qui reconnaissent et se lient aux modifications d'histones via des domaines protéiques spécialisés. Ces facteurs peuvent aussi bien être des enzymes modifiant elles-mêmes les histones que des complexes protéiques capables de remodeler la chromatine. Ces enzymes de modification sont classées selon les réactions qu'elles catalysent (Figure13). Enfin, les complexes de remodelage de la chromatine peuvent remplacer certaines histones (H2A et H3) par des variants (H2A.X, H2A.Z, CENH3) impliqués dans divers processus spécifiques dans la cellule (Talbert and Henikoff, 2010).

1.3.2- Marques positivement corrélées à la méthylation de l'ADN : H3K9me3 et H3K36me3

Les premières évidences montrant que les modifications d'histones pourraient contrôler la méthylation de l'ADN viennent d'études chez le champignon *Neurospora crassa* où la méthylation globale de l'ADN dépend de la tri-méthylation de H3K9 (Tamaru et al., 2001). Chez les plantes, la marque H3K9me3 est aussi importante pour la mise en place d'une partie de la méthylation de l'ADN, notamment au niveau des rétro-transposons (Jackson et al., 2002 ; Malagnac et al., 2002). Chez les mammifères, les deux enzymes Suv39h1 et Suv39h2 sont responsables de la marque H3K9me3 au niveau de l'hétérochromatine péricentromérique et sont indispensables pour en maintenir la stabilité (Peters et al., 2001 ; Lehnertz et al., 2003). La protéine HP1 (heterochromatin protein 1) servirait de lien entre H3K9me3 et la méthylation de l'ADN. Il existe trois variants de cette protéine, HP1 α , HP1 β et HP1 γ , qui interagissent tous via leur chromo-domaine avec H3K9me3 et via leur « chromoshadow » domaine avec Dnmt1, Dnmt3A et Dnmt3B (Bannister et al., 2001 ; Lehnertz et al., 2003 ; Fuks et al., 2003 ; Smallwood et al., 2007). SETDB1, une autre histone méthyltransférase permettant la tri-méthylation

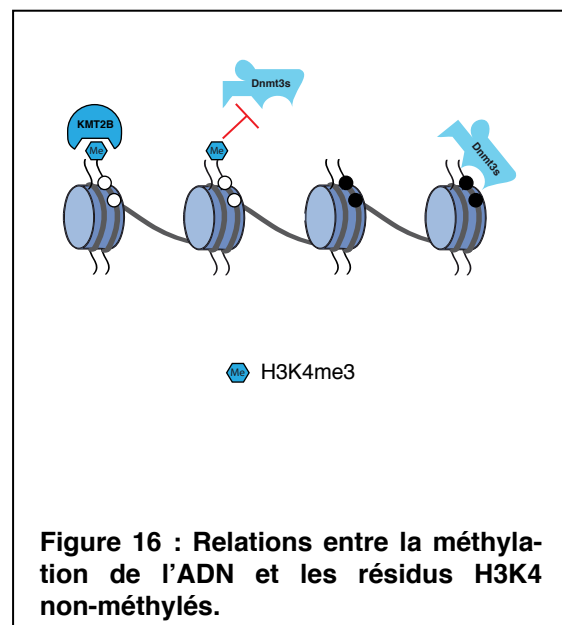
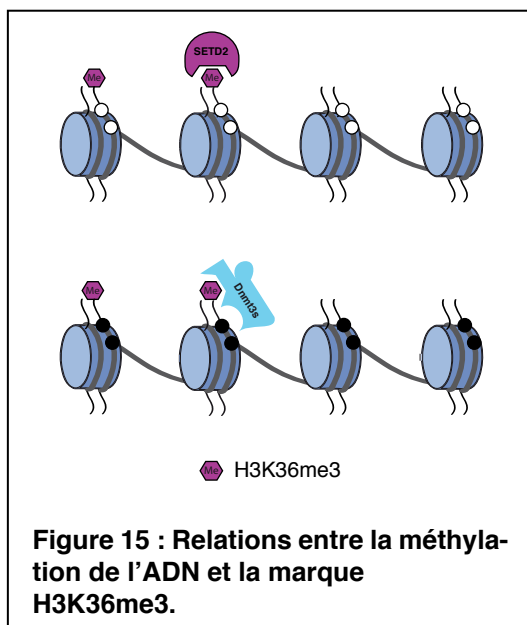
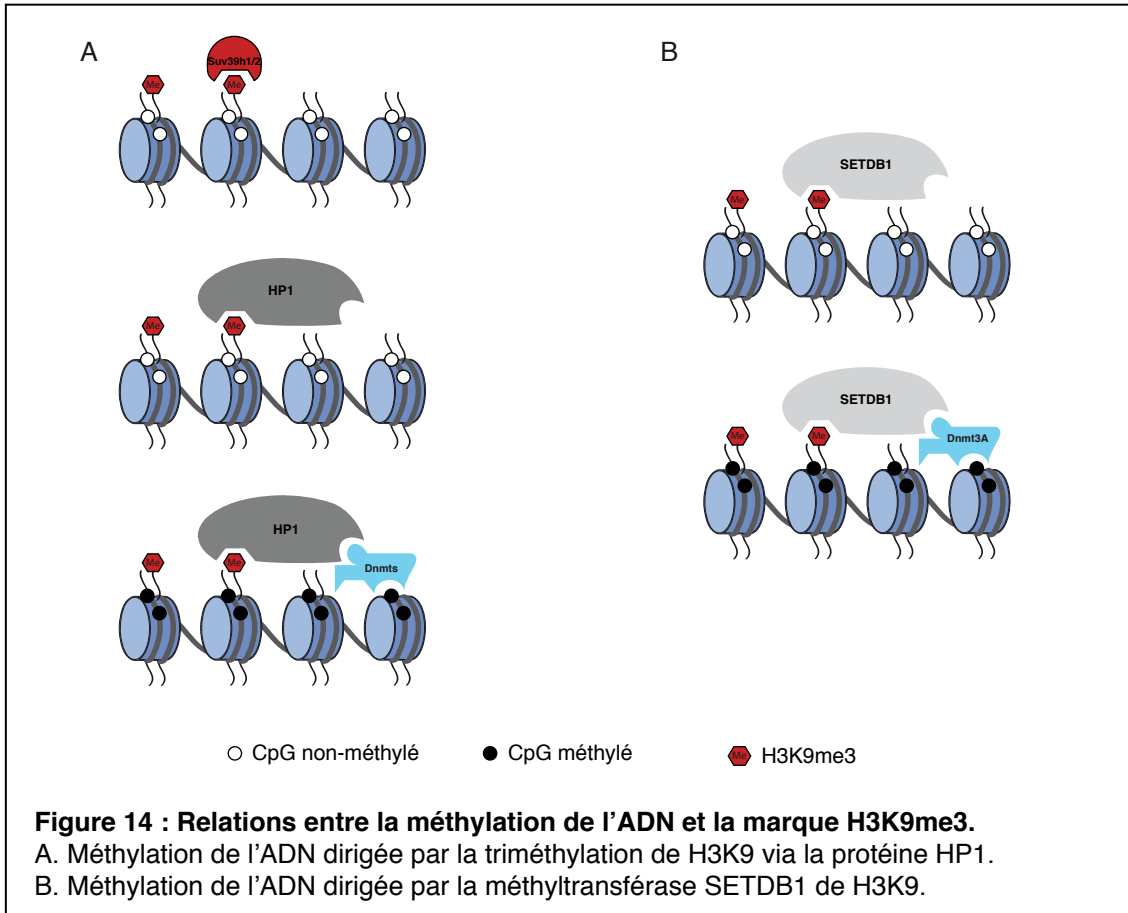


de H3K9 sur l'éuchromatine, a été détectée en complexe avec l'ADN méthyltransférase *de novo* Dnmt3A (Li et al., 2006). La marque H3K9me3 répressive, via la protéine HP1 ou seule, permettrait le recrutement des Dnmt au niveau de régions qui doivent rester silencieuses apportant un second niveau de « verrouillage » (Figure14).

Comme nous l'avons déjà vu précédemment, une autre marque a aussi été étudiée pour sa relation positive avec la méthylation de l'ADN : il s'agit de la marque H3K36me3 que l'on retrouve dans le corps des gènes transcrits (Hawkins et al., 2010). Il a été proposé que la méthylation de l'ADN dans le corps des gènes empêcherait l'initiation d'évènements de transcription aberrants et un rôle similaire est proposé pour la marque H3K36me3 (Bird, 1997 ; Carrozza et al., 2005). Cette marque est établie par l'histone méthyltransférase SETD2 qui forme un complexe avec l'ARN polymérase II hyperphosphorylée pendant l'élongation (Edmunds et al., 2008 ; Sun et al., 2005 ; Yoh et al., 2008). Dnmt3A et 3B ont dans leur domaine N-terminal un domaine PWWP impliqué dans la reconnaissance spécifique de la marque H3K36me3 (Dhayalan et al., 2010 ; Baubec et al., 2015). De plus, l'interaction entre le domaine PWWP et H3K36me3 stimule l'activité de Dnmt3A pour méthyler l'ADN (Dhayalan et al., 2010). Cette même interaction permet de cibler Dnmt3B et donc la méthylation de l'ADN au niveau du corps des gènes (Baubec et al., 2015 ; Morselli et al., 2015). Toutes ces études montrent que le dépôt de la marque H3K36me3 au cours de la transcription permet le recrutement des Dnmt3A et 3B afin de méthyler le corps des gènes et potentiellement, préserver l'intégrité du génome (Figure15).

1.3.3- Marques antagonistes à la méthylation de l'ADN : H3K4me3 et H3K27me3

Les séquences du génome dépourvues de méthylation de l'ADN sont en majorité des promoteurs contenant des îlots CpG (Bestor et al., 1992 ; Rollins et al., 2006). Ces promoteurs ont des caractéristiques chromatinienne tel que la marque H3K4me3 lorsqu'ils sont transcriptionnellement actifs. On ne trouve jamais d'association entre H3K4me3 et méthylation de l'ADN (Weber et al., 2007 ; Meissner et al., 2008 ; Otani et al., 2009 ; Edwards et al., 2010). De plus, dans des cellules ES mutantes pour KMT2B, une H3K4 histone méthyltransférase, on observe un gain aberrant de méthylation de l'ADN au niveau de l'îlot CpG du promoteur du gène *MagohB* (Ladopoulos et al., 2013).



En accord avec cette relation mutuellement exclusive, il a été montré que le domaine ADD de Dnmt3A et Dnmt3L interagit avec la queue de l'histone H3 seulement quand la lysine 4 est déméthylée (Ooi et al., 2007 ; Otani et al., 2009 ; Zhang et al., 2010). Une telle interaction n'a pas été montrée entre Dnmt3B et les résidus non-méthylés H3K4, bien qu'il ait été observé une relation exclusive entre Dnmt3B et H3K4me3 (Morselli et al., 2015). Comme Dnmt3B possède aussi un domaine ADD, il est probable que Dnmt3B soit aussi exclue des sites enrichis en H3K4me3 durant le développement (Chedin, 2011). (Figure16).

Une autre marque est aussi corrélée de façon négative, mais de manière moins drastique, à la méthylation de l'ADN : c'est la tri-méthylation de H3K27 (Meissner et al., 2008). En effet, ces deux marques sont mutuellement exclusives au niveau des îlots CpG mais on peut les observer ensemble au niveau de régions pauvres en CpG (Brinkman et al., 2012). La marque H3K27me3 est catalysée par l'enzyme EZH2 (Enhancer of Zeste 2) du complexe PRC2 (polycomb repressive complex 2) et l'unité SUZ12 permet de cibler H3K27me3 dans le génome. Dans des cellules mutantes pour une ou l'autre des unités de PRC2, on observe une perte de H3K27me3 mais pas spécialement un gain de méthylation de l'ADN (Lindroth et al., 2008). Cette même étude suggère que la présence de H3K27me3 n'empêcherait pas la méthylation de l'ADN mais plutôt que la méthylation de l'ADN empêcherait la marque H3K27me3 de se répandre à des régions inappropriées (Lindroth et al., 2008). De plus, dans des cellules ES mutantes pour les 3 Dnmt (*Dnmt-TKO*), en absence de la méthylation de l'ADN, la marque H3K27me3 n'est plus proprement ciblée dans le génome et une perte d'enrichissement à des endroits précis du génome résulte en une dé-repression des gènes normalement réprimés par PRC2 (Brinkman et al., 2012 ; Reddington et al., 2013). Jusque récemment, aucun mécanisme n'avait pu être mis en évidence pour expliquer cette relation négative entre les deux marques. Des éléments de réponse ont été apportés par une étude montrant que FBXL10, une déméthylase de H3K4 et H3K36, empêcherait la méthylation de l'ADN aux promoteurs des gènes réprimés par PRC2 et donc enrichis en H3K27me3 (Boulard et al., 2015).

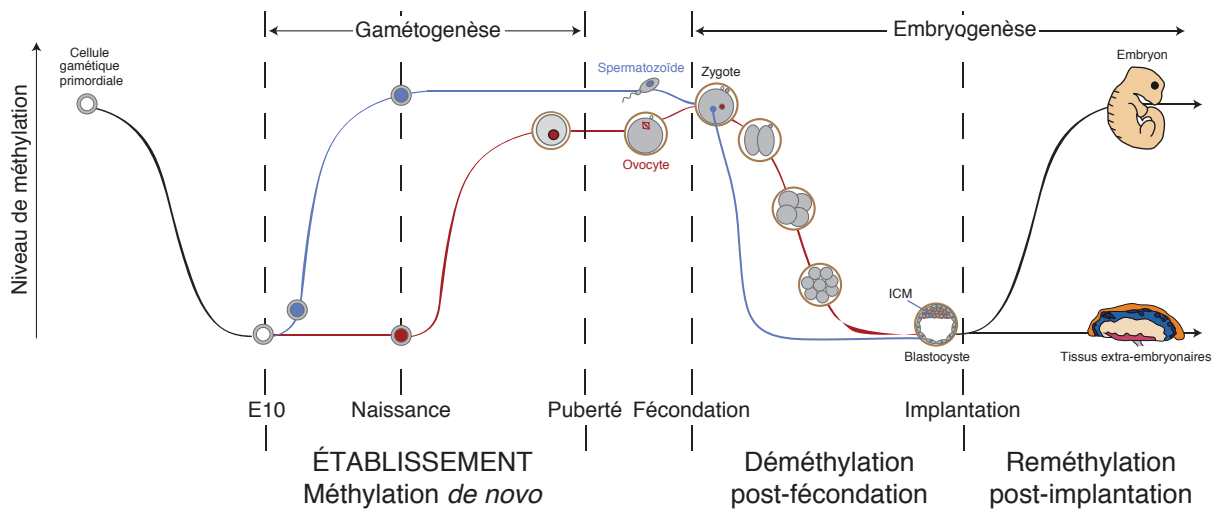


Figure 17 : Dynamique de la méthylation de l'ADN lors de la gamétogenèse et de l'embryogenèse. Les courbes rouges et bleues représentent respectivement les génomes maternel et paternel. L'établissement des marques de méthylation spécifiques du sexe se fait au cours de la gamétogenèse. Ces marques essentielles sont ensuite effacées après la fécondation et ré-établies lors de la reméthylation post-implantatoire que subit le génome de l'embryon précoce.

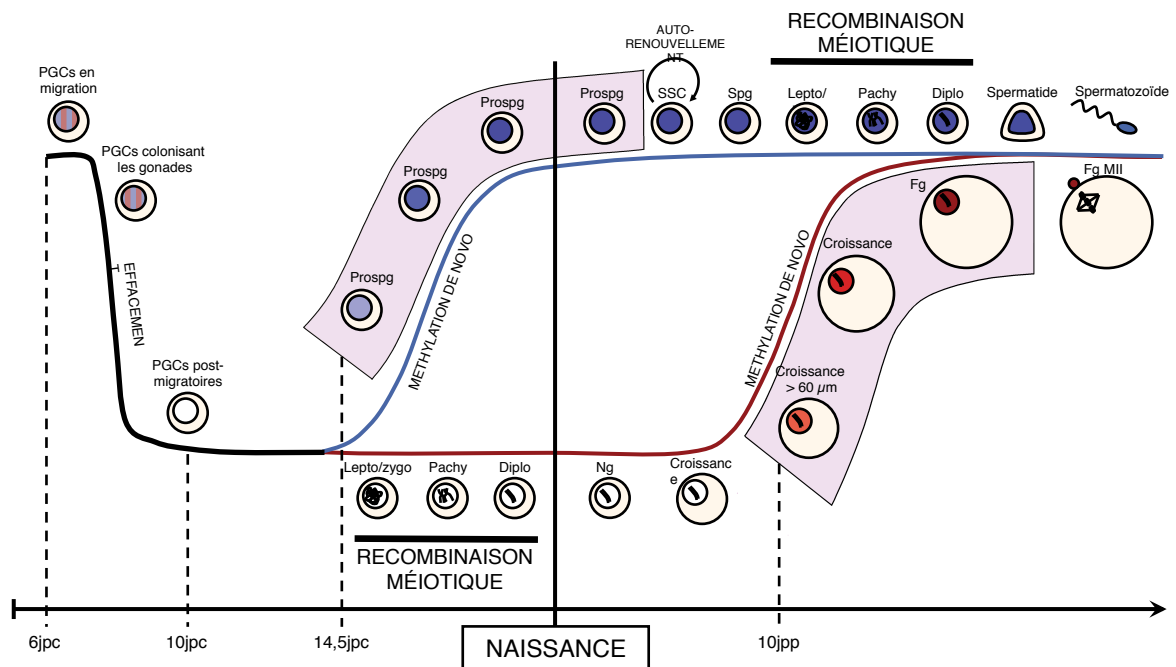


Figure 18 : Cinétique d'établissement des marques de méthylation lors de la gamétogenèse. L'expression de Dnmt3L est mise en évidence en violet (d'après Bourc'his and Proudhon, 2008).

1.4- La dynamique de méthylation du génome au cours du développement

Le génome des mammifères subit plusieurs vagues de déméthylation et re-méthylation *de novo* pendant le développement *in utero* (Reik et al., 2001). Tout d'abord, une première vague de déméthylation a lieu dans les cellules germinales primordiales (PGC) lors de leur migration vers les futures gonades autour de 10dpc chez la souris (Hajkova et al., 2002 ; Trasler, 2006). Ensuite, une vague de méthylation *de novo* permet la mise en place des marques de méthylation sexe-spécifiques dans les futurs gamètes mâles et femelles (Schaefer et al., 2007 ; Bourc'his and Proudhon, 2008). Puis, quelques heures après la fécondation, le génome paternel est déméthylé de façon active, alors que le génome maternel perd progressivement sa méthylation par un mécanisme passif couplé à la réplication de l'ADN jusqu'au stade blastocyste. Après l'implantation, les deux génomes parentaux sont de nouveau méthylés *de novo*. (Figure17&18).

1.4.1- Déméthylation du génome dans les PGC

La déméthylation du génome des PGC, qui a lieu de manière similaire entre les deux sexes, permet d'atteindre le niveau physiologique de méthylation connu le plus bas : 7% de CpG méthylés (Popp et al., 2010). Très peu de séquence échappent à ce processus, mais on peut citer comme exception les rétrotransposons de type IAP (Bourc'his and Bestor, 2004 ; Kato et al., 2007). Cependant, le mécanisme de déméthylation impliqué reste controversé : actif ou passif ou les deux (Popp et al., 2010 ; Kagiwada et al., 2013 ; Hackett et al., 2013 ; Ohno et al., 2013). En faveur de la déméthylation active, on peut avancer le temps réduit pendant lequel elle intervient et l'observation de cassures simple brin, qui pourraient signer l'activité de la voie de réparation de l'ADN suite à l'excision de base (BER : Base Excision Repair) lors du remplacement de cytosines méthylées par des cytosines non méthylées (Hajkova et al., 2010). Les protéines AID (Activation-Induced cytidine Deaminase) et TGD (Thymine DNA Glycosylase), qui sont impliquées dans le mécanisme d'excision de base, joueraient aussi un rôle dans cette déméthylation (Popp et al., 2010 ; Guibert and Weber, 2013). De plus, la conversion des méthylcytosines (5mC) en hydroxyméthylcytosines (5hmC) par les protéines TET (Ten Eleven Translocation) pourrait aussi entrer en jeu (Hackett et al., 2013 ; Yamaguchi et al., 2013 ; Guibert and Weber, 2013) (Figure19). Enfin, une déméthylation incomplète de l'ADN est observée dans les PGC mutantes pour les

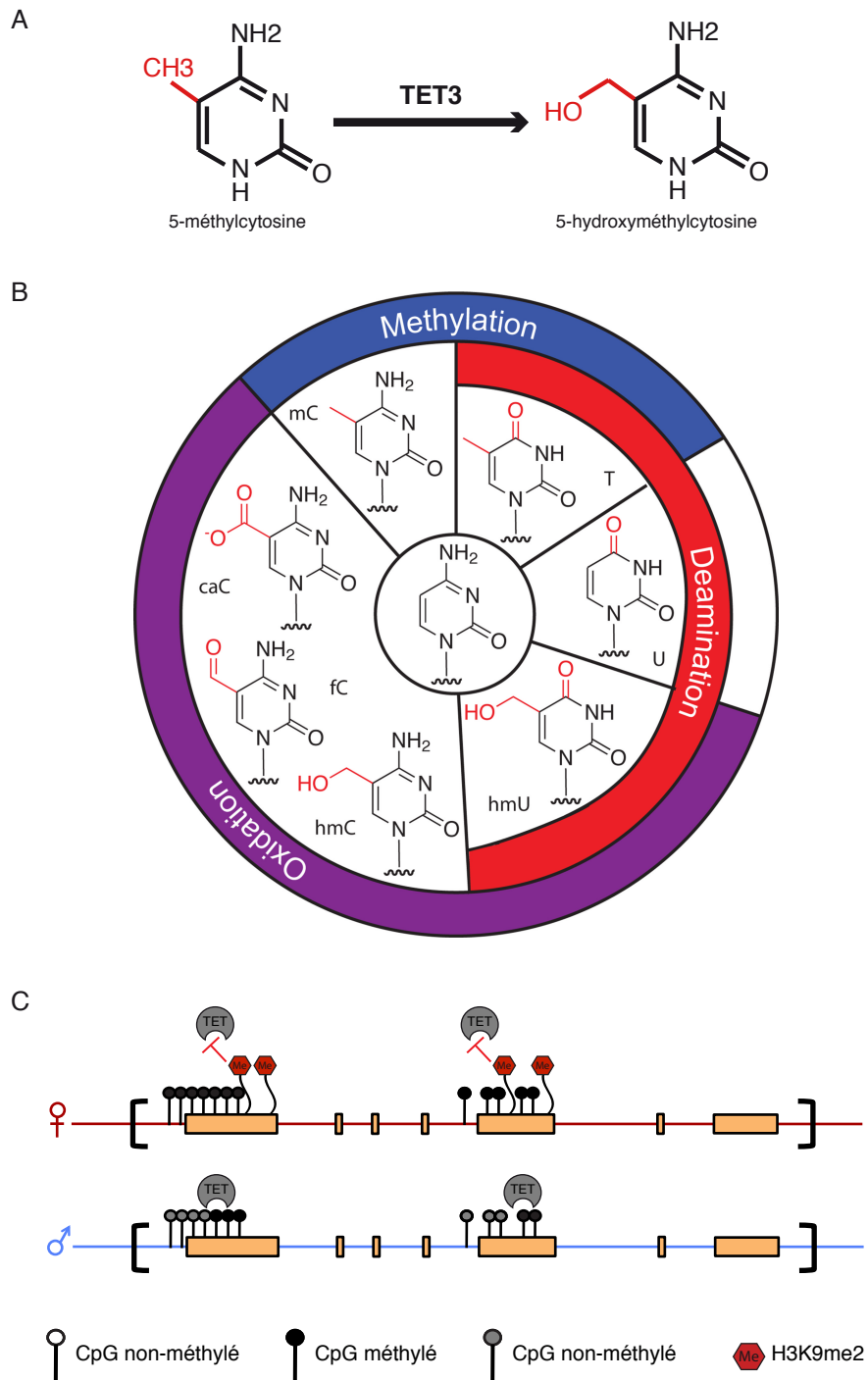


Figure 19 : La carte des cytosines du génome.

A. Réaction d'oxydation de la 5-méthylcytosine en 5-hydroxyméthylcytosine par la protéine TET3.

B. Dans le génome, les cytosines peuvent être modifiées par des réactions de déamination, d'oxydation ou de déméthylation afin de générer une série d'analogues. Ces modifications des cytosines influencent l'expression des gènes et l'identité cellulaire. Parmi ces analogues, les modifications enzymatiques engendrent la 5-méthylcytosine (5mC), la 5-hydroxyméthylcytosine (5hmC), la 5-formylcytosine (5-fC), la 5-carboxycytosine (5caC), la 5-hydroxyméthyluracil (5hmU) ainsi que l'uracil (U) et la thymine (T). (d'après Nabel et al., 2012).

C. Mécanisme de protection de la méthylation de l'ADN par la protéine PGC7 dans le génome maternel contre l'hydroxyméthylation.

protéines TET, AID ou TGD (Hajkova et al., 2010 ; Popp et al., 2010 ; Cortellino et al., 2011 ; Hackett et al., 2013).

1.4.2- Mise en place de la méthylation dans les gamètes

La méthylation *de novo* germinale est principalement dépendante de Dnmt3A stimulée par Dnmt3L dans les deux sexes (Bourc'his et al., 2001 ; Bourc'his and Bestor, 2004 ; Kaneda et al., 2004 ; La Salle et al., 2004 ; Kato et al., 2007). Cependant, il existe un dimorphisme sexuel important dans le niveau de méthylation, dans les séquences ciblées par cette méthylation et dans la chronologie d'acquisition (Schaefer et al., 2007 ; Bourc'his and Proudhon, 2008) (Figure 18). Plusieurs études ont montré que le génome du spermatozoïde était beaucoup plus méthylé que celui de l'ovocyte (80% versus 54%) (Monk et al., 1987 ; Smallwood et al., 2011 ; Kobayashi et al., 2012 ; Smith et al., 2012). Dans le sperme, la méthylation de l'ADN est enrichie au niveau des régions inter-géniques et des séquences répétées, notamment les rétrotransposons de type IAP (Intracisternal A Particle) et LINE-1 (Long Interspersed Elements) (Bourc'his and Bestor, 2004 ; Kato et al., 2007). A l'inverse dans l'ovocyte, les régions intra-géniques, y compris les CGI associés à des promoteurs intra-géniques, sont les cibles de la méthylation de l'ADN (Smallwood et al., 2011 ; Kobayashi et al., 2012). Cette différence de cible permet d'expliquer la plus forte méthylation du sperme par rapport à l'ovocyte, les rétrotransposons comptant pour la moitié du génome des mammifères (Rollins et al., 2006).

Dans la gamétogenèse mâle, la méthylation de l'ADN est mise en place avant la naissance. Juste après la vague de déméthylation, les cellules germinales mâles entrent en quiescence et s'arrêtent en phase G0/G1 de la mitose au stade prospermatogonie autour de 14dpc. La méthylation *de novo* de l'ADN commence à ce stade de façon simultanée dans toutes les cellules et se finit quelques jours après la naissance lorsque ces cellules deviennent des cellules souches spermatogoniales (Li et al., 2004 ; La Salle and Trasler, 2006). Plusieurs mécanismes sont proposés pour expliquer la spécificité de la méthylation germinale mâle. La présence de domaines riches en H3K4me3 permettrait l'exclusion de la méthylation spécifiquement au niveau de certaines séquences, telles que les régions soumises à empreinte génomique maternelle (Henckel et al., 2012). Au contraire la méthylation pourrait être activement guidée sur les

rétrotransposons par des petits ARN régulateurs particuliers, les piRNA (piwi-interacting RNAs). Les rétrotransposons, transcrits dans les PGC lorsque la méthylation de l'ADN a été effacée, serviraient de substrats pour produire ces piRNA via l'activité des protéines PIWI et guideraient ensuite la méthylation de l'ADN, probablement indirectement, sur les promoteurs des rétrotransposons par homologie de séquence (Aravin et al., 2008 ; Aravin and Bourc'h, 2008).

Dans la gamétogenèse femelle, l'acquisition de la méthylation se fait après la naissance (Lucifero et al., 2002 ; Lucifero et al., 2004). Après avoir subi la déméthylation de leur génome, les PGC femelles entrent en méiose mais s'arrêtent avant la naissance au stade diplotène de la prophase de méiose I. La méthylation *de novo* est alors établie lorsque les ovocytes entrent dans leur phase de croissance juste avant l'ovulation, à partir de la puberté. Les ovocytes méthylés sont ensuite de nouveau bloqués en métaphase de méiose II jusqu'à la fécondation (Smallwood et al., 2011 ; Shirane et al., 2013). Du fait de la localisation intra-génique préférentielle de la méthylation ovocytaire, cette méthylation pourrait être guidée par la transcription dans le corps des gènes via la marque H3K36me3 (Smallwood et al., 2011). La déméthylation de H3K4me3 par l'enzyme KDM1B au niveau de certaines régions spécifiques pourrait aussi guider la méthylation (Ciccone et al., 2009). Enfin, Dnmt1 pourrait aussi jouer un rôle dans la mise en place de cette méthylation *de novo* en « rebouchant les trous », sur des sites hémi-méthylés laissés par Dnmt3A (Shirane et al., 2013).

1.4.3- Déméthylation de la méthylation dans l'embryon après la fécondation

Au moment de la fécondation, les deux génomes parentaux arrivent avec leurs marques de méthylation différentielles héritées de leur passage par la gamétogénèse mâle ou femelle. Juste après la fécondation, une vague de déméthylation drastique du génome a lieu et la majorité des marques établies pendant la gamétogénèse sont perdues pendant le développement pré-implantatoire de l'embryon. Ainsi, dans le blastocyste, le niveau estimé de méthylation est de 20% (Kobayashi et al., 2012, Smith et al., 2012, Smallwood et al., 2011). En comparaison avec les PGC, cette déméthylation post-fécondation n'est donc pas aussi extrême et on connaît des exceptions notoires à la déméthylation embryonnaire : les régions soumises à empreinte génomique parentale et

les rétrotransposons de la famille des IAP (Bourc'his and Bestor, 2004 ; Kato et al., 2007).

Les deux génomes parentaux ne suivent pas la même cinétique de déméthylation. La méthylation du pronucleus paternel est effacée en quelques heures après la fécondation (Oswald et al., 2000 ; Santos et al., 2002) (Figure17). La réduction du niveau de 5mC s'accompagne d'un enrichissement en 5hmC, signe d'un mécanisme de déméthylation actif catalysé par les enzymes TET et notamment l'enzyme TET3 (Iqbal et al., 2011 ; Wossidlo et al., 2011 ; Gu et al., 2011) (Figure19A). L'oxydation des 5mC en 5hmC permettrait ensuite soit une déméthylation passive par perte des 5hmC au cours de la réplication, soit une déméthylation active par excision des 5hmC (Guibert and Weber, 2013). Cependant, l'observation de cassures doubles brins dans le pronucleus mâle laisse penser qu'un mécanisme actif, via la voie de réparation suite à l'excision de base (BER), est impliqué ici aussi, comme dans les PGC (Wossidlo et al., 2010). De plus, les 5hmC peuvent être encore oxydées en 5fC (5-formylcytosine) puis en 5caC (5-carboxylcytosine) par les enzymes TET, et la glycosylase TDG (Thymine DNA Glycosylase), impliquée dans les mécanismes du BER été reconnue pour se lier et exciser les 5caC dans les cellules ES (He et al., 2011) (Figure19B). Des études ont aussi permis de montrer que ces trois marques étaient présentes dans l'embryon pré-implantatoire et que leur niveau était réduit au cours de la réplication de l'ADN, argument en faveur d'un mécanisme de réparation suite à l'excision de base (Inoue et al., 2011 ; Inoue and Zhang, 2011).

Les marques de méthylation du pronucleus maternel sont quant à elles maintenues dans les heures qui suivent la fécondation (Rougier et al., 1998 ; Santos et al., 2002) (Figure17). La méthylation maternelle est protégée pendant ces quelques heures par des facteurs qui se lient aux modifications d'histones présentes sur le génome maternel méthylé, notamment les résidus H3K9me2/me3 (Arney et al., 2002 ; Santos et al., 2005). La protéine PGC7/STELLA se lie aux résidus H3K9me2 et ainsi protège les 5mC présentes sur le génome maternel contre l'action de conversion en 5hmC par la protéine TET3 (Nakamura et al., 2012 ; Szabo and Pfeifer, 2012) (Figure19C). A partir du stade 2-cellules, on observe cependant une perte graduelle du niveau de méthylation par déméthylation passive du génome maternel (Rougier et al., 1998 ; Santos et al., 2002 ; Smith et al., 2012). Cette perte passive de 5mC reposerait sur la séquestration de Dnmt1 dans le cytoplasme; l'absence de Dnmt1 au niveau des

fourches de réplication entraînerait cette perte progressive et passive des marques de méthylation (Mertineit et al., 1998 ; Howell et al., 2001).

1.4.4- Etablissement de la méthylation dans l'embryon après l'implantation

Les deux génomes parentaux sont de nouveau méthylés *de novo* autour de l'implantation qui survient environ au 4.5dpc chez la souris ; le génome de l'embryon à 9.5dpc est globalement méthylé, avec un niveau similaire de 60/80% trouvé dans les cellules somatiques (Smith et al., 2009 ; Borgel et al., 2010 ; Auclair et al., 2014) (Figure17). La méthylation de l'ADN n'a lieu que dans le génome de l'épiblaste, les tissus extra-embryonnaires (futur placenta) restent quant à eux globalement hypométhylés (Chapman et al., 1984 ; Monk et al., 1987). La mise en place de cette méthylation post-implantatoire dépend de Dnmt3A et Dnmt3B, avec cependant une contribution majeure de Dnmt3B (Auclair et al., 2014).

Deux hypothèses ont été proposées pour tenter d'expliquer les mécanismes de ciblage de la méthylation embryonnaire. Une des hypothèses propose un ciblage actif des Dnmt *de novo* sur des séquences spécifiques. Les différents phénotypes, et les différences de séquences présentant un défaut de méthylation, observés chez les souris mutantes pour chacune des Dnmt *de novo* soutiennent cette hypothèse (Okano et al., 1999). Par exemple, les embryons déficients pour Dnmt3B présentent un défaut de méthylation au niveau des séquences d'ADN satellite mineur et des gènes de la lignée germinale spécifiquement (Okano et al., 1999 ; Borgel et al., 2010). L'autre hypothèse, à l'inverse, est que le génome de l'embryon post-implantatoire serait méthylé par défaut et que toutes les cytosines seraient donc méthylées sans discrimination de séquences, ce qui semble être soutenu par le fort taux de méthylation des CpG (60/80%). Seules certaines séquences spécifiques du génome, comme les régions soumises à empreinte génomique, échapperaient à cette méthylation par des mécanismes de protection (voir Introduction section 2.3.2).

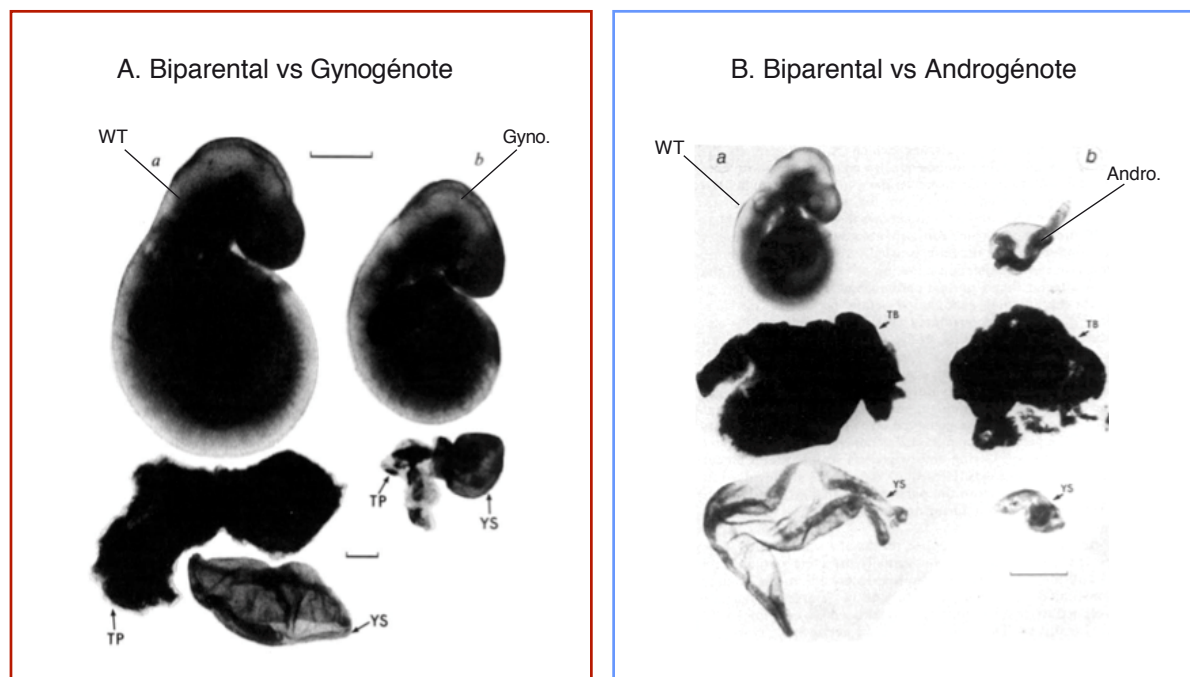


Figure 20 : Non équivalence fonctionnelle des génomes parentaux.

Cette propriété a été mise en évidence, en 1984, par la production d'embryons diploïdes uniparentaux par transfert de noyaux gamétiques. Les gynogénètes (Gyno.) sont issus de la réunion de deux génomes maternels et les androgénètes (Andro.) de deux génomes paternels. Les deux types d'embryons présentent une létalité embryonnaire à 9,5jpc associée à des phénotypes en miroir. Les gynogénètes développent des tissus embryonnaires et très peu d'annexes embryonnaires alors que les androgénètes présentent une hyperprolifération extra-embryonnaire et pratiquement pas de tissus embryonnaires. (d'après Barton et al., 1984 et Surani et al., 1984).

WT : embryons sauvages, TP : trophoblaste, YS : sac vitellin.

2- L’empreinte génomique parentale

La partie introductive précédente révèle que la méthylation de l’ADN est une marque abondante dans le génome et qu’elle subit des remaniements importants au cours du développement. En association avec des modifications d’histones et des facteurs de transcription, elle permet de contrôler l’expression du génome. Il existe une catégorie bien particulière de gènes sous dépendance étroite de la méthylation de l’ADN : ce sont les gènes soumis à empreinte génomique (GSE).

L’empreinte parentale ou empreinte génomique est un processus épigénétique qui fait référence à une catégorie de gènes chez les mammifères placentaires dont la mémoire parentale est maintenue tout au long de la vie. Mon projet de thèse a visé à identifier de nouveaux gènes soumis à empreinte génomique. Afin de bien comprendre quels ont été nos critères de recherche pour trouver de nouveaux GSE, je décris dans cette partie les caractéristiques génétiques, épigénétiques et fonctionnelles de ces gènes et de leurs régions de contrôle, les ICR (Imprinting Control Regions).

2.1- Découverte, définition et rôles de l’empreinte génomique parentale

2.1.1- Découverte de l’empreinte génomique parentale

Les premières évidences de l’existence de l’empreinte génomique parentale viennent d’expériences montrant la nécessité de fournir un génome maternel et un génome paternel pour le développement de l’embryon chez la souris (Barton et al., 1984 ; Surani et al., 1984 ; McGrath and Solter, 1984). Cette non-équivalence des génomes parentaux a été démontrée par des expériences de transfert nucléaire chez la souris : deux pronuclei paternels ou deux pronuclei maternels transférés dans un ovocyte fécondé énucléé produisent des embryons diploïdes uniparentaux qui meurent à mi-gestation, autour de 9dpc (Figure20). Les gynonogénotes, issus de la réunion de deux génomes maternels, développent des tissus embryonnaires mais très peu d’annexes embryonnaires. Les androgénotes, composés de deux génomes paternels, développent à l’inverse très peu de tissus embryonnaires et présentent une hyperprolifération extra-embryonnaire. Ces phénotypes en miroir illustrent le fait que chaque génome parental a une identité fonctionnelle propre et qu’une empreinte doit

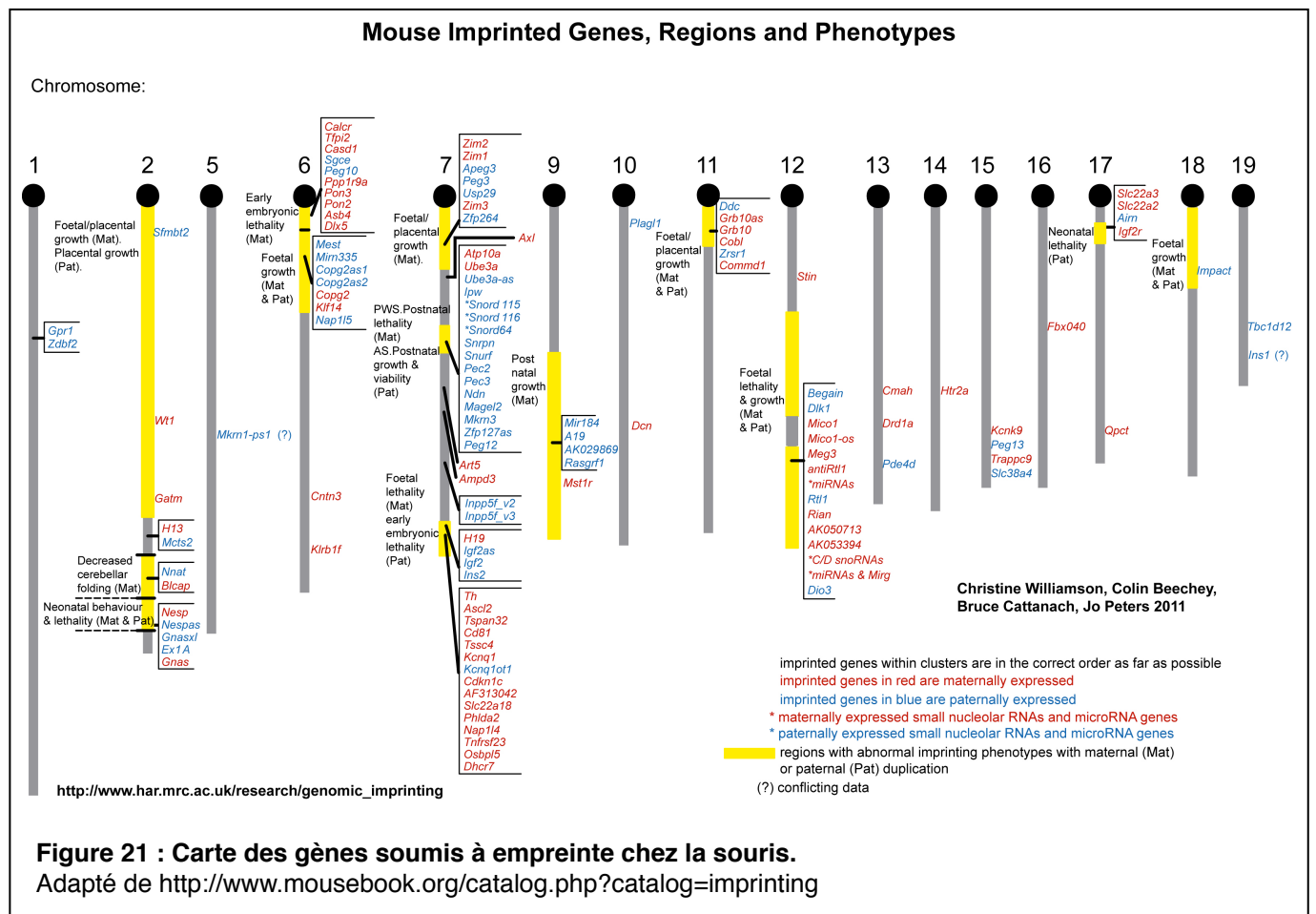


Figure 21 : Carte des gènes soumis à empreinte chez la souris.
 Adapté de <http://www.mousebook.org/catalog.php?catalog=imprinting>

marquer chacun d'entre eux. Cette empreinte est de plus de nature non-génétique, puisque les souris utilisées pour ces expériences étaient parfaitement consanguines (« inbred »).

Une autre étude a confirmé l'impact de l'origine parentale de chaque génome sur le développement de l'embryon (Cattanach and Kirk, 1985). Les produits issus de disomies uniparentales (UPD) - où deux copies d'un chromosome ou d'un bout de chromosome sont héritées seulement de la mère ou du père - ont démontré chez la souris que certaines régions du génome ne fonctionnaient pas de la même manière en fonction de leur origine parentale. C'est alors qu'est née l'hypothèse que certains loci spécialisés pourraient être responsables de phénotypes dits à effets parentaux. Plusieurs de ces phénotypes sont associés à des retards ou excès de croissance, aussi bien de l'embryon que du placenta. Différents symptômes liés à des unidisomies parentales ont aussi été observés chez l'humain (Horsthemke and Buiting, 2008).

Le premier locus mis en évidence comme étant soumis à empreinte est le locus *Igf2/H19*, qui présente deux gènes ayant une expression mono-parentale à partir de l'allèle maternel, *Igf2r* (*Insulin-like Growth Factor 2 Receptor*) et *H19*, et un gène ayant une expression mono-allélique paternelle, *Igf2* (*Insulin-like Growth Factor 2*) (Barlow et al., 1991 ; DeChiara et al., 1991 ; Bartolomei et al., 1991). Plusieurs groupes ont essayé de comprendre par quel mécanisme cette expression mono-allélique était contrôlée et la méthylation de l'ADN a été proposée comme étant la marque responsable de ces effets parentaux (Surani et al., 1984 ; Reik et al., 1987 ; Swain et al., 1987). D'autres études ont confirmé cette hypothèse et ont permis d'identifier les régions différenciellement méthylées (DMR) associées aux trois premiers gènes soumis à empreinte (Bartolomei et al., 1993 ; Ferguson-Smith et al., 1993 ; Stoger et al., 1993). Il a ensuite été montré que la méthylation de l'ADN était indispensable au maintien de l'expression mono-allélique et que la perte de cette marque entraînait une expression ou une répression biallélique des GSE (Brandeis et al., 1993 ; Li et al., 1993).

2.1.2- Le point sur l'empreinte génomique parentale aujourd'hui : définitions et faits

De nos jours, par le biais de découvertes fortuites ou de criblages dirigés, plus d'une centaine de gènes ont été répertoriés comme étant soumis à empreinte chez la souris et l'humain (Williamson et al., 2011) (Figure 21). Les GSE sont souvent organisés

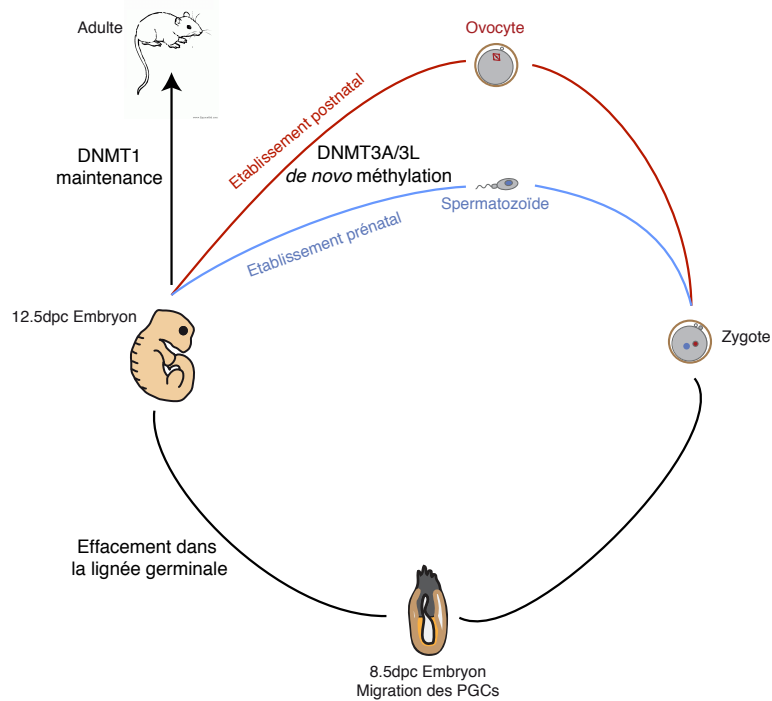


Figure 22 : Cycle de l’empreinte.

La méthylation des ICRs est héritée des gamètes, mise en place d’une empreinte maternelle dans l’ovocyte et mise en place d’une empreinte paternelle dans les spermatozoïdes. Ce statut de méthylation différentielle est maintenu tout au long de la vie malgré les remaniements drastiques de méthylation qui caractérisent le génome de l’embryon précoce.

en cluster autour de régions de contrôle de l’empreinte appelées ICR (Imprinting Control Regions) ou gDMR (germline Differentially Methylated Regions) qui contrôlent leur expression en *cis*. Ces régions sont marquées différemment par la méthylation de l’ADN au cours de la gamétogenèse lorsque les génomes sont physiquement séparés : les allèles paternels sont marqués dans le sperme et les allèles maternels dans l’ovocyte (Surani et al., 1984). Chez la souris, 24 gDMR ont été identifiées présentant un fort dimorphisme sexuel dans leur répartition puisque 21 d’entre elles sont méthylées dans l’ovocyte (ICR maternelle) et seulement 3 dans le sperme (ICR paternelle) (Reik and Walter, 2001 ; Schulz et al., 2010 ; Proudhon et al., 2012 ; Duffié et al., 2014). Un locus est dit soumis à empreinte maternelle ou paternelle en fonction de la provenance de la méthylation de son ICR, mais les GSE qu’elle contrôle peuvent aussi bien être exprimés à partir de l’allèle maternel que de l’allèle paternel.

Les ICR suivent une dynamique de méthylation au cours du développement, c’est ce que l’on appelle le cycle de l’empreinte (Figure22). La méthylation des ICR est établie dans les gamètes mâle et femelle avant la fécondation, en même temps que la méthylation du reste du génome. Les allèles méthylés des ICR sont protégés de la vague de déméthylation globale du génome pendant le développement pré-implantatoire. De la même manière, les allèles déméthylés des ICR sont aussi protégés de la vague de re-méthylation *de novo* du génome après l’implantation (voir Introduction section 2.3). Dans les PGC, les marques de méthylation sont effacées puis ré-établies dans les gamètes en fonction du sexe de l’individu, à savoir une empreinte maternelle dans les ovocytes des individus femelles et une empreinte paternelle dans les spermatozoïdes des individus mâles.

2.1.3- Rôles des gènes soumis à empreinte génomique

Afin de comprendre le rôle de l’empreinte génomique parentale chez les mammifères, l’étude de la fonction des GSE connus semble une bonne piste. Cependant, ils sont impliqués dans des catégories ontologiques moléculaires très variées, allant du facteur de croissance (*Igf2*) au suppresseur de croissance neuronale (*Ndn*) ou encore au transporteur de solutés (famille des *Slc*) en passant par des fonctions pro-apoptotiques (*Zac1*, *Cdkn1c*). En revanche, leur profil d’expression tissulaire est informatif, dans le sens où les GSE sont surtout exprimés en vie *in utero* (embryon et annexes

Gène à expression maternelle	FONCTION	Gène à expression paternelle
- Igf2r	Défauts de croissance embryonnaire, placentaire et/ou postnatale	+ Igf2
- Gnas		+ GnasXL
- Tssc3/lpl		+ Peg1/MEST
- Mash2		+ Peg3/Pw1
- Grb10/meg1		+ Rasgrf1
-/+ Cdkn1c		+ Dlk1
Nesp	Défauts de comportement ou neurologiques	+ Peg1/Mest
Ube3a		+ Peg3/Pw1
Kcnq1		+ Rasgrf1
H19	Pas de phenotype évident	Snrpn/Snurf
Slc22a2		Peg12/Frat3
Slc22a3		Ins2

Tableau 2 : La fonction des GSE déterminée par invalidation génétique.
(d'après d'Epigenetics © 2006 CSHL press)

embryonnaires) et dans les tissus neuraux après la naissance (Babak et al., 2015). De même, les phénotypes observés suite à l'invalidation de plusieurs d'entre eux montrent un rôle prédominant de l'empreinte au cours du développement embryonnaire, mais aussi sur la croissance et le comportement des individus après la naissance (Tableau2). L'empreinte génomique parentale servirait alors à modérer la croissance pendant le développement et les échanges entre la mère et le fœtus via le placenta, mais aussi d'un point de vue cognitif après la naissance (Smith et al., 2006 ; Isles et al., 2006 ; Barlow and Bartolomei, 2007 ; Proudhon and Bourc'his, 2010a).

Une grande majorité des GSE joue un rôle sur la croissance embryonnaire. Une partie d'entre eux est exprimée à partir de l'allèle paternel et tendent à promouvoir la croissance alors que l'autre partie, exprimée à partir de l'allèle maternel, tend à la réprimer. C'est de là qu'a émergé la théorie du conflit des génomes parentaux qui propose que le père désire une descendance et donc un fœtus qui ingère beaucoup de ressources afin d'être plus fort que ses frères qui pourraient venir d'un autre père et une mère qui essaye d'empêcher que le fœtus ne lui prenne trop de ressources afin qu'il en reste pour ses frères et elle même. Cette théorie est parfaitement illustrée par le gène *Igf2r* exprimé à partir de l'allèle maternel et qui interagit de façon négative avec le gène *Igf2* à expression paternelle afin de limiter la croissance de l'embryon (Haig and Graham., 1991).

Une autre partie des GSE est spécifiquement exprimée dans le cerveau après la naissance et agissent dans des processus neurologiques (Barlow and Bartolomei, 2014). Ces gènes réguleraient les interactions entre la mère et ses petits, notamment les comportements importants pour l'allaitement et l'attachement. Un des exemples les plus connus est celui du gène *Peg3* (*Paternally Expressed Gene 3*), qui est exprimé à partir de l'allèle paternel dans l'hypothalamus des petits et de la mère, et qui permettrait d'apprendre à téter aux nouveau-nés et à la mère d'accepter la tétée (Curley et al., 2004). Un autre exemple plus récent est celui du gène *Grb10* exprimé à partir de l'allèle maternel et qui contrôle l'afflux de nutriment dans la glande mammaire chez la mère et régule en même temps la demande pour ces nutriments chez les petits (Cowley et al., 2014).

Imprinted Locus	Chrom	Localisation of Germline DMR	Parental origin of methylation	ICR ?	Associated imprinted genes and non-coding RNA	References (germline DMR/ICR)
<i>Zdbf2</i>	1	10 kb upstream <i>Zdbf2</i> Exon1	Paternal	ND		Kobayashi <i>et al.</i> (2009)
<i>Mcts2</i>	2	<i>Mcts2</i> Promoter/Exon1	Maternal	ND	<i>H13</i>	Wood <i>et al.</i> (2007)
<i>Gnas</i>	2	<i>GnasXL/Nespas</i> promoter and <i>Gnas</i> 1A promoter (domain containing two germline DMRs)	Maternal	ICR (<i>GnasXL/NespasDMR</i>)	<i>Gnasxl</i> ; <i>Nesp</i> ; <i>Nespas</i> ; <i>Gnas exon1A</i>	Coombes <i>et al.</i> (2003), Liu <i>et al.</i> (2000), and Williamson <i>et al.</i> (2004, 2006)
<i>Peg10</i>	6	<i>Peg10/Sgce</i> promoter	Maternal	ND	<i>Sgce</i> ; <i>Ppp1r9a</i> ; <i>Pon1</i> ; <i>Pon2</i> ; <i>Pon3</i> ; <i>Asb4</i> ; <i>Calcr</i> ; <i>Tpfl2</i>	Ono <i>et al.</i> (2003)
<i>Mest</i> (<i>Peg1</i>)	6	<i>Mest</i> promoter/exon1	Maternal	ND	<i>Klf14</i> ; <i>Copg2</i> ; <i>Copg2as</i> ; <i>Mim335</i>	Lucifero <i>et al.</i> (2004)
<i>Nap1L5</i>	6	<i>Nap1L5</i> promoter/exon1	Maternal	ND		Woods <i>et al.</i> (2007)
<i>Peg3</i>	7	<i>Peg3</i> promoter/exon1	Maternal	ND	<i>Zim1</i> ; <i>Zim2</i> ; <i>Zim3</i> ; <i>Usp29</i> ; <i>Zfp264</i>	Kim <i>et al.</i> (2003)
<i>Snrpn</i>	7	<i>Snrpn</i> promoter/exon1	Maternal	ICR	<i>Atp10a</i> ; <i>Ube3a</i> ; <i>Snurf</i> ; <i>Ndn</i> ; <i>Magel2</i> ; <i>Mkri3</i> ; <i>Pec2</i> ; <i>Pec3</i> ; <i>Peg12</i> ; <i>Ube3a-as</i> ; <i>Zfp127as</i> ; <i>SnoRNas</i>	Shemer <i>et al.</i> (1997) and Bielinska <i>et al.</i> (2000)
<i>Inpp5f</i>	7	<i>Inpp5f-V2</i> promoter/exon1	Maternal	ND	<i>Inpp5f-V2</i> ; <i>Inpp5f-V3</i>	Wood <i>et al.</i> (2007)
<i>Igf2</i>	7	2 kb upstream of <i>H19</i> (<i>H19DMR</i>)	Paternal	ICR	<i>H19</i> ; <i>Ins2</i>	Tremblay <i>et al.</i> (1997) and Thorvaldsen <i>et al.</i> (1998)
<i>Kcnq1</i>	7	<i>Kcnq1-ot 1</i> promoter (<i>Kvdmr1</i>)	Maternal	ICR	<i>Th</i> ; <i>Ascl2</i> ; <i>Tspan32</i> ; <i>Cd81</i> ; <i>Tssc4</i> ; <i>Kcnq1-ot1</i> ; <i>Cdkn1c</i> ; <i>Slc22a18</i> ; <i>Phlda2</i> ; <i>Nap114</i> ; <i>Osbpl5</i>	Fitzpatrick <i>et al.</i> (2002) and Yatsuki <i>et al.</i> (2002)
<i>Rasgrf1</i>	9	30 kb upstream of <i>Rasgrf1</i> exon1	Paternal	ICR	<i>A19</i> ; <i>Mir184</i>	Shibata <i>et al.</i> (1998) and Yoon <i>et al.</i> (2002)
<i>Plagl1</i> (<i>Zac1</i>)	10	<i>Plagl1</i> promoter/exon1	Maternal	ND		Smith <i>et al.</i> (2002)
<i>Grb10</i>	11	<i>Grb10</i> CpG island2 - Brain-specific promoter	Maternal	ICR	<i>Ddc</i> ; <i>Cobl</i> ; <i>Grb10as</i>	Arnaud <i>et al.</i> (2003) and Shiura <i>et al.</i> (2009)
<i>U2af1-rs1</i> (<i>Zrsr1</i>)	11	<i>U2af1-rs1</i> promoter/exon1	Maternal	ND	<i>Commf1</i>	Wood <i>et al.</i> (2007)
<i>Dlk1</i>	12	13 kb upstream of <i>Gtl2</i> (IG-DMR)	Paternal	ICR	<i>Begain</i> ; <i>Dio3</i> ; <i>Rtl1</i> ; <i>Gtl2</i> ; <i>Rian</i> ; <i>C/D SnoRNAs</i> ; <i>Mirg</i> ; <i>miRNAs</i>	Lin <i>et al.</i> (2003) and Takada <i>et al.</i> (2002)
<i>KcnK9</i>	15	<i>Peg13</i> promoter	Maternal	ND	<i>Peg13</i>	Ruf <i>et al.</i> (2007)
<i>Slc38aA</i>	15	<i>Slc38a4</i> promoter	Maternal	ND		Chotalia <i>et al.</i> (2009)
<i>Igf2r</i>	17	<i>Airn</i> promoter	Maternal	ICR	<i>Airn</i> ; <i>Slc22a2</i> ; <i>Slc22a3</i>	Stoger <i>et al.</i> (1993) and Wutz <i>et al.</i> (2001)
<i>Impact</i>	18	<i>Impact</i> promoter/exon1	Maternal	ND		Okamura <i>et al.</i> (2000)

Tableau 3 : Liste des gDMRs identifiées dans le génome de la souris.
Les loci soumis à empreinte tirent leur nom du gène codant le plus proche de la gDMR.
(d'après Arnaud, 2010)

2.2- Les régions de contrôle de l'empreinte génomique

Les régions de contrôle de l'empreinte sont appelées ICR (Imprinting Control Regions), ou gDMR (germline Differentially Methylated Regions) puisque la méthylation des ces séquences est mise en place dans la lignée germinale. Pour qu'une gDMR soit définie comme une ICR, il faut que la délétion génétique de cette région affecte les profils d'expression des gènes présents à ce locus ; plusieurs ICR connues aujourd'hui ont ainsi été validées par des approches fonctionnelles (Tableau3). Il existe aussi des DMR secondaires ou somatiques, les sDMR (somatic Differentially Methylated Regions), dont la méthylation est mise en place au cours du développement post-fécondation, souvent en conséquence de la transcription des GSE au sein du locus. Des profils de méthylation différentiels sur chaque allèle parental ne sont pas la seule caractéristique des ICR. Comme détaillé ci-dessous, d'autres facteurs discriminants sont associés aux ICR, comme la nature de leur séquence nucléotidique et les modifications d'histones associées.

2.2.1- Caractéristiques génétiques et épigénétiques générales des ICR

Une caractéristique des ICR est leur densité en CpG, au moins pour les ICR maternelles (voir plus bas) (Figure23A). Mais cela ne suffit pas à les distinguer du reste du génome puisque d'autres régions possèdent la même densité en CpG que les ICR (Schulz et al., 2010). L'espacement entre les CpG a été proposé comme une signature possible des ICR. Des modèles moléculaires du complexe Dnmt3A-Dnmt3L montrent que la distance optimale pour la mise en place de la méthylation serait de 8-10pb entre deux motifs CpG (Jia et al., 2007). Cette périodicité a été reportée pour 12 ICR maternelles ; cependant elle n'a pas été retrouvée dans toutes les ICR. De plus, elle ne semble pas spécifique aux ICR (Jia et al., 2007 ; Tomizawa et al., 2011).

Les ICR étant protégées des vagues de déméthylation et re-méthylation du génome embryonnaire, des motifs de liaison pour des facteurs impliqués dans leur protection pourraient les différencier du reste du génome. Le problème est qu'il y a sûrement plusieurs facteurs impliqués dans ces phénomènes de protection et donc les motifs de liaison seront sûrement différents d'une ICR à l'autre. Un motif de liaison commun à toutes les ICR a quand même été identifié, le motif hexaséquentiel TGCCGC de liaison à la protéine Zfp57 qui protège les allèles méthylés de la vague de

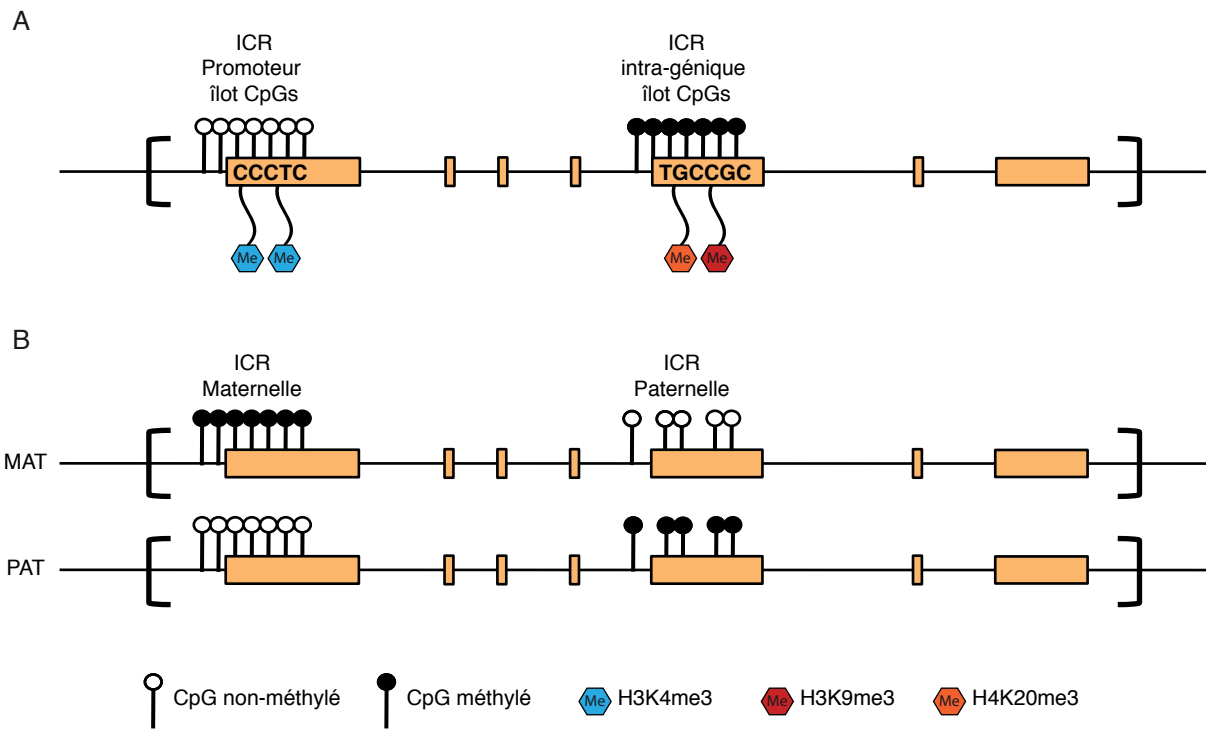


Figure 23 : Caractéristiques épigénétiques des ICRs.

A. Caractéristiques communes à toutes les ICRs.

B. Différences entre les ICRs maternelles et paternelles.

déméthylation du génome (Quenneville et al., 2011). Un autre motif CCCTC permettant la liaison de la protéine CTCF, qui protégerait les allèles déméthylés de la reméthylation, a aussi été identifié (Stamatoyannopoulos et al., 2012) (Figure23A). Malheureusement, ces deux motifs de liaison sont aussi retrouvés dans d'autres séquences du génome et ne permettent donc pas de distinguer exclusivement les ICR. La présence d'îlots CpG et des motifs de liaison à CTCF et Zfp57 sont donc des indices de la présence d'une ICR, mais ne suffisent pas à les identifier.

Une autre caractéristique des ICR est leur profil de modifications d'histones particulier : on y trouve en effet une combinaison de marques actives et de marques répressives, illustrant l'état différentiel des deux allèles parentaux (McEwen and Ferguson-Smith, 2010). Plusieurs études de loci soumis à empreinte ont montré que les allèles méthylés réprimés des ICR arboraient une méthylation des résidus H3K9 et H4K20, alors que les allèles non-méthylés actifs étaient enrichis pour les marques de méthylation en H3K4 et d'acétylation en H3K9 (Fournier et al., 2002 ; Li et al., 2004 ; Yang et al., 2003 ; Regha et al., 2007). D'une manière générale, les ICR peuvent être reconnues par l'enrichissement simultané en H3K4me3, H3K9me3 et H4K20me3 (Figure23A). Au niveau des promoteurs de certains GSE, une autre marque H3K27me3 a été observée (McEwen and Ferguson-Smith, 2010). Cette marque serait cependant plus liée à des modulations tissulaires ou développementales de l'expression des GSE plutôt qu'au statut empreinté proprement dit.

2.2.2– Les ICR maternelles

Les gDMR/ICR maternelles connues (20) ont une méthylation héritée de l'ovocyte. Ces gDMR sont très riches en CpG et elles co-localisent avec des « supers » îlots CpG (Schulz et al., 2010). Elles sont généralement associées aux promoteurs de gènes, codants pour une protéine ou produisant des ARN non-codants (Kobayashi et al., 2006 ; Schulz et al., 2010) (Figure23B). En plus de la méthylation de l'ADN et des modifications d'histones, d'autres mécanismes sont impliqués dans la régulation des loci soumis à empreinte maternelle. Mon projet de thèse visant à identifier de nouvelles gDMR/ICR maternelles, je vais décrire les mécanismes d'action présents au niveau de ces ICR maternelles.

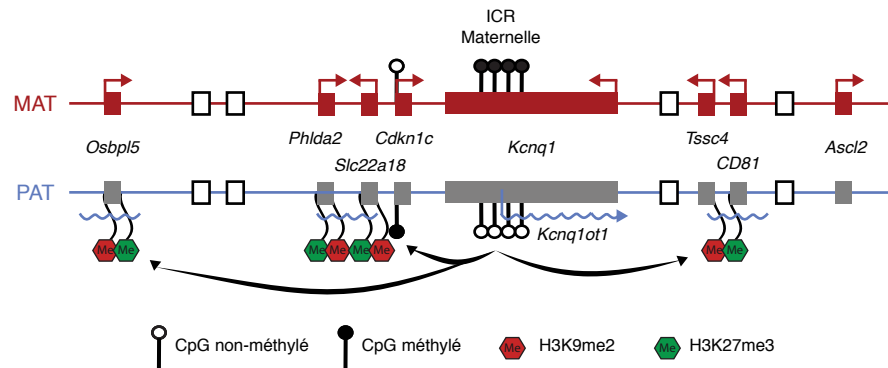


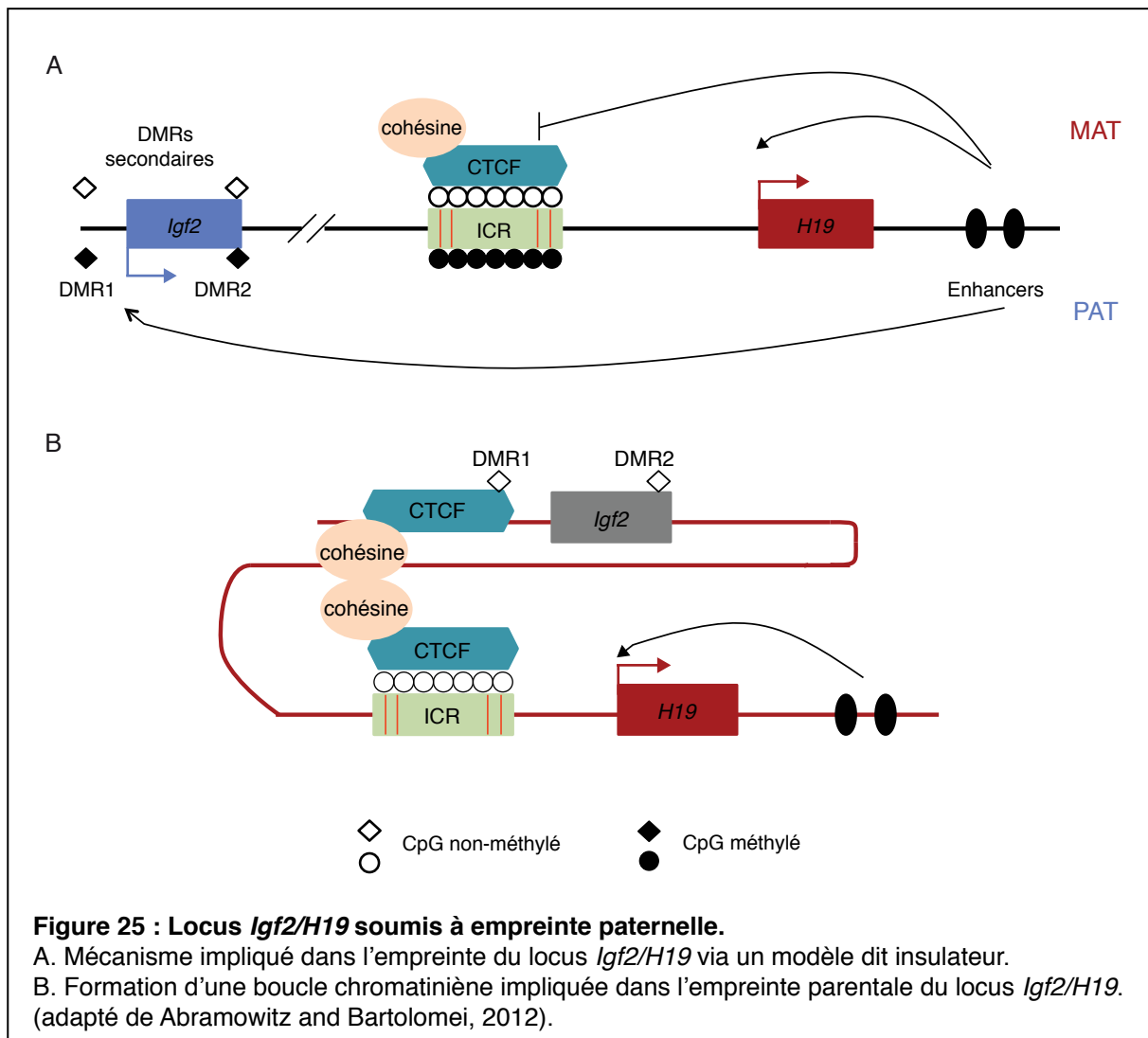
Figure 24 : Locus *Kcnq1* soumis à empreinte maternelle.

L'utilisation de la transcription d'un long ARN non-codant afin de réprimer les gènes situés sur le même allèle *en cis* fait partie de ces mécanismes. Un des locus soumis à empreinte maternelle utilisant ce mécanisme est celui du cluster de *Kcnq1* sur le chromosome 7 chez la souris. Ce cluster contient 8 gènes codant pour des protéines, exprimés à partir de l'allèle maternel dont *Kcnq1* et un long ARN non-codant *Kcnq1ot1*, exprimé à partir de l'allèle paternel en anti-sens du gène *Kcnq1* (Smilinich et al., 1999) (Figure 24). L'ICR maternelle, KvDMR, est localisée dans l'intron 10 du gène *Kcnq1* et est associée au promoteur de *Kcnq1ot1*. La méthylation de KvDMR sur l'allèle maternel entraîne la répression de l'ARN non-codant et l'expression des gènes sur cet allèle. En revanche, sur l'allèle paternel non-méthylé, l'expression de *Kcnq1ot1* entraîne la répression de tous les gènes codants (Pandey et al., 2008). La délétion de KvDMR ou l'interruption prématurée de la transcription de *Kcnq1ot1* entraîne une expression biallélique des 8 gènes codants de la région (Fitzpatrick et al., 2002 ; Mancini-Dinardo et al., 2006). Comme pour le locus d'*Igf2r*, la transcription d'un ARN non-codant et/ou le transcrit en lui-même sont requis pour réprimer la transcription des autres gènes du locus. Des modifications d'histones sont ici impliquées, l'ARN non-codant *Kcnq1ot1* recrute G9a au niveau des promoteurs des gènes codants et la marque H3K9me2 y est déposée (Lewis et al., 2004 ; Wagschal et al., 2008). *Kcnq1ot1* interagit aussi avec EZH2 et permettrait le dépôt de H3K27me3 (Umlauf et al., 2004 ; Terranova et al., 2008 ; Zhao et al., 2010 ; Kaneko et al., 2014).

D'autres mécanismes concernent plus particulièrement les gDMR/ICR maternelles intragéniques. Tout d'abord, la gDMR peut servir de promoteur pour un transcrit alternatif initié à partir de l'allèle paternel non-méthylé, comme c'est le cas pour la DMR de *Cdh15* (voir Résultats section 2.1) (Proudhon et al., 2012). Deuxièmement, l'ICR peut contenir un signal d'arrêt de la transcription. Dans ce cas, l'absence de méthylation sur l'allèle paternel permettrait la liaison de certains facteurs sur l'ICR qui induiront un arrêt précoce de la transcription à partir de ce même allèle, comme il a été reporté au locus *Mcts2* (Wood et al., 2008).

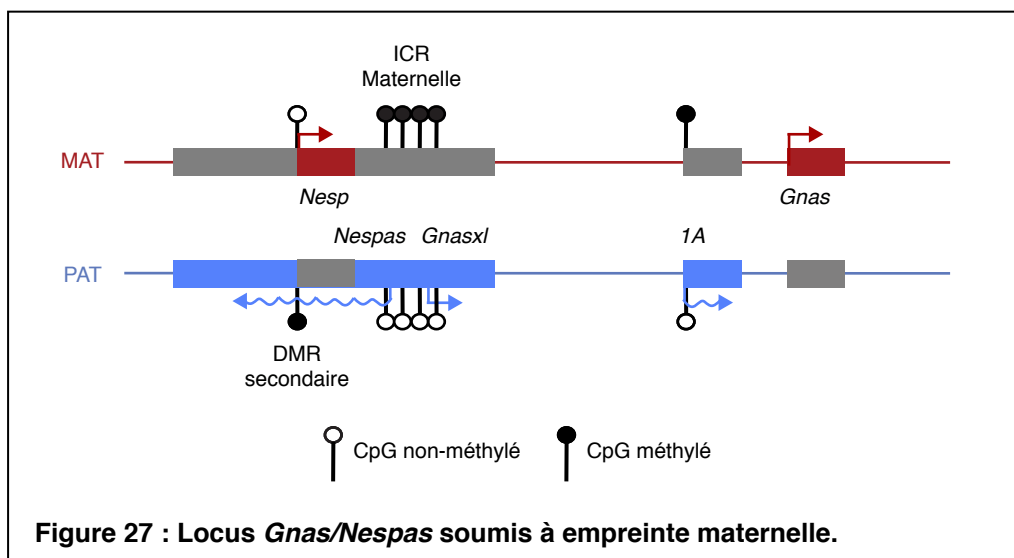
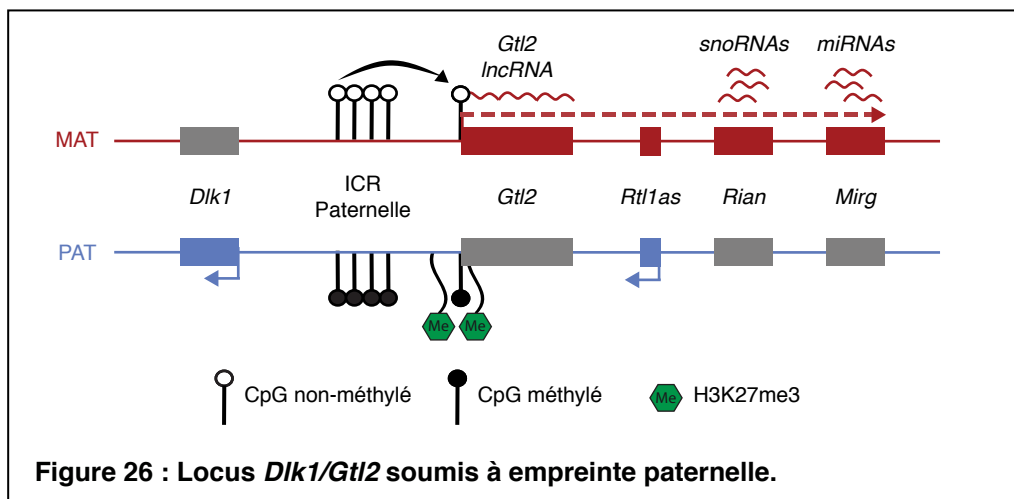
2.2.3– Les ICR paternelles

On connaît à ce jour seulement 3 ICR paternelles. Elles sont associées aux loci *Igf2/H19*, *Rasgrf1* et *Dlk1/Gtl2*. En rapport avec la spécificité générale de la distribution



de la méthylation dans les gamètes mâles et femelles, les ICR paternelles se distinguent par une localisation inter-génique et une plus faible densité en CpG comparées aux ICR maternelles, qui sont riches en CpG et associées en général à des promoteurs (Kobayashi et al., 2006 ; Schulz et al., 2010) (Figure23B). Différents mécanismes sont impliqués dans la régulation des loci soumis à empreinte paternelle : un modèle dit « insulateur » impliquant la fixation de CTCF au niveau des ICR d'*Igf2/H19* et de *Rasgrf1*, et un modèle impliquant l'activation de la transcription d'ARN non-codants au locus *Dlk1/Gtl2* (Edwards and Ferguson-Smith, 2007 ; Ideraabdullah et al., 2008 ; MacDonald, 2011 ; Ferguson-Smith, 2011 ; Sanli and Feil, 2015).

Le locus *Igf2/H19* sur le chromosome 7 chez la souris contient 3 gènes soumis à empreinte : 2 gènes codant pour les protéines *Igf2* et *Ins* exprimés à partir de l'allèle paternel et un ARN non-codant *H19* exprimé à partir de l'allèle maternel. Il existe trois DMR à ce locus, toutes méthylées sur l'allèle paternel : l'ICR/gDMR, située en amont du promoteur d'*H19*, et les deux DMR secondaires d'*Igf2*, à savoir la DMR1 en amont du promoteur qui agit comme un répresseur et DMR2 dans l'exon 6 du gène qui agit comme un enhancer (Bartolomei et al., 1993 ; Constância et al., 2000 ; Murrell et al., 2001). L'expression mono-allélique des gènes *H19* et *Igf2* dépend de la méthylation paternelle de ces DMR en association avec des modifications d'histones. La mise en place des marques de méthylation au niveau des DMR par Dnmt3A et Dnmt3L dans la lignée germinale mâle est nécessaire, leur délétion entraînant une expression biallélique des gènes *H19* et *Igf2* (Tucker et al., 1996 ; Kaneda et al., 2004). L'ICR de *H19* contient 4 sites de fixation CCCTC au facteur CTCF qui peut se lier seulement sur l'allèle maternel déméthylé. Lorsque CTCF est lié, il agit comme un insulateur et bloque l'accès du promoteur d'*Igf2* aux enhancers présents en aval de *H19* sur l'allèle maternel. La méthylation de l'ICR de *H19* sur l'allèle paternel empêche CTCF de se lier et le promoteur d'*Igf2* peut alors se lier aux enhancers (Figure25A). CTCF joue aussi un rôle dans l'interaction entre les différentes parties d'un chromosome (Filippova, 2008). Les protéines CTCF peuvent recruter les protéines cohésines et le complexe CTCF-cohésine initie alors la création d'une boucle chromosomique. Sur l'allèle maternel, le complexe CTCF-cohésine est lié à l'ICR de *H19* et permet d'isoler le gène *H19* afin de maintenir son expression sur l'allèle maternel et de renforcer la répression d'*Igf2* par la création d'un domaine répressif (Murrell et al., 2004 ; Kurukuti et al., 2006 ; Wendt et al., 2008 ; Parelho et al., 2008) (Figure25B).



Le locus *Dlk1/Gtl2* sur le chromosome 12 chez la souris s'étend sur presque 1Mb et contient 3 gènes soumis à empreinte codant pour les protéines *Dlk1*, *Dio3* et *Rtl1* exprimés à partir de l'allèle paternel. Sur l'allèle maternel, un long ARN non-codant *Gtl2* et des micro-ARNs sont exprimés à partir des gènes *Gtl2*, *antiRtl1*, *C/D SnoRNAs* et *Mirg*. L'ICR paternelle de ce locus, située en amont du promoteur de *Gtl2*, agirait quand elle est déméthylée comme un enhancer et activerait l'expression de ces ARN sur l'allèle maternel, notamment via le recrutement de l'ARN polymérase II (Kota et al., 2014) (Figure 26). Le long ARN non-codant *Gtl2*, exprimé à partir de l'allèle maternel, interagit avec EZH2, ce qui pourrait de ce fait méthyliser les résidus H3K27 au niveau du promoteur de *Gtl2* sur l'allèle paternel afin d'en réprimer l'expression (Zhao et al., 2010 ; Kaneko et al., 2014 ; Kota et al., 2014). Un deuxième mode de régulation a aussi été observé : les micro-ARN produits à partir de l'allèle maternel non-méthylé interfèrent avec les ARN messagers des gènes exprimés à partir de l'allèle paternel et ciblent leur dégradation (Seitz et al., 2003 ; Lin et al., 2003 ; Saito et al., 2006 ; da Rocha et al., 2008). Il a été proposé que ces micro-ARN viennent en fait du long transcrit non-codant contrôlé par le promoteur de *Gtl2* (Schuster-Gossler et al., 1998 ; Seitz et al., 2004). D'autre part, les micro-ARN générés par le gène *antiRtl1* sur l'allèle maternel régulent le niveau d'expression du gène *Rtl1* sur l'allèle paternel (Davis et al., 2005).

2.2.4– Les DMR secondaires

Les DMR secondaires ou DMR somatiques (sDMR) acquièrent leur méthylation au cours du développement post-fécondation, toujours sous le contrôle de l'activité d'une gDMR/ICR (Edwards and Ferguson-Smith, 2007). On pense que la transcription des GSE dirige la méthylation allèle-spécifique sur ces DMR secondaires, notamment via le recrutement des ADN méthyltransférase *de novo* par la marque H3K36me3 associée à l'élongation de la transcription (Dhayalan et al., 2010 ; Latos et al., 2012). La transcription à travers le locus permettrait le maintien d'un état ouvert de la chromatine qui permettrait à la machinerie de la méthylation d'accéder à l'ADN.

Un exemple d'acquisition de la méthylation sur une sDMR est observé au niveau du locus soumis à empreinte maternelle *Gnas/Nespas* (Liu et al., 2005 ; Proudhon et al., 2012). Ce locus est situé sur le chromosome 2 chez la souris et contient 5 gènes soumis à empreinte : les gènes codant pour les protéines *Gnas*, *Gnasx1* et *Nesp* et 2 gènes

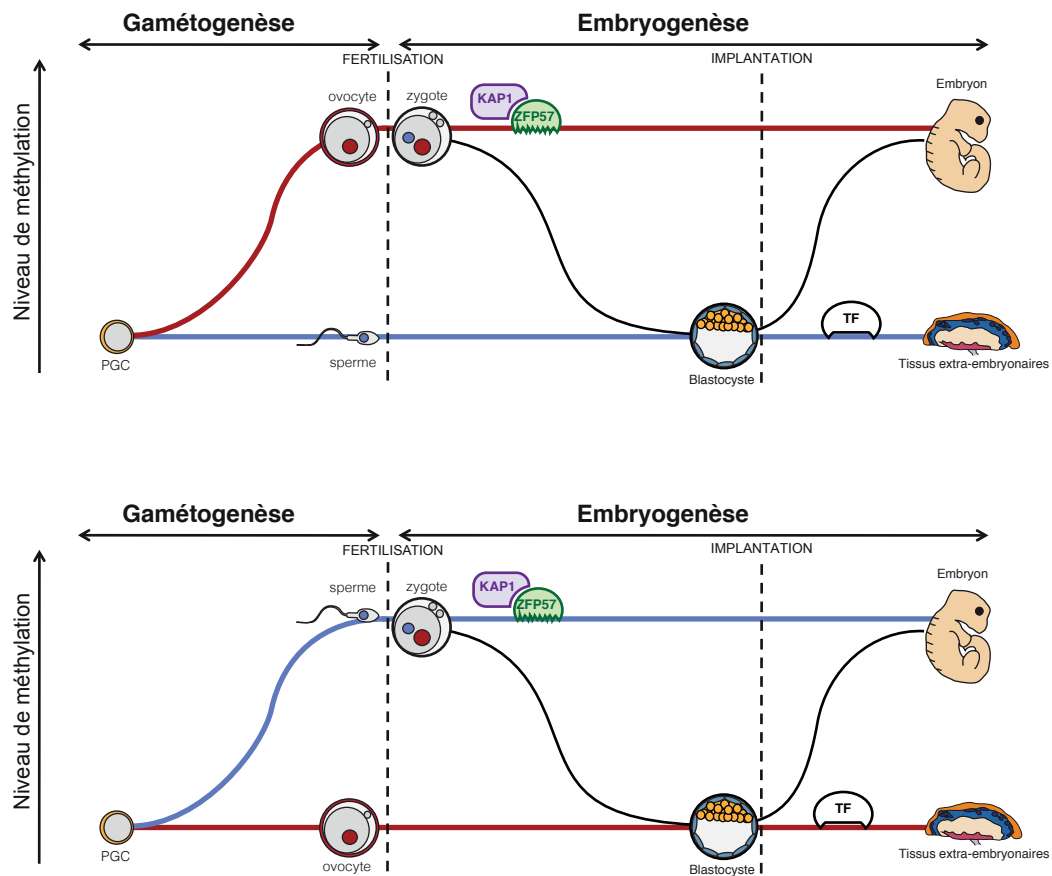


Figure 28 : Protection des allèles méthylés et déméthylés des ICRs.

A. Protection des allèles d'une ICR maternelle.

B. Protection des allèles d'une ICR paternelle.

produisant les ARN non-codants *Nespas* et *1A*. Le locus contient deux gDMR maternelles, une qui comprend les promoteurs des gènes *Nespas* et *Gnasx1* et qui est l'ICR du locus, et l'autre qui contient le promoteur du gène *1A* (Coombes et al., 2003 ; Williamson et al., 2006). Une troisième DMR de type secondaire est située au niveau du promoteur du gène *Nesp*, elle est quant à elle méthylée sur l'allèle paternel. La transcription à partir de l'allèle paternel de l'ARN non-codant *Nespas* induit le dépôt de la méthylation sur cette DMR secondaire dans les tissus somatiques. En absence de méthylation sur l'ICR maternelle, *Nespas* devient biallélique et la méthylation au niveau de la DMR secondaire a lieu aussi sur les deux allèles (Liu et al., 2005) (Figure27).

2.3- Protection des profils de méthylation différentiels des ICR

Les ICR ont la propriété unique de maintenir leur méthylation différentielle après la fécondation dans l'individu nouvellement formé, gardant ainsi en mémoire leur origine parentale sur chaque allèle, malgré les remaniements drastiques de méthylation qui caractérisent le génome de l'embryon précoce (Duffié and Bourc'his, 2013) (Figure28). Comment les ICR protègent leurs allèles méthylés et leurs allèles non-méthylés de ces vagues de déméthylation et re-méthylation *de novo* et quels acteurs entrent en jeu sont les deux questions du chapitre suivant.

2.3.1- Protection des allèles méthylés contre la déméthylation après la fécondation

Juste après la fécondation, le génome subit une vague de déméthylation durant laquelle les allèles méthylés des ICR sont à risque. Comment les allèles méthylés sont maintenus et quels sont les acteurs moléculaires impliqués dans la protection des ces allèles contre la déméthylation commencent de mieux en mieux à être connus. La protéine PGC7/STELLA est impliquée dans la protection des allèles méthylés au stade 1-cellule mais agirait seulement au niveau des ICR paternelles (Nakamura et al., 2007) (voir Introduction section 1.4.3). En revanche, la protéine Zfp57 protège les marques de méthylation à la fois des ICR maternelles et paternelles, comme le montre l'analyse d'embryons mutants maternel et zygotique de *Zfp57* (Li et al., 2008). La liaison de Zfp57 sur les ICR permet le recrutement d'un complexe hétérochromatique organisé autour de KAP1 (Krüppel-associated protein), aussi connu sous le nom de TRIM28, et qui contient

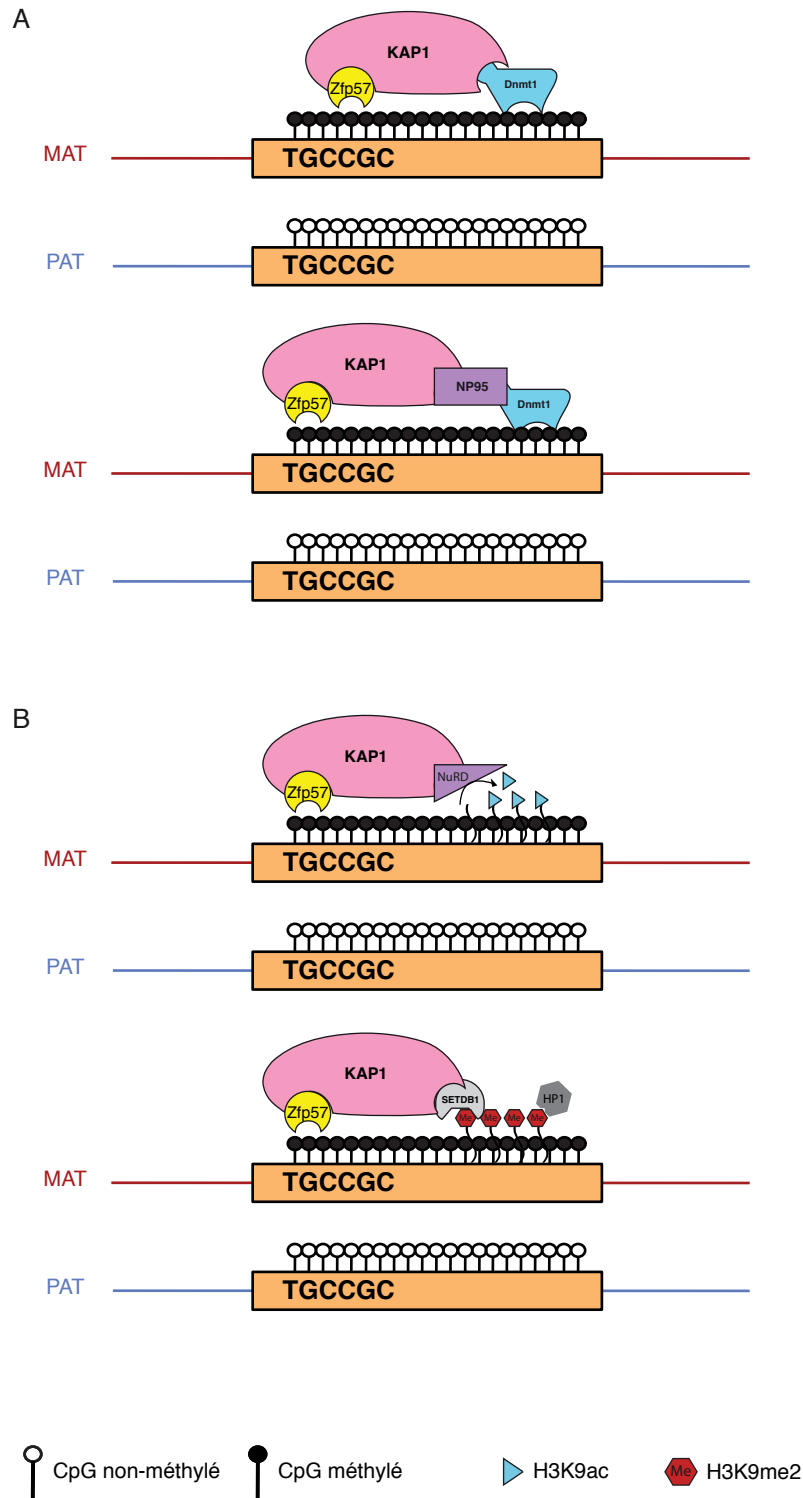


Figure 29 : Mécanismes de protection des ICRs méthylées.

A. Maintien de la méthylation au niveau des ICRs méthylées par Dnmt1.

B. Maintien de la méthylation par un mécanisme de protection des CpGs méthylés par H3K9me2 et HP1.

différents facteurs de remodelage de la chromatine. KAP1 est nécessaire au maintien de la méthylation via Zfp57, comme le suggère l'hypométhylation des ICR dans les embryons *Kap1*^{-/-} et dans les cellules ES exprimant une forme mutée de la protéine Zfp57 dans son domaine KRAB d'interaction avec KAP1 (Messerschmidt et al., 2012 ; Zuo et al., 2012). Dans les cellules ES *Dnmt-TKO*, la liaison du complexe KAP1/Zfp57 sur les ICR est perdue, indiquant que la liaison du complexe est dépendante de la méthylation (Quenneville et al., 2011). Des analyses de la séquence des sites de liaison de Zfp57 déterminés par ChIP-seq ont permis de révéler un motif hexamérique TGCCGC qui ne peut être lié par Zfp57 seulement quand le CpG interne est méthylé (Quenneville et al., 2011 ; Messerschmidt et al., 2012 ; Liu et al., 2012). Le complexe KAP1/Zfp57 se lie donc aux ICR par la combinaison d'une information génétique et épigénétique (Figure29A&B).

Parmi les composants associés au complexe KAP1, on retrouve l'H3K9 méthyltransférase SETDB1, la protéine HP1 et le complexe de remodelage des nucléosomes et de dé-acétylation des histones NuRD (Schultz et al., 2001 ; Schultz., 2002 ; Iyengar and Farnham, 2011). Dans les cellules ES, KAP1 a aussi été montré comme interagissant avec les ADN méthyltransférases Dnmt1, Dnmt3A et Dnmt3B, ainsi que la protéine Np95 qui interagit avec Dnmt1 (Quenneville et al., 2011 ; Zuo et al., 2012). Le recrutement de KAP1 sur les ICR permettrait donc ainsi le maintien des allèles méthylés pendant le développement pré-implantatoire, soit par une forte densité locale de l'enzyme de maintenance Dnmt1, ou par méthylation *de novo* réitérative par Dnmt3A/Dnmt3B contrecarrant la déméthylation passive. Cependant, les souris mutantes pour Dnmt3A et Dnmt3B ne présentent pas de défaut de méthylation au niveau des ICR, suggérant qu'elles ne sont pas directement responsables du maintien de la méthylation de ces régions (Hirasawa et al., 2008). Par contre, les souris mutantes pour Dnmt1 présentent une déméthylation complète des ICR, indiquant que ce serait Dnmt1 qui serait responsable du maintien de la méthylation au niveau des ces régions (Hirasawa et al., 2008). Selon un scénario admis, la protéine Zfp57 recruterait au niveau des ICR KAP1, qui elle-même interagit avec Dnmt1 directement ou via Np95 afin de maintenir la méthylation de ces ICR (Figure29A).

L'histone H3K9 méthyltransférase SETDB1, qui catalyse la réaction de triméthylation et dont l'activité favorise la liaison de la protéine HP1 sur les promoteurs des gènes réprimés, est associée à la protéine KAP1 (Schultz et al., 2002 ; Quenneville et

al., 2011). Le complexe NuRD de remodelage des nucléosomes, qui catalyse la réaction de dé-acétylation des résidus H3K9 notamment, est lui aussi associé à la protéine KAP1 (Schultz et al., 2001). Les marques H3K9me3 et H3K9ac ont été observées au niveau des ICR maternelles et paternelles mais en absence de KAP1, on observe une perte des marques H3K9me3 et un fort enrichissement en H3K9ac (Kacem and Feil, 2009 ; Quenneville et al., 2011) (Figure 29B).

2.3.2– Protection des allèles non-méthylés contre la méthylation après l'implantation

Pendant la vague de reméthylation génomique post-implantatoire, ce sont les allèles non-méthylés des ICR qui doivent être protégés. Comment ces allèles sont capables de maintenir cet état déméthylé est intrigant et aujourd'hui encore peu documenté. La façon la plus simple d'expliquer le maintien de cet état implique la présence de facteurs protecteurs au niveau des allèles déméthylés, qui pourraient agir au niveau de la chromatine en association avec des modifications d'histones.

Comme vu précédemment, d'une manière générale, les séquences du génome qui sont intrinsèquement réfractaires à la méthylation sont les îlots CpG situés en régions promotrices de gènes transcriptionnellement actifs (Bestor et al., 1992). Ces régions portent des caractéristiques répulsives pour la méthylation de l'ADN, telles que le variant d'histone H2A.Z et les marques H3K4me3 (Meissner et al., 2008 ; Edwards et al., 2010 ; Vavouri and Lehner, 2012) (voir Introduction section 1.3.3). Mais une étude a montré que dans des cellules ES mutantes pour Cfp1, une protéine responsable de la localisation de H3K4me3, les îlots CpG ne présentent pas spécialement de gain de méthylation (Clouaire et al., 2012). Il est donc évident que d'autres mécanismes entrent en jeu dans la protection des allèles déméthylés des ICR.

La transcription en elle-même peut jouer un rôle dans la protection des allèles déméthylés (Bird, 2002). Par exemple, au locus soumis à empreinte *Igf2r*, l'élongation de la transcription de l'ARN non-codant *Airn* est nécessaire pour empêcher la méthylation de l'ICR maternelle, qui est un îlot CpG (Stricker et al., 2008). En accord avec ce principe, les loci résistants à un gain de méthylation aberrant dans des cancers sont ceux qui présentent un fort niveau de transcription ainsi que la liaison de l'ARN Pol II, comparés à des loci susceptibles (Takeshima et al., 2009). De plus, des mutations dans les motifs de liaison aux facteurs de transcription ou l'abolition de la marque H3K4me3, qui induisent

par conséquent un arrêt de la transcription, entraînent un gain de méthylation au niveau des promoteurs de ces transcrits (Ladopoulos et al., 2013).

La séquence de l'ADN serait donc importante, notamment pour la liaison de facteurs de transcription, dans la protection des allèles non-méthylés. Une étude basée sur la comparaison des séquences d'îlots CpG qui maintiennent un état déméthylé et ceux qui regagnent de la méthylation au moment de l'implantation a mis en évidence 23 motifs, dont les motifs des facteurs de transcription Sp1 et STAT1 (Straussman et al., 2009). Empêcher la liaison du facteur de transcription Sp1 en créant des mutations dans son motif de liaison entraîne un gain de méthylation au niveau des promoteurs ciblés habituellement par Sp1 (Brandeis et al., 1994 ; Macleod et al., 1994). Un autre exemple est celui du facteur CTCF qui se lie aux allèles maternels non-méthylés au niveau de l'ICR du locus *Igf2/H19* et les protège d'un gain de méthylation (Schoenherr et al., 2003 ; Rand et al., 2004 ; Pant et al., 2004). CTCF se lie aussi au niveau des ICR des locus *Rasgrf1* et *Dlk1/Gtl2* et pourrait être impliqué dans la protection des allèles maternels déméthylés au niveau de ces loci (Yoon et al., 2005 ; Lin et al., 2011).

Il existe cependant une catégorie de gDMR/ICR maternelles découverte récemment qui ne résiste pas à cette vague de re-méthylation *de novo* du génome : les gDMR dites transitoires, qui ont fait l'objet d'une partie de mes travaux de thèse (Proudhon et al., 2012 ; Duffié et al., 2014) (voir Résultats section 2). Pour l'instant, leur nombre est trop restreint pour pouvoir identifier des motifs nucléotidiques et/ou des caractéristiques chromatinienne qui permettraient de les distinguer des autres gDMR/ICR qui elles sont protégées contre la re-méthylation et qui persistent tout au long de la vie.

2.4- Evolution de l'empreinte génomique parentale

Les gènes soumis à empreinte ont évolué spécifiquement chez les mammifères vivipares, à savoir les thériens. Aucun GSE n'a été identifié jusqu'à présent chez les mammifères ovipares (les monotrèmes) tels que l'ornithorynque. L'expression mono-allélique des gènes est une façon *a priori* risquée de réguler son génome pour l'organisme, puisqu'une mutation délétère sur l'allèle exprimé ne pourra pas être compensée par l'autre allèle. Il doit donc y avoir une pression évolutive importante pour

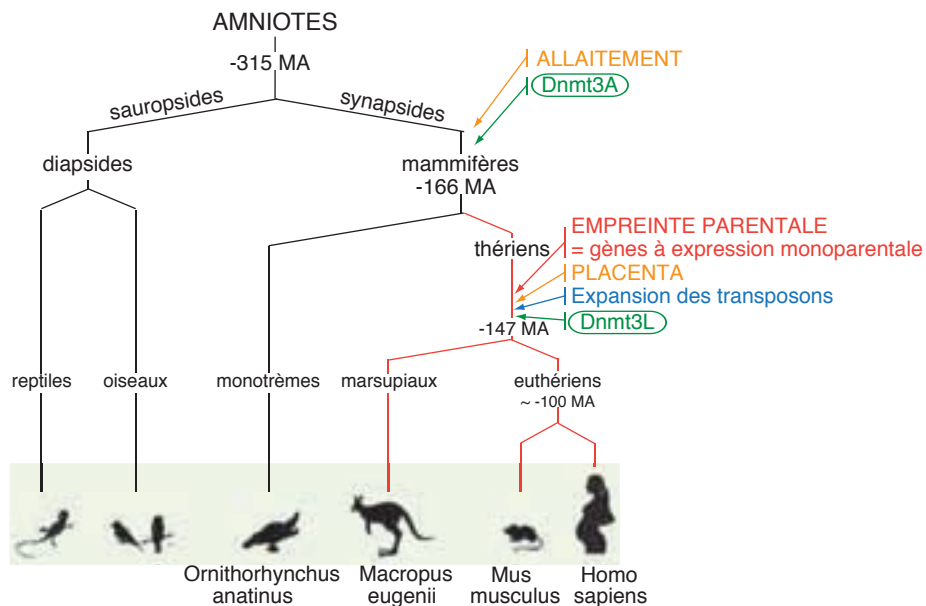


Figure 30 : Apparition de l’empreinte parentale au cours de l’évolution.

Trois groupes de mammifères existent : les monotrèmes et les thériens, composés eux-mêmes des marsupiaux et des euthériens. Tous ont recours à l’allaitement, tandis que le placenta est l’apanage des thériens. L’empreinte parentale est apparue avant la divergence des thériens, coïncidant avec l’acquisition du placenta, de Dnmt3L et de l’expansion des transposons. MA = millions d’années.

(d’après Proudhon and Bourc’his, 2010a).

maintenir ce mode de régulation. La question est donc de savoir quand, comment et pourquoi l’empreinte génomique parentale est apparue.

En terme de datation, l’empreinte génomique parentale serait donc apparue il y a plus de 150 millions d’années chez l’ancêtre des mammifères thériens, qui incluent les mammifères placentaires et les marsupiaux (Figure 30). Compte-tenu du rôle de la méthylation germinale dans le contrôle de l’empreinte, un prérequis nécessaire à son évolution semble être la disponibilité d’une machinerie de méthylation *de novo* dans les gamètes. De manière intéressante, l’apparition des GSE est corrélée à celle du gène *Dnmt3L*, ce co-facteur de stimulation de la méthylation *de novo* du génome des gamètes (Yokomine et al., 2006). Compte-tenu du rôle indispensable de *Zfp57* dans la maintenance de la méthylation liée à l’empreinte après la fécondation, il serait intéressant de vérifier sa présence ou non dans le génome des monotrèmes, non pourvus d’empreinte génomique.

Les GSE jouent un rôle très important sur la croissance de l’embryon et sur les échanges entre la mère et le fœtus via le placenta *in utero*, mais aussi après la naissance. La mère a besoin de contrôler tous ces échanges et les GSE seraient donc un moyen de le faire. Les spécificités reproductives des mammifères thériens, basées sur l’allaitement et le développement *in utero* via le placenta, auraient donc un rôle central. D’ailleurs chez les plantes, l’empreinte génomique parentale a évolué de façon convergente, chez les espèces dotées d’endosperme qui joue un rôle de tissu nourricier pour la graine, comme le placenta pour l’embryon de mammifère (Feil and Berger, 2007). En accord avec ce rôle crucial du placenta, on observe une expansion du nombre des GSE chez les euthériens spécifiquement par rapport aux marsupiaux, dont le placenta est beaucoup plus rudimentaire.

On connaît aujourd’hui trois scénarios possibles d’apparition de nouveaux GSE dans le génome (Proudhon and Bourc’h, 2010a). Premièrement, les gènes étaient déjà présents dans le génome ancestral et c’est l’apparition d’une ICR différenciellement méthylée dans la région qui a engendré une expression soumise à empreinte de ceux-ci, comme au niveau du locus *Mest/Peg1*. Deuxièmement, l’apparition d’une ICR coïncide avec des réarrangements chromosomiques autour de celle-ci entraînant l’apparition de cluster soumis à empreinte. C’est le cas du cluster *Snrpn/Ube3A* (Rapkins et al., 2006). Troisièmement, le gène dérive d’un transposon ou de l’activité d’un transposon et acquiert un statut soumis à empreinte dès son insertion ailleurs dans le génome. Cette

dernière catégorie de gènes inclue des retrogènes, notamment *Inpp5f_v2*, *Mcts2*, *Nap1l5* et *U2af1-rs1*, qui ont évolué récemment et ont tous un orthologue sur le chromosome X (Wood et al., 2007 ; McCole et al., 2011).

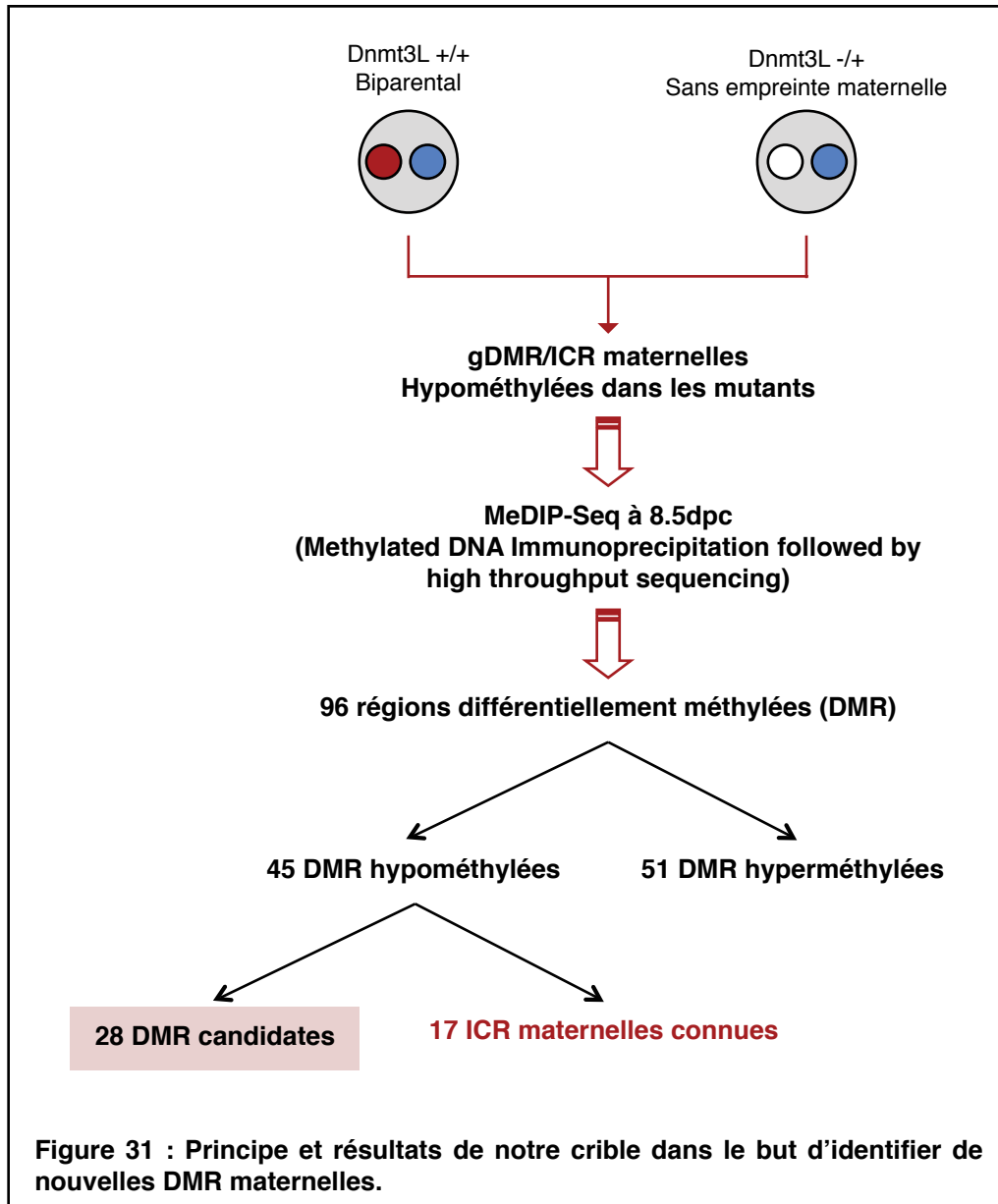
La plupart des gènes soumis à empreinte chez la souris le sont aussi chez l'humain bien qu'il existe des exceptions. En effet, certains GSE sont espèce-spécifiques, ce qui montre bien que l'empreinte génomique parentale est un processus évolutif en marche. C'est notamment le cas des gènes *Cdh15* et *Socs5* qui ne sont pas soumis à empreinte chez l'Homme et dont nous avons caractérisé au laboratoire le statut soumis à empreinte maternelle chez la souris (voir Résultats sections 2.1 et 3.2).

3- Les différents types de cribles développés pour l'identification de nouveaux GSE

Dès la découverte des premiers GSE en 1991, des tentatives de cribles ont été mises en place. Les premiers cribles historiques se sont appuyés sur des modèles de souris portant des disomies uniparentales et sur les embryons parthénogénotes et androgénotes. Tous ces modèles sont particulièrement adaptés, puisqu'ils sont basés sur la recherche de gains et de pertes de méthylation ou d'expression qui vont du simple au double. Du fait des caractéristiques phénotypiques des androgénotes (tissus extra-embryonnaires hyperplasiques) et des parthénogénotes (développement exclusif des tissus embryonnaires), il existe beaucoup de différences de méthylation et d'expression tissu-spécifiques entre ces embryons. Les cribles qui s'appuient sur ces modèles contiennent donc beaucoup de faux positifs, qui sont des gènes dérégulés mais sans lien avec l'empreinte génomique parentale. Par la suite, les récentes avancées des techniques de séquençage ainsi que le développement d'outils bioinformatiques adaptés ont permis de développer des cribles indépendamment de ces modèles de déficience drastique en empreinte génomique et d'obtenir, *a priori*, des listes plus robustes de candidats. Notre laboratoire a participé à ces efforts. Malheureusement, seul un petit nombre de ces candidats a ensuite été validé par des expériences complémentaires.

3.1- Les cribles basés sur les caractéristiques de séquence des ICR

Très peu de cribles ont été réalisés sur la base des caractéristiques de séquence des ICR, ces caractéristiques étant communes à d'autres régions du génome non soumises à empreinte. Une étude ancienne a cependant utilisé un modèle statistique basé sur des caractéristiques de séquence telles que la présence d'éléments répétés, de motifs de liaison à des facteurs de transcription ou encore la densité en CpG pour développer un algorithme d'identification de nouvelles régions soumises à empreinte chez la souris (Luedi et al., 2005). Parmi les 600 régions candidates identifiées, 16 d'entre elles ont été examinées de plus près et testées par des expériences complémentaires : seulement une s'est révélée arborer une expression mono-allélique au niveau du gène *Kcnk9* mais sans pour autant que l'îlot CpG de son promoteur ne



présente de méthylation différentielle (Ruf et al., 2007). Une étude similaire chez l'humain a permis d'identifier plus de 150 régions potentiellement soumises à empreinte et de confirmer cette expression mono-allélique du gène *KCNK9* en plus d'un autre gène, *DLGAP2* (Luedi, et al., 2007). La protéine Zfp57 se liant aux gDMR/ICR méthylées afin de les protéger de la vague de déméthylation du génome, il serait intéressant de regarder les régions arborant ce motif de liaison pour trouver de nouveaux candidats parmi les listes de candidats déjà établies avec d'autres critères. De manière encore plus informative, il serait intéressant de réaliser des ChIP-seq de Zfp57 sur des fonds alléliques hybrides.

3.2- Les cribles basés sur la méthylation de l'ADN des ICR

Une des caractéristiques essentielles des GSE étant la méthylation génomique différentielle de leurs ICR, plusieurs cribles se sont donc basés sur cette propriété. Ces études utilisent des méthodes immunologiques (MeDIP), chimiques (traitement au bisulfite) ou enzymatiques (enzymes de restriction sensibles à la méthylation) suivie d'hybridation sur puces ou de séquençage haut débit pour étudier la méthylation du génome (Strichman-Almashanu et al., 2002 ; Smith et al., 2003 ; Choufani et al., 2011 ; Nakabayashi et al., 2011). La technique de traitement au bisulfite de sodium de l'ADN permet de convertir par déamination les cytosines non-méthylées en uraciles (puis thymidines après PCR) alors que les cytosines méthylées sont résistantes à ce traitement, ce qui permet de les distinguer (Frommer et al., 1992). L'utilisation d'enzymes de restriction sensibles (*HhaI*, *HpaII*) ou dépendantes (*McrBC*) de la méthylation permet aussi de distinguer les séquences méthylées et non-méthylées. Les études sur les puces d'ADN sont biaisées car les chercheurs ont centré leurs études sur les îlots CpG et les promoteurs et de ce fait, sont sûrement passés à côté d'autres loci qui n'étaient pas représentés sur les puces. Les nouvelles techniques de séquençage à haut débit apportent un moyen d'étudier le génome entier de façon non-biaisé.

Notre laboratoire a conduit une étude comparative des méthylomes d'embryons sauvages et d'embryons sans empreinte maternelle *Dnmt3L*^{-/+} par MeDIP-Seq (Figure31). Ce crible nous a permis d'identifier 45 régions soumises à empreinte maternelle candidates parmi lesquelles on retrouve les 17 ICR maternelles déjà connues.

Parmi les 28 DMR candidates restantes, quatre d'entre elles ont été validées, portant le nombre connu de gDMR/ICR de 20 à 24. Ces quatre gDMR sont associées aux gènes *Cdh15*, *Nhlrc1*, *Zfp777* et *Zfp787*. Ce crible nous a aussi permis de mettre en évidence une nouvelle forme d'empreinte génomique, les gDMR transitoires dont le locus *Gpr1/Zdbf2* fait partie. Enfin, une nouvelle gDMR candidate associée au gène *Socs5* a attiré mon attention et la caractérisation de cette gDMR a occupé la fin de ma thèse. De manière intéressante, nous avons mis en évidence des gDMR/ICR dont la méthylation est variable selon les tissus et selon le stade du développement, en contraste avec les ICR « canoniques » qui étaient déjà connues et dont l'empreinte génomique est ubiquitaire.

Une étude de WGBS réalisée dans l'ovocyte et le sperme a permis d'identifier 1,600 CpG différentiellement méthylés entre ces deux gamètes, soit 646 régions de méthylation germinale différentielle (ou gDMR : 568 gDMR dans l'ovocyte et 78 dans le sperme) (Kobayashi et al., 2012). La moitié de ces CpG serait résistante à la vague de déméthylation du génome, maintenant environ 50% de méthylation dans le blastocyste. Parmi eux, on retrouve les DMR associées aux gènes *Cdh15*, *Zfp777*, *Zfp787*, *Gpr1* et *Socs5* que nous avons trouvés dans notre propre criblage (voir Résultats sections 2.1, 2.2 et 3.2). Une autre étude de WGBS réalisée dans le cortex préfrontal de souris hybrides adultes a permis d'identifier 23 nouvelles DMR potentielles, dont 9 seulement sont en dehors de loci soumis à empreinte connus (Xie et al., 2012). Parmi ces 9 régions candidates, on retrouve les régions associées au gène *Nhlrc1* et au locus *Gpr1/Zdbf2*. Enfin, la dernière étude de WGBS la plus récente réalisée dans 17 tissus adultes a permis d'identifier plus de 300,000 régions différentiellement méthylées de façon tissu-spécifique (tsDMRs), indépendamment du statut d'empreinte génomique cependant (Hon et al., 2013). La région associée au gène *Socs5* en fait partie.

Le traitement bisulfite de l'ADN entraîne beaucoup de changements C vers T dans le génome et comme énormément de polymorphismes exprimés (SNPs) sont des Cytosines versus des Thymines, de potentiels gènes candidats peuvent échapper à ces études basées sur la conversion bisulfite de tissus hybrides. De plus, le choix du tissu et du stade de développement auxquels sont faites ces études, sont très importants. Par exemple les 23 régions découvertes dans le cortex préfrontal sont peut-être spécifiques de ce tissu et ne seront pas soumises à empreinte dans les 17 autres tissus de la dernière étude.

3.3- Les cribles basés sur l'expression mono-allélique des GSE

Une autre caractéristique des GSE est leur expression mono-allélique à partir d'un des deux allèles parentaux. La comparaison quantitative sur des puces des profils d'expression d'embryons parthénogénètes et androgénètes a mis en évidence plus de 2,000 transcrits différentiellement exprimés (Nikaido et al., 2003). Une étude complémentaire utilisant la technique de pyroséquençage de SNP dans des embryons hybrides réciproques, a finalement validé trois d'entre eux comme ayant une expression soumise à empreinte (Ruf et al., 2006). Le séquençage de cDNA produit à partir d'embryons à 9,5dpc issus de croisements hybrides réciproques a révélé que la majorité des nouveaux GSE candidats découverts se trouvaient dans des régions soumises à empreinte déjà connues. Trois nouveaux GSE ont tout de même été identifiés, dont *Zdbf2* (Babak et al., 2008). D'autres études de RNA-Seq réalisées dans le placenta murin ou humain ont permis d'identifier de très nombreux candidats (1088 et 495 respectivement) (Okada et al., 2012 ; Das et al., 2013). Cependant, le problème posé par l'infiltration maternelle dans ce tissu ne permet pas d'être certain des résultats obtenus, surtout si les gènes sont exprimés à partir de l'allèle maternel (Proudman and Bourc'his, 2010b ; Wang et al., 2011 ; Okada et al., 2012). En utilisant une technique de transplantation d'embryons de la lignée CH3/HeJ dans des femelles pseudo-gestantes de la lignée différente C57Bl/6J, il est possible de distinguer les gènes exprimés exclusivement dans le placenta grâce à la présence de SNP entre les deux lignées (Okada et al., 2012).

Une des études de RNA-Seq alléliques les plus connues est celle réalisée par le groupe de Catherine Dulac à partir de différentes parties du cerveau de souris hybrides issues de croisements réciproques, en stade embryonnaire et adulte. La surprise est venue du très grand nombre de transcrits (1,300) potentiellement soumis à empreinte (Gregg et al., 2010). Ce nombre très élevé de candidats a attiré la curiosité et a posé la question de l'étendue de l'empreinte génomique dans le cerveau et notamment des mécanismes qui pourraient agir dans ce tissu. Cependant, une seconde analyse de ces données de RNA-seq avec des outils bioinformatiques plus poussés combinée à des étapes de validation de ces GSE candidats, a révélé que la majorité d'entre eux étaient des faux positifs (DeVeale et al., 2012). Il est à noter que parmi les GSE candidats de

cette étude (Gregg et al., 2010), nous avons tout de même validé l'un d'entre eux, le gène *Cdh15* (voir Résultats section 2).

L'étude la plus récente en terme de criblage de GSE par RNA-seq allélique a été réalisée sur 33 tissus murins hybrides à différents stades du développement et 45 tissus humains (Babak et al., 2015). Cette étude a permis de mettre en évidence 74 nouveaux GSE candidats chez la souris et 17 chez l'homme. Mais une étude complémentaire par pyroséquençage des candidats murins a révélé que seulement 12 présentaient une expression mono-allélique, dont 7 faisaient partie de régions soumises à empreinte déjà connues. Néanmoins, cette étude exhaustive représente un atlas important de documentation de la spécificité tissulaire d'expression des GSE connus, en termes de quantité mais aussi d'allélisme, chez la souris et l'homme. Elle démontre notamment de manière non biaisée la plus grande prédilection de gènes soumis à empreinte à être exprimés en vie *in utero* (embryon et placenta) et dans les tissus neuraux par rapport à n'importe quel autre tissu post-natal, en rapport avec la fonction supposée de l'empreinte génomique dans le développement et les liens sociaux (Babak et al., 2015).

En conclusion, tous les cribles basés sur la méthylation différentielle des ICR ou l'expression mono-allélique des GSE ont au final identifié très peu de candidats. De ce fait, beaucoup s'accordent à dire que la majorité des loci soumis à empreinte ainsi que leurs transcrits ont été identifiés. Cependant, malgré leur nombre limité, les GSE sont impliqués dans de nombreuses situations pathologiques chez l'Homme, en réponse à une surexpression ou à une répression, illustrant l'importance d'un contrôle précis de la dose de ces gènes. L'apparition d'anomalies dans la lignée germinale ou très précocement au cours du développement donne lieu à des pathologies congénitales, comme les syndromes de Beckwith-Wiedemann et de Prader-Willi/Angelman (Hirasawa and Feil, 2010). Si ces anomalies apparaissent dans des tissus somatiques, elles peuvent provoquer le développement de cancers (Feinberg et al., 2002). Evaluer le nombre réel de gènes soumis à empreinte, comprendre leur mode de régulation et déterminer leur impact sur la physiologie et les phénotypes représentent ainsi un challenge important en biologie. De même, l'identification de nouvelles classes de gènes qui divergent de la définition classique des GSE nous permettra de mieux comprendre la fonction de l'empreinte génomique parentale dans le génome des mammifères et son importance évolutive.

RESULTATS

1- Contexte et questions abordées

Lorsque j'ai commencé mon projet de thèse en 2011, on connaissait environ cent gènes soumis à empreinte génomique, répertoriés chez la souris et généralement conservés chez l'homme. Une vingtaine d'ICR contrôlant cette centaine de GSE avait de plus été identifiée. La grande majorité d'entre elles (17) était connue comme étant contrôlée par méthylation maternelle, apposée dans l'ovocyte, contre seulement 3 sous contrôle paternel par méthylation dans le spermatozoïde (Schulz *et al.*, 2010). Bien que le nombre théorique de GSE était situé de façon informelle entre 100 et 200, les méthodes de cartographie non biaisées par séquençage haut débit- soit des profils de méthylation, soit des profils alléliques de transcription- n'en étaient qu'à leur début (Babak *et al.*, 2008 ; Gregg *et al.*, 2010 ; Xie *et al.*, 2012). Le nombre réel de gènes soumis à empreinte, la réalité de ce biais de contrôle maternel dominant et la variation possible du statut d'empreinte au cours de la vie étaient encore des notions incertaines. La question posée au début de mes recherches était ainsi la suivante : combien de GSE et d'ICR existe-t-il réellement dans le génome des mammifères ? Et plus particulièrement combien d'ICR maternelles ?

L'objectif général de mon projet de thèse était ainsi d'**identifier de nouvelles régions de contrôle de l'empreinte génomique maternelle, de suivre la stabilité de leur statut empreinté au cours du développement et de la différenciation cellulaire, et d'analyser leur impact sur la régulation génique, via la méthylation de l'ADN**. Mon étude a été principalement basée sur le modèle murin, avec un souci d'étude de la conservation des phénomènes observés chez l'Homme.

2- Identification de nouvelles DMR maternelles germinales non-ubiquitaires

Pour identifier et caractériser de nouvelles ICR maternelles, je me suis basée sur une étude comparative des profils de méthylation génomique d'embryons murins sauvages et *Dnmt3L*^{-/+}, réalisée dans le laboratoire par Charlotte Proudhon, une précédente étudiante en thèse (Introduction Figure 31) (Proudhon et al., 2012). Ce criblage était basé sur la technique de MeDIP-Seq (Methylated DNA Immunoprecipitation) qui consiste en une immunoprécipitation de la fraction méthylée du génome suivie de séquençage à haut débit. Les embryons *Dnmt3L*^{-/+} sont dépourvus de méthylation d'origine maternelle, car issus de la fécondation d'ovocytes mutants *Dnmt3L*^{-/-} incapables de méthyliser leur génome (Bourc'his *et al.*, 2001). Ce défaut de méthylation maternelle entraîne une létalité embryonnaire systématique à 10dpc, en association avec une dérégulation de l'expression des gènes soumis à empreinte maternelle. En dehors des régions de contrôle de l'empreinte maternelle, les autres régions du génome sont normalement méthylées. En comparant le méthylome des embryons sauvages avec celui des embryons mutants à 8,5dpc, les régions qui présentaient un défaut de méthylation chez les mutants sont les régions dépourvues de méthylation héritées maternellement et donc nos régions candidates. Ces DMR hypométhylées peuvent être nommées DMR germinales (gDMR) puisque leur méthylation est établie dans l'ovocyte, mais on ne peut pas les caractériser d'ICR à ce stade, puisque leur étude fonctionnelle par délétion génétique n'a pas été encore réalisée.

Un nombre de 45 DMR candidates ont été identifiées par ce crible, parmi lesquelles ont été retrouvées les 17 ICR maternelles connues, validant l'efficacité de cette approche et son aptitude potentielle à mettre en évidence de nouvelles régions de contrôle de l'empreinte maternelle. Nous disposons ainsi de 28 régions DMR maternelles candidates, ce qui représente un petit nombre et gage de la spécificité du criblage (Introduction Figure 31). Au cours d'une première étude (Proudhon et al., 2012), quatre d'entre elles ont été validées, portant le nombre de DMR maternelles connues de 17 à 21. Cette première publication, à laquelle j'ai participé, a permis de mettre en avant l'existence de régions de contrôle de l'empreinte dont le statut de

méthylation peut être variable en fonction du tissu (DMR tissu-spécifique) ou du stade considéré (DMR transitoire), en opposition aux DMR/ICR ubiquitaires qui étaient précédemment connues, et dont le statut de méthylation allélique parental est globalement maintenu dans tous les tissus et tout au long de la vie. Une deuxième publication, à laquelle j'ai aussi contribué (Duffié et al., 2014), s'est ensuite plus précisément focalisée sur une DMR transitoire, associée au locus *Gpr1/Zdbf2*. Mes contributions à ces travaux sont résumées ci-dessous.

2.1- Notion d'empreinte génomique tissu-spécifique : le cas de *Cdh15*

J'ai participé à la caractérisation d'une DMR tissu-spécifique associée au gène de la M-cadhérine (*Cdh15*) (localisé en 8qE1), et qui coïncide avec un îlot CpG intragénique, entre les exons 9 et 12. Cette DMR maintient une méthylation spécifique de l'allèle maternel seulement dans certains tissus, dont l'hypothalamus et le muscle, et regagne de la méthylation sur l'allèle paternel conduisant à une perte d'empreinte parentale dans d'autres tissus, comme le cervelet et le foie. De manière intéressante, bien que la région associée soit conservée chez l'homme, le statut d'empreinte génomique ne l'est pas.

Mon travail a consisté à étudier la fonction de cette DMR intragénique. J'ai ainsi montré que la DMR de *Cdh15* contient un site d'initiation de la transcription d'un transcrit alternatif plus court, produit spécifiquement à partir de l'allèle paternel dans l'hypothalamus, du fait de la méthylation présente sur l'allèle maternel. Ce transcrit n'existe pas dans le cervelet, en relation avec l'état complètement méthylé de la DMR sur les deux allèles parentaux dans ce tissu. Le statut de méthylation soumis à empreinte de la DMR de *Cdh15* étant conservé dans les MEF (Mouse Embryonic Fibroblasts), j'ai pu étudier les marques chromatiniennes associées par des expériences de ChIP-qPCR. J'ai ainsi pu montrer que l'allèle paternel non méthylé contient exclusivement la marque H3K4me2, signe de son activité transcriptionnelle. L'allèle maternel en revanche est enrichi en H4K20me3, signe de répression de la transcription. De façon surprenante, la marque H3K9me2 s'est avérée présente à un niveau similaire sur les deux allèles contrairement aux ICR classiques où cette marque n'est présente que sur l'allèle réprimé. La présence de marques d'histones opposées de type H3K4me et H4K20me illustre bien l'activité opposée des deux allèles au niveau de la gDMR de *Cdh15*, mais la

marque H3K9me normalement associée à l'allèle réprimé, ne donne pas de distinction allélique à ce locus, tout au moins dans les MEF.

2.2- Notion d'empreinte génomique transitoire : le cas de *Gpr1/Zdbf2*

L'étude des DMR candidates des gènes *Zfp777* (6qB2.3) et *Zfp787* (7qA1) a permis de mettre en avant le concept d'empreinte génomique transitoire (Proudhon et al., 2012). Ces DMR présentent une méthylation maternelle spécifique de la fécondation jusqu'au stade blastocyste, mais regagnent de la méthylation sur l'allèle paternel lors de la vague de méthylation *de novo* du génome au moment de l'implantation de l'embryon. Le statut empreinté n'existe donc que pour quelques jours (4 chez la souris) pour ces gènes. Aucune évidence d'expression mono-allélique n'a été trouvée pour ces deux gènes, même pendant cette période de développement précoce, sûrement du fait de la position en 3'UTR de ces DMR. En revanche, comme expliqué ci-dessous, l'expression allélique du locus *Gpr1/Zdbf2* est régulée par une forme d'empreinte génomique transitoire.

Nous avons voulu explorer plus en détail le locus *Gpr1/Zdbf2* car une DMR candidate associée à cette région présentait une hyperméthylation dans les embryons mutants *Dnmt3L*^{-/+} comparés aux embryons sauvages : cette caractéristique est attendue pour une DMR secondaire sous le contrôle d'une gDMR/ICR maternelle (Proudhon et al., 2012). Ce locus avait précédemment été décrit comme soumis à empreinte paternelle (Hiura et al., 2010), or les ICR paternelles ne sont pas affectées par une perte de méthylation maternelle chez les embryons *Dnmt3L*^{-/+}. Ces contradictions nous ont intrigués et nous avons voulu comprendre ce qui se passait au niveau de ce locus en terme de dynamique de méthylation au cours du développement.

Le locus *Gpr1/Zdbf2* (1qC2) s'est avéré être un locus très complexe en terme d'empreinte génomique parentale, avec plusieurs régions de méthylation parentale différentielle dont le statut de méthylation varie au cours du développement : une DMR maternelle germinale transitoire qui coïncide avec le promoteur alternatif d'une forme longue du transcrit *Zdbf2* (*Zinc Finger DBF-Type Containing 2*) appelée *Liz*, et une DMR paternelle établie somatiquement (ou secondairement) en position intergénique, en amont du promoteur canonique de *Zdbf2*. J'ai analysé le statut de modifications

d'histones et leur enrichissement allélique à ces différentes DMR dans des cellules ES et dans un tissu différencié, le cerveau néo-natal. Cette analyse a révélé de manière surprenante que ce locus ne présente pas de combinaison classique de marques activatrices et répressives associée aux DMR/ICR, malgré une méthylation différentielle et une expression mono-allélique de *Liz* et *Zdbf2*.

Par contre, j'ai mis en évidence un bloc de la marque répressive H3K27me3 s'étendant de la DMR paternelle secondaire au promoteur canonique de *Zdbf2*. Cet enrichissement en H3K27me3 est conservé dans différents types cellulaires, comme le révèle l'analyse de données publiques de ChIP-Seq (Mikkelsen *et al.*, 2007 ; Marks *et al.*, 2012). Le promoteur de *Zdbf2*, qui est un îlot CpG, est enrichi en H3K27me3 sur l'allèle maternel comparé à l'allèle paternel (70% versus 30%), en concordance avec une expression paternelle spécifique de *Zdbf2*. Cependant, aucune distinction parentale n'a été observée au niveau de la DMR paternelle somatique pour cette marque, malgré le statut de méthylation génomique différentiel de cette région. En utilisant une technique de ChIP suivie de séquençage de la méthylation après traitement au bisulfite (ChIP-bisulfite), j'ai confirmé la présence de H3K27me3 sur les deux allèles parentaux, méthylés et non-méthylés de la DMR paternelle. La méthylation de l'ADN et de H3K27 sont donc présentes conjointement sur le même allèle. Ce résultat peut apparaître surprenant, car la méthylation de l'ADN et la méthylation de H3K27 sont considérées comme mutuellement exclusives (Tanay *et al.*, 2007). Cependant, des études récentes ont montré que cette exclusion était vraie au niveau des îlots CpG mais que ces deux marques répressives pouvaient coexister dans des régions pauvres en CpG (Brinkman *et al.*, 2012), comme c'est le cas pour la DMR somatique paternelle du locus *Gpr1/Zdbf2*.

La majorité des gènes soumis à empreinte chez la souris sont conservés chez l'Homme. Ainsi, la méthylation paternelle spécifique de la DMR secondaire du locus *GRP1/ZDBF2* et l'expression paternelle spécifique de *ZDBF2* ont été décrites précédemment chez l'Homme (Kobayashi *et al.*, 2009 ; Hiura *et al.*, 2010). Un transcrit (Aceview nehura.cAug10) était de plus prédit comme débutant dans une région riche en CpG, similaire au transcrit *Liz* chez la souris. J'ai pu montrer que le locus *GPR1/ZDBF2* est bien conservé chez l'humain au niveau de la dynamique de méthylation de l'ADN et de l'expression des différents transcrits de *LIZ*. Comme chez la souris, le locus *GPR1/ZDBF2* semble donc bien soumis à une empreinte transitoire maternelle relayée par une empreinte secondaire paternelle chez l'Homme.

Articles :

Proudhon C.*, Duffié R.*, Ajjan S., Cowley M., Iranzo J., Carbajosa G., Saadeh H., Holland M. L., Oakey R.J., Rakyan V.K., Schulz R. and Bourc'his D. (2012), Protection against *de novo* methylation is instrumental in maintaining parent-of-origin methylation inherited from the gametes. *Molecular Cell*, 47(6): 909-20.

*These authors contributed equally to this work

Duffié R., Ajjan S., Greenberg M.V., Zamudio N., Escamilla del Arenal M., Iranzo J., Okamoto I., Barboux S., Fauque D. and Bourc'his D. (2014). The *Gpr1/Zdbf2* locus provides new paradigms for transient and dynamic genomic imprinting in mammals. *Genes & Dev.* 28: 463-78.

Article 1

Protection against *de novo* methylation is instrumental
in maintaining parent-of-origin methylation
inherited from the gametes.

Proudhon C.*, Duffié R.*, Ajjan S., Cowley M., Iranzo J., Carbajosa G., Saadeh H., Holland M.
L., Oakey R.J., Rakyan V.K., Schulz R. and Bourc'his D.

*These authors contributed equally to this work

Protection against De Novo Methylation Is Instrumental in Maintaining Parent-of-Origin Methylation Inherited from the Gametes

Charlotte Proudhon,^{1,4} Rachel Duffié,^{1,4} Sophie Ajjan,¹ Michael Cowley,² Julian Iranzo,¹ Guillermo Carbajosa,³ Heba Saadeh,² Michelle L. Holland,³ Rebecca J. Oakey,² Vardhman K. Rakyan,³ Reiner Schulz,² and Déborah Bourc'h^{1,*}

¹INSERM U934/CNRS UMR3215, Institut Curie, 75248 Paris Cedex 05, France

²Department of Medical and Molecular Genetics, King's College London, London WC2R 2LS, England, UK

³Blizard Institute of Cell and Molecular Science, Queen Mary University of London, London E1 4NS, England, UK

⁴These authors contributed equally to this work

*Correspondence: deborah.bourchis@curie.fr

<http://dx.doi.org/10.1016/j.molcel.2012.07.010>

Open access under CC BY license.

SUMMARY

Identifying loci with parental differences in DNA methylation is key to unraveling parent-of-origin phenotypes. By conducting a MeDIP-Seq screen in maternal-methylation free postimplantation mouse embryos (*Dnmt3L*^{-/+}), we demonstrate that maternal-specific methylation exists very scarcely at midgestation. We reveal two forms of oocyte-specific methylation inheritance: limited to preimplantation, or with longer duration, *i.e.* maternally imprinted loci. Transient and imprinted maternal germline DMRs (gDMRs) are indistinguishable in gametes and preimplantation embryos, however, de novo methylation of paternal alleles at implantation delineates their fates and acts as a major leveling factor of parent-inherited differences. We characterize two new imprinted gDMRs, at the *Cdh15* and *AK008011* loci, with tissue-specific imprinting loss, again by paternal methylation gain. Protection against demethylation after fertilization has been emphasized as instrumental in maintaining parent-of-origin methylation inherited from the gametes. Here we provide evidence that protection against de novo methylation acts as an equal major pivot, at implantation and throughout life.

INTRODUCTION

Fertilization ensures the propagation of genetic and epigenetic information from one generation to the next. In mammals, epigenetic and long-lasting effects inherited in a parent-of-origin manner are known as genomic imprinting (Barlow, 2011). The main epigenetic mark that ensures their transmission and effects is DNA methylation. Methylation marks at imprinted loci are established in a sex-specific manner during gametogenesis, at genomic loci referred to as germline differentially methylated regions (gDMRs). After fertilization, these gDMRs act in *cis* to

control the monoallelic and parent-specific expression of a subset of genes, the imprinted genes. Germline DMRs can affect imprinted expression in a variety of ways, including promoter control of protein-coding and noncoding RNAs, regulation of transcription elongation, and long distance insulator activities. The allelic differences of gDMRs also include a typical chromatin signature, consisting of both repressive and permissive histone marks (McEwen and Ferguson-Smith, 2010). Disruption of imprinted expression upon genetic deletion of a gDMR is ultimate proof that it functions as an imprinting control region (ICR).

To date, around 120 imprinted genes have been identified in mouse and human. They are under the control of 20 identified gDMRs/ICRs, 17 of which are methylated in the oocyte (maternal gDMRs), and only three in sperm (paternal gDMRs). Parental ICRs are also sexually dimorphic in terms of CpG content and genomic localization: maternal ICRs are CpG island (CGI) promoters, while paternal ICRs are relatively CpG poor and intergenic. Evolutionary reasons for these discrepancies may be linked to the different developmental kinetics of male and female gametogenesis (Bourc'h and Bestor, 2006; Schulz et al., 2010). Importantly, maternal ICRs have a dominant role in early development, regulating biological pathways related to the establishment of the fetomaternal interface (Schulz et al., 2010).

It is becoming increasingly clear that the acquisition of sex-specific methylation patterns extends beyond imprinted regions in gametes (Kobayashi et al., 2012a; Smallwood et al., 2011; Smith et al., 2012). CpG islands are more prone to being methylated in the oocyte than in sperm and globally, about a thousand CGIs may be specifically methylated in the oocyte genome, exceeding the number of known maternal ICRs by far. Imprinted and nonimprinted methylation is likely to be established in the same way in the oocyte, under the control of the de novo DNA methyltransferase DNMT3A and its cofactor DNMT3L, and in a transcription-dependent manner (Bourc'h et al., 2001; Chotalia et al., 2009; Kaneda et al., 2004; Kobayashi et al., 2012a; Smallwood et al., 2011).

Instead of specific targeting mechanisms for their establishment in gametes, what truly distinguishes ICRs from the rest of the genome is their treatment after fertilization, in the wake of the extensive methylation changes inherent to early mammalian development. Before implantation, methylated alleles of ICRs

are resistant to the genome-wide erasure of gametic methylation that coincides with the acquisition of embryonic pluripotency. Specific *trans* acting factors have been identified as critical in maintaining ICR methylation during this period, such as binding of the KRAB (Krüppel-associated box-containing) zinc finger protein system, which involves Zfp57 and the heterochromatin inducer KAP1/TRIM28 (Li et al., 2008; Mackay et al., 2008; Quenneville et al., 2011; Zuo et al., 2012). Following this critical period, parental differences in ICR methylation are thought to persist throughout life, with no stage- and tissue-specificity, although this aspect of imprinting has not been fully addressed.

To gain insight into the extent of gametic methylation inheritance, we performed a genome-wide screen for gDMRs in the mouse postimplantation embryo. Considering their larger number, we specifically looked for maternally transmitted gDMRs, by comparative methylation profiling of wild-type embryos and embryos lacking oocyte-inherited methylation, using MeDIP-Seq (Methylated DNA Immunoprecipitation followed by high throughput sequencing). We exploited the well-characterized *Dnmt3L* mutant system, in which postimplantation *Dnmt3L*^{-/-} embryos generated by fertilization of *Dnmt3L*^{-/-} oocytes completely lack maternal imprints, while methylation patterns at paternal ICRs and repeats are unaltered (Bourc'his et al., 2001; Schulz et al., 2010).

Our approach proved to be highly sensitive and efficient at identifying regions of oocyte-inherited methylation. We found all 17 known maternal gDMRs/ICRs and less than thirty new candidates, revealing that loci that maintain maternal-specific methylation are rare in the postimplantation embryo. We characterized two new maternally imprinted gDMRs, located within the *Cdh15* gene and at the promoter of *AK008011*, a retroposed pseudogene. Further analyses enabled us to demonstrate that inherited maternal gDMRs can exist in a permanent or transient state after fertilization, and that avoidance of de novo methylation during embryo or tissue differentiation plays a key role in the permanency of parent-of-origin methylation inherited from the gametes.

RESULTS

A Genome-wide Screen for Regions of Oocyte-Inherited Methylation by MeDIP-Seq

We generated DNA methylation profiles by MeDIP-Seq for pools of 8.5 dpc (days post coitum) embryos of either a wild-type or *Dnmt3L*^{-/-} genotype. For each pool, two independent MeDIP preparations were sequenced. We obtained 41 M and 32.7 M distinct and uniquely alignable reads for wild-type and *Dnmt3L*^{-/-} embryos, respectively (Supplemental Information). For a 1000 bp sliding window size and a permissive false discovery rate (FDR) threshold of < 50%, 398 differentially methylated regions (DMRs) were identified genome-wide, 163 of which were hypomethylated and 235 hypermethylated in *Dnmt3L*^{-/-} embryos (Table S1). This suggests that overall, wild-type and *Dnmt3L*^{-/-} postimplantation embryos have similar methylation profiles and differ only at a small set of discrete loci.

The relatively small number of DMRs between wild-type and *Dnmt3L*^{-/-} embryos was not due to a lack of sensitivity of our MeDIP-Seq approach, as all of the 17 known maternal ICRs

were identified at an FDR threshold of only 5%. Furthermore, when ranked by confidence score ($-10\log_{10}$ of FDR), the top 13 ranks were occupied by known maternal ICRs (Figure 1A). The highest level of significance was obtained for the *Peg13* ICR (FDR < 10^{-27}), then for the *KvDMR* ICR (FDR < 10^{-25}), which regulates the promoter of the *Kcnq1ot1* noncoding RNA, and the lowest for the *Peg10/Sgce* ICR (FDR < 0.04) (Figures 1B, S1A, and S1B). Of note, the *Peg3* and *Snrpn* ICRs, which are prone to regaining methylation in some *Dnmt3L*^{-/-} progeny (Arnaud et al., 2006), were identified as highly significant hypomethylated DMRs (FDR < 0.02) (Figures S1C and S1D).

Unexpectedly, our screen uncovered hypermethylated DMRs, which gain methylation in *Dnmt3L*^{-/-} embryos. Among them, we found known secondary imprinted DMRs, which acquire methylation in somatic tissues as a consequence of maternal ICR control. For example, the *Gnas* locus contains two maternal ICRs, encompassing the *Gnas* *ex1A* (FDR < 10^{-18}) and the *Nespas* promoters (FDR < 10^{-20}) (Figure 1C). Paternal *Nespas* transcription induces the paternal methylation in *cis* of a secondary, somatically acquired DMR at the *Nesp* promoter. In the absence of maternal germline methylation, *Nespas* expression becomes biallelic and *Nesp* methylation occurs on both alleles (Liu et al., 2005). Our MeDIP-Seq approach unambiguously detected hypermethylation at the *Nesp* secondary DMR in *Dnmt3L*^{-/-} embryos (FDR < 0.02) (Figure 1C). Our MeDIP-Seq screen also proved to be highly specific: the three genetically confirmed paternal ICRs (*H19-Igf2*, *Dlk1-Gtl2* and *Rasgrf1*) that acquire methylation in the male germline were not identified as DMRs (Figure S2A). In addition, sequences that acquire methylation specifically in the embryo, such as CGI promoters of germline expressed genes (Borgel et al., 2010), also showed similar profiles between wild-type and *Dnmt3L*^{-/-} embryos (Figure S2B). In summary, evidence from known positive and negative controls demonstrates that our MeDIP-Seq screen accurately identified regions of oocyte-dependent methylation in the embryo.

To prioritize our candidate DMRs, we applied certain stringency filters, based on systematic genomic features of known maternal ICRs. Sequences that acquire methylation in oocytes tend to be CG-rich and among them, maternal ICRs have an observed to expected CpG ratio > 0.5 (Schulz et al., 2010). Given this fact and due to the functional link between CG density and DNA methylation-mediated transcriptional control (Weber et al., 2007), we excluded DMR candidates that contained fewer than 10 CpGs and that had a ratio of < 0.213 (the median across all identified DMRs). Additionally, we found that all known maternal ICRs and their associated secondary somatic DMRs have < 25% repeat sequence content. Given this relatively repeat-free nature and the difficulty of accurately assigning the genomic origin of short sequencing reads that originate from repeats, we further excluded candidates with a repeat content > 25%. These filters reduced the DMR number to 96, 47 hypomethylated and 49 hypermethylated, which showed a dispersed distribution throughout the mouse genome (Figure 1D).

Improved Definition of Known Imprinted Loci and Identification of New gDMRs

We first used our MeDIP-Seq data to improve the genetic map of germline or secondary DMRs for known imprinted loci that have

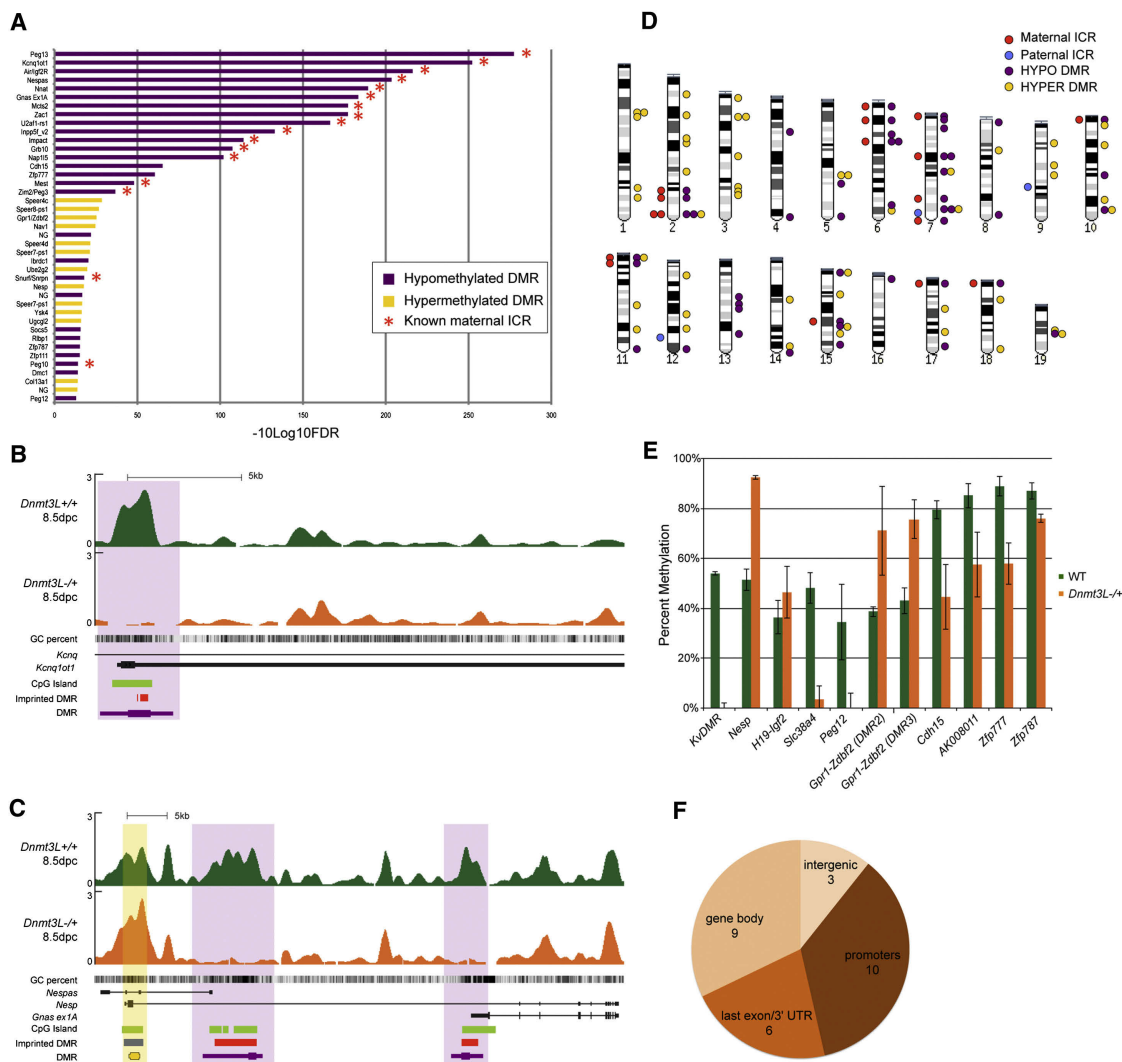


Figure 1. MeDIP-Seq Screen for the Identification of Oocyte-Dependent DMRs, Using 8.5 dpc WT and *Dnmt3L*^{-/-} Embryos

(A) DMRs with an FDR of up to 5% are ranked from top to bottom in order of statistical significance ($-\log_{10}(\text{FDR})$). The 17 known maternal ICRs are labeled with red asterisks. NG means "no gene," according to UCSC annotation.

(B) MeDIP-Seq profile of the *Kcnq1ot1* locus controlled by KvDMR, a known maternal ICR (red). The tracks depict the MeDIP-Seq profiles of 8.5 dpc WT embryos and *Dnmt3L*^{-/-} embryos, which are highly similar except for a the hypomethylated KvDMR (purple). Genes are oriented 5' to 3', and the y axis scale expresses the number of fragments per million mapped fragments.

(C) MeDIP-Seq profile of the *Gnas* locus, controlled by two known maternal ICRs, which are hypomethylated (purple). This locus also contains a secondary somatic DMR, hypermethylated (yellow) in *Dnmt3L*^{-/-} embryos.

(D) Mouse karyotype with the positions of 47 hypomethylated and 48 hypermethylated candidate DMRs in *Dnmt3L*^{-/-} embryos, and the previously known ICRs.

(E) MSRE-qPCR validation of methylation. Error bars show the standard deviation from three independent digestions.

(F) Transcript position of 28 hypomethylated DMRs, which represent new potential maternal gDMRs.

not been fully documented. Two hypomethylated DMRs coincided with the promoters of the *Slc38a4* (FDR < 0.5) and *Peg12* (FDR < 0.05) genes (Figures S1E and S1F), and were confirmed to be hypomethylated in *Dnmt3L*^{-/-} embryos by MSRE-qPCR assays (Methylation-Sensitive Restriction Enzyme coupled with quantitative PCR) (Figure 1E). Paternal-specific expression and maternal-specific methylation had been previously reported at these loci in somatic tissues (Kobayashi et al., 2002; Smith et al., 2003). Reduced-Representation

Bisulfite Sequencing (RRBS) confirmed here that the *Slc38a4* DMR is indeed methylated in oocyte and hypomethylated in sperm and can be categorized as a genuine maternal gDMR (Figure S3A). However, the *Peg12* DMR was unmethylated in both oocyte and sperm, suggesting that it is not a gDMR. Our screen also led to a reassessment of the imprinted *Gpr1-Zdbf2* locus, which was originally characterized as a fourth imprinted region controlled by paternal methylation (Hiura et al., 2010; Kobayashi et al., 2009). We identified two hypermethylated

Table 1. Maternal gDMR Candidates: Hypomethylated, Single-Copy, and Relatively CpG-Rich DMRs

Chr	Start	End	Closest Transcript	CGI	CpG Content	Sperm Methylation ^a	Liver Methylation ^b
8	125387861	125390344	Cdh15	CGI21235	0.37	−5.1	0.1334
6	47974007	47975979	Zfp777	CGI17361	0.56	−5.3	0.6911
13	54209856	54211139	Sfxn1		0.25	−0.9	<i>0.5000</i>
17	87524084	87525819	Socs5	CGI9617	0.43	−3.4	0.0020
7	86519538	86521097	Rlbp1		0.23	−0.9	0.1334
7	6083480	6084890	Zfp787	CGI18282	0.64	−4.7	0.6964
7	24992450	24993377	Zfp111	CGI18528	0.45	−2.7	<i>0.7500</i>
15	76010966	76012080	Plec1	CGI7215	0.68	−1.5	<i>1.0000</i>
11	102057005	102057998	Hdac5		0.30	−1.2	<i>0.0078</i>
13	60557950	60559042			0.24	−1.1	<i>1.0000</i>
16	20530221	20531293	Dvl3	CGI8073	0.27	−1.9	<i>0.3438</i>
4	150993001	150994022	Camta1		0.33	−2.6	<i>1.0000</i>
4	53727006	53728024	Fcmd	CGI14304	0.89	−3.5	<i>0.8750</i>
15	102047274	102048271	Itgb7		0.26	−1.3	<i>0.7500</i>
5	106629403	106630408	nenese		0.66	−4.7	<i>0.3125</i>
10	122303419	122304463	Ppm1h		0.26	−1.4	0.0384
7	148034494	148035458	Odf3		0.26	−1.0	<i>1.0000</i>
7	147267611	147268611	Drd1ip	CGI19894	0.34	−2.2	<i>1.0000</i>
15	11250512	11251415	Adamts12		0.29	−1.4	<i>1.0000</i>
13	66815007	66816022	2410141K09Rik		0.57	−2.9	<i>1.0000</i>
14	122056331	122057326	Dock9		0.32	−0.8	<i>1.0000</i>
8	12262778	12263827			0.21	−1.1	<i>1.0000</i>
12	118489501	118490705	Ptpn2		0.30	1.3	0.5000
19	45385459	45386455	sneefar		0.27	−0.8	<i>0.7734</i>
6	125660898	125661895	Tmem16b	CGI18014	0.28	−1.0	<i>1.0000</i>
11	115748842	115749776	Myo15b		0.22	−1.2	0.5000
10	74869979	74870980	Upb1		0.28	−2.2	<i>1.0000</i>
13	96588183	96589194	Iqgap2		0.27	−0.8	0.1938

Information shown: genomic coordinates of the DMR as determined by USeq, closest transcript (RefSeq), CGI reference number (Illingworth et al., 2010), observed/expected CpG ratio, methylation status in sperm, and evidence for maternal methylation in hybrid fetal liver.

^alog₂ of fold-change relative to wildtype embryos: negative values are indicative of sperm hypomethylation.

^bone-tailed binomial test p values: entries with p > 0.2 are indicative of low degree of evidence for maternal-specific methylation; italics highlight entries where the test was underpowered.

DMRs (FDR < 0.01 and 0.07), which overlap with two originally described paternal gDMRs (DMR2 and DMR3) (Figures 1A and S1G). However, as discussed earlier, paternal gDMRs, such as the *H19-Igf2* DMR, are unaltered in our screen. The methylation status of the *Gpr1-Zdbf2* DMRs in *Dnmt3L*^{−/−} embryos is instead reminiscent of a secondary, somatic DMR similar to the *Nesp* DMR, as validated by MSRE-qPCR (Figure 1E) and bisulfite sequencing (data not shown), and was recently independently confirmed (Kobayashi et al., 2012b).

In our search for additional novel maternal gDMRs, we focused on 28 hypomethylated, single-copy and relatively CpG-rich candidate DMRs (Table 1), which importantly do not belong to known imprinted regions. All but three of the candidate gDMRs were within a transcription unit, among which ten were located to an annotated promoter, six overlapped with the last exon/3′ UTR, and nine were within a gene body (Table S1 and Figure 1F). Contrary to known maternal ICRs, all of which coin-

cide with promoter-associated CGIs, only nine out of 28 candidates overlapped with a CGI and only five of those were associated with an annotated promoter. To evaluate which candidate gDMRs may constitute regions of bona fide oocyte-specific methylation, we interrogated publicly available CGI methylation data in the mouse oocyte (Kobayashi et al., 2012a): all but two CGI candidates from our screen were found methylated. We further integrated MeDIP-Seq data from mouse C57Bl6/J sperm (Table 1), and found only one of the candidates to be methylated in sperm, confirming that sequences methylated in the oocyte are usually not methylated in sperm (Kobayashi et al., 2012a; Smallwood et al., 2011; Smith et al., 2012). As an indication of long-term maintenance of maternal-specific methylation, we integrated MeDIP-Seq data from fetal 17.5 dpc hybrid mouse liver from C57Bl6/J and PWD/PhJ strain crosses. In these samples, parental allele-specific sequencing reads were counted at known SNPs between the parental strains

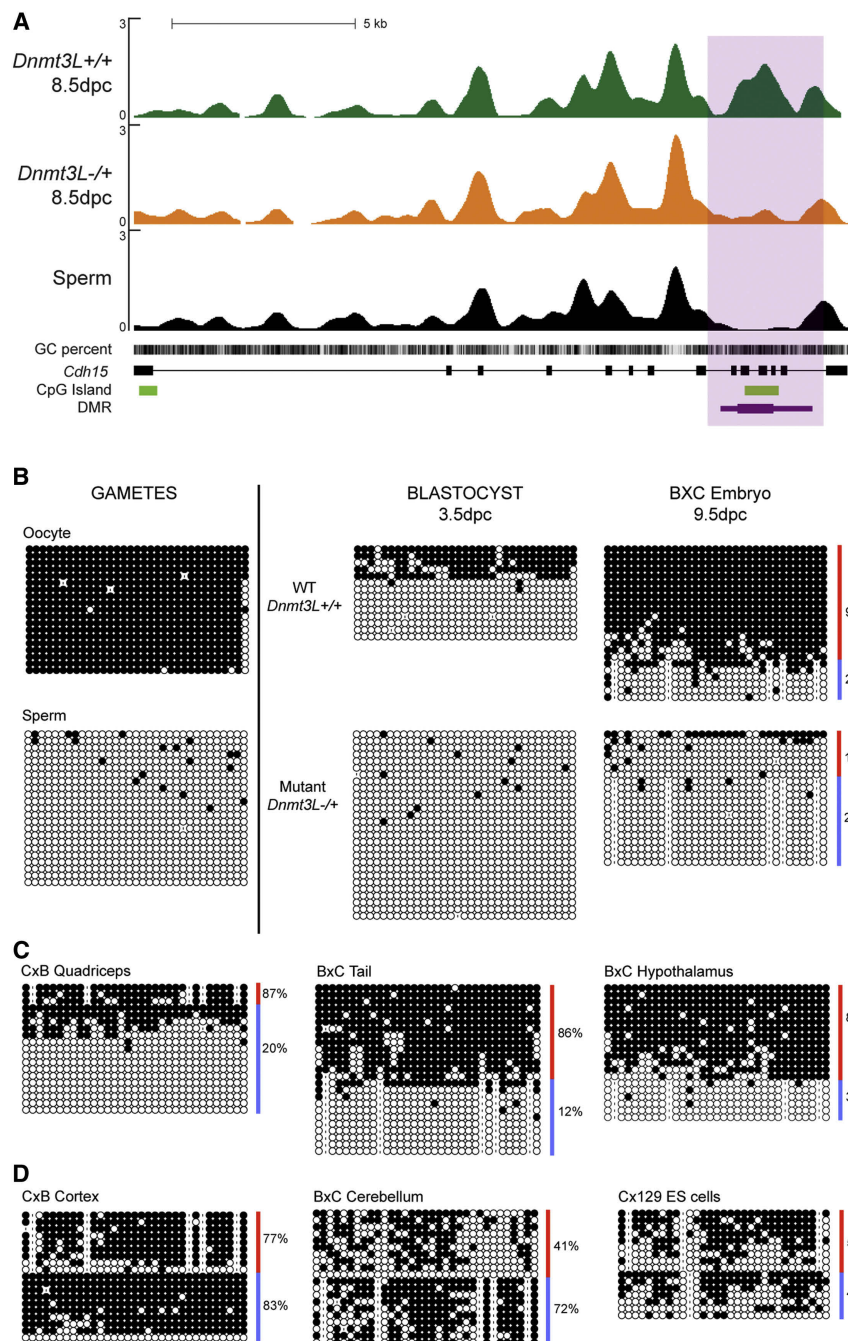


Figure 2. Identification and Methylation Analysis of the *Cdh15* DMR

(A) MeDIP-Seq profile of the *Cdh15* locus. Genes are oriented 5' to 3'. The MeDIP-Seq tracks show an intragenic CGI with hypomethylation in *Dnmt3L*^{-/-} embryos and sperm compared to WT embryos (thick part of purple bar: highest statistical confidence).

(B) Developmental analysis of *Cdh15* DMR methylation by bisulfite sequencing. Red and blue lines delineate maternal and paternal alleles.

(C) Maternal-specific methylation is maintained in various hybrid adult tissues.

(D) The parental specificity of the DMR is lost in cortex, cerebellum and ES cells, by methylation acquisition on paternal alleles (blue). Black circle: methylated CpG, white circle: unmethylated CpG, dash: absent CpG corresponding to strain-specific SNPs or, rarely, sequencing errors. Mouse strains: B = C57Bl6/J, C = CAST/Ei, 129 = 129 Sv.

We went on to study four candidate maternal gDMRs in more detail, chosen for their high level of significance in our screen and their association with a CGI, a systematic feature of currently known maternal ICRs. Importantly, MSRE-qPCR and bisulfite sequencing confirmed their hypomethylation in 8.5 dpc *Dnmt3L*^{-/-} embryos and in sperm (Figure 1E and data not shown). Three are located in gene bodies toward the 3' end of the respective canonical RefSeq transcript (*Cdh15*, *Zfp777* and *Zfp787*). The fourth candidate overlaps with the promoter of *AK008011*, a mono-exonic retrogene.

The *Cdh15* DMR Controls the Paternal- and Tissue-Specific Expression of an Intragenic Transcript

The *Cdh15* DMR (ranked 14th; FDR < 0.02) spans exons 10 to 12 of the *Cdh15* gene (Figures 1A and 2A), which maps to distal chromosome 8 (8qE2) and encodes the M-cadherin protein, a cell-adhesion protein linked to muscle and

(Supplemental Information). For six candidate gDMRs, we found evidence for maternal-specific methylation maintenance (binomial $p < 0.2$) (Table 1). Of note, the lack of evidence for maternal-specific methylation from liver was mostly due to a paucity of SNPs (7/28 DMRs) or low read depth over an existing SNP (14/28 DMRs). Using bisulfite sequencing, we further assessed four of these candidates. None of these showed maternal-specific methylation in fetal liver (Figure S4), confirming that very few new maternal gDMRs persisting after implantation are left to be uncovered.

cerebellum (Padilla et al., 1998; Rose et al., 1995). By bisulfite sequencing of exon 11, we showed that this DMR fulfills the three developmental criteria of a maternally imprinted gDMR (Figures 2B and S3B): (1) methylation acquisition in the oocyte but not in sperm, (2) maintenance of maternally methylated alleles prior to implantation, as revealed by the lack of methylated alleles in maternal-imprint free *Dnmt3L*^{-/-} blastocysts compared to wild-type blastocysts, and (3) protection of the paternally unmethylated alleles after implantation, as shown in 9.5 dpc embryos derived from C57Bl6/J and CAST/Ei strains. Moreover, we

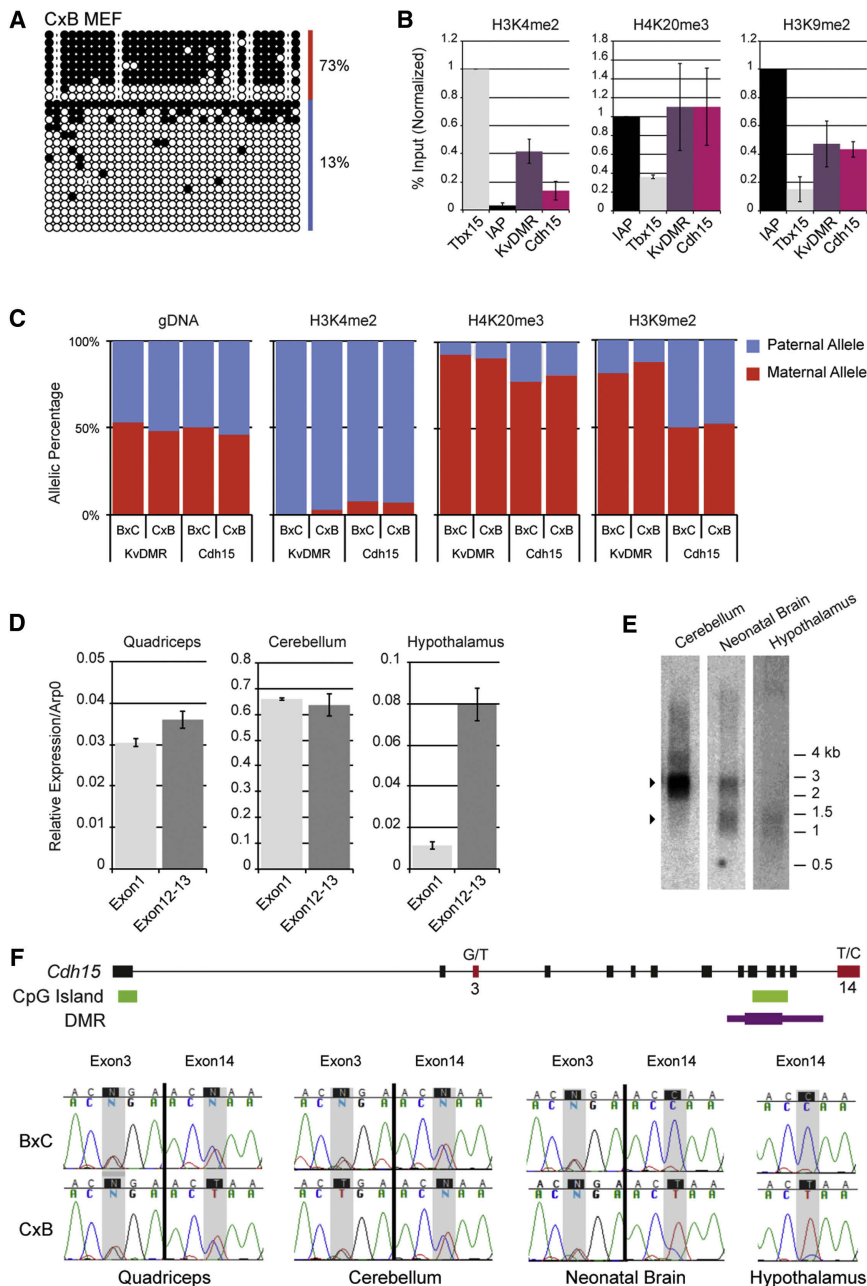


Figure 3. Chromatin and Expression Analysis at the *Cdh15* Locus

(A and B) In MEFs, the *Cdh15* DMR (A) globally maintains maternal-specific DNA methylation, (B) shows combined enrichment of permissive and repressive histone marks. Relative enrichments were calculated from ChIP-qPCR experiments as Input %, and normalized to *Tbx15* promoter for H3K4me2, and to IAP 5'LTR for H4K20me3 and H3K9me2. Error bars show the standard deviation from three biological replicates.

(C) Permissive and repressive marks show opposite allelic enrichment by ChIP-pyrosequencing, on reciprocal BxC and CxB MEFs. Genomic DNA (gDNA) was used to exclude assay-specific biases.

(D) RT-qPCR assay shows equal measurement of expression upstream (exon 1) and downstream (exons 12–13) of the DMR in quadriceps and cerebellum. A 10-fold higher expression is detected downstream in the hypothalamus.

(E) Northern blot analysis identifies a full 3 kb transcript in cerebellum, a shorter 1–1.5 kb transcript in adult hypothalamus, and both forms in neonatal brain.

(F) RT-PCR sequencing tracks of the allelic expression status of the main and short *Cdh15* transcripts. SNP nucleotides (red) are indicated in the B then C order.

paternal alleles (binomial $p = 0.133$) (Table 1). However, in ES cells, adult cortex and cerebellum, parental specificity was lost due to acquisition of methylation on the paternal alleles (Figure 2D). The intragenic *Cdh15* DMR is therefore conserved during adulthood, but in a tissue-specific manner.

We next investigated the chromatin state of the *Cdh15* DMR by immunoprecipitation (ChIP). We measured the quantity and allelic specificity of three marks associated with imprinted DMRs (H3K4me2, H3K9me2 and H4K20me3), in MEFs (mouse embryonic fibroblasts), which globally maintain maternal-specific DNA methylation (Figure 3A). The *Cdh15* DMR was found enriched in repressive

demonstrated that the maternal allele is unable to regain methylation in *Dnmt3L*^{+/+} postimplantation embryos, confirming the obligate passage through the female germline to imprint this locus.

In adult tissues and cells derived from C57Bl6/J by CAST/Ei or 129 Sv by CAST/Ei crosses, the methylated status of maternal alleles was consistently maintained. In quadriceps, tail and hypothalamus, methylation differences between parental alleles were highly significant (Fisher's exact $p < 10^{-11}$), although a minority of paternal alleles tended to regain methylation in quadriceps (Figure 2C). MeDIPSeq analysis of fetal hybrid liver DNA also showed higher methylation of maternal compared to

H3K9me2 and H4K20me3, at a level similar to the typical maternal ICR KvDMR (Figure 3B). Permissive H3K4me2 marks were also found at this locus. We assayed the allele-specificity of these marks by pyrosequencing, exploiting SNPs between the C57Bl6/J and CAST/Ei strains (Figure 3C). H3K4me2 was associated with the paternal allele, while H4K20me3 was enriched on the maternal allele. In contrast to KvDMR, for which H3K9me2 is maternally enriched, this mark was equally distributed on both parental alleles at the *Cdh15* DMR. This shows that the *Cdh15* DMR harbors opposite allelic states of histone modifications in MEFs, with respect to H3K4 and H4K20 methylation.

Cdh15 is highly expressed in satellite cells of skeletal muscles and granular cells of the cerebellum (Cornelison and Wold, 1997; Rose et al., 1995). Moreover, evidence for paternal-specific expression was recently reported in adult hypothalamus (Gregg et al., 2010). By using quantitative RT-PCR, we readily detected *Cdh15* expression in quadriceps, cerebellum and hypothalamus, with the strongest detection in cerebellum (Figure 3D). Expression measurements were equal upstream and downstream of the DMR (exon 1 versus exons 12–13) for the quadriceps and the cerebellum, suggesting the existence of a transcript elongating from the 5' canonical promoter throughout the coding unit. However, in the hypothalamus, 10-fold greater expression was measured downstream of the DMR, suggesting the existence of a transcript originating intragenically. Northern blot analysis confirmed the production of a single full-length transcript in the cerebellum, around the expected 3 kb size, while the hypothalamus specifically expressed a shorter version of *Cdh15*, which could be detected with a probe spanning exons 9–14 (Figure 3E), but not exons 5–9 (data not shown). Both transcripts were present in neonatal brains.

The allelic status of *Cdh15* expression was determined in reciprocal BxC and CxB crosses: while biallelic expression was found upstream and downstream of the DMR in quadriceps and cerebellum, only paternal expression of the short *Cdh15* transcript was detected in hypothalamus (Figure 3F). In neonatal brains, a switch from biallelic to monoallelic expression was observed at the DMR. Further allelic mapping by RT-PCR revealed that the short paternal transcript originates between exons 9 and 10, 5' of the *Cdh15* DMR, a region that showed maternal-specific methylation in neonatal brain (Figures S5A and S5B). The *Cdh15* DMR probably corresponds to an intragenic promoter, specifically active in brain cell-types detectable at birth and in the hypothalamus at adulthood. Its differential methylation correlates with differential allelic transcription in these cell types.

The *Cdh15* DMR defines chromosome 8 as a new chromosome harboring an imprinted locus. To determine the extent of *Cdh15* DMR control, we measured the allelic expression of the three closest neighboring genes (*Acsf3*, *AK040202* and *Ankrd11*), by an RT-PCR pyrosequencing-based approach. We did not detect imprinted expression for these genes, in a bank of reciprocal hybrid tissues including embryonic, fetal, neonate and adult stages (Figure S6). It is therefore highly probable that the *Cdh15* DMR does not control an imprinted cluster.

Of clinical interest, the human *CDH15* gene has been associated with intellectual disability (Bhalla et al., 2008). The *CDH15* gene has a similar genomic organization to its mouse homolog, notably with an intragenic CGI spanning exons 9 to 12. We analyzed the imprinted status of this locus in human fetal liver, a tissue we find to maintain maternal-specific methylation in mouse. Unexpectedly, the 5' part of the CGI was completely methylated, while the 3' part that includes the region homologous to the sequence we analyzed in mouse was completely unmethylated (Figure S5C). Lack of methylation was confirmed in lymphocyte and placental DNA (data not shown). Our study does not support a conservation of imprinting for the *CDH15* locus in humans but rather points to a bipartite CGI.

The *AK008011* DMR Is a Tissue-Specific Imprinted gDMR at a Mouse Pseudogene

The second DMR we focused on maps to a CGI located 5' of *AK008011*, an intronless gene (Figure 4A). It was likely generated via the retrotransposition of a *Coro1c* mRNA (Coronin, Actin binding protein 1c located on chromosome 5) to a region 1.5 kb downstream of *Nhlrc1* on chromosome 13qA5, an event that occurred specifically in the mouse lineage (Kent et al., 2003). We uncovered this small DMR through a 500 bp sliding window analysis (FDR < 0.35) (Table S1), while no DMR was identified at the *Coro1c* locus. Bisulfite-based methylation analysis revealed 1) methylation acquisition in oocyte but not in sperm, 2) protection of maternally methylated alleles prior to implantation and 3) protection of paternally unmethylated alleles after implantation (Figure 4B). This locus therefore behaves as a typical maternal imprinted gDMR during the critical window around fertilization and implantation. However, in adult life, this gDMR becomes tissue-specific. While maternal-specific methylation is properly maintained in tail and fibroblasts (Figure 4C), the quadriceps, cortex and liver show dense methylation (over 60%) of both paternal and maternal alleles (Figure 4D). This finding again questions the view of the permanency of imprinted gDMRs throughout life.

In an attempt to investigate the impact of this gDMR on allelic expression, we designed primers that specifically distinguish *AK008011* from *Coro1c* mRNA. However, we could not detect expression in tissues where the gDMR is conserved. The high rate of nucleotide divergence between mouse strains suggests a low selective pressure on this gene, which may be a pseudogene: 40 SNPs are referenced at *AK008011*, including 9 nonsynonymous ones, versus 6 synonymous changes at the transcribed region of *Coro1c* (MGI and dbSNP build 128). Further examination of the closest gene, *Nhlrc1*, did not reveal a bias in parental expression in any tissue from our hybrid bank (data not shown), suggesting that the *AK008011* gDMR does not have long-range imprinting effects.

Zfp777 and *Zfp787* DMRs Are Transient Maternal gDMRs

The last two hypomethylated DMRs we validated (FDR < 2%) map to CGIs located in the last exon of the *Zfp777* (6qB2.3) and *Zfp787* (7qA1) genes, which encode zinc finger proteins (Figures 5A and S7A). As is typical for maternal gDMRs, we found methylation acquisition specifically in the oocyte, and protection of maternally methylated alleles prior to implantation (Figures 5B, S3B, and S7B). However, paternal alleles of these DMRs undergo de novo methylation at implantation, so that both parental alleles displayed similar levels of methylation at 9.5 dpc as well as in 17.5 dpc fetal liver (Table 1). Contrary to imprinting-associated gDMRs, which show lifelong parental differences at least in some tissues, these DMRs may be categorized as transient gDMRs. Interestingly, in *Dnmt3L*^{-/-} embryos, paternal and maternal alleles were equally methylated at 9.5 dpc, suggesting no differential treatment of the two alleles (Figures 5B and S7B). The global methylation per parental allele was slightly lower than age-matched wild-type embryos, likely as a consequence of a postimplantation developmental delay (Bourc'his et al., 2001). Examination of various normal adult

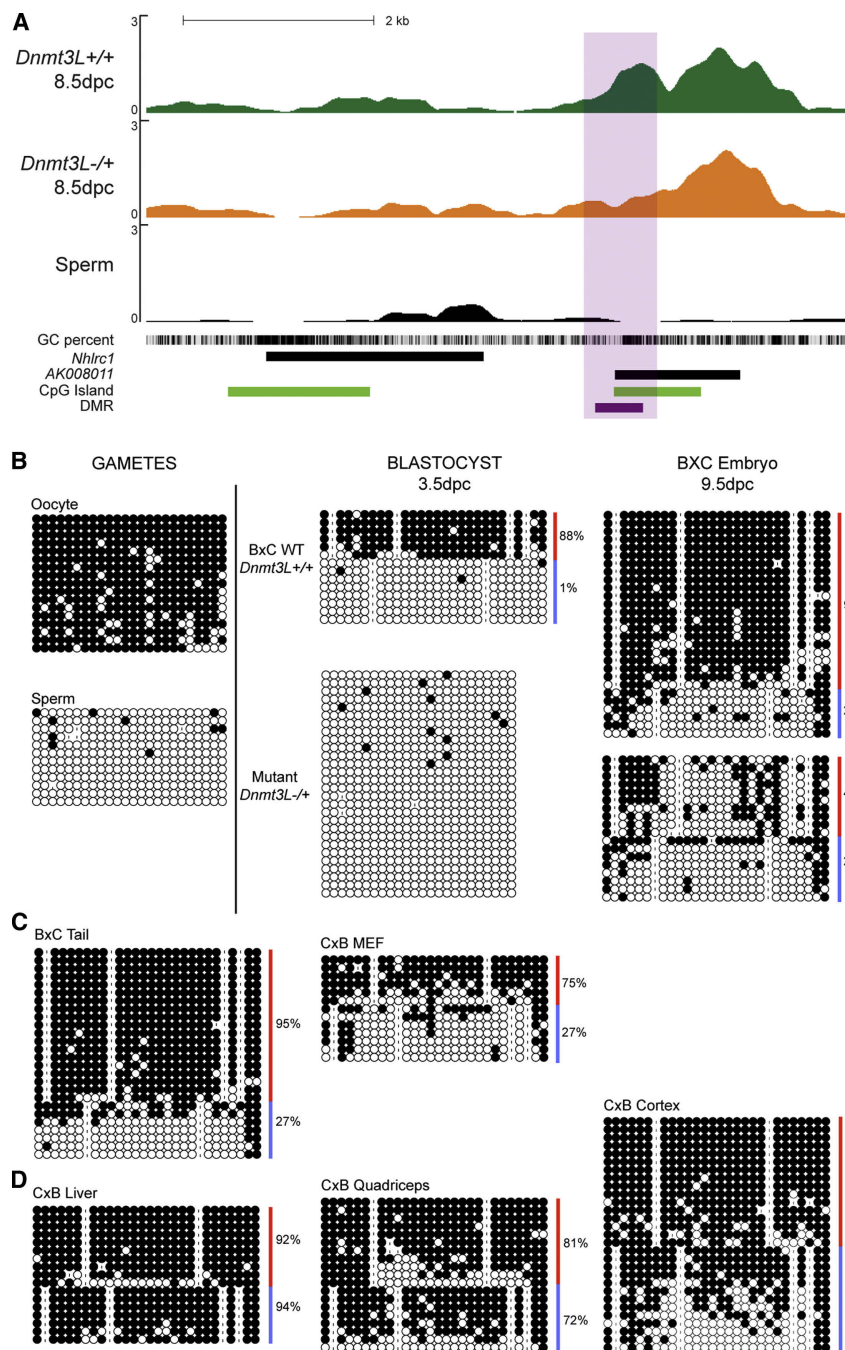


Figure 4. Identification and Methylation Analysis of the AK008011 DMR

(A) MeDIP-Seq profile of the AK008011 locus. (B) Developmental analysis of DNA methylation of this locus by bisulfite sequencing. (C) Maternal-specific methylation is maintained in tail and MEFs. (D) The parental specificity is lost in liver, quadriceps and cortex, by methylation acquisition on paternal alleles (blue).

lated (Figures 5D and S7C). This was observed in ES cells that were isolated and grown in conditions optimal for “ground-state” pluripotency cells (2i medium) (Nichols et al., 2009), and in ES cells cultured in classical medium (data not shown). Similar to previous findings (Borgel et al., 2010; Dean et al., 1998), our results demonstrate that ES cells do not necessarily maintain the allelic status of sequences that are differentially methylated in the blastocyst, and specifically, may not be a suitable cellular model for studying transient gDMRs.

When allelic expression patterns were measured in hybrid reciprocal tissues, we found no parental bias in *Zfp777* and *Zfp787* expression, even in the preimplantation blastocyst, where parental methylation differences still exist (Figures 5E and S7D). Our results suggest that the methylation located in the 3' end of the *Zfp777* and *Zfp787* genes may not functionally impinge on their expression.

DISCUSSION

In mammals, the oocyte and sperm genomes harbor distinct methylation patterns, as a result of different kinetics and constraints exerted on gamete production in the two sexes. The inheritance of parent-specific methylation at fertilization provides the opportunity for differential allelic regulation in the progeny, with genomic imprinting as the

most durable form of parent-specific regulation of gene expression. Our present work demonstrates that the total number of maternal germline DMRs persisting throughout development and adulthood is very limited, in line with current estimates for the number of known ICRs. From this study, it can be concluded that genomic imprinting is an unusual form of regulation in mammals.

Recent genome-wide studies have highlighted preimplantation demethylation as a major determinant of gametic methylation clearance (Borgel et al., 2010; Kobayashi et al., 2012a;

tissues by bisulfite sequencing confirmed complete methylation later in life (Figures 5C and S7C). Further analysis of the last exon of *ZNF777* in human postimplantation tissues confirmed the existence of methylated alleles only (data not shown).

The observation of parent-specific marks at the blastocyst stage prompted us to analyze the allelic methylation of these transient gDMRs in ES cells. Bisulfite analysis showed that ES cells do not reproduce the parental methylation differences of their biological progenitors; the *Zfp777* DMR was biallelically methylated, while the *Zfp787* DMR was biallelically hypomethy-

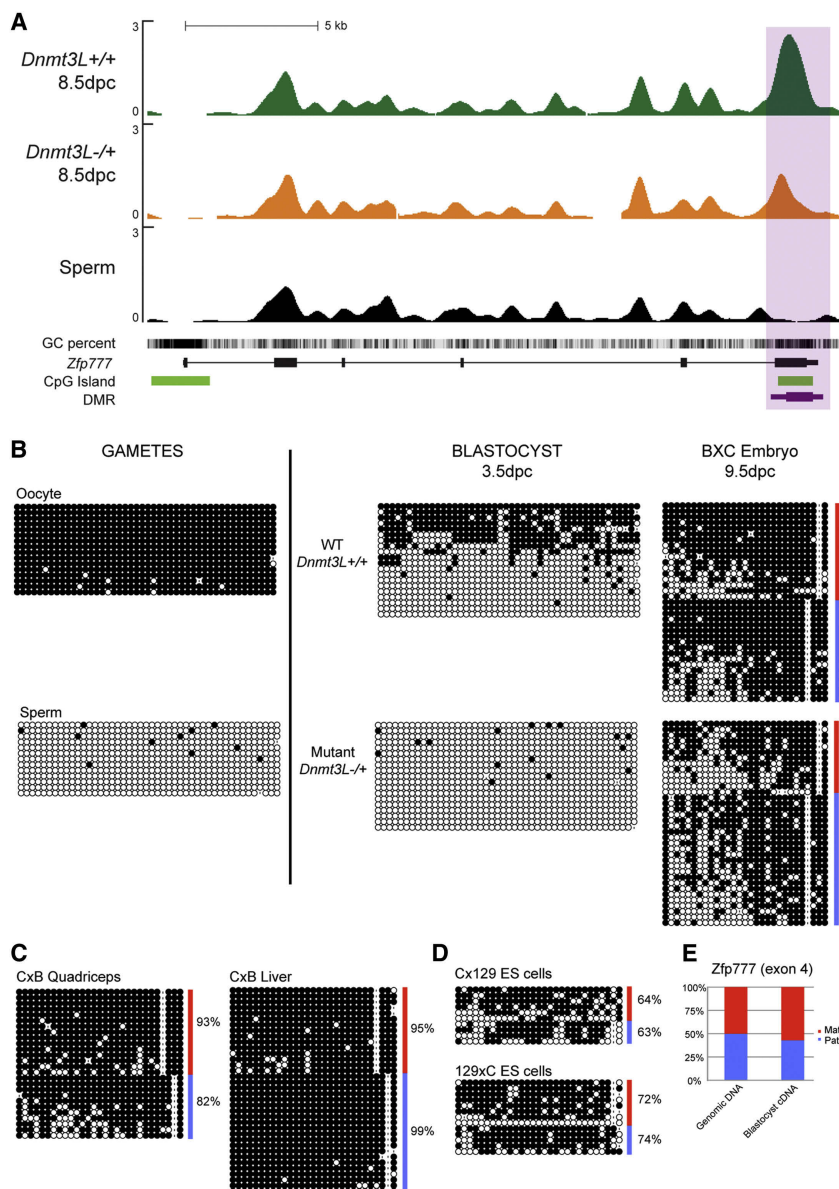


Figure 5. Identification and Methylation Analysis of the *Zfp777* DMR

(A) MeDIP-Seq profile of the *Zfp777* locus. (B) Contrary to imprinted gDMRs, DNA methylation is gained on paternal alleles at implantation and parental alleles exhibit similar methylation levels both in WT and *Dnmt3L*^{-/-} 9.5 dpc embryos. (C) Adult tissues show a fully methylated pattern. (D) Parental alleles are similarly methylated in ES cells. (E) RT-PCR pyrosequencing analysis of a SNP located in the 3' UTR shows biallelic expression of *Zfp777* in hybrid blastocysts at 3.5 dpc.

with our unfiltered candidate list (*AK008011* and *Casc1*). Interestingly, loss of parent-specific marks by de novo methylation may not be restricted to the new imprinted loci we describe: indeed, at traditionally known ICRs, loss of differential methylation has been sporadically reported in normal adult cells, occurring by methylation gain, rather than loss (Fang et al., 2012; Ferrón et al., 2011).

Our work increases the number of known imprinted gDMRs to 23, including two new loci to be referenced. While the *Cdh15* DMR is associated with parent-specific expression, the *AK008011* DMR may not be functional, showing that imprinted gDMRs may not necessarily be selected for a role in gene regulation. Previous studies had alluded to a possible imprinted status of *Cdh15*. Analysis of chromosome 8 duplications led to the characterization of a region of complete maternal methylation and intermediate paternal methylation in embryos and neonates (Kelsey et al., 1999). While no parent-specific *Cdh15* expression was found in embryos, a recent analysis reported paternal-specific expression of

Smallwood et al., 2011). We reveal here that de novo methylation plays an equally important role in leveling parental methylation differences inherited from the gametes (Figure 6). *Zfp777* and *Zfp787* DMRs lose their maternal specificity early, by paternal methylation acquisition at implantation. *Cdh15* and *AK008011* DMRs are protected at implantation, but nevertheless gain paternal methylation later, during tissue differentiation. The permanency and universality of imprinted gDMRs was a commonly held notion in genomic imprinting. The tissue-specificity of *Cdh15* and *AK008011* gDMRs revisits this notion and highlights the limitation of studies performed on a specific adult tissue for identifying new imprinted gDMRs. In this regard, a recent genome-wide screen performed on adult mouse cortex uncovered nine candidate regions of parent-of-origin methylation (Xie et al., 2012), of which only two candidates overlapped

three SNPs confined to exons 12 to 14, in adult hypothalamus (Gregg et al., 2010). Our study resolves the *modus operandi* of this locus, by the identification of a maternal gDMR that maps to *Cdh15* exons 10–12, which is maintained in a tissue-specific manner and controls the paternal expression of a short alternative transcript in neonatal brain and adult hypothalamus. The *Cdh15* DMR may be a docking site for transcription factors expressed in specific brain cell types, whose binding/activity is impaired by maternal DNA methylation.

Cdh15/M-cadherin is an adhesion protein that mediates cell-to-cell interactions. Homozygous *Cdh15* null mice are viable, and show no apparent defects in skeletal muscle and cerebellum, likely due to compensation from other cadherins (Hollnagel et al., 2002). Moreover, there is no evidence of parent-of-origin effects in these mutant mice. However, the

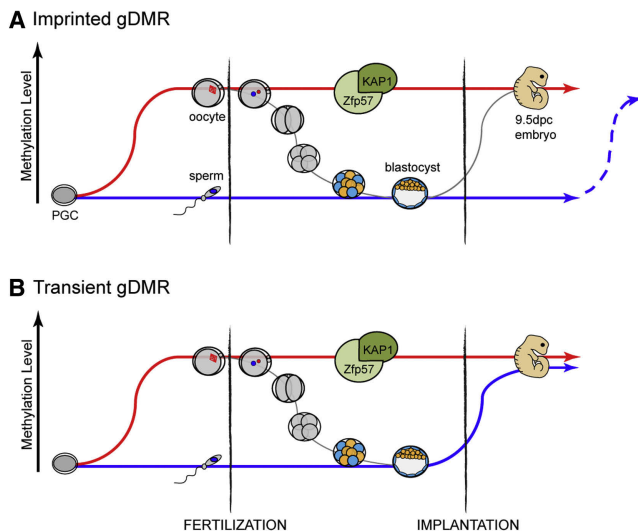


Figure 6. Different Fates of Oocyte-Inherited Methylation

(A and B) Maternal alleles (red line) of (A) imprinted and (B) transient gDMRs acquire methylation in oocytes and are protected against genome-wide (gray line) demethylation during preimplantation development. These two types of gDMRs contain Zfp57 binding sites and physically interact with Zfp57/KAP1. However, while unmethylated paternal alleles (blue line) of imprinted gDMRs are protected against de novo methylation at implantation, transient gDMRs are permissive to this process. Imprinted gDMRs can also gain methylation on paternal alleles later during life, in a tissue-specific manner (dotted blue line).

corresponding deletion targets exons 1 to 4, and therefore, should not impair the production of the short imprinted *Cdh15* transcript. Interestingly, a similar 5' truncated form of cadherin with altered adhesion activity has been described in specific neurons of the chick embryo (Shirabe et al., 2005). Provided that the short imprinted *Cdh15* transcript is translated, it may likewise exert specific functions in mammalian hypothalamic cells, related to cell communication, polarization and shaping.

By identifying both imprinted and transient gDMRs, our screen highlights that these two types of gDMRs are indistinguishable in gametes and early embryos (Figure 6). Recruitment of KAP1 through Zfp57 binding was shown to be required for the maintenance of methylated alleles of ICRs (Li et al., 2008; Mackay et al., 2008; Quenneville et al., 2011; Zuo et al., 2012). By in silico analysis we found that all the maternal gDMRs we validated (*Slc38a4*, *Cdh15*, *AK008011*, *Zfp777* and *Zfp787*) contain some hexanucleotide motifs for Zfp57 binding (Table S1). Moreover, Zfp57 and KAP1 are enriched at these sites in published ES cell ChIP-Seq data (Table S1) (Quenneville et al., 2011). Our study therefore shows that the presence of Zfp57 motifs cannot be used as a hallmark of genomic imprinting, as it is also found at transient gDMRs. However, it is likely to specify all genomic sequences that maintain gametic methylation during preimplantation development. Interestingly, four Zfp57 binding motifs exist at the intragenic CGI of the human *CDH15* locus. While we found no evidence of imprinting, we cannot exclude that this locus is a true maternal gDMR in human, either transient or tissue-specific.

Methylation gain at implantation is what discriminates transient from lifelong imprinted gDMRs. The former are permissive,

while the latter are refractory to this process. *Zfp777*, *Zfp787*, and *Cdh15* DMRs are all intragenic sequences, for which a strong positive correlation has been reported between DNA methylation and transcriptional read through from the host gene (Ball et al., 2009; Chotalia et al., 2009). Paternal de novo methylation at transient *Zfp777* and *Zfp787* gDMRs may therefore be facilitated by ongoing transcription from these genes at implantation. Conversely, at the imprinted *Cdh15* gDMR, low levels of transcription from the main upstream promoter, local enrichment in H3K4 methylation and transcription factor occupancy may protect from de novo methylation (Lienert et al., 2011; Ooi et al., 2007). The same rules would apply later during life with tissue formation. The DMR is conserved in tissues where it acts as an active promoter for the short paternal *Cdh15* transcript (hypothalamus and neonatal brain), and is potentially protected by transcription factors and/or H3K4me marks. However, in tissues where the short transcript is not expressed, different methylation states are observed and seem to correlate with the expression status of the long canonical *Cdh15* transcript.

Our screen designed at 8.5 dpc was effective at identifying tissue-specific imprinted gDMRs, because it was performed at a time when they are still universal. It also identified transient gDMRs during their remethylation process. Although the transient gDMRs we found do not seem to affect expression, presumably because of their 3' position, other transient gDMRs may regulate the transcriptome of the peri-implantation embryo. Notably, the parental specificity of these methylated sequences should be lost upon somatic nuclear transfer, resulting in two methylated alleles instead of one during preimplantation development. Furthermore, as for imprinted gDMRs, transient gDMRs may be sensitive to assisted reproductive technologies, involving stimulation of oocyte production and preimplantation embryo culture.

EXPERIMENTAL PROCEDURES

MeDIP-Seq

MeDIP-Seq was performed on three pooled litters with the *Dnmt3L*^{+/+} (WT) or *Dnmt3L*^{-/-} genotype. All MeDIP and sequencing library preparations were performed in parallel. Additionally, MeDIP-Seq was performed on three independent C57Bl6/J sperm samples and twelve pools of three livers each of 17.5 dpc fetal hybrid C57Bl6/J and PWD/PhJ mice. MeDIP enrichment and preparation of paired-end sequencing libraries were then performed as described (Down et al., 2008), using monoclonal anti-5-methylcytosine antibody (Eurogentech) and magnetic anti-mouse beads (Dynabeads) for immunoprecipitation. All libraries were sequenced using an Illumina GA2x instrument.

DNA Methylation Analyses

For MSRE-qPCR, the methylation-dependent restriction enzyme *McrBC* was used, and methylation percentages were calculated according to (Oakes et al., 2009). Values represent the average of three independent digestion experiments, performed on DNA from 8.5 dpc litters of eight embryos. For bisulfite conversion, DNA was treated with the EpiTect kit (QIAGEN). BiQ Analyzer software was used for sequence alignments (Bock et al., 2005) and clones with identical patterns of conversion were removed from the final pileup.

ACCESSION NUMBERS

Raw MeDIP-Seq read and alignment data are available from GEO with accession GSE32687.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, three tables, Supplemental Experimental Procedures, and Supplemental References and can be found with this article online at <http://dx.doi.org/10.1016/j.molcel.2012.07.010>.

ACKNOWLEDGMENTS

We are grateful to Simon Quenneville, Edith Heard, Anne-Valérie Gendrel, and Elphège Nora for insightful discussions. We thank Alex Meissner for sharing RRBS data, Sandrine Barbaux for providing human DNA samples, and Philippe Clerc and Philip Avner for providing the hybrid ES cells. This research was supported by a European Young Investigator Award (EURYI) and the Fondation Schlumberger (D.B.), a BBSRC grant (BB/G00711X/1) (V.K.R. and G.C.), and a Wellcome Trust Grant (085448/Z/08/Z) (R.J.O.). C.P. was supported by a fellowship from ARC, R.D. by the Institut Curie, S.A. by the Cancéropôle Ile-de-France, and R.S. and M.L.H. by RCUK.

Received: February 20, 2012

Revised: June 1, 2012

Accepted: July 6, 2012

Published online: August 16, 2012

REFERENCES

- Arnaud, P., Hata, K., Kaneda, M., Li, E., Sasaki, H., Feil, R., and Kelsey, G. (2006). Stochastic imprinting in the progeny of *Dnmt3L*^{-/-} females. *Hum. Mol. Genet.* 15, 589–598.
- Ball, M.P., Li, J.B., Gao, Y., Lee, J.H., LeProust, E.M., Park, I.H., Xie, B., Daley, G.Q., and Church, G.M. (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.* 27, 361–368.
- Barlow, D.P. (2011). Genomic imprinting: a mammalian epigenetic discovery model. *Annu. Rev. Genet.* 45, 379–403.
- Bhalla, K., Luo, Y., Buchan, T., Beachem, M.A., Guzauskas, G.F., Ladd, S., Bratcher, S.J., Schroer, R.J., Balsamo, J., DuPont, B.R., et al. (2008). Alterations in *CDH15* and *KIRREL3* in patients with mild to severe intellectual disability. *Am. J. Hum. Genet.* 83, 703–713.
- Bock, C., Reither, S., Mikeska, T., Paulsen, M., Walter, J., and Lengauer, T. (2005). BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics* 21, 4067–4068.
- 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.
- Bourc'his, D., and Bestor, T.H. (2006). Origins of extreme sexual dimorphism in genomic imprinting. *Cytogenet. Genome Res.* 113, 36–40.
- 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.
- Chotalia, M., Smallwood, S.A., Ruf, N., Dawson, C., Lucifero, D., Frontera, M., James, K., Dean, W., and Kelsey, G. (2009). Transcription is required for establishment of germline methylation marks at imprinted genes. *Genes Dev.* 23, 105–117.
- Cornelison, D.D., and Wold, B.J. (1997). Single-cell analysis of regulatory gene expression in quiescent and activated mouse skeletal muscle satellite cells. *Dev. Biol.* 191, 270–283.
- Dean, W., Bowden, L., Aitchison, A., Klose, J., Moore, T., Meneses, J.J., Reik, W., and Feil, R. (1998). Altered imprinted gene methylation and expression in completely ES cell-derived mouse fetuses: association with aberrant phenotypes. *Development* 125, 2273–2282.
- Down, T.A., Rakyan, V.K., Turner, D.J., Flicek, P., Li, H., Kulesha, E., Gräf, S., Johnson, N., Herrero, J., Tomazou, E.M., et al. (2008). A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat. Biotechnol.* 26, 779–785.
- Fang, F., Hodges, E., Molaro, A., Dean, M., Hannon, G.J., and Smith, A.D. (2012). Genomic landscape of human allele-specific DNA methylation. *Proc. Natl. Acad. Sci. USA* 109, 7332–7337.
- Ferrón, S.R., Charalambous, M., Radford, E., McEwen, K., Wildner, H., Hind, E., Morante-Redolat, J.M., Laborda, J., Guillemot, F., Bauer, S.R., et al. (2011). Postnatal loss of *Dlk1* imprinting in stem cells and niche astrocytes regulates neurogenesis. *Nature* 475, 381–385.
- Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G.P., Haig, D., and Dulac, C. (2010). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* 329, 643–648.
- Hiura, H., Sugawara, A., Ogawa, H., John, R.M., Miyauchi, N., Miyazaki, Y., Horiike, T., Li, Y., Yaegashi, N., Sasaki, H., et al. (2010). A tripartite paternally methylated region within the *Gpr1-Zdbf2* imprinted domain on mouse chromosome 1 identified by meDIP-on-chip. *Nucleic Acids Res.* 38, 4929–4945.
- Hollnagel, A., Grund, C., Franke, W.W., and Arnold, H.H. (2002). The cell adhesion molecule M-cadherin is not essential for muscle development and regeneration. *Mol. Cell. Biol.* 22, 4760–4770.
- Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.* 6, e1001134.
- 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.
- Kelsey, G., Bodle, D., Miller, H.J., Beechey, C.V., Coombes, C., Peters, J., and Williamson, C.M. (1999). Identification of imprinted loci by methylation-sensitive representational difference analysis: application to mouse distal chromosome 2. *Genomics* 62, 129–138.
- Kent, W.J., Baertsch, R., Hinrichs, A., Miller, W., and Haussler, D. (2003). Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc. Natl. Acad. Sci. USA* 100, 11484–11489.
- Kobayashi, S., Kohda, T., Ichikawa, H., Ogura, A., Ohki, M., Kaneko-Ishino, T., and Ishino, F. (2002). Paternal expression of a novel imprinted gene, *Peg12/Frat3*, in the mouse 7C region homologous to the Prader-Willi syndrome region. *Biochem. Biophys. Res. Commun.* 290, 403–408.
- Kobayashi, H., Yamada, K., Morita, S., Hiura, H., Fukuda, A., Kagami, M., Ogata, T., Hata, K., Sotomaru, Y., and Kono, T. (2009). Identification of the mouse paternally expressed imprinted gene *Zdbf2* on chromosome 1 and its imprinted human homolog *ZDBF2* on chromosome 2. *Genomics* 93, 461–472.
- Kobayashi, H., Sakurai, T., Imai, M., Takahashi, N., Fukuda, A., Yayoi, O., Sato, S., Nakabayashi, K., Hata, K., Sotomaru, Y., et al. (2012a). Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet.* 8, e1002440.
- Kobayashi, H., Sakurai, T., Sato, S., Nakabayashi, K., Hata, K., and Kono, T. (2012b). Imprinted DNA methylation reprogramming during early mouse embryogenesis at the *Gpr1-Zdbf2* locus is linked to long cis-intergenic transcription. *FEBS Lett.* 586, 827–833.
- Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., and Ferguson-Smith, A.C. (2008). A maternal-zygotic effect gene, *Zfp57*, maintains both maternal and paternal imprints. *Dev. Cell* 15, 547–557.
- Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F., and Schübeler, D. (2011). Identification of genetic elements that autonomously determine DNA methylation states. *Nat. Genet.* 43, 1091–1097.
- Liu, J., Chen, M., Deng, C., Bourc'his, D., Nealon, J.G., Erlichman, B., Bestor, T.H., and Weinstein, L.S. (2005). Identification of the control region for tissue-specific imprinting of the stimulatory G protein α -subunit. *Proc. Natl. Acad. Sci. USA* 102, 5513–5518.
- Mackay, D.J., Callaway, J.L., Marks, S.M., White, H.E., Acerini, C.L., Boonen, S.E., Dayanikli, P., Firth, H.V., Goodship, J.A., Haemers, A.P., et al. (2008). Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in *ZFP57*. *Nat. Genet.* 40, 949–951.

- McEwen, K.R., and Ferguson-Smith, A.C. (2010). Distinguishing epigenetic marks of developmental and imprinting regulation. *Epigenetics Chromatin* 3, 2.
- Nichols, J., Silva, J., Roode, M., and Smith, A. (2009). Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. *Development* 136, 3215–3222.
- Oakes, C.C., La Salle, S., Trasler, J.M., and Robaire, B. (2009). Restriction digestion and real-time PCR (qAMP). *Methods Mol. Biol.* 507, 271–280.
- Ooi, S.K., 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.
- Padilla, F., Broders, F., Nicolet, M., and Mege, R.M. (1998). Cadherins M, 11, and 6 expression patterns suggest complementary roles in mouse neuromuscular axis development. *Mol. Cell. Neurosci.* 11, 217–233.
- Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P.V., Grimaldi, G., Riccio, A., and Trono, D. (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.
- Rose, O., Grund, C., Reinhardt, S., Starzinski-Powitz, A., and Franke, W.W. (1995). Contactus adherens, a special type of plaque-bearing adhering junction containing M-cadherin, in the granule cell layer of the cerebellar glomerulus. *Proc. Natl. Acad. Sci. USA* 92, 6022–6026.
- Schulz, R., Proudhon, C., Bestor, T.H., Woodfine, K., Lin, C.-S., Lin, S.-P., Prissette, M., Oakey, R.J., and Bourc'his, D. (2010). The parental non-equivalence of imprinting control regions during mammalian development and evolution. *PLoS Genet.* 6, e1001214.
- Shirabe, K., Kimura, Y., Matsuo, N., Fukushima, M., Yoshioka, H., and Tanaka, H. (2005). MN-cadherin and its novel variant are transiently expressed in chick embryo spinal cord. *Biochem. Biophys. Res. Commun.* 334, 108–116.
- 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.
- Smith, R.J., Dean, W., Konfortova, G., and Kelsey, G. (2003). Identification of novel imprinted genes in a genome-wide screen for maternal methylation. *Genome Res.* 13, 558–569.
- 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.
- 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.
- Xie, W., Barr, C.L., Kim, A., Yue, F., Lee, A.Y., Eubanks, J., Dempster, E.L., and Ren, B. (2012). Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* 148, 816–831.
- Zuo, X., Sheng, J., Lau, H.T., McDonald, C.M., Andrade, M., Cullen, D.E., Bell, F.T., Iacovino, M., Kyba, M., Xu, G., and Li, X. (2012). Zinc finger protein ZFP57 requires its co-factor to recruit DNA methyltransferases and maintains DNA methylation imprint in embryonic stem cells via its transcriptional repression domain. *J. Biol. Chem.* 287, 2107–2118.

**Protection against De Novo Methylation
Is Instrumental in Maintaining Parent-of-Origin
Methylation Inherited from the Gametes**

Charlotte Proudhon, Rachel Duffié, Sophie Ajjan, Michael Cowley, Julian Iranzo, Guillermo Carbajosa, Heba Saadeh, Michelle L. Holland, Rebecca J. Oakey, Vardhman K. Rakyan, Reiner Schulz, and Déborah Bourc'his

Supplemental Information Inventory

Supplemental Figure/Table	Related main Figure/Table	Information
Figure S1	Figures 1A, 1B, 1C, 1D	MeDIP-Seq profiles of known imprinted loci controlled by maternal ICRs, as positive controls of the sensitivity of the screen
Figure S2	Figures 1A 1D, and 1E	MeDIP-Seq profiles of known paternal ICRs and germline genes, as controls of the specificity of the screen
Figure S3	Figures 1A, 1E, 2B and 5B	Germline methylation profiles of candidate maternal gDMRs
Figure S4	Table 1	Investigation of allelic methylation in 17.5dpc fetal liver
Figure S5	Figure 2, 3E, 3F	Further characterization of <i>Cdh15</i> imprinting in mouse and human
Figure S6	Figure 3	Lack of imprinted expression of genes in the vicinity of <i>Cdh15</i>
Figure S7	Figure 5 and Table 1	Identification and methylation analysis of a second transient maternal gDMR, associated with <i>Zfp787</i>
Table S1	Figures 1, 2, 4, 5, 6, and Table 1	DMRs identified by MeDIP-Seq between <i>Dnmt3L</i> ^{-/-} and wildtype 8.5dpc embryos
Table S2	Figures 1, 2, 3, 4, 5	Primer list
Table S3	Figures 3B, 3C	Antisera

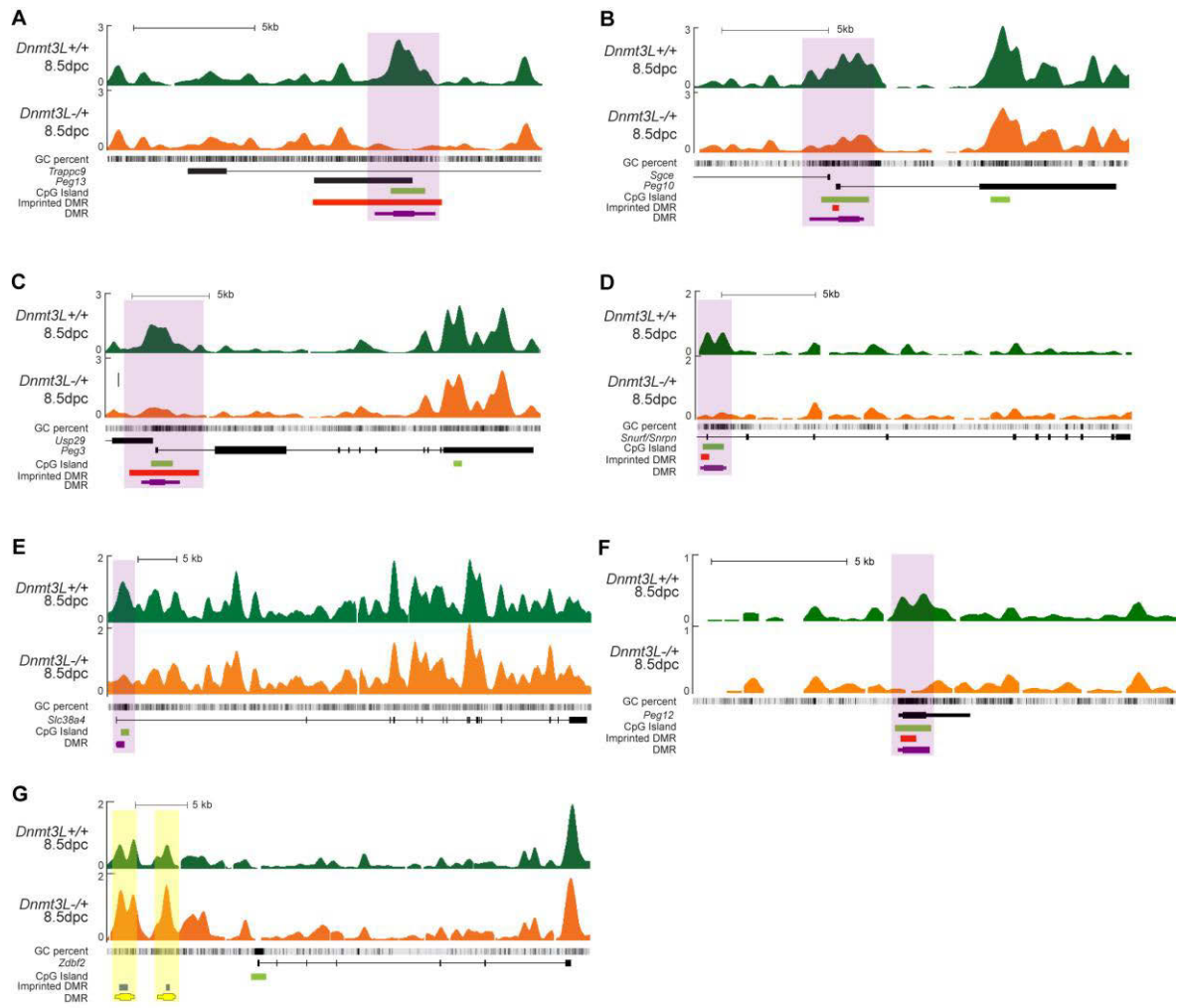


Figure S1. MeDIP-Seq profiles of known imprinted loci controlled by maternal ICRs, as positive controls of the sensitivity of the screen

Known maternal ICRs associated with (A) *Peg13*, the highest ranking locus of the screen, and with (B) *Peg10/Sgce*, the lowest ranking ICR of the screen. (C) *Peg3* ICR, and (D) *Snrpn/Snurfl* ICR, two loci that were reported to sporadically regain methylation in *Dnmt3L*^{-/+} embryos, also show significant hypomethylation by MeDIP-Seq. The screen allowed the improved mapping of regions of Dnmt3L-dependent methylation in partially documented imprinted loci, such as (E) *Slc38a4* and (F) *Peg12*. (G) Two hypermethylated DMRs (yellow), located upstream of the *Zdbf2* gene, coincide with two regions previously reported to be paternally imprinted gDMRs (blue). Their gain of methylation in *Dnmt3L*^{-/+} embryos instead indicates that they are secondary somatic DMRs, under the control of a primary maternal gDMR.

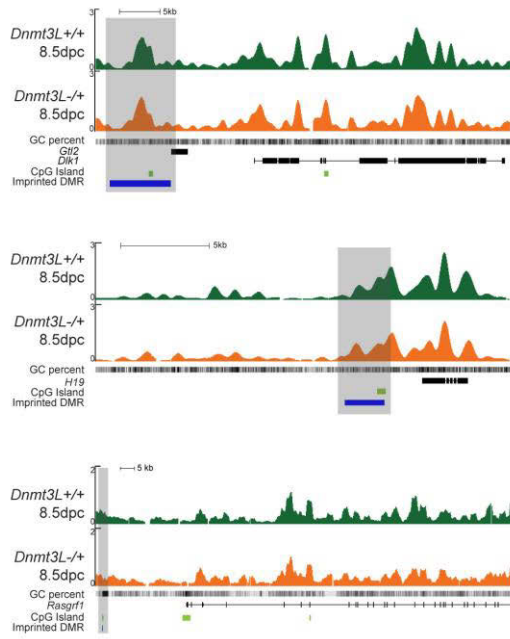
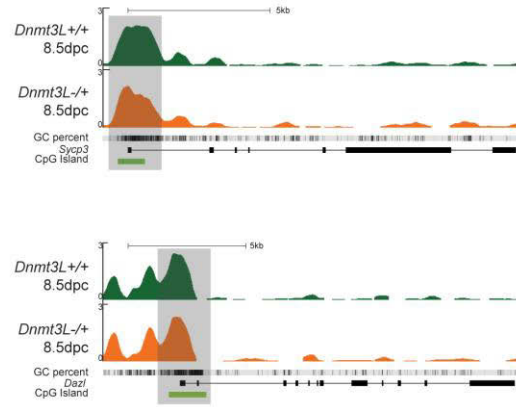
A**B**

Figure S2. MeDIP-Seq profiles of known paternal ICRs and germline genes, as controls of the specificity of the screen

Methylation at these loci is not altered in *Dnmt3L*^{-/+} embryos (gray). (A) Known paternal ICRs (blue), associated with *Dlk1-Gtl2* (top), *H19-Igf2* (middle), and *Rasgrf1* (bottom) loci. These gDMRs depend on sperm-inherited DNA methylation. (B) Promoters of germline-expressed genes, associated with the *Sycp3* gene (top) and the *Dazl* gene (bottom). These regions acquire methylation somatically, in the embryo at the time of implantation, and do not depend on oocyte-inherited methylation.

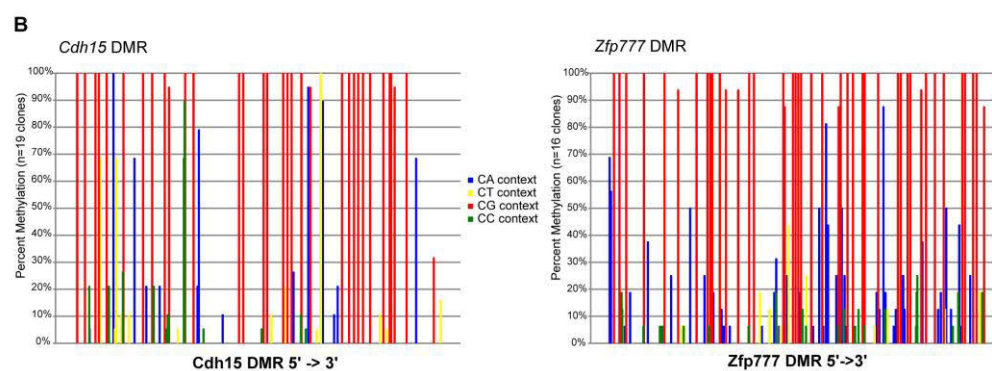
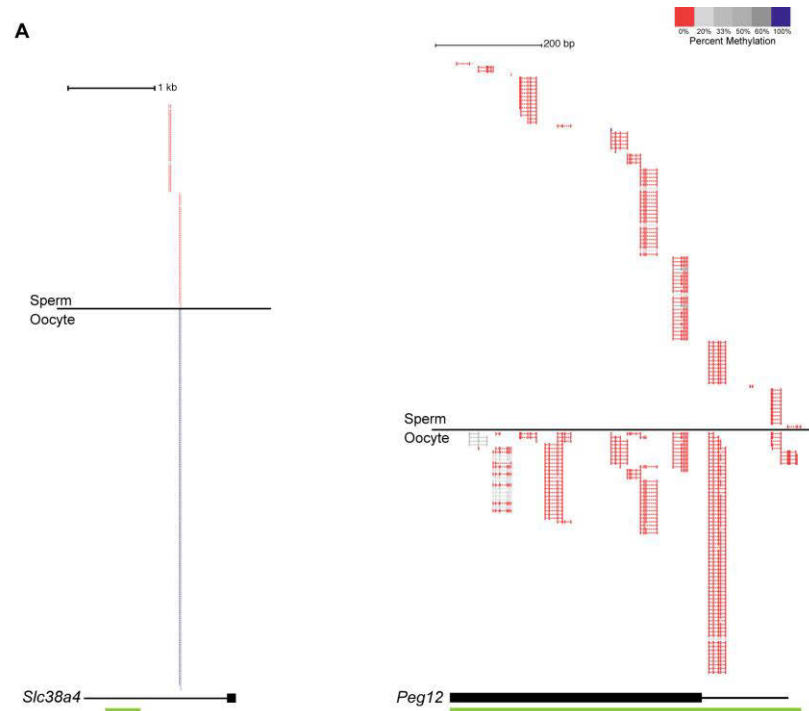


Figure S3. Germline methylation profiles of candidate maternal gDMRs

RRBS data showing that (A) the *Slc38a4* DMR is specifically methylated in oocytes but not in sperm, and therefore classifies as a maternal gDMR, while (B) the *Peg12* DMR is unmethylated both in oocytes and sperm and is therefore not a gDMR. Blue=methylated, red=unmethylated, green=CpG island. (C) Detection of oocyte non-CG methylation by bisulfite sequencing at the imprinted maternal gDMR of *Cdh15* and at the transient maternal gDMR of *Zfp777*.

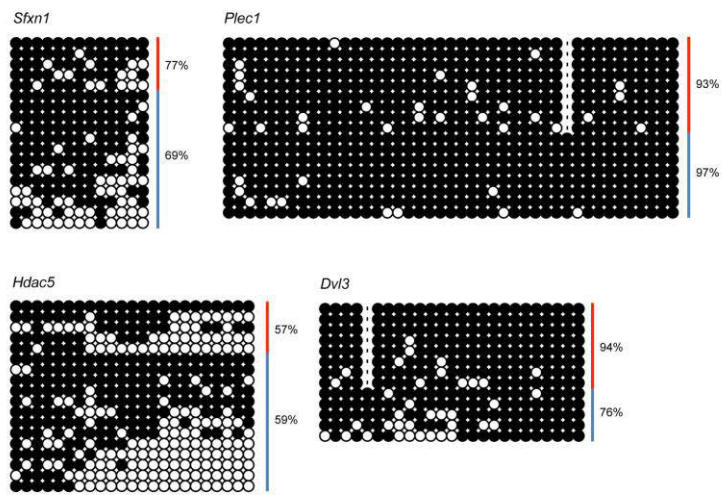


Figure S4. Investigation of allelic methylation in 17.5dpc fetal liver

DNA from BxC crosses was used for bisulfite sequencing of four maternal gDMR candidates (*Sfxn1*, *Plec1*, *Hdac5* and *Dvl3*), which failed to reach statistical significance because of low read count in fetal liver MeDIP-Seq data. All of them showed equal methylation on maternal and paternal alleles, demonstrating that maternal-specific methylation is not maintained at these loci in this post-implantation tissue.

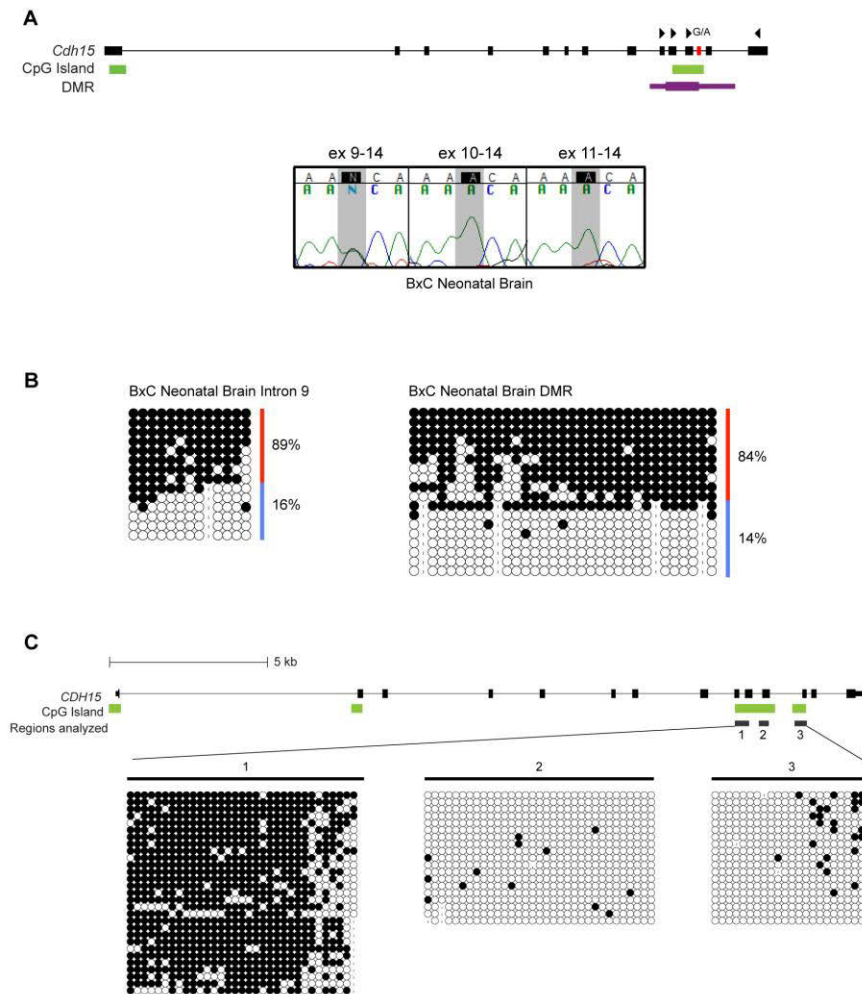


Figure S5. Further characterization of *Cdh15* imprinting in mouse and human

(A) Mapping of the short *Cdh15* transcript in mouse by RT-PCR sequencing reveals a switch from biallelic to monoallelic expression between exons 9 and 10, placing the origin of the short *Cdh15* imprinted transcript around intron 9. Amplifications were performed on BxC and CxB neonatal brain RNA, where both the canonical and the short *Cdh15* transcript isoforms are expressed, using a reverse primer in exon 14 and forward primers in exons 9, 10 or 11. The A/G SNP is located in exon 12 (red). Arrow heads indicate primer positions. (B) In mouse neonatal brains, maternal-specific methylation is observed both at the putative TSS of the short imprinted *Cdh15* transcript (intron 9) and at the DMR. (C) In humans, the region corresponding to the mouse *Cdh15* DMR also overlaps with a large CGI. We analyzed three parts of this CGI in fetal liver DNA. Region 1 was found globally methylated, with no obvious allelic specificity, as seen with an informative SNP. The most 3' part of Region 1 showed a decrease in methylation. Regions 2 and 3 located downstream were found completely unmethylated, again with no evidence of allelic specificity, although no SNP was found in these amplified regions.

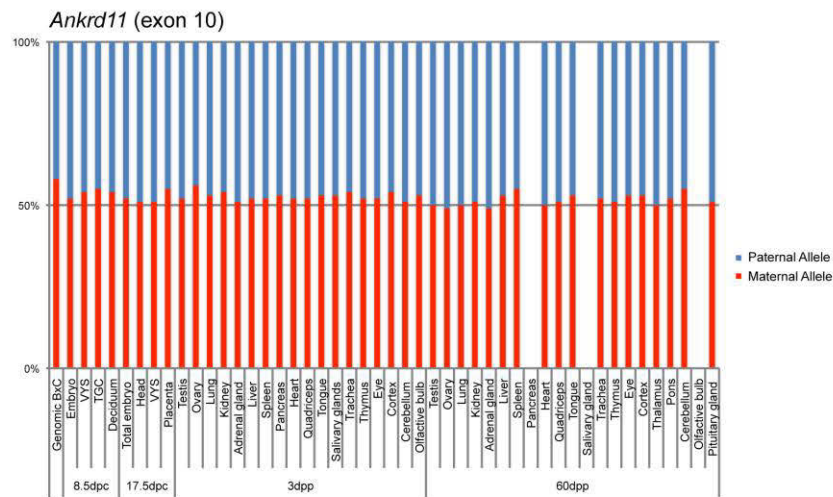
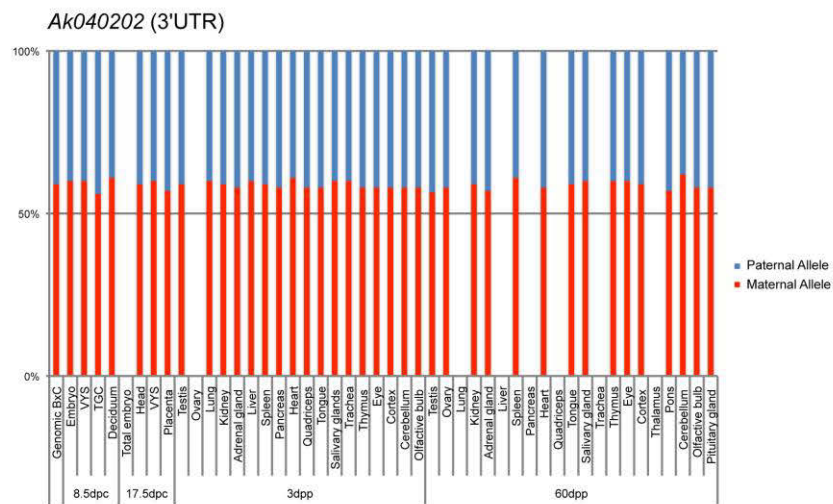
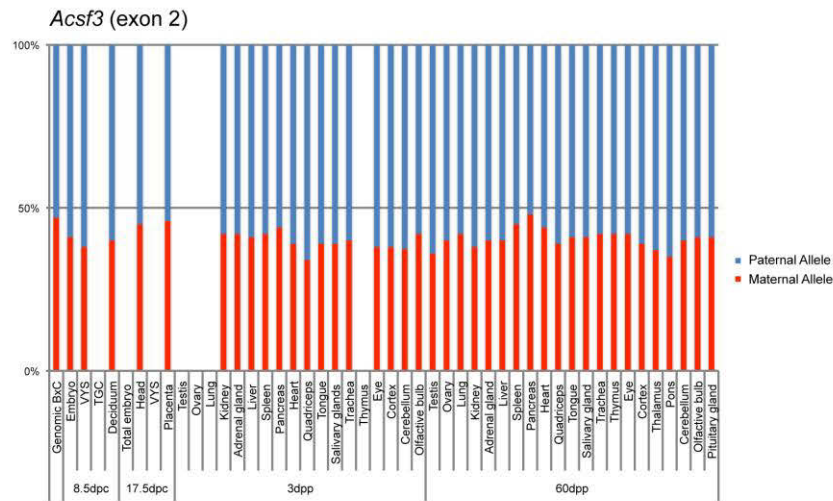


Figure S6. Biallelic expression of *Asf3*, *Ankrd11* and *AK040202*, located in the vicinity of *Cdh15*

RT-PCR pyrosequencing was performed on a bank of 47 tissues from reciprocal BxC and CxB crosses. Results are presented for one orientation of the cross only. Locations of SNPs are indicated after each gene name. Missing bars indicate no detected expression.

Figure S7. Identification and methylation analysis of the transient maternal gDMR associated with *Zfp787*

(A) MeDIP-Seq profile of the *Zfp787* locus. Genes are oriented 5' to 3'. The hypomethylated DMR maps to a CGI in the last exon of the *Zfp787* gene. (B) Developmental analysis of *Zfp787* DMR methylation by bisulfite sequencing. This region acquires its methylation in oocytes but not in sperm, maintains differential methylation during preimplantation, but is completely unmethylated in *Dnmt3L*^{-/+} blastocysts. Contrary to imprinted gDMRs, methylation is gained on paternal alleles at implantation and parental alleles exhibit similar methylation levels at 9.5dpc, both in WT and *Dnmt3L*^{-/+} embryos. (C) Adult hybrid tissues (tail and liver) show a fully methylated pattern. In opposition to the blastocyst, both parental alleles are unmethylated in ES cells. (D) A RT-PCR pyrosequencing assay was designed to study a SNP located in *Zfp787* exon 3. The *Zfp787* gene is biallelically expressed throughout development and adulthood.

Table S1. DMRs identified by MeDIP-Seq between *Dnmt3L*^{-/-} and wildtype 8.5dpc embryos

DMRs were identified using a sliding window approach (See Material and Methods). We carried out the analysis twice, using two different window sizes: 1000bp and 500bp. Each window size is reported on a separate sheet (named “1000bp” and “500bp”). *Columns A-F*: percent of the DMR covered by a particular repeat class (Satellite, Sine, Line, etc.). *Columns G & H*: total percent retrotransposon and repetitive sequence content, respectively. Red highlights indicate high repeat content (>25% and >50%). *Columns I-L & K*: overlaps of the DMR with (AceView) transcripts, split into promoters, 3'UTR, exons and introns, and overlap with known (red, blue) or validated (purple, yellow) gDMRs/ICRs. *Columns N-T*: DMR log₂ fold-change in methylation (cyan: hypomethylation; magenta: hypermethylation), FDR, genomic coordinates (Q & T: start and end of entire region meeting FDR<50% threshold; R & S: start and end of 1Kbp/500bp sub-window where the most differential methylation was observed (thick part of bar representing DMRs in Figures) and on which the FDR calculation is based). *Columns U-X*: counts of CGs, Cs and Gs within DMR, and the corresponding observed over expected #CG ratio. *Columns Y-AE*: overlaps of DMR with ChIP-seq peaks for Zfp57+Kap1+Setdb1 reported in (Quenneville et al., 2011), midpoint distance between DMR and peak, peak genomic coordinates, number of occurrences of Zfp57 binding motif within DMR. *Columns AF-AM*: genomic position of SNP between C57Bl6/J and PWD/PhJ used for binomial test, distance of SNP to DMR midpoint, number of reads overlapping SNP, number of reads from reference allele (C57Bl6/J), number of reads from alternative allele (PWD/PhJ), one-sided binomial test p-value for H1: better than 50% chance of maternal methylation, one-sided binomial test p-value for H1: better than 50% chance of paternal methylation, position and allele counts for all SNPs within DMR. *Columns AN-AQ*: log₂ fold-change in methylation for DMRs between sperm and wildtype 8.5dpc embryos that most overlap with DMRs between *Dnmt3L*^{-/-} and wildtype embryos, FDR for sperm vs wildtype DMRs, percent of *Dnmt3L*^{-/-} vs wildtype DMR covered by sperm vs wildtype DMRs, log₂ fold-change and FDR for all sperm vs wildtype DMRs that overlap the *Dnmt3L*^{-/-} vs wildtype DMRs. *Columns AR-AT*: distance to closest exon showing allelic expression bias in data from (Gregg et al., 2010), parental allele with higher expression, binomial test confidence score (-10log₁₀ p-value). *Columns AU-BA*: distance to closest promoter investigated for histone modifications by (Mikkelsen et al., 2007), gene name, RefSeq accession, promoter class (H/I/LCP = high/intermediate/low CpG content), observed modifications in ESCs, NPCs and MEFs. *Columns BB-BO*: overlapping peaks of histone modifications called by (Mikkelsen et al., 2007) using two different methods (BB-BF: HMM method; BG-BO: sliding window method). *Columns BP-BR*: distance to closest (Hiura et al., 2010) DMR candidate by MeDIP-chip (promoter array), p-value and fold-change as reported. *Columns BS-CN*: distance to closest CGI reported in (Illingworth et al., 2010), followed by annotation from the original publication. *Columns CO-CW*: Average percent methylation of overlapping CGIs, identified by CGI ID from (Illingworth et al., 2010) and/or genomic coordinates, with sufficient BS-seq data in wildtype oocytes, *Dnmt3L*^{-/-} oocytes, sperm, blastocysts and mouse ES samples, as reported in Kobayashi et al., 2012a.

Table S2. Primer list

Bisulfite		
Sfxn1-outer	Fwd Rev	5'- TTTGGGAGTTTTGAGTATTTGAG -3' 5'- ATAACCTCCTCTCTAAAAATAAATTC -3'
Sfxn1-inner	Fwd Rev	5'- TTTGGGAGTTTTGAGTATTTGAG -3' 5'- CTCATCTAAAATTATTATTCAAAATC -3'
Plec1-outer	Fwd Rev	5'- TTTTGTGGTTAGAAGTATTTTTAT -3' 5'- ACCCTAAATAAACTAAACTACAA -3'
Plec1-inner	Fwd Rev	5'- GATTTTAGTTAGTTTAGTTTTAG -3' 5'- CCAAATACAAATAACCTAAAAA -3'
Dvl3-outer	Fwd Rev	5'- GGATTAGTTGTAGAGTTATTGAG -3' 5'- ACTCCTAAAAAACCATAACTAC -3'
Dvl3-inner	Fwd Rev	5'- GTAGTTTATTTTGGTTGGAGATT -3' 5'- CCATAACTTAAATATATATAATAACTA -3'
Hdac5-outer	Fwd Rev	5'- AGATAGAGAGGAAGGGAAAGT-3' 5'- TCCTCATCCCACTCTAAACCT-3'
Hdac5-inner	Fwd Rev	5'- GAGGGTGGAGAGGGGTAGAT-3' 5'- TTCTCTACCAACTACAACCCTTA-3'
Cdh15-outer	Fwd Rev	5'- TTATGTTTTTGTATTGGTTTTTTTTT -3' 5'- AAAATAAAAAAAAAAACCCAAAACAAA
Cdh15-inner	Fwd Rev	5'- TTAGTTGAATTTTAGAGTATTAGATT -3' 5'- AAATAAAACCAAAAACCTTCACTCAC -3'
Cdh15 intron9-outer	Fwd Rev	5'- TAGAAGATTGGTTGTAAGTGGA-3' 5'- ACCTATCTCTTAAATCTCAAAATTA-3'
Cdh15 intron9-inner	Fwd Rev	5'- GTTGGTATAGAGTTATTATTTTAG-3' 5'- ACCTATCTCTTAAATCTCAAAATTA-3'
AK008011-gSNP-outer	Fwd Rev	5'- GTAGAGGGAGTATTAGGAAGG -3' 5'- CACATCATCAAACCTCTATAACAAA -3'
AK008011-gSNP-inner	Fwd Rev	5'- GTTATTATAGTATTGATTATTTTTTA -3' 5'- AGAATTTAATTATTGGGTTTTTTG -3'
Zfp777-BI6-outer	Fwd Rev	5'- TAAATGTTTTGAGTGTGATAGTAG -3' 5'- CCTTACACCTAACTACACATCTAA -3'
Zfp777-BI6-inner	Fwd Rev	5'- AGTTAAGTTTGATTAAGTATTAGATT -3' 5'- CCTTACACCTAACTACACATCTAA -3'
Zfp777-gSNP-outer	Fwd Rev	5'- ATAGATTGTAAATGAAGGAGA -3' 5'- CACCCACATTTCATAAACATA -3'
Zfp777-gSNP-inner	Fwd Rev	5'- GTGGATTATTAGATTGATATG -3' 5'- CACCCACATTTCATAAACATA -3'
Zfp787-BI6-outer	Fwd Rev	5'- AGATGTGTGTAGGTTTTTTATTAA -3' 5'- AAAACCATATACCTACTCAAAATA -3'
Zfp787-BI6-inner	Fwd Rev	5'- AGATGTGTGTAGGTTTTTTATTAA -3' 5'- AAAACTCCAACCTCATACAACACC -3'
Zfp787-gSNP-outer	Fwd Rev	5'- TTGGTGGGTAAATAGTTTAGTT -3' 5'- ATATAACAAAACTTCAACCACT -3'
Zfp787-gSNP-inner	Fwd Rev	5'- ATTTGGTTTTGATAGTTTTGGG -3' 5'- ATATAACAAAACTTCAACCACT -3'
hCDH15-R1-outer	Fwd Rev	5'- TTTTGGGGTAAATTTAGATTTTATT -3' 5'- TAACCTAACTAAACTCCAATTC -3'

hCDH15-R1-inner	Fwd	5'- TTTTGTATTAGTTATTTTAAGGATTA -3'
	Rev	5'- AGTTATATTAAGGTTTAGGTTTTTT -3'
hCDH15-R2-outer	Fwd	5'- TGGGGTAGGGTTATTTATTG -3'
	Rev	5'- TCACTCACCCAACAACAAAA -3'
hCDH15-R2-inner	Fwd	5'- TGGGGTAGGGTTATTTATTG -3'
	Rev	5'- CCCAACTAAAACCTATACC -3'
hCDH15-R3-outer	Fwd	5'- TTTATTAGTTAAGTTTTTTTTTTTATAA -3'
	Rev	5'- AGGAGTTTGTATTTGAGATTTTAA -3'
hCDH15-R3-inner	Fwd	5'- TTTATTAGTTAAGTTTTTTTTTTTATAA -3'
	Rev	5'- GGAGGAGGATTAGGTGAGG -3'
hZNF777-outer	Fwd	5'- GTGATTAAGATAGAGGAATAAG -3'
	Rev	5'- GTATTATTTGTTGGAGTATTAG -3'
hZNF777-inner	Fwd	5'- ATATGAGGTTAGTATGTATTAG -3'
	Rev	5'- TTTAAGTTATTTAGTTTTGGTAG -3'

ChIP-qPCR

Cdh15DMR-ChIP	Fwd	5'- GAGAGCAAACGCTGAACGTC -3'
	Rev	5'- GGACATTCTGGGACCCCTTT -3'
KvDMR-ChIP	Fwd	5'- AACTCCGAATAAGCAGCCTTC -3'
	Rev	5'- TGGAGTACGTGTTGTTATGTGG -3'
Tbx15-ChIP	Fwd	5'- TCCCCCTTCTCTTGTGTCAG -3'
	Rev	5'- CGGAAGCAAGTCTCAGATCC -3'
IAP-ChIP	Fwd	5'- CTCCATGTGCTCTGCCTTCC -3'
	Rev	5'- CCCCGTCCCTTTTTTAGGAGA -3'

MSRE-qPCR

KvDMR-MSRE	Fwd	5'- AACTCCGAATAAGCAGCCTTC -3'
	Rev	5'- TGGAGTACGTGTTGTTATGTGG -3'
H19-MSRE	Fwd	5'- ACATTCACACGAGCATCCAGG -3'
	Rev	5'- GCTCTTTAGGTTTGGCGCAAT -3'
Nesp-MSRE	Fwd	5'- AGCGCAAGGAGGAAAACAG -3'
	Rev	5'- ATTAGTGACGCCGATGG -3'
Slc384a-MSRE	Fwd	5'- TCCCTCTCCTCTGAAGTCCTC -3'
	Rev	5'- GAGACTGCTTCCACATGGT -3'
Peg12-MSRE	Fwd	5'- TGTGCCCTCCATCACAATC -3'
	Rev	5'- CAAAGCTTCCCGCTCACTC -3'
Zdbf2_DMR1-MSRE	Fwd	5'- AACTTTAGAGCCCCAGAAGG -3'
	Rev	5'- CCATCCGACAATTCAAAATC -3'
Zdbf2_DMR2-MSRE	Fwd	5'- ACCCTGGTCCGACCTTATGT -3'
	Rev	5'- GGGAAGTCCCATCCTTTAGC -3'
Cdh15DMR-MSRE	Fwd	5'- GAGAGCAAACGCTGAACGTC -3'
	Rev	5'- CCAGCACAATGACCAAGTGC -3'
AK008011-MSRE	Fwd	5'- ATGCCGGAAGTTGCTCTG -3'
	Rev	5'- GGATGGATGGAGGCTTCTC -3'
Zfp777-MSRE	Fwd	5'- TGAAGCTCTTCTCGCACTCC -3'
	Rev	5'- CCTTCATTTGCAGTCTGTGC -3'
Zfp787-MSRE	Fwd	5'- GGACACACGAAGGGCTTTA -3'
	Rev	5'- GCAGCTTCACACAGAGCAAG -3'

Northern

Cdh15Probe 5-6	Fwd	5'- CTTCTGTCACCAGGGCTGAG -3'
	Rev	5'- GGGCATTGTCGTTGATGTC -3'
Cdh15Probe 8-9	Fwd	5'- GGACTATGAGAGCCGTGAGC -3'
	Rev	5'- CTCTGTACCAGCCGTCCTTC -3'

Cdh15Probe 9-14	Fwd	5'- AAGACTGGCTGCAAGTGGAC -3'
	Rev	5'- CTCCCAGGCTGGACAGAAT -3'

Pyrosequencing

Cdh15ex11 (ChIP)	Fwd	5'- (biot)CGAGAGCAAACGCTGAAC -3'
	Rev	5'- CCCAAGCTGACGCCTACAC -3'
	Seq	5'- CAGGCAAGTGCCATC -3'
KvDMR	Fwd	5'- AACGGAGCCCCCTCACTCT -3'
	Rev	5'- (biot)GACCCCTGAGCTTTGTAGCC -3'
	Seq	5'- CATTAAAACAGCTACCACAT -3'
Nhlrc1-ex1	Fwd	5'- GCGTGGTAGCAGGTGAGA -3'
	Rev	5'- (biot)CCTGCCGAGCCTGTGACA -3'
	Seq	5'- AGCCCAGGAGCTCCA -3'
Nhlrc1-3'UTR	Fwd	5'- (biot)ATGTGCTGGGGCAAGATCTA-3'
	Rev	5'- CTGCCTGCTAGCGTCATGG -3'
	Seq	5'- CGTCATGGGCATATGAG -3'
Acsf3-ex2	Fwd	5'- ACTTGGCACTGCCCTTCA -3'
	Rev	5'- (biot)TGGTAGGCAGGAGGCTATGT -3'
	Seq	5'- TGATTCCCAGAAGACAC -3'
AK040202-3'UTR	Fwd	5'- GGAATGGGCTGGCTTCTGAA -3'
	Rev	5'- (biot)GTGGTCACCTAGGCTGCTTTCTTA -3'
	Seq	5'- GGCTTCTGAAAACAGTG -3'
Ankrd11-ex10	Fwd	5'- (biot)TCTCGAGGTCTTTCTGGGACAG -3'
	Rev	5'- ATAAAGATGAGCAACGGGAACG -3'
	Seq	5'- GGCAAGAAAGACAGCAG -3'
Zfp777-3'UTR	Fwd	5'- GGTCTAAGGCCCTGGGACA -3'
	Rev	5'- (biot)ACGGGAGGGAAGGAGGGTT -3'
	Seq	5'- CCAGACTGAGCTCAGTG -3'
Zfp787-ex3	Fwd	5'- CACTCTGAGCAGGCATATGG -3'
	Rev	5'- (biot)TACCGACTGCGGCAAGAC -3'
	Seq	5'- GCCGGTGCTGCACCA -3'
Zfp787-3'UTR	Fwd	5'- (biot)GGGGGGCAGAAAAAATAATG -3'
	Rev	5'- TGCATCAAAAGGGCTTCC -3'
	Seq	5'- GAATTTTCTTTAATCATAC -3'

RT-PCR

Cdh15-ex3SNP	Fwd	5'- GTGTCTGAGAACCACAAACGC -3'
	Rev	5'- CCTCTGAACACATCCTGTAGG -3'
Cdh15-ex14SNP	Fwd	5' CGTGCCTCCCTACGACACAG -3'
	Rev	5'- GGGCGACTAAGGGTGGCTTC -3'
Cdh15-ex8	Fwd	5'- TCCAGTGTTCAGAGAACCC -3'
Cdh15-ex9	Fwd	5'- AAGACTGGCTGCAAGTGGAC -3'
Cdh15-ex10	Fwd	5'- GCTGAACCCAGAGTACCAG -3'
Cdh15-ex11	Fwd	5'- AGCCACCCAGCAGCGAGAG -3'
Cdh15-ex13	Rev	5'- ACTGATGAAGTTGGCAATGT -3'
Cdh15-ex14	Rev	5'- CTCCCAGGCTGGACAGAAT -3'

RT-qPCR

Cdh15-ex1qPCR	Fwd	5'- GTCACTCAGCCCAATGCTCT -3'
	Rev	5'- GAGCAGAGCAGAACCCATTG -3'
Cdh15-ex12-13qPCR	Fwd	5'- TGTTCATGGCCTACAAGAG -3'
	Rev	5'- CGCAGCTGGTTTATGTCGTA -3'
Arp0	Fwd	5'- TCCAGAGGCACCATGAAATT -3'
	Rev	5'- TCGCTGGCTCCACCTT -3'

Table S3. Antisera List

	Furnisher	Reference
H3	Abcam	ab1791
H3K4me2	Abcam	ab32356
H3K4me3	Active Motif	39159
H3K9me2	Abcam	ab1220
H3K9me3	Millipore	07-442
H4K20me3	Active Motif	39180

Supplemental Experimental Procedures

Isolation of gametes, embryos and tissues

Dnmt3L^{+/+} and *Dnmt3L*^{-/-} embryos were obtained at 8.5dpc from a C57Bl6/J background by natural mating, respectively from *Dnmt3L*^{+/+} and *Dnmt3L*^{-/-} females crossed with wildtype males (Bourc'his et al., 2001). Blastocysts were obtained after superovulation and natural mating. Hybrid embryos at 9.5dpc were obtained by crossing Bl6 females with CAST/E1 males. Sperm was collected from the epididymus of 5 week-old C57Bl6/J males. Wildtype oocytes were collected at the MII stage from superovulated B6CBAF1/J females. For allelic expression studies, 47 tissues from reciprocal crosses between C57Bl6/J and CAST/Ei (BxC and CxB) were collected at 8.5dpc, 17.5dpc, 3dpp and 60dpp. All procedures using animals were reviewed and approved by the Institut Curie Animal Care and Use Committee.

Sequencing details and alignment

The *Dnmt3L*^{-/-} embryos, wildtype embryos and sperm samples were 2x 36bp paired-end sequenced using one lane per sample, except for wildtype embryo sample #1 for which two lanes were used. The liver samples were 12-way multiplexed and 50bp single-end sequenced using 7 lanes. Reads were aligned to the mm9 reference genome using novoalign (<http://www.novocraft.com/main/index.php>) except for the sperm samples that were aligned using bowtie (Langmead et al., 2009).

DMR identification

DMRs were identified between 1) *Dnmt3L*^{-/-} and wildtype embryo samples, and 2) sperm and wildtype embryo samples, using USeq and DeSeq (Anders and Huber, 2010; Nix et al., 2010). The input alignment data were limited to the unique (default

novoalign/bowtie definitions) and primary alignments of properly paired reads that were not flagged by Picard (<http://picard.sourceforge.net/>) as potential PCR duplicates and had a minimum mapping quality of 13. Thus, USeq identified DMRs using 32.7M (across 2 samples) versus 41M (2 samples) data points (fragment midpoints) in comparison 1), and 86.7M (3 samples) versus 41M (2 samples) data points in 2).

DMR annotation

The DMRs determined by USeq in the comparison (*Dnmt3L*^{-/+} versus wildtype embryos) were annotated with the overlapping DMRs from the comparison (sperm versus wildtype embryos), and the parental allele-specific reads counts for the contained SNPs, if any. The read counts were augmented with binomial test results for significant deviation from the expected 1:1 ratio between the parental alleles. Additional third party data were added based on coordinate overlaps and/or maximum distance thresholds using PostgreSQL (<http://www.postgresql.org/>). These data and their sources are: repeat content (RepeatMasker: <http://www.repeatmasker.org/>), transcripts (Thierry-Mieg and Thierry-Mieg, 2006), CpG islands (Illingworth et al., 2010), known imprinted gDMRs/ICRs (WAMIDEX: (Schulz et al., 2008), evidence for allele-specific gene expression in mouse brain (Gregg et al., 2010), histone modifications at promoters in mouse embryonic stem cells, fibroblasts and neural progenitor cells (Mikkelsen et al., 2007), MeDIP-chip data comparing mouse parthenogenic and androgenic samples (Hiura et al., 2010), Zfp57/KAP1/SetB1 ChIP-Seq peaks (Quenneville et al., 2011), and average percent methylation of CGIs in wildtype oocytes, *Dnmt3L*^{-/-} oocytes, sperm, blastocysts and mouse ES samples (Kobayashi et al. 2012a).

Allele-specific read count statistics for liver samples

Positional and allele information for approximately 3.9M SNPs between the C57BL/6 and PWD/PhJ strains was obtained from Mouse Genome Informatics (MGI) at the Jackson laboratories (<http://www.informatics.jax.org/>). Read pileups over all SNP sites were generated using samtools (Li et al., 2009) and turned into parental allele-specific counts. The observed read counts for the maternal and the paternal alleles of each candidate gDMR (Table 1 and Supplementary Table S1) were tested for consistency with the null hypothesis (H0) of a read being equally likely to originate from either the maternal or the paternal allele (binomial distribution with $P=0.5$) versus the alternative hypothesis (H1) that the chance of a read coming from the maternal allele is higher ($P>0.5$). The p-value of the test (not to be confused with P) expresses the chance of making a type 1 error when rejecting H0 in favor of H1. A p-value of 1 was entered in Table 1 and Supplementary Table S1 if there was no informative SNP and hence, no data. Similarly,

instances where read coverage of the SNP was <10 (insufficient power to reject H₀ at $\alpha=0.05$ even for a 9:1 observed skew in the number of reads toward maternal methylation), are highlighted in italics in Table 1.

DNA methylation analysis by MSRE-PCR and bisulfite sequencing

Genomic DNA was isolated by standard procedures, except for sperm, where a modified extraction method was used to eliminate somatic cells present in the epididymus semen (Jeffreys et al., 1994). For MSRE-qPCR, 1 μ g of DNA was mixed with restriction buffer and then split in two tubes, one corresponding to the digested experiment with 20U of the methylation-sensitive restriction enzyme *McrBC* and twice the amount of recommended GTP, and one to the undigested control with water instead. Samples were incubated at 37°C for 5h, heat inactivated at 65°C for 20 min and 1ng was used for real-time PCR using the SYBR Green technology (Applied Biosystems). For the bisulfite experiment, PCR amplicons were designed to include single nucleotide polymorphisms to infer the parental origin of alleles. The equivalent of at least 300 diploid cells was used per PCR. For oocytes and blastocysts, a method based on agarose bead embedding was used. Methylation analysis of the paternally imprinted *H19* DMR was systematically performed on DNA of pools of oocytes and sperm to verify the lack of somatic contamination (data not shown). For sequencing, we cloned PCR fragments with the pCR2.1 Topo TA cloning kit (Invitrogen). Around 30 clones were analyzed per experiment. All PCR primers are listed in Supplemental Table S2.

ChIP quantitative and allelic analysis

ChIP was performed on chromatin after formaldehyde crosslinking from MEFs according to previously published protocols (Navarro et al., 2005). Details of the antisera used are provided as Supplemental Table S3. Quantitative analysis of immunoprecipitated DNA was determined in real-time PCR assays using the SYBR Green technology (Applied Biosystems). The fold enrichment of each target sequence was calculated as 2 to the power of the cycle threshold (Ct) difference between Input and antibody Bound fractions (% Input), normalized to those of the *Tbx15* housekeeping gene or to IAP sequences. Values given represent the average of three ChIP assays performed on independent chromatin preparations, one from BxC MEFs and two from CxB MEFs.

Allelic analyses were carried out by pyrosequencing with the PyroMark Q24 Pyrosequencer according to the manufacturer's instructions (Qiagen), after amplification using a HotStart Taq Polymerase (Promega). The relative level of the two parental alleles in each antibody bound fraction was quantified by the PyroMark Q24 1.0.10 Software

(Qiagen) using the allele quantification method. Assay-specific biases were controlled on Input DNA and H3 bound DNA.

RNA expression analysis

Total RNA was extracted from embryos and tissues (Trizol, Invitrogen). Northern blots were hybridized with PCR-generated probes in PerfectHyb buffer (Sigma). RNA integrity and loading were verified by probing the membrane with a β -actin probe (data not shown). For RT-PCR analyses, RNA was DNase-treated (Qiagen) prior to reverse-transcription with random hexamers (Superscript III, Invitrogen). Quantitative analysis of RNA expression was determined in real-time PCR assays using SYBR Green (Applied Biosystems), relative to *Rrm2* and *Arp0* genes as normalizers. Allelic analyses were carried out either by regular sequencing (Big Dye v3.1) or by pyrosequencing with the PyroMark Q24 Pyrosequencer, as described above. Assay-specific biases were controlled on BxC genomic DNA.

Supplemental References

- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol* 11, R106.
- Hiura, H., Sugawara, A., Ogawa, H., John, R. M., Miyauchi, N., Miyanari, Y., Horiike, T., Li, Y., Yaegashi, N., Sasaki, H., *et al.* (2010). A tripartite paternally methylated region within the Gpr1-Zdbf2 imprinted domain on mouse chromosome 1 identified by meDIP-on-chip. *Nucleic Acids Res* 38, 4929-4945.
- Jeffreys, A. J., Tamaki, K., MacLeod, A., Monckton, D. G., Neil, D. L., and Armour, J. A. (1994). Complex gene conversion events in germline mutation at human minisatellites. *Nat Genet* 6, 136-145.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079.
- Navarro, P., Pichard, S., Ciaudo, C., Avner, P., and Rougeulle, C. (2005). Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: implications for X-chromosome inactivation. *Genes Dev* 19, 1474-1484.
- Nix, D. A., Di Sera, T. L., Dalley, B. K., Milash, B. A., Cundick, R. M., Quinn, K. S., and Courdy, S. J. (2010). Next generation tools for genomic data generation, distribution, and visualization. *BMC Bioinformatics* 11, 455.
- Thierry-Mieg, D., and Thierry-Mieg, J. (2006). AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol* 7 Suppl 1, S12 11-14.

Article 2

The *Gpr1/Zdbf2* locus provides new paradigms
for transient and dynamic genomic imprinting
in mammals.

Duffié R., Ajjan S., Greenberg M.V., Zamudio N., Escamilla del Arenal M., Iranzo J.,
Okamoto I., Barbaux S., Fauque D. and Bourc'his D.

The *Gpr1/Zdbf2* locus provides new paradigms for transient and dynamic genomic imprinting in mammals

Rachel Duffié,¹ Sophie Ajjan,¹ Maxim V. Greenberg,¹ Natasha Zamudio,¹ Martin Escamilla del Arenal,¹ Julian Iranzo,¹ Ikuhiro Okamoto,^{1,4} Sandrine Barbaux,² Patricia Fauque,³ and Déborah Bourc'h^{1,5}

¹INSERM U934/CNRS UMR3215, Institut Curie, 75005 Paris, France; ²INSERM U1016/CNRS UMR8104, Institut Cochin, 75005 Paris, France; ³Hôpital de Dijon, Laboratoire de Biologie de la Reproduction, Université de Bourgogne, 21000 Dijon, France

Many loci maintain parent-of-origin DNA methylation only briefly after fertilization during mammalian development: Whether this form of transient genomic imprinting can impact the early embryonic transcriptome or even have life-long consequences on genome regulation and possibly phenotypes is currently unknown. Here, we report a maternal germline differentially methylated region (DMR) at the mouse *Gpr1/Zdbf2* (DBF-type zinc finger-containing protein 2) locus, which controls the paternal-specific expression of long isoforms of *Zdbf2* (*Liz*) in the early embryo. This DMR loses parental specificity by gain of DNA methylation at implantation in the embryo but is maintained in extraembryonic tissues. As a consequence of this transient, tissue-specific maternal imprinting, *Liz* expression is restricted to the pluripotent embryo, extraembryonic tissues, and pluripotent male germ cells. We found that *Liz* potentially functions as both *Zdbf2*-coding RNA and *cis*-regulatory RNA. Importantly, *Liz*-mediated events allow a switch from maternal to paternal imprinted DNA methylation and from *Liz* to canonical *Zdbf2* promoter use during embryonic differentiation, which are stably maintained through somatic life and conserved in humans. The *Gpr1/Zdbf2* locus lacks classical imprinting histone modifications, but analysis of mutant embryonic stem cells reveals fine-tuned regulation of *Zdbf2* dosage through DNA and H3K27 methylation interplay. Together, our work underlines the developmental and evolutionary need to ensure proper *Liz/Zdbf2* dosage as a driving force for dynamic genomic imprinting at the *Gpr1/Zdbf2* locus.

[Keywords: genomic imprinting; DNA methylation; epigenetics; mouse development]

Supplemental material is available for this article.

Received October 12, 2013; revised version accepted January 22, 2014.

The process of germline differentiation produces mature gametes with sexually dimorphic morphologies, functions, and epigenomes. In mammals, the sperm and oocyte genomes display notably different DNA methylation density and distribution (Smallwood et al. 2011; Kobayashi et al. 2012a; Smith et al. 2012) despite reliance on the same enzymatic machinery formed by the de novo DNA methyltransferase Dnmt3A and its cofactor, Dnmt3L (Bourc'h et al. 2001; Bourc'h and Bestor 2004; Kaneda et al. 2004). Intergenic regions and transposable repeats are densely methylated in sperm, while DNA methylation preferentially occurs inside genes in the oocyte, with deeper density at CpG islands (CGIs) coinciding with intragenic promoters. Genetic and genome-wide studies have suggested a role for active transcription

in oocyte DNA methylation whereby transcription promotes DNA methylation deposition in gene bodies (Chotalia et al. 2009; Kobayashi et al. 2012a). In sperm, small RNAs are instead responsible for targeting methylation to transposons (Aravin and Bourc'h 2008).

Given the heritable nature of DNA methylation, gametic sex-specific patterns are transmitted upon fertilization; this creates tremendous opportunity for parental asymmetry in the progeny (Duffié and Bourc'h 2013). However, as part of the early embryonic program, sequential events of DNA demethylation and remethylation remodel most gametically inherited methylation. Only a handful of regions endowed with favorable nucleotide sequence and chromatin environments are capable

⁴Present address: Department of Anatomy and Cell Biology, Kyoto University, Kyoto 606-8501, Japan.

⁵Corresponding author

E-mail deborah.bourc'h@curie.fr

Article is online at <http://www.genesdev.org/cgi/doi/10.1101/gad.232058.113>.

© 2014 Duffié et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genesdev.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported), as described at <http://creativecommons.org/licenses/by-nc/3.0/>.

of maintaining parent-of-origin DNA methylation after fertilization. These include regions subject to genomic imprinting. Around 20 imprinting control regions (ICRs) have been identified to date, and these dictate the parent-specific monoallelic expression of ~130 genes in *cis* (Schulz et al. 2008).

All known ICRs harbor binding motifs for ZFP57, a zinc finger protein that recruits the KRAB-associated protein 1 (KAP1)-dependent heterochromatin complex (Schultz et al. 2002; Quenneville et al. 2011). This property confers the ability of ICRs to maintain methylation on one parental allele during the genome-wide demethylation that accompanies preimplantation development (Li et al. 2008; Quenneville et al. 2011; Messerschmidt et al. 2012; Zuo et al. 2012). What protects the unmethylated allele of ICRs from global de novo methylation after implantation has not been determined precisely; occupancy by *trans*-acting factors and/or features linked to active transcription such as permissive histone modifications and R-loop structures are likely to be involved (Ooi et al. 2007; Lienert et al. 2011; Ginno et al. 2012). Once these critical periods of intense DNA methylation remodeling have reached completion, canonical ICRs perpetuate as parental differentially methylated regions (DMRs) in a life-long manner.

We recently uncovered new forms of genomic imprinting, which exist only transiently during preimplantation development or specifically in certain tissues during adulthood (Proudton et al. 2012). The underlying regions all have ZFP57-binding motifs and are protected against DNA demethylation after fertilization. However, they are susceptible to processes of de novo methylation that occur in the post-implantation embryo or during tissue differentiation and consequently lose imprinting by gain of DNA methylation. Transient genomic imprinting is intriguing: Whether short-term differential methylation is only a byproduct of the divergent sperm and oocyte methylomes or could functionally impact mammalian genome regulation is currently unknown.

Here we report the case of a region of transient germline DMRs (gDMRs) inherited from the oocyte at the mouse *Gpr1/Zdbf2* (DBF-type zinc finger-containing protein 2) locus (1qC2) (Kobayashi et al. 2009). We found that this maternal gDMR coincides with a promoter, which initiates transcription of long isoforms of *Zdbf2* (*Liz*) in a paternal-specific manner but is shut down upon DNA methylation gain after implantation. Importantly, *Liz* transient expression is associated with acquisition of a paternal somatic DMR (sDMR) and paternal-specific transcription of *Zdbf2* from its canonical promoter, both of which are stably maintained for the rest of life. We further reveal the ubiquitous availability of *Liz/Zdbf2* transcripts throughout development, the fine-tuned regulation of *Zdbf2* promoter activity by interplay between DNA methylation and histone modifications, and the conservation of *GPR1/ZDBF2* gene regulation in humans. In conclusion, we demonstrate for the first time the potential for short-term and long-term effects of transient genomic imprinting on mammalian genome regulation.

Results

Two gDMRs at the Gpr1/Zdbf2 locus exhibit dynamic allele-specific methylation during development

The *Gpr1/Zdbf2* locus was first described to be paternally imprinted, with three intergenic paternally methylated DMRs (DMR1, DMR2, and DMR3) that are located between 8.5 and 16 kb upstream of the *Zdbf2* transcription start site (TSS) and can be considered as a single entity (Fig. 1A; Hiura et al. 2010). However, recent work, including our own, suggested that this locus may be primarily under maternal imprinting control (Kobayashi et al. 2012b; Proudton et al. 2012). Accordingly, we found that the originally described paternal DMR lacks important hallmarks of all ICRs identified so far: TGCCGC motifs for ZFP57 binding and local ZFP57/KAP1 enrichment, as seen in available embryonic stem (ES) cell chromatin immunoprecipitation (ChIP) sequencing (ChIP-seq) data (Quenneville et al. 2011). Instead, ZFP57/KAP1 enrichment was found at a CGI in the second intron of the closest neighboring gene, *Gpr1* (G protein-coupled receptor 1), and localized ~65 kb upstream of the paternal DMR (Supplemental Fig. S1A). An *in silico* search allowed us to uncover three ZFP57 recognition motifs at this intragenic CGI.

To resolve the parental origin of imprinting control at the *Gpr1/Zdbf2* locus, we analyzed DNA methylation by bisulfite cloning/sequencing through mouse development. Crosses between the C57Bl6/J (B) and CAST/Ei (C) mouse strains permitted parental distinction based on the presence of single nucleotide polymorphisms (SNPs). Investigation of *Dnmt3L*^{-/-} embryos derived from DNA methylation-free *Dnmt3L*^{-/-} oocytes further allowed us to infer the role of maternally inherited DNA methylation (Bourc'his et al. 2001). Reciprocal gDMRs were identified, with the paternal DMR harboring sperm-specific methylation (paternal gDMR) and the ZFP57/KAP1 site harboring oocyte-specific methylation (maternal gDMR) (Fig. 1A; Supplemental Fig. S1B). Of note, the maternal gDMR bisulfite amplicon includes the three ZFP57 sequence motifs, which are conserved among our different mouse strains. In embryonic day 3.5 (E3.5) blastocysts, DNA methylation remained at maternal alleles of the maternal gDMR, likely due to protection by the ZFP57/KAP1 complex. In contrast, the paternal gDMR underwent demethylation and loss of parent specificity.

After implantation, the two parental gDMRs were subjected to de novo DNA methylation. At the maternal gDMR, gain of methylation on the paternal allele abolished maternal specificity by E9.5 (Fig. 1A; Supplemental Fig. S1B). Full biallelic methylation was consistently maintained in all somatic tissues examined (fibroblasts, kidney, muscle, cortex, pituitary gland, and liver) (Fig. 1B; data not shown). This classifies this locus as a transient maternal gDMR, only present during preimplantation development (Proudton et al. 2012). As previously observed for transient gDMRs, de novo methylation occurred indiscriminately on both parental alleles in *Dnmt3L*^{-/-} embryos, which lacked methylation at the

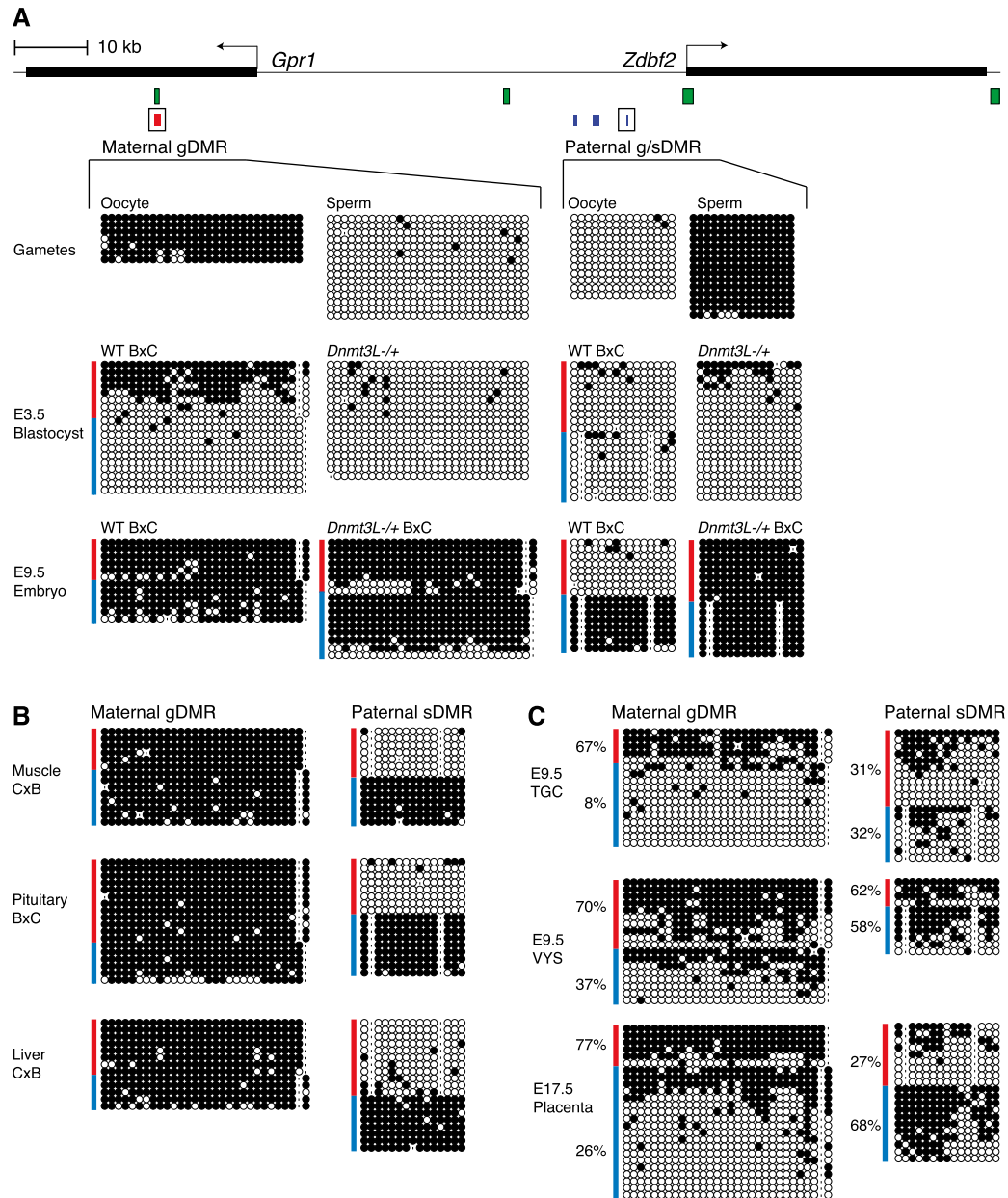


Figure 1. DNA methylation profiling at two parental DMRs of the *Gpr1/Zdbf2* locus during mouse development. (A) Mouse *Gpr1/Zdbf2* locus indicating CGIs (green bars) and positions of maternal DMRs (red bar) and tripartite paternal DMRs (blue bars) (DMR1, DMR2, DMR3 from left to right). DNA methylation was analyzed by bisulfite cloning/sequencing in gametes and embryos of wild-type (WT) and maternal methylation-free *Dnmt3L*^{-/+} background (A), adult somatic tissues (B), and extraembryonic tissues (C). The maternal DMR is a transient gDMR: It acquires methylation in oocytes, maintains maternal-specific DNA methylation during preimplantation development, and is lost by methylation gain after implantation, except in extraembryonic tissues. Of note, CpGs in positions 2, 5, and 7 of the amplified maternal gDMR amplicon are embedded into hexanucleotidic ZFP57-binding motifs. The paternal DMR is methylated in sperm and lost in the blastocyst but is re-established at implantation and provides paternal-specific information in embryonic and adult somatic tissues, qualifying as both gDMR and sDMR (g/sDMR). Results for paternal DMR3 are depicted here; similar patterns were obtained with DMR2 (data not shown). (Red lines) Maternal alleles; (blue lines) paternal alleles; (white circles) unmethylated CpG; (black circles) methylated CpG; (dash) absent CpG due to SNP or, rarely, sequencing error. Mouse strains used were as follows: C57Bl6/J (designated B) and CAST/Ei (designated C).

maternal gDMR before implantation (Fig. 1A; Supplemental Fig. S1B; Proudhon et al. 2012).

Remarkably, from a totally methylation-free status, the paternal DMR returned to paternal-specific methylation

after implantation (Fig. 1A; Supplemental Fig. S1B). This classifies this paternal DMR as both a gDMR established in sperm and a sDMR established post-fertilization. Paternal-specific methylation at the sDMR was extremely stable,

as it was maintained in all adult tissues examined (Fig. 1B; data not shown). Interestingly, in the *Dnmt3L*^{-/+} context, both parental alleles were targeted for de novo methylation, and the locus was fully methylated by E9.5. This demonstrates that paternal sDMR specificity depends on maternally inherited DNA methylation. Finally, the two CGIs flanking the paternal DMR—one encompassing the *Zdbf2* promoter and the other at an intergenic location ~25 kb upstream—were consistently devoid of DNA methylation throughout development and adulthood in both the wild-type and *Dnmt3L*^{-/+} contexts (Supplemental Fig. S1C).

De novo methylation after implantation globally spares the extraembryonic lineage (Monk et al. 1987). Accordingly, we found that the maternal gDMR was maintained in a preimplantation, blastocyst-like pattern in extraembryonic tissues: Preferential maternal methylation and paternal hypomethylation were observed in trophoblast giant cells (TGCs) and visceral yolk sacs (VYSs) at E9.5 and later, during gestation, in E17.5 placenta (Fig. 1C). Some de novo methylation occurred at the paternal sDMR, but while the paternal allele reached full methylation in the embryo, the density was lower in extraembryonic tissues (30%–60%), and maternal alleles were also targeted, suggesting relaxed imprinted methylation at the paternal sDMR in these tissues.

In summary, two DMRs at the *Gpr1/Zdbf2* locus are targeted by de novo DNA methylation in a parent-of-origin manner during gametogenesis. During preimplantation, only the maternal gDMR is maintained, placing the locus under maternal imprinting. In post-implantation embryos, a switch in parent-of-origin information occurs in the wake of embryonic de novo DNA methylation: The paternal DMR is somatically reestablished at E6.5, just before the maternal gDMR disappears by methylation gain (Supplemental Fig. S1B). Once acquired, the paternal sDMR provides lifelong, paternal specificity to the locus. Extraembryonic tissues are an exception, where maternal specificity is maintained at the maternal gDMR.

Characterization of *Liz*, a transcript linking the two DMRs in ES cells

We next explored ES cells as a model system to investigate the molecular events that underlie the DNA methylation dynamics at the *Gpr1/Zdbf2* locus in embryos. Early passages of ES cells derived in “ground-state” pluripotency 2i conditions (Ying et al. 2008) showed faithful maintenance of blastocyst methylation patterns; i.e., maternal-specific methylation of the maternal gDMR and lack of methylation at the paternal sDMR (Fig. 2A). However, the maternal gDMR rapidly lost DNA methylation upon passage (above passage 5 [P5]) or differentiation, a feature we systematically observed in several independent 2i lines. Furthermore, biallelic methylation was acquired at the paternal sDMR upon differentiation, as observed by day 4 (D4) of retinoic acid (RA) treatment (Fig. 2A). These data underscore, as in *in vivo* embryos, the dependence of the paternal sDMR methyl-

ation upon the maternal gDMR methylation status. ES cells grown in classic serum conditions showed even more aberrant methylation patterns, already demonstrating full paternal sDMR methylation at undifferentiated states (D0) (Supplemental Fig. S2A). In conclusion, no ES cell model is likely to fully recapitulate the epigenetic changes that occur *in vivo* at the *Gpr1/Zdbf2* locus. However, despite loss of imprinting, we were able to use 2i ES cells as a model system to understand what may signal de novo methylation at the sDMR during embryonic differentiation.

Many sDMRs have been proposed to gain methylation as a result of their position along the path of transcripts controlled by promoter-associated gDMRs (John and Lefebvre 2011). Transcriptome reconstruction from RNA sequencing (RNA-seq) data shows evidence of a plethora of transcriptional activity arising from the maternal gDMRs (extending antisense to *Gpr1* and sense to *Zdbf2*) and present in ES cells but not in differentiated cell types (Guttman et al. 2010). Using RT-PCR and 3' and 5' RACE in 2i ES cells, we confirmed the presence of a TSS within the maternal gDMR, identified additional exons and splicing events, and demonstrated that these transcripts can be polyadenylated (Fig. 2B). While the shortest variants stopped before or within the paternal sDMR domain, the longest extended over and linked the *Gpr1* and *Zdbf2* genes, covering ~100 kb in genomic distance. These long transcripts shared multiple exons with *Zdbf2*, which are of relatively small sizes (110 base pairs [bp] on average), except the large 12-kb exon 7 and 3' untranslated region (UTR). This led us to postulate that the maternal gDMR coincides with an alternative promoter for *Zdbf2* transcription, located ~73 kb upstream of the annotated RefSeq *Zdbf2* TSS. We termed these extended transcripts *Liz*, for long isoforms of *Zdbf2*. *Liz* spliced variants contain at least two 5' exons that do not overlap with *Zdbf2*, and *Zdbf2* exons 1 and 2 were never incorporated in *Liz* transcripts (Fig. 2B).

Northern blotting with a probe specific for *Liz* detected a unique band, slightly larger than 12 kb (Fig. 2C), indicative of a long spliced variant extending over the last exon of *Zdbf2*, not significantly different in size from canonical *Zdbf2* mRNA. RT-qPCR confirmed that spliced forms of *Liz* were fivefold to 10-fold more abundant than unspliced forms, with ratios similar to protein-coding transcripts (data not shown). Comparison with a probe recognizing both *Zdbf2* and *Liz* revealed dynamic expression changes over RA-induced ES cell differentiation. In undifferentiated D0 and early D2 differentiating ES cells, *Liz* was the dominant transcript. As *Liz* levels declined in abundance by differentiation D4, the canonical form of *Zdbf2* became up-regulated, suggesting a switch from *Liz* to *Zdbf2* promoter usage. Using primers specific to *Liz* or *Zdbf2*, we confirmed the reverse abundance of *Liz* and *Zdbf2* RNAs before and after differentiation in multiple differentiation conditions (Fig. 2D; data not shown).

In ES cell ChIP-seq data sets, the *Liz* TSS is occupied by RNA polymerase II (Pol II) and the pluripotency transcription factor Sox2, which could explain pluripotency-associated *Liz*

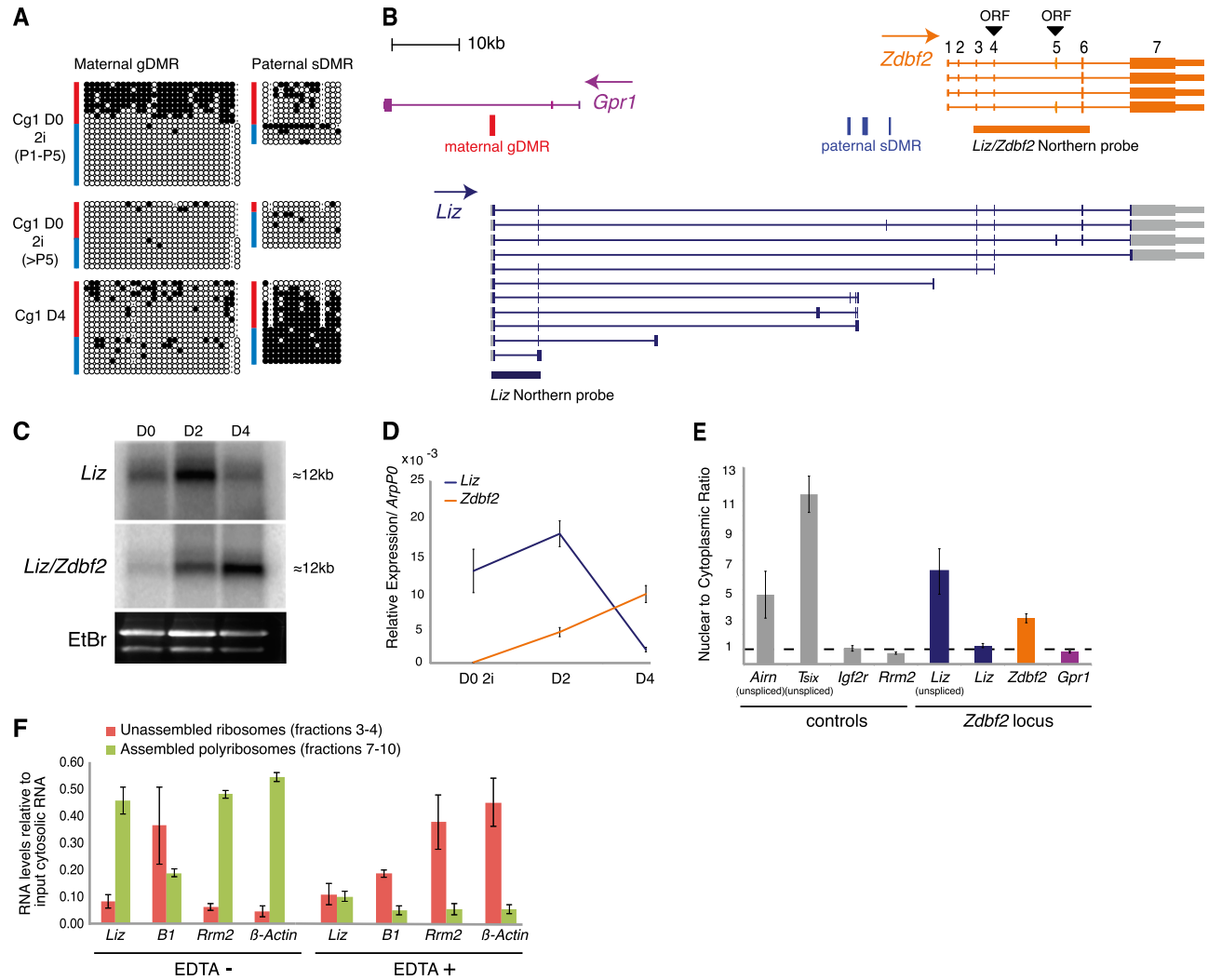


Figure 2. Identification of *Liz* transcripts in ES cells. (A) Bisulfite analysis of maternal gDMR and paternal sDMR methylation in 2i-derived ES cells (Cg1, Cx129 hybrid background) in undifferentiated states (D0) before and after five passages (P5) and after differentiation (D4). (B) Transcription map of the mouse *Gpr1/Zdbf2* locus, with positions of putative ORFs for *Zdbf2* translation. Transcripts originating from the maternal gDMR were characterized by sequencing of RT-PCR and RACE products from 2i ES cell RNA. Many transcripts incorporate *Zdbf2* exons, called long isoforms of *Zdbf2* (*Liz*) (blue). (C) Northern blot analysis over ES cell differentiation using a probe specific to *Liz* and a probe that recognizes *Liz* and *Zdbf2* (probe positions in B). (D) RT-qPCR quantification of *Liz* (*Liz* exons 1 and 2) or *Zdbf2* (*Zdbf2* exons 1–3) transcripts confirms reverse abundance of these transcripts over differentiation. Error bars indicate SEM of two individual ES cell lines: Cg1 and gC6. (E) RT-qPCR quantification of nuclear-to-cytoplasmic RNA enrichment. Unless otherwise specified, primers detecting spliced RNA forms (spanning introns) were used. Error bars are as above. (F) RT-qPCR quantification of spliced *Liz* levels after sucrose gradient fractionation of cytosolic RNA without or with EDTA treatment (EDTA– and EDTA+). Unassembled ribosomal fractions (3 and 4) and assembled polyribosomal fractions (7–10) were determined by 28S/18S ratio measurement. Error bars indicate SEM of two biological 2i ES cell grown replicates.

expression (Supplemental Fig. S2C; Mikkelsen et al. 2008). Importantly, using an RT-PCR pyrosequencing assay exploiting strain-specific SNPs, we found *Liz* to be dominantly paternally expressed in early passage 2i ES cells (Supplemental Fig. S2B), which harbor maternal-specific methylation at the maternal gDMR/*Liz* promoter (Fig. 2A). Upon passages, as methylation was lost, *Liz* expression became biallelic, and this coincided with biallelic methylation at the paternal sDMR later upon differentiation (Figs. 1A, 2A; Supplemental Fig. S2B). As the paternal sDMR lies within the path of *Liz* transcription, we

hypothesize that *Liz* mediates the effect of the maternal gDMR on the downstream paternal DMR methylation via either the act of transcription or the RNA product itself.

Liz and *Zdbf2* transcripts have similar coding potential

Liz was previously reported to be a long intergenic RNA with no coding potential (GenBank accession no. AB777270) (Guttman et al. 2010, 2011; Kobayashi et al. 2012b). While this may be true for the short isoforms that

do not extend into the *Zdbf2* gene, we found the long isoforms containing multiple *Zdbf2*-derived exons to be the most abundant isoforms in ES cells. ZDBF2 is a protein of unknown function. The C2H2-type zinc finger motif at its N terminus is the only annotated domain (pfam07535) that relates to the budding yeast DBF4 (dumbbell-forming protein 4) protein. Two putative ORFs exist within the RefSeq mouse *Zdbf2* sequence, initiated in exon 4 or exon 5 (Fig. 2B), and have similarly high coding potential calculator (CPC) scores (23.1 vs. 23.2) (Kong et al. 2007). Although the main initiating codon is reported to be in exon 5 (University of California at Santa Cruz [UCSC] and Ensembl genome browsers), the corresponding ORF lacks the complete zinc finger motif (UniProt Q5S200). Importantly, a large fraction of canonical *Zdbf2* and nearly all *Liz* spliced transcripts skip exon 5; this suggests that the exon 4 initiating codon is frequently accessible, which would potentiate translation of a 2498-amino-acid-long protein containing the conserved DBF4 zinc finger (Supplemental Fig. S2D). Importantly, *Liz* and canonical *Zdbf2* exon 4-derived ORF scores are nearly identical (22.9 vs. 23.1), making it likely that *Liz* is an alternative source of zinc finger-containing *Zdbf2* protein during development, with a different 5' UTR region.

To further characterize the translational properties of *Liz*, we performed subcellular fractionation of RNA from D2 differentiating ES cells. The known noncoding *Tsix* and *Airn* RNAs showed high nuclear to cytoplasmic ratios in RT-qPCR assays, in contrast to the protein-coding *Igf2r* and *Rrm2* RNA controls, which were dominantly cytoplasmic, thus potentiating their translation (Fig. 2E; Seidl et al. 2006; Nora et al. 2012). Using *Liz*-specific primers, we found unspliced forms of *Liz* to be retained in the nucleus, while spliced forms were preferentially exported to the cytoplasm, suggesting potential ribosome association (Fig. 2E). Importantly, *Liz* cytoplasmic enrichment was higher compared with canonical spliced *Zdbf2* at this stage of differentiation. Using sucrose gradient fractionation of cytosolic RNAs, we demonstrated that spliced *Liz* RNAs are indeed bound to translating ribosomes (Fig. 2F). The *Liz* polyribosome-association profile was similar to that of the protein-coding *Rrm2* and β -actin RNAs, while the SINEs B1 retrotransposon non-coding transcripts were preferentially found in unassembled ribosome fractions. Moreover, *Liz* RNA showed a clear redistribution toward unassembled ribosomes upon EDTA treatment, which disrupts polysomes, further confirming the specificity of the polyribosome association and the possibility that *Liz* is efficiently translated in ES cells. In summary, while *Liz* RNA may have cis-acting function on DNA methylation, its conserved coding potential and polyribosome association strongly suggest that it may dually function as a message for the *Zdbf2* protein.

Liz and *Zdbf2* promoter usage is dynamically regulated during embryonic development

We hypothesized that the switch from *Liz* to canonical *Zdbf2* transcription observed upon ES cell differentiation

would also occur in vivo during early embryonic development. RT-qPCR analysis revealed that the *Zdbf2* mRNA is abundantly produced in the oocyte and transmitted upon fertilization (Fig. 3A). While the maternal *Zdbf2* transcript was rapidly degraded, *Liz* was activated and accumulated in the pluripotent preimplantation embryo from the eight-cell stage to the blastocyst stage. As in ES cells, the long exon 4-containing form was produced during this period. In the E6.5 epiblast, *Liz* levels sharply dropped, and canonical *Zdbf2* transcripts became the prevailing form in post-implantation embryos, consistent with what was observed in vitro in differentiating ES cells. Therefore, *Zdbf2*-related transcripts are continuously present in the early embryo but originate from distinct sources: first from the canonical *Zdbf2* promoter, then from *Liz*, and finally, from *Zdbf2* again.

To visualize *Liz* allelic expression on a single-cell level, we performed RNA fluorescent in situ hybridization (FISH) of nascent *Liz* transcripts with an intronic BAC probe, spanning the intergenic region 3' of *Gpr1* and 5' of the *Zdbf2* promoter. Single pinpoints indicative of mono-allelic expression were visible in blastomeres of E3.5 blastocysts from a pure C57Bl6/J background (Fig. 3B). *Liz* expression was not lineage-restricted, as we observed predominant monoallelic signals in both the inner cell mass (ICM) and the trophectoderm (TE). Determination of allelic specificity by RT-PCR pyrosequencing further revealed that *Liz* is paternally expressed in blastocysts (Fig. 3C), concordant with maternal methylation of the *Liz* promoter at this stage (Fig. 1A). The repressive effect of DNA methylation was further demonstrated by biallelic expression of *Liz* in maternal methylation-free *Dnmt3L*^{-/+} E3.5 blastocysts (Fig. 1A), as seen by RNA FISH and RT-pyrosequencing. Biallelic expression resulted in a doubling of *Liz* mRNA levels (Fig. 3B,C).

After implantation, in E8.5 embryos, as paternal sDMR methylation is acquired (Figs. 1A, 3A), we also found *Zdbf2* to be paternally expressed (Fig. 3D). We validated the positive correlation between intergenic paternal sDMR methylation and the canonical *Zdbf2* promoter activity with the observation of biallelic and twofold increase of *Zdbf2* transcripts in E8.5 *Dnmt3L*^{-/+} embryos (Fig. 3D), where paternal sDMR methylation is biallelic (Fig. 1A; Supplemental Fig. S1B). Moreover, reactivation of the normally silent maternal allele of *Zdbf2* in the *Dnmt3L*^{-/+} background is strong evidence that the hypomethylated state of the *Liz* promoter and/or *Liz* expression before implantation is associated with *Zdbf2* promoter activation after implantation.

Liz expression occurs in extraembryonic tissues and during spermatogenesis

We reasoned that *Liz* should also exist in male germ cells and extraembryonic tissues, two contexts outside of the preimplantation window where the *Liz* promoter is totally or partially DNA methylation-free (Fig. 1A). We successfully amplified *Liz* mRNA by RT-qPCR in testes and placenta in levels similar to 2i ES cells but not in any

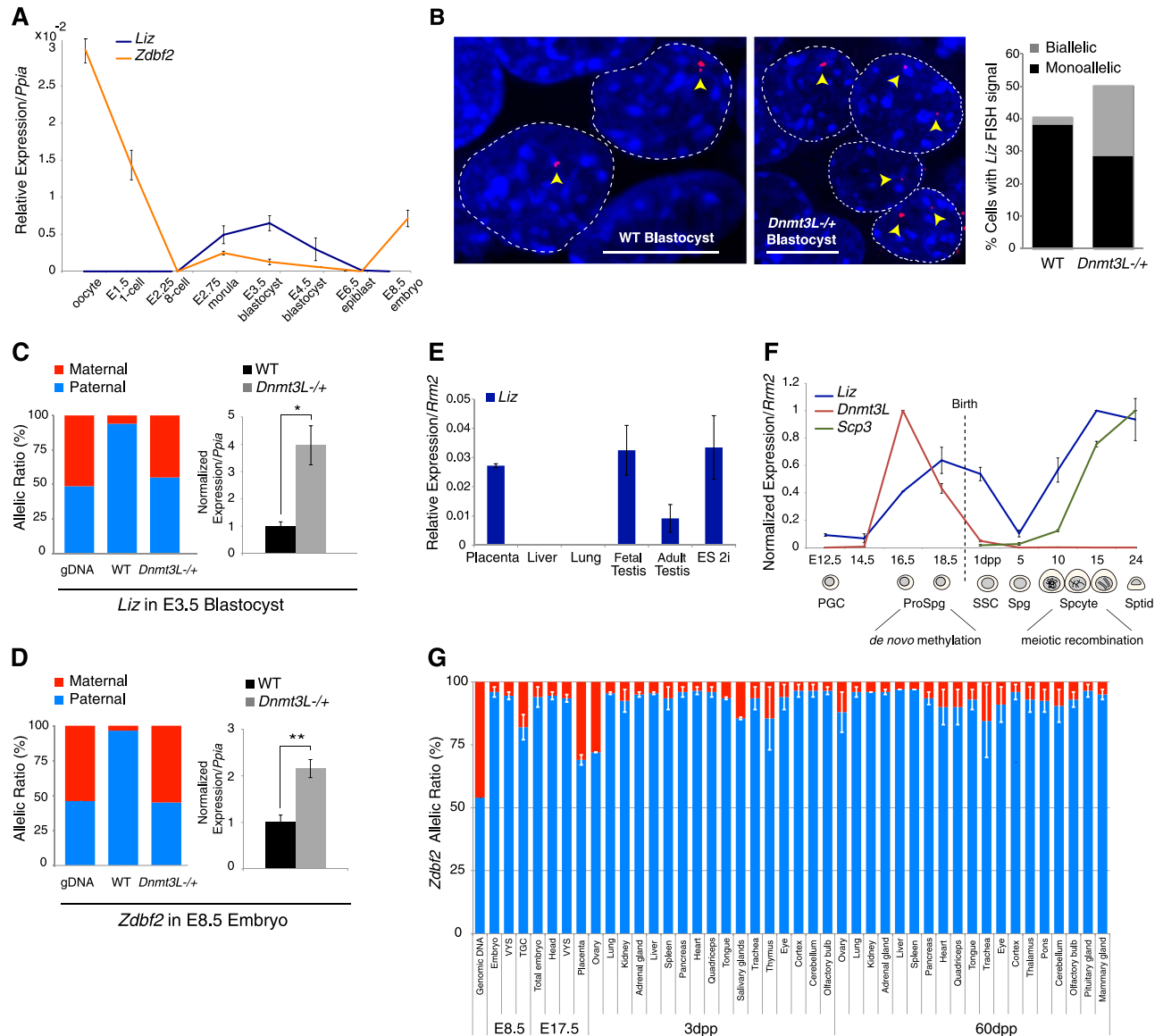


Figure 3. *Liz* and *Zdbf2* expression is dynamically regulated during mouse embryonic development. (A) *Liz*- and *Zdbf2*-specific detection by RT-qPCR during mouse embryonic development. Error bars indicate SEM of technical replicates in pooled oocytes or embryos. (B) RNA FISH detection of nascent *Liz* transcripts in wild-type (WT) and *Dnmt3L*^{-/-} E3.5 blastocysts. (Left panel) Representative microscopy images. Yellow arrows point to *Liz* FISH signals (red). Dashed white lines indicate nucleus borders delineated by DAPI staining (blue). Bar, 5 μ m. (Right panel) Cell counting of biallelic and monoallelic *Liz* signals. Wild type, *n* = 25 blastocysts/1488 cells; *Dnmt3L*^{-/-}, *n* = 13 blastocysts/659 cells. (C, left panel) RT-PCR pyrosequencing detects *Liz* paternal-specific expression in E3.5 BxC wild-type blastocysts and biallelic expression in *Dnmt3L*^{-/-} blastocysts. Genomic DNA (gDNA) was used to exclude assay-specific biases. (Right panel) RT-qPCR shows increased *Liz* levels in *Dnmt3L*^{-/-} blastocysts. Error bars indicate SEM technical replicates of pooled blastocysts. (*) *P* < 0.03, Student's *t*-test. (D) Same analyses as in C, for *Zdbf2* in E8.5 wild-type and *Dnmt3L*^{-/-} embryos. (**) *P* < 0.002, Student's *t*-test. (E) RT-qPCR analysis of *Liz* expression in various mouse tissues. Error bars indicate SEM of technical replicates. (F) RT-qPCR analysis of *Liz* expression over spermatogenesis. Levels were measured in testes at nine different ages, from embryonic (E) to postnatal stages (days post-partum [dpp]), to follow representative male germ cell types appearing during the course of the first spermatogenesis wave, as indicated below the graph. *Dnmt3L* and *Scp3* expression peaks denote periods of de novo DNA methylation and meiotic recombination, respectively. Highest values for each transcript are set to 1. Error bars are as above. (G) Stability of *Zdbf2* paternal-specific expression measured by RT-PCR pyrosequencing in tissues collected in embryonic E8.5 and E17.5 stages and after birth at 3 and 60 dpp. Error bars indicate SEM in reverse cross BxC and CxB tissues.

other tested somatic or postnatal tissue (Fig. 3E). Because we were able to amplify long *Liz* transcripts that extend over the *Zdbf2* RefSeq gene and include exon 4, we concluded that, like in early embryonic cells, *Liz* mRNA can serve as a source of *Zdbf2* protein in the placenta and testis.

As a maternally imprinted gDMR, the *Liz* promoter does not gain methylation in the male germline. Accordingly, *Liz* expression was biallelic in fetal and adult testes (Supplemental Fig. S3A). Interestingly, *Liz* expression exhibited developmentally regulated dynamics throughout

spermatogenesis (Fig. 3F). In fetal testes, *Liz* transcripts accumulated from E14.5 to E18.5, concomitant with the wave of male germline de novo DNA methylation (Kato et al. 2007) and in particular with germline methylation of the paternal DMR (Hiura et al. 2010). After birth, *Liz* transcription increased at the onset of meiosis. Considering the consistently fully methylated status of the *Liz* promoter in all somatic cell types, *Liz* expression in the testis is likely to emanate from germ cells. This was confirmed in sorted primordial germ cells from fetal testes (Supplemental Fig. S3B). Additionally, in *Dnmt3L* mutant adult testes, which exhibit a severe germ cell depletion phenotype (Bourc'his and Bestor 2004), *Liz* expression was dramatically reduced, behaving as typical germ cell markers such as *Vasa* (Supplemental Fig. S3C). In extraembryonic tissues, maternal-specific methylation is globally maintained at the *Liz* promoter. Accordingly, a strong paternal bias in *Liz* expression was found in hybrid E9.5 VYSs and TGCs and E17.5 placenta (Fig. 1A; Supplemental Fig. S3D). *Liz* expression therefore occurs in preimplantation embryos, extraembryonic tissues throughout gestation, and male germ cells, all of which are types in which the *Liz* promoter is not completely obstructed by DNA methylation. Besides these specific contexts, we only observed use of the *Zdbf2* promoter.

Stability and extent of imprinted control by the paternal sDMR

As *Zdbf2* monoallelic expression occurs without allele-specific DNA methylation at its TSS (Supplemental Fig. S1C), we questioned the stability of *Zdbf2* paternal-specific expression throughout life. To this end, we conducted quantitative RT-PCR pyrosequencing using a bank of 50 hybrid tissues from reciprocal BxC and CxB crosses collected at fetal, neonatal, and adult stages. According to microarray profiling data (BioGPS), *Zdbf2* is most highly expressed in brain tissues. We found paternal *Zdbf2* expression bias to be extremely stable throughout life and in a tissue-wide manner, independent of *Zdbf2* expression level (Fig. 3G).

As many imprinted DMRs exert their control over large gene clusters (Barlow 2011), we further assessed the imprinted status of the three most proximal genes relative to *Liz/Zdbf2*. We first focused on the *Gpr1* gene (45 kb 5' of the paternal sDMR), which houses the transient maternal gDMR and was formerly reported to have kidney-specific paternal expression (Hiura et al. 2010). Despite an extensive search, we could not find evidence of parent-specific expression in kidney or other somatic tissues, in extraembryonic tissues where *Gpr1* expression levels are the highest (BioGPS), or in late blastocysts, where we detected a slight peak of *Gpr1* expression (Supplemental Fig. S3E–H). Biallelic expression was also consistently observed tissue- and stage-wide for *Eef1b2* (87 kb 5' of the sDMR) and *Adam23* (182 kb 3' of the sDMR) despite a report of moderately biased paternal expression in fetal brains for the latter (Supplemental Fig. S3E; DeVeale et al. 2012). Contrary to previous conclu-

sions, our findings suggest that the two regions of differential DNA methylation of the *Gpr1/Zdbf2* locus, the maternal gDMR and the paternal sDMR, are exclusively linked to the local regulation of *Zdbf2* imprinted expression in both its long (*Liz*) and canonical forms.

LIZ expression and coding potential are evolutionarily conserved in humans

As an indication of functional relevance of *Liz*, we checked its conservation from mice to humans. Paternal-specific methylation at the paternal g/sDMR and paternal *ZDBF2* expression were previously described in humans, showing imprinting at the syntenic *GPR1/ZDBF2* locus (2q33.3) (Kobayashi et al. 2009; Hiura et al. 2010). We designed RT-qPCR primers that mapped to a predicted transcript (Aceview nehura.cAug10), which arises from a relatively CpG-rich region within *GPR1* intron 2 and is antisense to *GPR1* (similar to mouse *Liz*) but is interrupted before the *ZDBF2* RefSeq sequence (Fig. 4A). As in mice, we detected this transcript in the placenta and testis but not in somatic adult tissues (Fig. 4B). Unlike in mice, however, we failed to measure significant expression in different human ES lines.

We confirmed the production of short truncated versions of *LIZ* in both the placenta and fetal testis by RT-PCR (Fig. 4A,C). Importantly, primers anchored in the first putative exon of *LIZ* and the last exon of *ZDBF2* amplified long multiexonic *LIZ* variants in placenta but not in testes. These extended over the full *ZDBF2* RefSeq sequence, a feature that we confirmed by the detection of a >12-kb band by Northern blotting, which indicates the frequent inclusion of the large last exon/3' UTR of *ZDBF2* (Supplemental Fig. S4A). As in mice, ATG codons exist in exons 4 and 5 of *ZDBF2*, but only exon 4 can potentiate a zinc finger-containing *ZDBF2* ORF (Supplemental Fig. S2D). Exon 4 is present in the majority of *LIZ* forms (Fig. 4A); we concluded that *LIZ* also has the coding potential for a functional *ZDBF2* protein in humans, at least in placenta.

Bisulfite-based methylation analysis around the putative *LIZ* promoter revealed absence of methylation in sperm, full allelic methylation in somatic tissues (liver), and differentially methylated alleles in placenta, indicative of a DMR, although we could not infer parental origin due to the lack of informative SNPs in our samples (Fig. 4D; Supplemental Fig. S4B). Therefore, DNA methylation at the putative human *LIZ* promoter shows developmental dynamics similar to the mouse *Liz* promoter, reminiscent of a tissue-specific maternal gDMR maintained in extraembryonic lineages only. However, unlike in mice, the *LIZ* promoter was fully methylated in several human ES cell lines (Fig. 4E; Supplemental Fig. S4B), in agreement with the more advanced, post-implantation-like state of human versus mouse ES cells (McEwen et al. 2013). These results explain our ability to detect *LIZ* transcription in the testis and placenta but not in human ES cells.

Finally, we analyzed the developmental dynamics of DNA methylation at the paternal DMR in humans. The

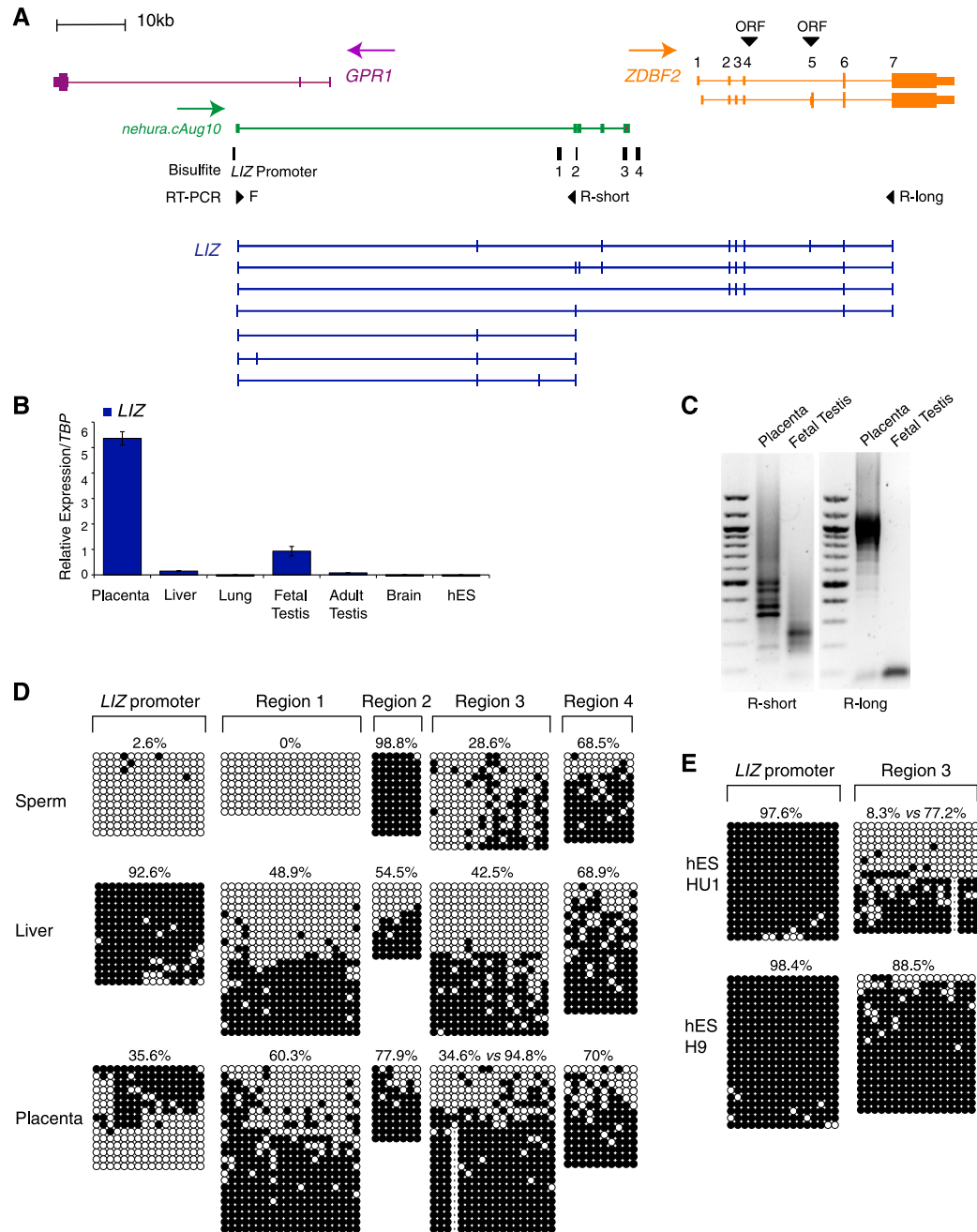


Figure 4. DNA methylation and expression at the human *GPR1/ZDBF2* locus. (A) Map of the *GPR1/ZDBF2* locus indicating RT-PCR primers designed for a predicted transcript (green) that detect multiple *LIZ*-like transcripts (blue) in the placenta, including isoforms sharing exons with *ZDBF2* and incorporating *ZDBF2* putative ORFs. (B) RT-qPCR analysis of *LIZ* in human tissues and ES cells. Error bars indicate SEM of technical replicates for tissues and HU1 and H9 biological replicates for ES cells. (C) RT-PCR of short and long forms of *ZDBF2* in placenta and fetal testes. (D) Bisulfite-based DNA methylation analysis in the sperm, liver, and placenta of the *Liz* promoter and four intergenic regions (regions 1–4) upstream of the *ZDBF2* promoter (primer positions in A). (E) DNA methylation in HUES1 and H9 ES cells.

25-kb intergenic region upstream of the *ZDBF2* promoter exhibits low conservation across 32 eutherian mammals (PhyloP) (Pollard et al. 2010) compared with the entire *GPR1/ZDBF2* locus (Supplemental Fig. S4C). This indicates poor evolutionary constraint over the paternal sDMR sequence, except for a 1-kb region, which maps 3'

to the short *BC028329* EST and shares 74% homology between mice and humans. We investigated four regions within the putative paternal DMR locus, including the conserved *BC0288329* sequence (region 4) and a region of relative homology with the mouse paternal DMR2 (58%, region 3) (Fig. 4A). In the liver and placenta, we consistently

found two distinct populations of alleles, methylated and unmethylated, indicative of a single extended DMR or several discrete DMRs in the region (Fig. 4D). In human ES cells, while allelic specificity was conserved in HUES1 and H1, H9 showed complete DNA methylation (Fig. 4E; Supplemental Fig. S4B). This may reflect either loss of imprinting or genetic loss of maternal chromosome 2. In sperm from two fertile men, only one sub-region (region 2) was 100% methylated, as expected for a paternal gDMR (Fig. 4D; Supplemental Fig. S4B); the paternal DMR of the *GPR1/ZDBF2* locus is therefore more constrained in the germline than in somatic tissues, a feature also manifested in publically available whole-genome bisulfite sequencing data (Molaro et al. 2011). The other regions ranged from 0% (region 1) to variable methylation (30%–60%, regions 3 and 4) with extensive allelic variability.

Although pluripotent embryonic states were not accessible due to the advanced epigenetic features of human ES cells, the human *GPR1/ZDBF2* locus seems to conserve the same DNA methylation dynamics that we found in mice. Importantly, *LIZ* transcripts also have the potential to function as regulatory and protein-coding RNA species in humans.

Dynamic and unusual chromatin features at the imprinted Gpr1/Zdbf2 locus

To gain insight into the chromatin regulation of promoter usage and imprinting of the *Gpr1/Zdbf2* locus, we analyzed histone modification patterns in mouse undifferentiated 2i ES cells and in a differentiated tissue—the whole brain from neonates—by ChIP-qPCR. We found enrichment of histone H3 Lys 4 di- and trimethylation (H3K4me2 and H3K4me3) at the *Liz* promoter in ES cells, while the canonical *Zdbf2* promoter was preferentially enriched in the brain, reflecting differential promoter activity in these two cell types (Fig. 5A; Supplemental Fig. S5A,B). The intergenic CGI was occupied by these active marks in both ES cells and the brain, in agreement with its ubiquitous DNA methylation-free status.

Co-occurrence of H3K4 and H3K9 methylation is considered to be a hallmark of imprinted DMRs, in correlation with the differential allelic activity of these loci (McEwen and Ferguson-Smith 2010). In concordance with the rapid loss of imprinting in culture, we did not find this typical chromatin signature anywhere throughout the *Gpr1/Zdbf2* locus in ES cells. In the brain, the maternal gDMR was enriched in H3K9me3 only, reflecting the extinction and full DNA methylation of the *Liz* promoter upon differentiation. Surprisingly, the intergenic paternal sDMR also failed to display patterns of dual H3K4/H3K9 methylation, despite imprinted DNA methylation in the brain (sDMR2 and sDMR3) (Fig. 5A; Supplemental Fig. S5A,B). Instead, there was significant enrichment of H3K9me3 compared with the rest of the locus, but the level of H3K4 methylation was low and similar to random intergenic sites located 5' of the intergenic CGI border. Finally, co-occurrence of H3K4/H3K9 methylation was not found at the *Zdbf2* promoter despite consistent

monoallelic paternal activity in all somatic tissues. In summary, while histone modification patterns are consistent with the developmental promoter switch in *Zdbf2* transcription, the imprinted status of the *Gpr1/Zdbf2* locus is not apparent at the chromatin level.

Interestingly, at a region overlapping the paternal sDMR, we found a compact block of repressive H3K27me3 marks that was delineated by the intergenic CGI in 5' and the *Zdbf2* promoter CGI in 3' (Fig. 5A). This broad local H3K27me3 enrichment was constitutively observed in all cell types available in public ChIP-seq data sets (Supplemental Fig. S5A; Mikkelsen et al. 2007; Marks et al. 2012). However, we noted some developmental differences: The paternal sDMR was 50-fold more enriched in brain cells compared with 2i ES cells, and H3K27me3 marks extended into the *Zdbf2* gene in brain cells but not in ES cells (Fig. 5A).

Dual roles of DNA and H3K27 methylation on Zdbf2 expression

We then tested the allelism of H3K27me3 enrichment by ChIP followed by quantitative pyrosequencing of whole neonate brains from hybrid mouse crosses. At the *Zdbf2* promoter, which is constitutively devoid of DNA methylation, the maternal allele was substantially more enriched in H3K27me3 than the paternal allele (70/30 ratio) (Fig. 5B), in agreement with maternal-specific *Zdbf2* silencing in differentiated tissues. However, at the paternal sDMR, there was no parental distinction in H3K27me3 enrichment. Using sequential bisulfite sequencing after ChIP, we confirmed the presence of both paternally DNA methylated and maternally unmethylated alleles in H3- and H3K27me3-captured chromatin (Fig. 5C; Supplemental Fig. S5C); this indicates that, at least in brains, DNA and H3K27 methylation can co-occur on the paternal allele of the paternal sDMR. While DNA and H3K27 methylation are usually antagonistic at CGIs, recent studies have shown their compatibility in CG-poor regions (Tanay et al. 2007; Brinkman et al. 2012; Statham et al. 2012), which is the case for the paternal sDMR.

We functionally tested the respective role of DNA and H3K27 methylation on *Zdbf2* expression by comparing ES cell lines that are genetically deficient for the three active DNA methyltransferases (*Dnmt1*, *Dnmt3A*, and *Dnmt3B*; *Dnmt* triple knockout [TKO]) or the H3K27 methyltransferase *Ezh2* (*Ezh2* knockout [KO]) (Fig. 5D). The J1 ES cells, from which the *Dnmt*-TKO cells were derived (Tsumura et al. 2006), showed full DNA methylation at the paternal sDMR, as consistently observed for serum-grown ES cells (Supplemental Figs. S2A, S5D). In DNA methylation-free *Dnmt*-TKO J1 cells, RT-qPCR assays measured increased *Liz* transcript levels, confirming the repressive role of DNA methylation on the *Liz* promoter. In contrast, levels of canonical *Zdbf2* were reduced, arguing toward a positive role of the paternal sDMR methylation on *Zdbf2* promoter activity. Alternatively, higher *Liz* transcripts may also be responsible for reduced *Zdbf2* expression via cross-talk between the two promoters.

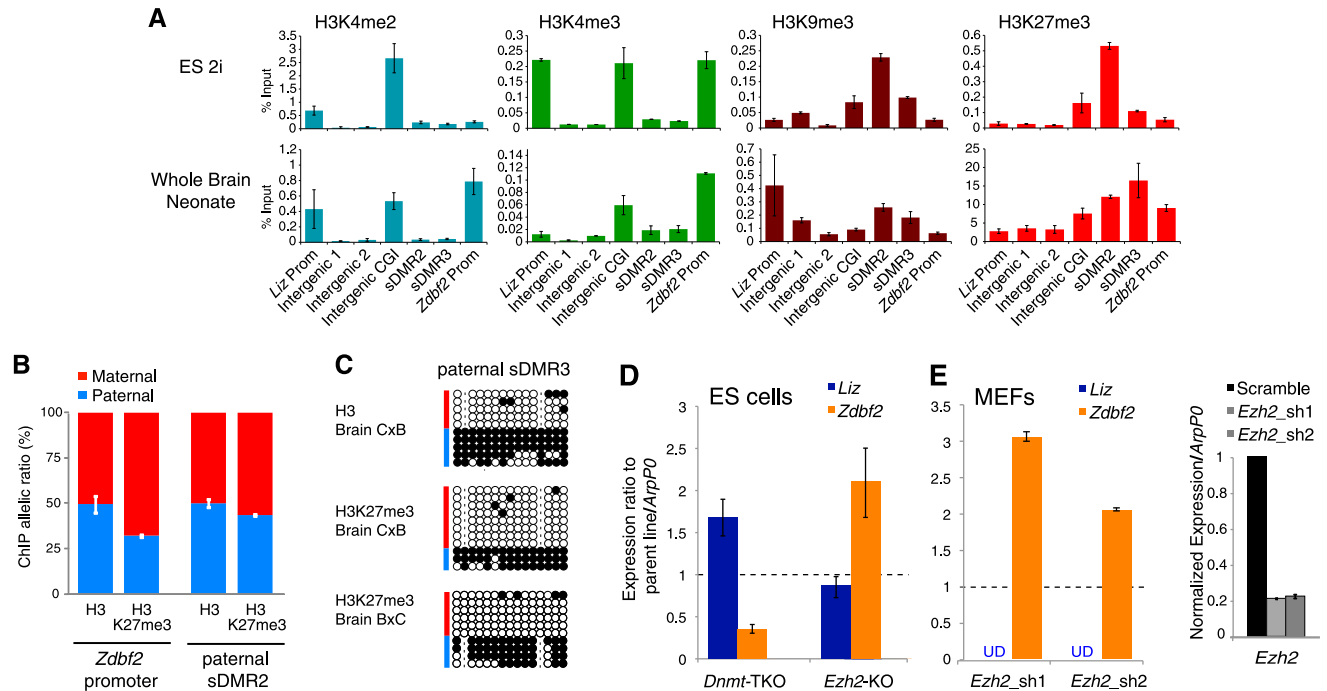


Figure 5. Interplay between histone modifications and DNA methylation at the mouse *Gpr1/Zdbf2* locus. (A) Histone modification profiling by ChIP-qPCR in 2i Cg1 ES cells and neonate whole brains (WBs) over the *Gpr1/Zdbf2* locus (primer positions in Supplemental Fig. S5A). Error bars indicate SEM of two biological replicates for ES cells and SEM from reverse cross BxC and CxB neonates for whole brains. (B) H3K27me3 allelic enrichment detected by ChIP pyrosequencing at the *Zdbf2* promoter and at the paternal sDMR in neonate brains. Error bars indicate SEM from reverse crosses BxC and CxB. (C) Bisulfite-based DNA methylation analysis of H3- and H3K27me3-bound chromatin at the paternal sDMR (DMR3) in neonate brains from CxB and BxC crosses. (D) RT-qPCR analysis of *Liz* and *Zdbf2* RNA levels in *Dnmt*-TKO and *Ezh2*-KO ES cells, normalized to wild-type (WT) parental cell lines. Error bars indicate SEM from two (*Dnmt*-TKO) and four (*Ezh2*-KO) biological replicates. (E, left panel) RT-qPCR analysis of *Liz* and *Zdbf2* levels after *Ezh2* shRNA-mediated knockdown in MEFs. *Liz* levels were undetectable (UD), as expected from *Liz* full DNA methylation in somatic cells. (Right panel) RT-qPCR analysis shows 80% depletion of *Ezh2* levels by two independent hairpins. Error bars indicate SEM from two biological replicates.

The effect of H3K27me3 marks was inferred after tamoxifen-induced deletion of the *Ezh2* gene in 2i-derived ES cells. Ten days after *Ezh2* deletion, EZH2 protein and H3K27me3 marks were barely detectable, while the paternal sDMR maintained low DNA methylation levels, comparable with the nondeleted parental 2i cell line (Supplemental Fig. S5D,E). Upon *Ezh2* knockout, we found a highly reproducible up-regulation of *Zdbf2* (Fig. 5D), which was not related to altered pluripotency state of these cells (normal *Oct4* and *Nanog* mRNA levels) (data not shown). This suggests that H3K27me3 acts by repressing the *Zdbf2* promoter. To determine whether H3K27me3 also tunes down *Zdbf2* expression in differentiated cells where the paternal sDMR is DNA methylated, we performed an *Ezh2* shRNA knockdown in immortalized hybrid mouse embryonic fibroblasts (MEFs) (Fig. 5E; Supplemental Fig. S5D,E). With two independent hairpins inducing 80% reduction of *Ezh2* mRNA, we observed a more than twofold increase in *Zdbf2* levels, confirming that H3K27me3 consistently down-regulates *Zdbf2* output in different cell types. Interestingly, we did not find a relaxation of imprinted *Zdbf2* expression in *Ezh2* knockdown MEFs by RT-PCR pyrosequencing (data not shown). Together, these

results imply that DNA and H3K27me3 methylation oppositely regulate the activity of the *Zdbf2* promoter, with DNA methylation likely acting on expression allelism, and H3K27me3 modulating expression levels in a direct or indirect manner. These two marks do not seem to be spatially competing: Loss of DNA methylation does not increase H3K27me3 levels at the paternal sDMR in *Dnmt*-TKO ES cells, as manifested in available ChIP-seq data sets (Brinkman et al. 2012), and H3K27me3 loss does not promote ectopic gain of DNA methylation at this locus in ES cells or MEFs (Supplemental Fig. S5C).

Discussion

Our study of the *Gpr1/Zdbf2* locus provides new important paradigms in mammalian genomic imprinting. It illustrates the functional relevance of transient imprinting for short- and long-term genome regulation. It also reveals the existence of dynamic genomic imprinting occurring via a parent-of-origin switch in DNA methylation. Together, our work demonstrates how intricate mechanisms can evolve to ensure proper gene dosage regulation through life despite the dramatic epigenetic reprogramming undergone by the embryonic genome.

Genomic imprinting is currently defined by epigenetic characteristics universally shared by all lifelong imprinted loci: (1) acquisition of a gDMR in one of the parental germline only, (2) an uninterrupted continuum of paternal- or maternal-specific methylation from gametes to adult progeny, and (3) the combination of both repressive and permissive histone marks. The *Gpr1/Zdbf2* locus breaks these rules: It contains two gDMRs of opposite parental origins, neither of which is permanent after fertilization, and they lack dual patterns of H3K4/H3K9 methylation. Importantly, the *Gpr1/Zdbf2* locus demonstrates a lifelong imprinted status (Fig. 6): The maternal gDMR provides maternal-specific methylation in the preimplantation embryo and in the placenta after implantation, while the paternal DMR is somatically re-established in the embryo after implantation and maintained in somatic tissues. Discovering whether the maternal DMR alone or both parental DMRs together act as ICRs (i.e., sequences regulating imprinted gene expression) will require genetic testing. Incidentally, our study raises the possibility that transient gDMRs may be tissue-specific gDMRs of the most extreme sort: those that diverge at the time of the first lineage differentiation and are maintained in extra-embryonic but not embryonic tissues. This implies that the number of imprinted gDMRs may be greater in the placenta than in the embryo.

Transcriptional events likely drive the cross-talk between the two parental DMRs, which allows a parent-of-origin switch in DNA methylation at the *Gpr1/Zdbf2* locus: The maternal gDMR hosts a starting site for transcripts that are paternally expressed in early embryos and 2i ES cells. Among them, the *Liz* spliced variants, which extend across the paternal DMR and incorporate

many *Zdbf2* exons, are the dominantly produced isoforms. *Liz* expression positively correlates with paternal DMR methylation in *cis*: Paternal-specific *Liz* expression is associated with paternal-specific sDMR methylation. When *Liz* transcription is biallelic, as in *Dnmt3L*^{-/-} embryos, cultured 2i ES cells, or fetal male germ cells, biallelic methylation is acquired. However, *Liz* expression is not sufficient: The height of *Liz* expression coincides with the lowest DNA methylation of the paternal sDMR in blastocysts. More likely, *Liz* expression primes paternal sDMR accessibility to de novo DNA methyltransferases for when they become available: after implantation or during ES cell differentiation. Interestingly, ChIP-seq data sets show evidence of H3K36me3 marks lining the path of *Liz* transcription in ES cells (Supplemental Fig. S5A; Mikkelsen et al. 2007). Given the ability of the PWWP domain of DNMT3 enzymes to interact with H3K36me3-modified tails (Dhayalan et al. 2010), changes in histone modifications likely provide intermediates between *Liz* expression and paternal DMR de novo methylation. Whether the act of *Liz* transcription or the *Liz* RNA per se is involved in this process will be an important question to address.

Chromatin remodeling properties have been traditionally attributed to noncoding RNAs, including at imprinted loci such as *Airn* or *Kcnq1ot1* (Pandey et al. 2008; Khalil et al. 2009; Latos et al. 2012). Accordingly, *Liz* was originally described as a long intergenic noncoding RNA (lincRNA) (Guttman et al. 2010, 2011) despite the lack of experimental proof (Kobayashi et al. 2012b, 2013). In contrast, we show here that *Liz* RNA can be polyadenylated, is efficiently spliced and exported to the cytoplasm, associates with translating polyribosomes, includes multiple exons of *Zdbf2*, and, most

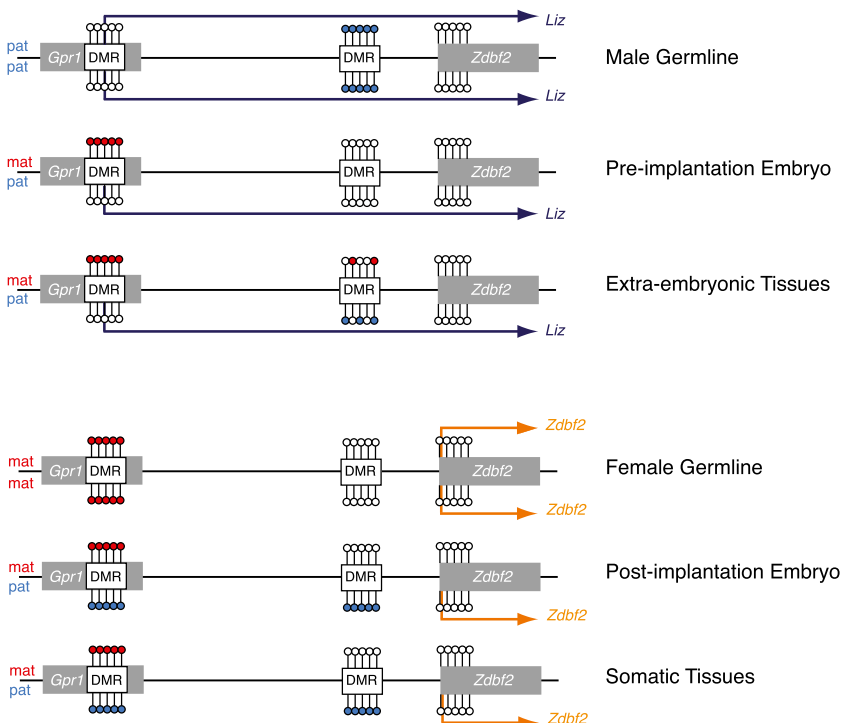


Figure 6. Model for developmental dynamics of promoter usage and parental DNA methylation at the mouse *Gpr1/Zdbf2* locus. *Zdbf2*-related transcripts are produced from alternative promoters: *Liz* (dark blue) is available in the male germline, preimplantation embryos, and extraembryonic tissues; canonical *Zdbf2* (orange) is expressed in the female germline, post-implantation embryos, and somatic tissues. Parent-of-origin methylation occurs in a lifelong manner: Maternal-specific (red) information is provided by the maternal DMR in preimplantation embryos and extra-embryonic tissues, and paternal-specific information (light blue) is carried by the paternal intergenic DMR in post-implantation embryos and somatic tissues. Sex-specific methylation occurs in the male and female germlines.

importantly, has the same coding potential as *Zdbf2* for an ORF, which is evolutionarily conserved in humans. Because *Liz* and canonical *Zdbf2* transcripts share the same putative ORF, it is impossible to distinguish their respective association efficiency and occupancy timing with the translation machinery by ribosome footprinting (Ingolia et al. 2011; Guttman et al. 2013). However, as *Liz* and *Zdbf2* vary by their 5' UTR regions, differential secondary structure or stability could entail additional levels of post-transcriptional regulation. In conclusion, we postulate that *Liz* likely functions as an alternative mRNA source for *Zdbf2* protein translation during development, depending on promoter availability (Fig. 6).

Nothing is known about *Zdbf2* protein function, except for the inclusion of a DBF4-like zinc finger when translation is initiated from exon 4. The eponymous DBF4 protein is an activator of the CDC7 serine/threonine kinase, which has a conserved role in DNA replication initiation from yeast to humans (Hughes et al. 2012). In yeast, the *dbf4/cdc7* complex was moreover linked to chromosome segregation during meiosis (Kovacikova et al. 2013; Le et al. 2013). Coincidentally, we found that *Liz* expression peaks at meiosis during mouse spermatogenesis, and canonical *Zdbf2* is highly transcribed in the oocyte, leading to the exciting possibility that *Zdbf2* may fulfill meiotic roles in mammals.

Alternative promoter usage is a common process of tissue- and stage-specific gene regulation, which may affect 40% of protein-coding genes in mouse and human genomes (Sun et al. 2011). Very little is known about the nature of the mutual regulation between alternative promoters. We found that the downstream *Zdbf2* promoter is turned on when the upstream *Liz* promoter is turned off upon differentiation: *Liz* transcription may prevent activity from the *Zdbf2* canonical promoter in a competitive model where the promoter located the most upstream is dominant over the one located the most downstream (Maunakea et al. 2010). However, our results further suggest that *Liz* activity prior to differentiation promotes *Zdbf2* promoter activation in *cis* after differentiation: *Liz* expression precedes *Zdbf2* expression from the same, paternal allele, and in cases of biallelic *Liz* expression, biallelic *Zdbf2* expression occurs. As discussed above, in addition to functioning as a protein-coding RNA, *Liz* may prepare *Zdbf2* promoter activity via chromatin changes, therefore classifying as a bifunctional RNA.

Finally, we identified an unusual interplay between DNA and H3K27 methylation that highlights the developmental complexity of *Liz* and *Zdbf2* regulation. Using *Dnmt3L*^{-/-} embryos, *Dnmt*-TKO ES cells, and 2i versus serum ES cells, we demonstrated the functional link between DNA methylation and transcriptional activity but with opposite effects on *Liz* and canonical *Zdbf2*. DNA methylation at the *Liz* promoter is negatively correlated with *Liz* transcription levels. In contrast, *Zdbf2* paternal-specific expression is extremely stable throughout somatic life yet does not rely on differential DNA methylation or H3K9 methylation at its promoter,

as is usually the case for imprinted genes. The only form of parent-specific DNA methylation found in the vicinity is at the intergenic paternal DMR located a few tens of kilobases upstream. Here, DNA methylation positively correlates with activity of the canonical *Zdbf2* promoter: When DNA methylation is biallelically gained, biallelic *Zdbf2* expression is observed. When DNA methylation is absent at the sDMR, *Zdbf2* expression is down-regulated. The other important chromatin component of *Zdbf2* regulation is linked to H3K27me3 marks, which occur on the maternal allele at the *Zdbf2* promoter specifically, while it densely accumulates on both maternal and paternal alleles over the paternal sDMR located upstream. Importantly, we found that upon deletion or knockdown of the H3K27 methyltransferase of the Polycomb complex (*Ezh2*), *Zdbf2* expression levels are increased. These results suggest that DNA methylation and H3K27 methylation coregulate *Zdbf2* expression, with intergenic DNA methylation linked to the active *Zdbf2* allele, while H3K27me3 down-regulates *Zdbf2* output.

Here we provide evidence for multiple conserved mechanisms acting at the *Gpr1/Zdbf2* locus for the sole regulation of *Zdbf2* dosage, which is the only gene to show imprinted expression over 400 kb. We show here that besides serving as a new paradigm of dynamic genomic imprinting, the *Gpr1/Zdbf2* locus can provide an important model system for dissecting the signals shaping DNA methylation patterns, the interplay between DNA and histone modifications, and alternative promoter usage.

Materials and methods

Additional materials and methods can be found in the Supplemental Material.

Isolation of mouse gametes, embryos, and tissues

The care and use of animals used here strictly adhered to European and National Regulation for the Protection of Vertebrate Animals used for Experimental and other Scientific Purposes (directive 86/609 and 2010/63). Mice were bred in the pathogen-free Animal Care Facility of the Institut Curie (agreement no. C 75-05-18). MII stage oocytes from superovulated females, epididymal sperm, preimplantation embryos, post-implantation embryos, extraembryonic tissues, and postnatal tissues were prepared as described previously (Proudhon et al. 2012). Purity of gamete samples was assessed by bisulfite-based DNA methylation analysis of maternally imprinted *KvDMR* and paternally imprinted *H19/Igf2* DMD. The *Mus musculus* C57Bl/6J strain was used in a wild-type or *Dnmt3L* mutant background (Bourc'his et al. 2001). Allelic information was obtained from crosses between C57Bl/6J and *Mus castaneus* CAST/Ei or *Mus molossinus* JF1 strains.

Bisulfite-based DNA methylation analysis

Genomic DNA was extracted according to Smith et al. (2009), except for sperm DNA (Jeffreys et al. 1994). The Epitect kit (Qiagen) was used for bisulfite conversion of DNA, except for limited quantities (oocytes and embryos from E3.5 to E6.5) where DNA was embedded in agarose beads before processing

for bisulfite conversion according to Proudhon et al. (2012). Nested or seminested PCR was performed using the primers listed in Supplemental Table S1. PCR amplicons were gel-purified (Qiagen) and subsequently cloned in pCR2.1 Topo-TA vector (Invitrogen) before Sanger sequencing of at least 30 clones. BiQ Analyzer software was used for sequence alignments (Bock et al. 2005).

Polysome fractionation

Sucrose gradient polysome fractionation was performed as described in Gu et al. (2012), with some modifications. ES cells grown in 15-mm plates in 2i medium (5×10^7 cells per plate) were incubated with 100 μ g/mL cycloheximide (Sigma) for 15 min to arrest ribosome movement on mRNAs. Cells were then incubated for 5 min on ice in lysis buffer (20 mM Tris-HCl at pH 7.5, 50 mM KCl, 10 mM MgCl₂, 1% Triton, 1 mM DTT, 40 U/mL RNase inhibitors [Promega], 100 μ g/mL cycloheximide), harvested by scraping off the dish, transferred to Eppendorf tubes, homogenized by three passages through a 26-gauge needle, and further incubated for 5 min on ice. Nuclei and insolubilized material were pelleted by two rounds of centrifugation at 13,000 rpm for 6 min. Clarified lysates were then kept as total cytosolic RNA fraction (input) or processed for fractionation by loading 1 mL onto an 11-mL linear 10%–50% sucrose gradient in 20 mM Tris-HCl (pH 7.5), 50 mM KCl, 10 mM MgCl₂, 1 mM DTT, and 100 μ g/mL cycloheximide prepared using the gradient master (Biocomp). A control experiment of polyribosome disruption was performed by adding 15 mM EDTA before sucrose fractionation. Twelve fractions were collected from top to bottom after ultracentrifugation at 36,000 rpm for 2 h at 4°C using an SW41 swinging-bucket rotor (Beckman Coulter). To determine the distribution of nonassembled ribosomal fractions and translation-engaged polyribosomes (without EDTA), 28S/18S ratios were measured after RNA extraction using the 2000 Bioanalyzer (Agilent). Nonassembled ribosomes have a 28S/18S ratio below or above 2, while assembled 80S ribosomes show 28S/18S ratios close to 2. Based on these measurements and RNA integrity, layers three and four were pooled as the nonassembled ribosome fraction, and layers seven to 10 were pooled as the polyribosome-loaded fraction.

RNA FISH

RNA FISH on preimplantation embryos was performed as previously described (Okamoto et al. 2004) using the intron-spanning BAC probe W11-2819H7 (BacPac Consortium at Children's Hospital Oakland Research Institute). The probe maps to chr1: 63266216–63306435 (UCSC mm9 alignment). Probe labeling was done with SpectrumRed-dUTP by nick translation (Vysis). Images were acquired using a spinning-disk confocal microscope (Zeiss) at 63 $\times$ magnification with 0.3 μ m between each image obtained in Z dimension. Images were analyzed using ImageJ software (NIH).

Acknowledgments

We thank Patricia Diabangouaya, Michel Wassef, Ronan Chaligné, Manuela Portoso, Elphège Nora, Veronica Rodilla, and Marius Walter for experimental help; Raphaël Margueron for the anti-EZH2 antibody and *Ezh2*-shRNAs; Claire Rougeulle for human ES material; Stéphane Viville for Oct4-GFP-sorted PGCs; the imaging facility Bioimaging Cell and Tissue Core Facility of the Institut Curie (PICT-IBiSA) at BDD; and the Institut Curie animal facility for excellent mouse husbandry. D.B.'s laboratory is part of the Laboratoire d'Excellence (LABEX) entitled DEEP (11-

LBX-0044). This research was supported by a European Young Investigator Award (EURYI) and the Fondation Schlumberger. R.D. and S.A. are the recipients of PhD fellowships from Institut Curie and Canceropôle Ile-de-France, respectively.

References

- Aravin AA, Bourc'his D. 2008. Small RNA guides for de novo DNA methylation in mammalian germ cells. *Genes Dev* **22**: 970–975.
- Barlow DP. 2011. Genomic imprinting: A mammalian epigenetic discovery model. *Annu Rev Genet* **45**: 379–403.
- Bock C, Reither S, Mikeska T, Paulsen M, Walter J, Lengauer T. 2005. BiQ Analyzer: Visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics* **21**: 4067–4068.
- Bourc'his D, Bestor TH. 2004. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**: 96–99.
- Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. 2001. Dnmt3L and the establishment of maternal genomic imprints. *Science* **294**: 2536–2539.
- Brinkman AB, Gu H, Bartels SJ, 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.
- Chotalia M, Smallwood SA, Ruf N, Dawson C, Lucifero D, Frontera M, James K, Dean W, Kelsey G. 2009. Transcription is required for establishment of germline methylation marks at imprinted genes. *Genes Dev* **23**: 105–117.
- DeVeale B, van der Kooy D, Babak T. 2012. Critical evaluation of imprinted gene expression by RNA-Seq: A new perspective. *PLoS Genet* **8**: e1002600.
- Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, Jeltsch A. 2010. THE DNMT3A PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J Biol Chem* **285**: 26114–26120.
- Duffié R, Bourc'his D. 2013. Parental epigenetic asymmetry in mammals. *Curr Top Dev Biol* **104**: 293–328.
- Ginno PA, Lott PL, Christensen HC, Korf I, Chedin F. 2012. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell* **45**: 814–825.
- Gu L, Zhang H, He J, Li J, Huang M, Zhou M. 2012. MDM2 regulates MYCN mRNA stabilization and translation in human neuroblastoma cells. *Oncogene* **31**: 1342–1353.
- Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C, et al. 2010. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol* **28**: 503–510.
- Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, et al. 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**: 295–300.
- Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES. 2013. Ribosome profiling provides evidence that large non-coding RNAs do not encode proteins. *Cell* **154**: 240–251.
- Hiura H, Sugawara A, Ogawa H, John RM, Miyauchi N, Miyazaki Y, Horiike T, Li Y, Yaegashi N, Sasaki H, et al. 2010. A tripartite paternally methylated region within the Gpr1-Zdbf2 imprinted domain on mouse chromosome 1 identified by meDIP-on-chip. *Nucleic Acids Res* **38**: 4929–4945.
- Hughes S, Elustondo F, Di Fonzo A, Leroux FG, Wong AC, Snijders AP, Matthews SJ, Cherepanov P. 2012. Crystal

- structure of human CDC7 kinase in complex with its activator DBF4. *Nat Struct Mol Biol* **19**: 1101–1107.
- Ingolia NT, Lareau LF, Weissman JS. 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**: 789–802.
- Jeffreys AJ, Tamaki K, MacLeod A, Monckton DG, Neil DL, Armour JA. 1994. Complex gene conversion events in germline mutation at human minisatellites. *Nat Genet* **6**: 136–145.
- John RM, Lefebvre L. 2011. Developmental regulation of somatic imprints. *Differentiation* **81**: 270–280.
- Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, Sasaki H. 2004. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429**: 900–903.
- Kato Y, Kaneda M, Hata K, Kumaki K, Hisano M, Kohara Y, Okano M, Li E, Nozaki M, Sasaki H. 2007. Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum Mol Genet* **16**: 2272–2280.
- Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, et al. 2009. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci* **106**: 11667–11672.
- Kobayashi H, Yamada K, Morita S, Hiura H, Fukuda A, Kagami M, Ogata T, Hata K, Sotomaru Y, Kono T. 2009. Identification of the mouse paternally expressed imprinted gene *Zdbf2* on chromosome 1 and its imprinted human homolog *ZDBF2* on chromosome 2. *Genomics* **93**: 461–472.
- Kobayashi H, Sakurai T, Imai M, Takahashi N, Fukuda A, Yayoi O, Sato S, Nakabayashi K, Hata K, Sotomaru Y, et al. 2012a. Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet* **8**: e1002440.
- Kobayashi H, Sakurai T, Sato S, Nakabayashi K, Hata K, Kono T. 2012b. Imprinted DNA methylation reprogramming during early mouse embryogenesis at the *Gpr1-Zdbf2* locus is linked to long *cis*-intergenic transcription. *FEBS Lett* **586**: 827–833.
- Kobayashi H, Yanagisawa E, Sakashita A, Sugawara N, Kumakura S, Ogawa H, Akutsu H, Hata K, Nakabayashi K, Kono T. 2013. Epigenetic and transcriptional features of the novel human imprinted lncRNA *GPR1AS* suggest it is a functional ortholog to mouse *Zdbf2linc*. *Epigenetics* **8**: 635–645.
- Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G. 2007. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res* **35**: W345–W349.
- Kovacikova I, Polakova S, Benko Z, Cipak L, Zhang L, Rumpf C, Miadokova E, Gregan J. 2013. A knockout screen for protein kinases required for the proper meiotic segregation of chromosomes in the fission yeast *Schizosaccharomyces pombe*. *Cell Cycle* **12**: 618–624.
- Latos PA, Pauler FM, Koerner MV, Senergin HB, Hudson QJ, Stocsits RR, Allhoff W, Stricker SH, Klement RM, Warczok KE, et al. 2012. Airn transcriptional overlap, but not its lncRNA products, induces imprinted *Igf2r* silencing. *Science* **338**: 1469–1472.
- Le AH, Mastro TL, Forsburg SL. 2013. The C terminus of *S. pombe* DDK subunit Dfp1 is required for meiosis-specific transcription and cohesin cleavage. *Biol Open* **2**: 728–738.
- Li X, Ito M, Zhou F, Youngson N, Zuo X, Leder P, Ferguson-Smith AC. 2008. A maternal-zygotic effect gene, *Zfp57*, maintains both maternal and paternal imprints. *Dev Cell* **15**: 547–557.
- Lienert F, Wirbelauer C, Som I, Dean A, Mohn F, Schubeler D. 2011. Identification of genetic elements that autonomously determine DNA methylation states. *Nat Genet* **43**: 1091–1097.
- Marks H, Kalkan T, Menafrà R, Denisov S, Jones K, Hofmeister H, Nichols J, Kranz A, Stewart AF, Smith A, et al. 2012. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**: 590–604.
- Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y, et al. 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466**: 253–257.
- McEwen KR, Ferguson-Smith AC. 2010. Distinguishing epigenetic marks of developmental and imprinting regulation. *Epigenetics Chromatin* **3**: 2.
- McEwen KR, Leitch HG, Amouroux R, Hajkova P. 2013. The impact of culture on epigenetic properties of pluripotent stem cells and pre-implantation embryos. *Biochem Soc Trans* **41**: 711–719.
- Messerschmidt DM, de Vries W, Ito M, Solter D, Ferguson-Smith A, Knowles BB. 2012. Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. *Science* **335**: 1499–1502.
- Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**: 553–560.
- Mikkelsen TS, Hanna J, Zhang X, Ku M, Wernig M, Schorderet P, Bernstein BE, Jaenisch R, Lander ES, Meissner A. 2008. Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**: 49–55.
- Molaro A, Hodges E, Fang F, Song Q, McCombie WR, Hannon GJ, Smith AD. 2011. Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* **146**: 1029–1041.
- Monk M, Boubelik M, Lehnert S. 1987. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* **99**: 371–382.
- Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, et al. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**: 381–385.
- Okamoto I, Otte AP, Allis CD, Reinberg D, Heard E. 2004. Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science* **303**: 644–649.
- 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.
- Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C. 2008. *Kcnq1ot1* antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* **32**: 232–246.
- Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. 2010. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* **20**: 110–121.
- Proudhon C, Duffie R, Ajjan S, Cowley M, Iranzo J, Carbajosa G, Saadeh H, Holland ML, Oakey RJ, Rakyan VK, et al. 2012. Protection against de novo methylation is instrumental in maintaining parent-of-origin methylation inherited from the gametes. *Mol Cell* **47**: 909–920.
- Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, Baglivo I, Pedone PV, Grimaldi G, Riccio A, et al. 2011. In embryonic stem cells, ZFP57/KAP1 recognize

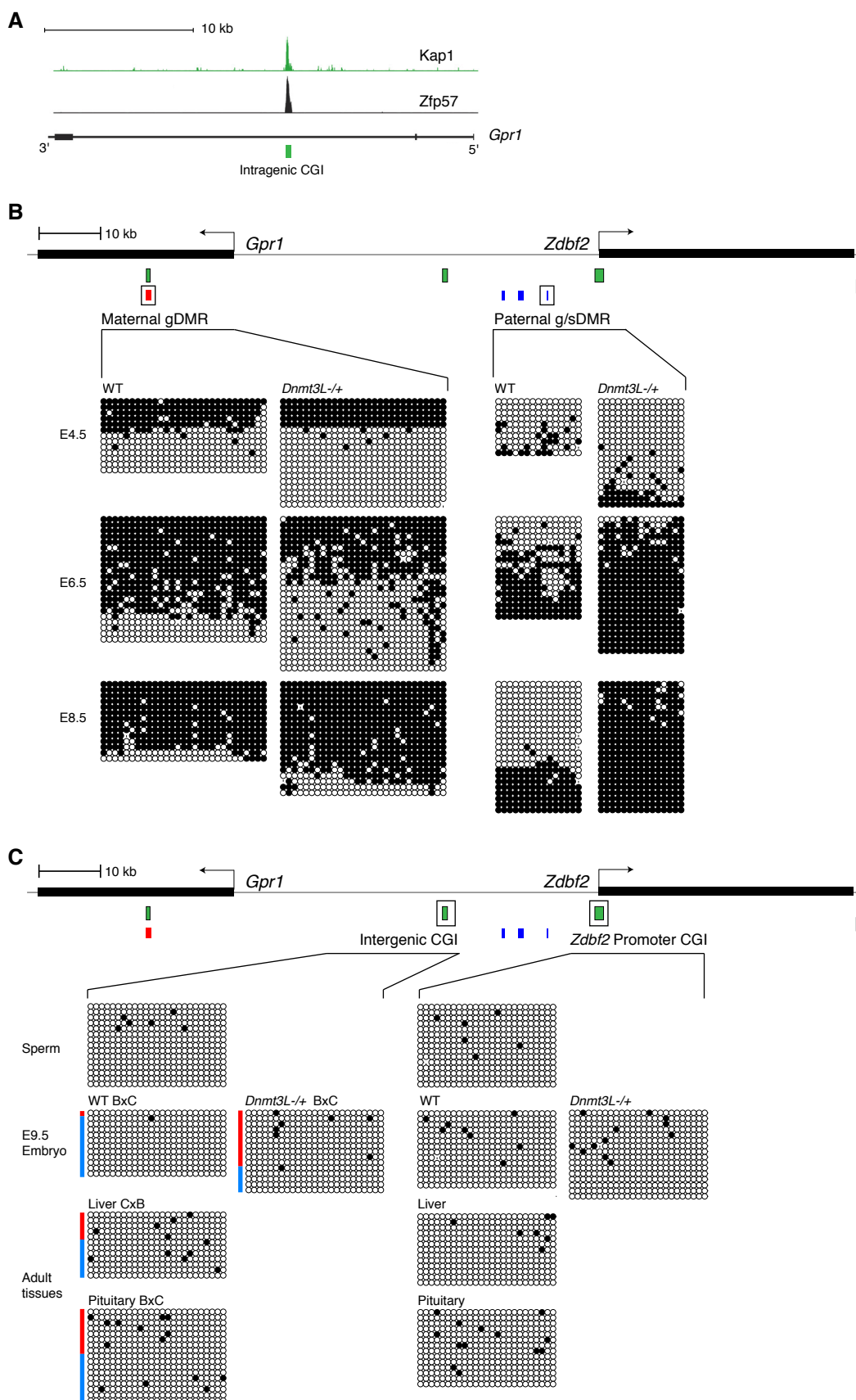
- a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol Cell* **44**: 361–372.
- Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ 3rd. 2002. SETDB1: A novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev* **16**: 919–932.
- Schulz R, Woodfine K, Menhenniott TR, Bourc'his D, Bestor T, Oakey RJ. 2008. WAMIDEX: A Web atlas of murine genomic imprinting and differential expression. *Epigenetics* **3**: 89–96.
- Seidl CI, Stricker SH, Barlow DP. 2006. The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export. *EMBO J* **25**: 3565–3575.
- Smallwood SA, Tomizawa S, Krueger F, Ruf N, Carli N, Segonds-Pichon A, Sato S, Hata K, Andrews SR, Kelsey G. 2011. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat Genet* **43**: 811–814.
- Smith ZD, Gu H, Bock C, Gnirke A, Meissner A. 2009. High-throughput bisulfite sequencing in mammalian genomes. *Methods* **48**: 226–232.
- 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.
- Statham AL, Robinson MD, Song JZ, Coolen MW, Stirzaker C, Clark SJ. 2012. Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. *Genome Res* **22**: 1120–1127.
- Sun H, Wu J, Wickramasinghe P, Pal S, Gupta R, Bhattacharyya A, Agosto-Perez FJ, Showe LC, Huang TH, Davuluri RV. 2011. Genome-wide mapping of RNA Pol-II promoter usage in mouse tissues by ChIP-seq. *Nucleic Acids Res* **39**: 190–201.
- Tanay A, O'Donnell AH, Damelin M, Bestor TH. 2007. Hyperconserved CpG domains underlie Polycomb-binding sites. *Proc Natl Acad Sci* **104**: 5521–5526.
- Tsumura A, Hayakawa T, Kumaki Y, Takebayashi S, Sakaue M, Matsuoka C, Shimotohno K, Ishikawa F, Li E, Ueda HR, et al. 2006. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**: 805–814.
- Ying QL, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, Cohen P, Smith A. 2008. The ground state of embryonic stem cell self-renewal. *Nature* **453**: 519–523.
- Zuo X, Sheng J, Lau HT, McDonald CM, Andrade M, Cullen DE, Bell FT, Iacovino M, Kyba M, Xu G, et al. 2012. Zinc finger protein ZFP57 requires its co-factor to recruit DNA methyltransferases and maintains DNA methylation imprint in embryonic stem cells via its transcriptional repression domain. *J Biol Chem* **287**: 2107–2118.

SUPPLEMENTAL INFORMATION

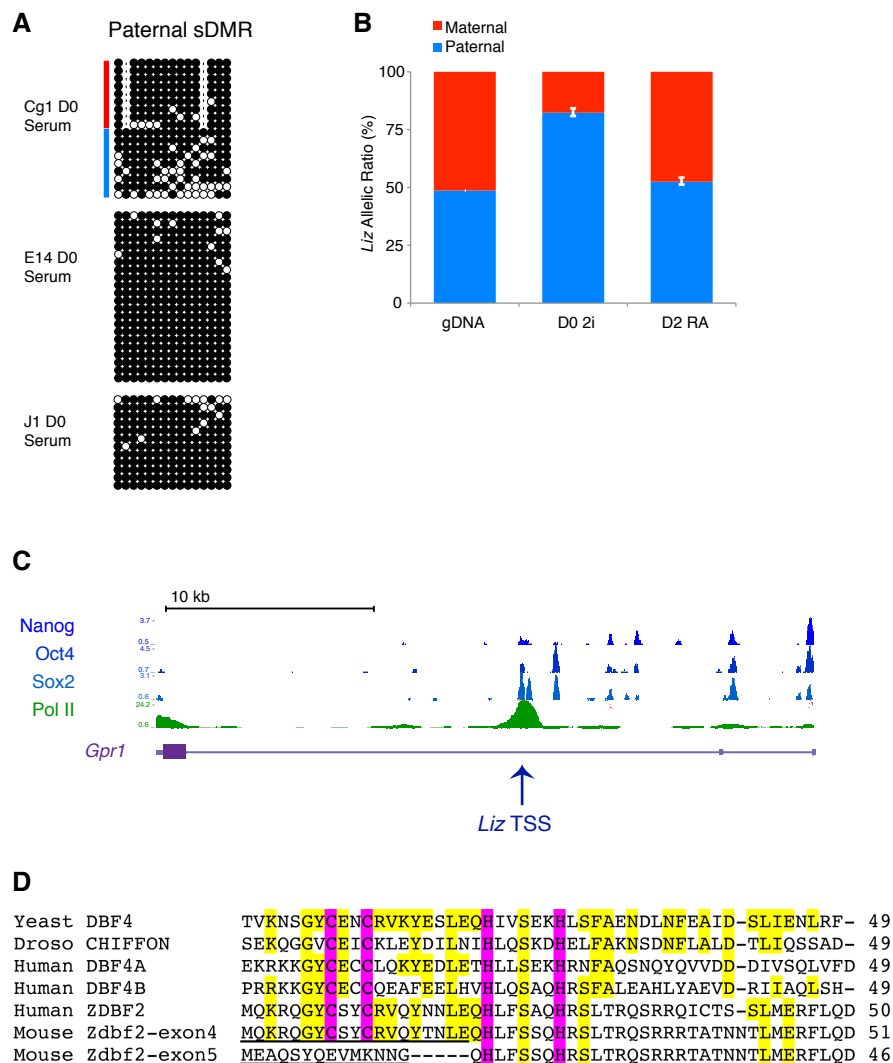
Supplemental Figures linked to the main manuscript

Supplemental Figure	Title	Relates to
Figure S1	DNA methylation profiling of the <i>Gpr1/Zdbf2</i> locus during mouse development	Figure 1
Figure S2	<i>Liz</i> and <i>Zdbf2</i> regulation in ES cells	Figure 2
Figure S3	Allelic expression at the <i>Gpr1/Zdbf2</i> locus	Figure 3
Figure S4	Expression, DNA methylation and conservation at the human <i>GPR1/ZDBF2</i> locus	Figure 4
Figure S5	Histone modifications at the mouse <i>Gpr1/Zdbf2</i> locus	Figure 5

GENESDEV/2013/232058
Duffié_FigureS1

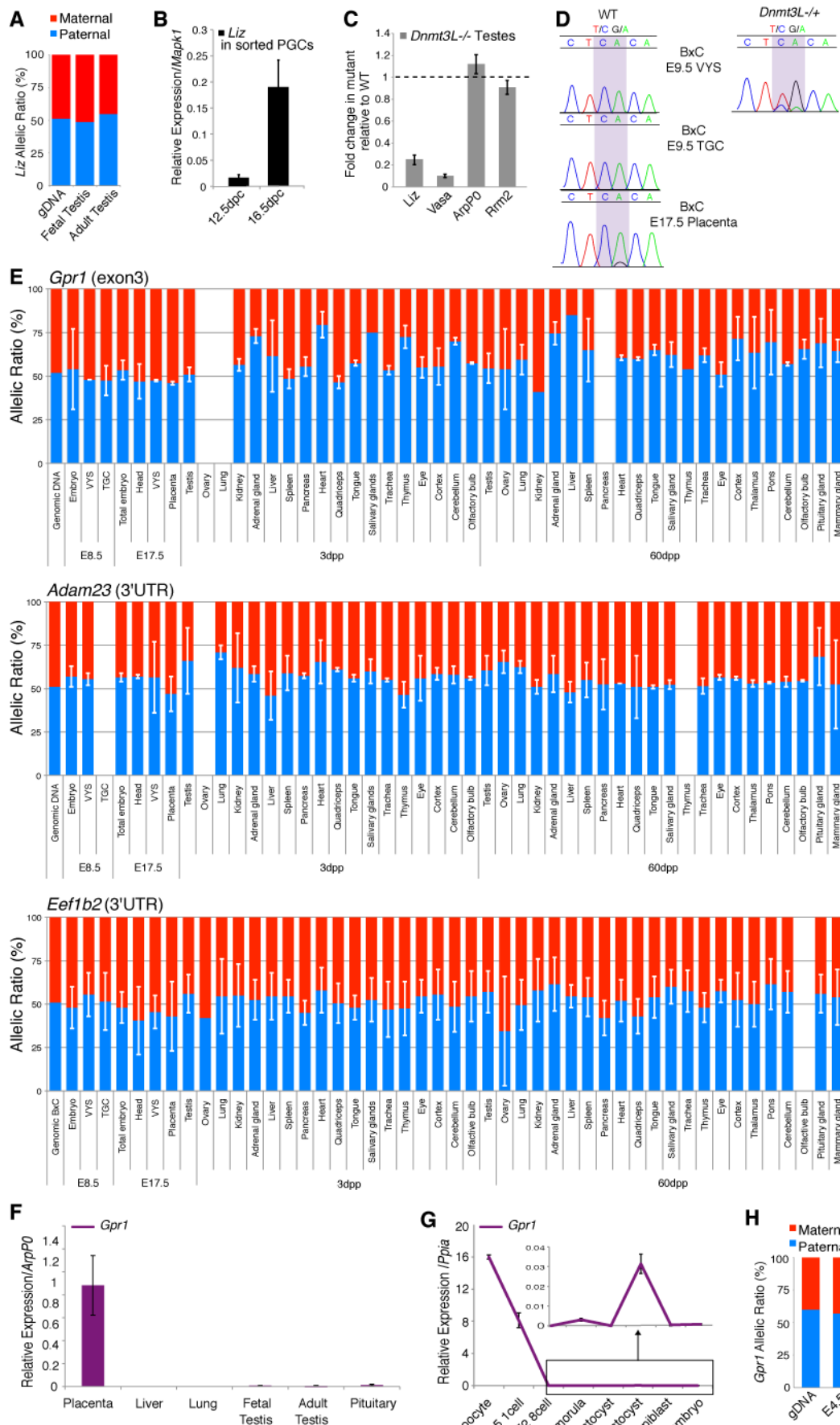


GENESDEV/2013/232058
Duffié_FigureS2

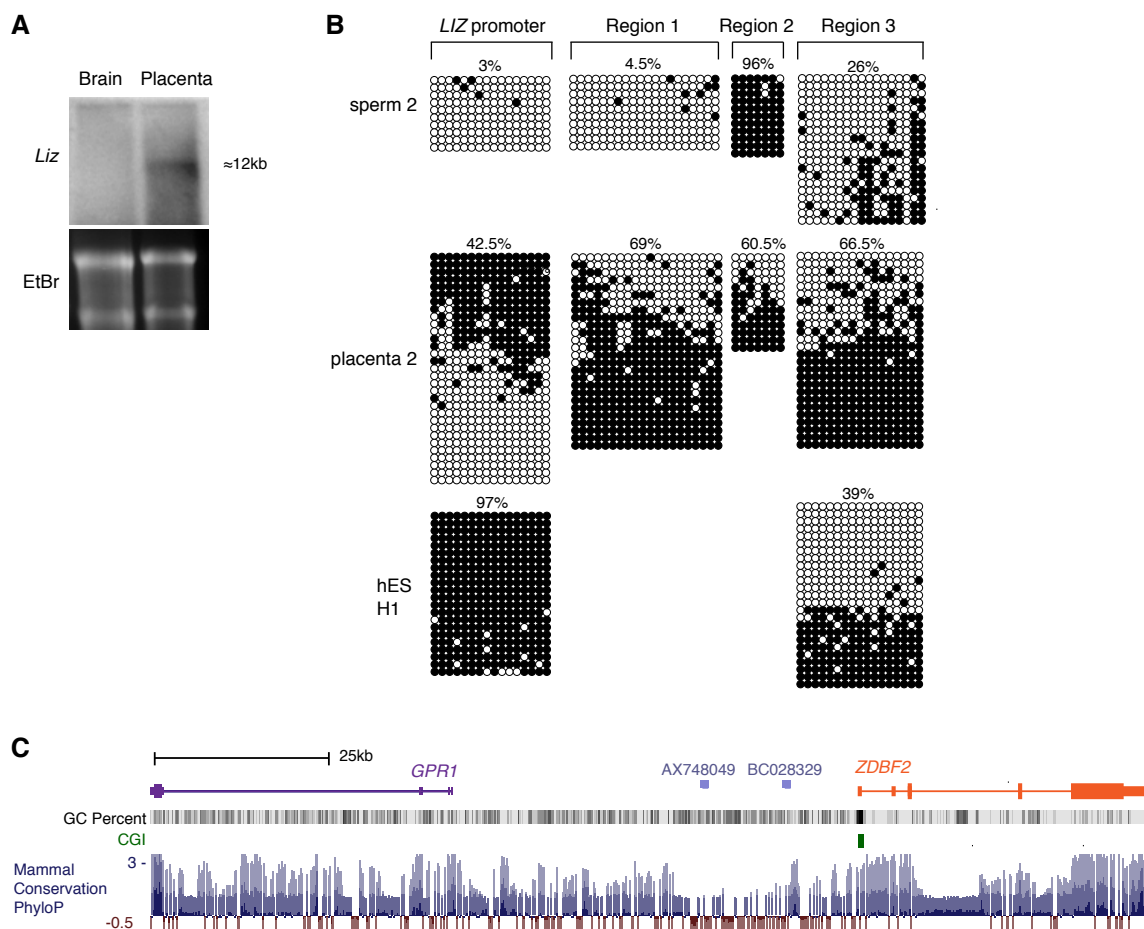


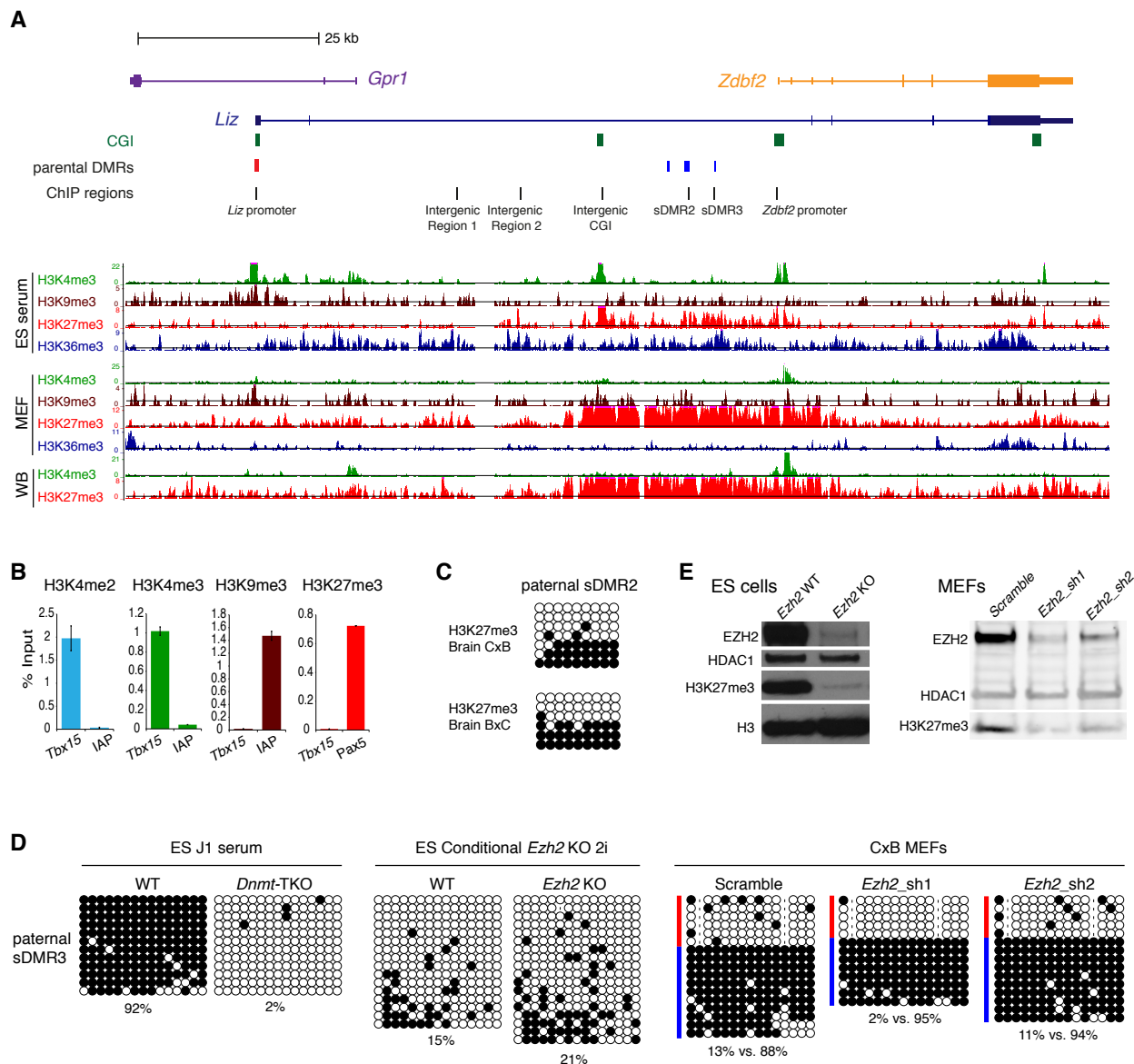
GENESDEV/2013/232058

Duffié_FigureS3



GENESDEV/2013/232058
Duffié_FigureS4





Supplemental Figure Legends

Supplemental Figure S1. DNA methylation profiling of the *Gpr1/Zdbf2* locus during mouse development. (A) Genome browser track of ChIP-Seq data (Quenneville et al. 2011) showing peaks of enrichment of ZFP57 and KAP1 at a CGI located in the second intron of *Gpr1*. (B) Bisulfite analysis of maternal and paternal DMR methylation in peri-implantation embryos of WT and *Dnmt3L*^{-/+} embryos of pure C57Bl6/J background. (C) Methylation analysis of the two CGIs flanking the paternal DMR, one in an intergenic position and the other coinciding with the *Zdbf2* promoter. Twenty-five CpGs are shown representative of the 51 (intergenic CGI) and 79 (*Zdbf2* promoter) CpGs amplified.

Supplemental Figure S2. *Liz* and *Zdbf2* regulation in ES cells. (A) Bisulfite analysis of paternal DMR methylation in different serum-grown ES cell lines. (B) RT-PCR pyrosequencing of *Liz* allelic expression in 2i Cg1 ES cells at early passage D0 and at D2 after RA differentiation. Genomic DNA (gDNA) was used to exclude assay-specific biases. (D) Genome browser track of ChIP-seq data of pluripotency factors and RNA polymerase II (RNA PolII) in serum-grown ES cells, showing peaks of enrichment around the *Liz* TSS (Mikkelsen et al. 2008). (D) Clustalw alignment of the 49AA zinc finger DBF4 motif of budding yeast with drosophila, human and mouse protein representatives. In mouse, the *Zdbf2* ORF initiated in exon 4 allows the production of the conserved DBF4 motif, while the exon 5-ORF lacks most of the conserved residues of the motif, including both cysteines of the C2H2 zinc finger. Conserved amino-acids are highlighted in yellow, cysteine and histidine residues in pink. Amino-acids specific to *Zdbf2* exon 4 and exon 5 are underlined.

Supplemental Figure S3. Developmental and allelic expression at the *Gpr1/Zdbf2* locus. (A) RT-PCR pyrosequencing assay of *Liz* in fetal and adult mouse testes. (B) RT-qPCR analysis of *Liz* in sorted primordial germ cells (PGCs) of fetal testes at 12.5 and 16.5dpc from Oct4-GFP transgenic mice. Error bars, SEM from two biological replicates. (C) RT-qPCR analysis of *Liz* in testes of 6weeks *Dnmt3L*^{-/-} males, which are characterized by an infertility phenotype due to massive germ cell depletion (Bourc'his and Bestor 2004). *Liz* downregulation was in the same order of magnitude as the germ-cell specific marker *Vasa*. Ubiquitous housekeeping *ArpP0* and *Rrm2* gene expression levels were not affected by the germ cell depletion. Values were normalized to *Beta-actin*. Error bars as above. (D) Sanger Sequencing of *Liz* in E9.5 TGC and VYS and E17.5 placenta, and in *Dnmt3L*^{-/+} E9.5 VYS from BxC hybrid embryos. (C) RT-PCR pyrosequencing assay in a bank of hybrid tissues

from four developmental time points of *Gpr1*, *Adam23* and *Eef1b2*. Missing values are due to inability to amplify in one or both crosses, likely related to very low expression levels. SNP positions are indicated in brackets. Error bars, SEM from reverse crosses, BxC and CxB. (D) RT-qPCR analysis of *Gpr1* expression in mouse tissues. Error bars, SEM of technical replicates. (E) RT-qPCR analysis over early embryonic development shows that *Gpr1* is highly expressed in the oocyte, degraded during preimplantation development, and slightly peaks in E4.5 blastocysts. Error bars, SEM of technical replicates on pooled embryos. (E) *Gpr1* was found to be biallelically expressed by RT-PCR pyrosequencing in BxJF1 E4.5 blastocysts.

Supplemental Figure S4. Expression, DNA methylation and conservation at the human *GPR1/ZDBF2* locus. (A) Northern blot analysis detects a 12kb *LIZ* product in human placenta but not in brain. Ethidium bromide staining shows equal RNA loading. (B) Bisulfite-based DNA methylation analysis of the *LIZ* promoter and three regions upstream of the *ZDBF2* TSS in additional human sperm, placenta, and ES (H1) samples. (C) Genome browser track of the human *GPR1/ZDBF2* locus showing PhyloP mammalian conservation scores. Low sequence-level conservation is observed over the 25kb region upstream of the *ZDBF2* promoter where the paternal DMR localizes in mouse, except for a peak corresponding to the *BC028329* EST.

Supplemental Figure S5. Histone modifications at the mouse *Gpr1/Zdbf2* locus. (A) At the top, map of the mouse *Gpr1/Zdbf2* locus showing regions examined by ChIP-qPCR in this study. At the bottom, genome browser tracks of publically available ChIP-seq data from serum-cultured ES cells, MEFs, and whole brain (WB) (Mikkelsen et al. 2007). (B) Positive and negative controls for ChIP-qPCR analyses of H3K4me2, H3K4me3, H3K9me3 and H3K27me3 enrichment, using the housekeeping gene *Tbx15*, retrotransposon IAP sequences, and the Polycomb-target *Pax3* gene. Data are shown for 2i ES cells, similar patterns were obtained for WB (data not shown). Error bars, SEM from two biological replicates. (C) DNA methylation analysis of H3K27me3-bound chromatin at the paternal sDMR2, in neonate WB of CxB and BxC crosses. (D) Western blots showing loss of EZH2 protein and global reduction of H3K27me3 levels in conditionally deleted 2i ES cells (left panel) and shRNA knockdown MEFs (right panel). HDAC1 and H3, loading controls. (E) Bisulfite-based DNA methylation analysis of the paternal sDMR (DMR3) in the cell types examined in Figure 5.

Supplemental Table S1 : Primer list**Bisulfite**

Maternal DMR	OF	5'-TGAGTATATTGTGGTTTATGTTATG-3'
	OR	5'-ACCTAACTCAATTCTAATAAAATAAC-3'
	IF	5'-AAGATTAAATTAAATTTAAATTATTGATG-3'
	IR	5'-CAAAAATAACCAATATTTCCAATAAAT-3'
Intergenic CGI	OF	5'-GAAATAGGAGGTTGGAGGAGTT-3'
	IF	5'-GTATATAGGAGTAGTAAATTTTGG-3'
	R	5'-TCCTTCCATTAATACCTAAAAATAA-3'
Paternal DMR (DMR3)	OF	5'-ATTTTATGTATTTATTAGTGAGTTG-3'
	OR	5'-CTAATCACAACCTCCCAATCAC-3'
	IF	5'-GAATTGAATTTTATATATTATTTTGG-3'
	IR	5'-CCTATATACAACCTACCTTACTT-3'
Paternal DMR (DMR2)	OF	5'-TAGTAATATGAAGATTTAGATTTAG-3'
	OR	5'-CTTAAC TACAAATACCTTTATTAC-3'
	IF	5'-ATTGTTGTAGGGTTTATAGTGG-3'
	IR	5'-CTACTTAACTCTTAAACAATCCAA-3'
<i>Zdbf2</i> CGI promoter	F	5'-GAAGGAGAGTAGGAGGTATTTA-3'
	OR	5'-AAACACCCACACTTATATAAAACA-3'
	IR	5'-TAATCTAAAATTATTAATAAACTCTCT-3'

RT-qPCR

<i>Liz</i> ex1-2	F	5'-CAGACAAAGGCCACAGTCAA-3'
	R	5'-GGGAAACCAGCACTGAATCT-3'
<i>Liz</i> ex2-4	F	5'-GAGAAAAGCCTCCTGTTCC-3'
	R	5'-TGAGCAGAAAAGAGCAAGCA-3'
<i>Liz</i> Unspliced	F	5'-CAGGCCCTGGAATGAGTAA-3'
	R	5'-AGCCACAAGCAGCCATAGT-3'
<i>Zdbf2</i> ex1-3	F	5'-CCGGAAGGAGAGCAGGAG-3'
	R	5'-GAGCAGAAAAGAGCAAGCA-3'
<i>Liz/Zdbf2</i> ex4	F	5'-GAGTGTTGAGGCCAGTAGC-3'
	R	5'-ACGGCAATAGCTGCAATACC-3'
<i>Liz/Zdbf2</i> ex6	F	5'-CACAGAAGCCTGACCAGACA-3'
	R	5'-TATATGGGTGGTGTGCGAGA-3'
<i>Liz/Zdbf2</i> ex6-7	F	5'-CTGCGACACCACCCATATAA-3'
	R	5'-AAAACCACCACCACTTCAGG-3'
<i>Gpr1</i> ex2-3	F	5'-TTCAAACACCACGACAGCTT-3'
	R	5'-TCCTTTGAGACTTCCATGATG-3'
<i>Ezh2</i>	F	5'-AATACATGTGCAGCTTTCTGTTC-3'
	R	5'-ACGAATTTTGTGGCCCTTTC-3'
<i>Airn</i>	F	5'-GGCTCAGCAAACAGCACCA-3'
	R	5'-GCCTGTGATTGCTCAGTTATTCC-3'
<i>Tsix</i> Unspliced	F	5'-GTGTGTCATAGCTCAAGAGG-3'
	R	5'-GGAGCCTAAACCTGTCTGTC-3'
<i>Igf2r</i>	F	5'-TGTGCAGTTACACATGGGAAGCT-3'
	R	5'-AGTGAGTCACCCACTGAGCG-3'
<i>SINEs B1</i>	F	5'-GTGGCGCACGCCTTTAATC-3'
	R	5'-GACAGGGTTTCTCTGTGTAG-3'
<i>ArpP0</i>	F	5'-TCCAGAGGCACCATGAAATT-3'
	R	5'-TCGCTGGCTCCACCTT-3'
<i>Beta-actin</i>	F	5'-AAGTGACGTTGACATCCG-3'
	R	5'-GATCCACATCTGCTGGAAGG-3'

<i>Ppia</i>	F	5'-TTACCCATCAAACCATTCCTTCTG-3'
	R	5'-AACCCAAAGAACTTCAGTGAGAGC-3'
<i>Rrm2</i>	F	5'-CCGAGCTGGAAAGTAAAGCG-3'
	R	5'-ATGGGAAAGACAACGAAGCG-3'
<i>MapK1</i>	F	5'-CCTTCAGAGCACTCCAGAAAGT-3'
	R	5'-ACAACACCAAAAAGGCATCC-3'

RT-PCR

<i>Liz</i>	F1	5'-CAGACAAAGGCCACAGTCAA-3'
	F2	5'-CAGTGCTGGTTTCCCTATCA-3'
	F3	5'-CTGGCATTTTCTCCCTACCC-3'
	F4	5'-AGCTGAAGCAAGATGGCTGT-3'
	R1	5'-AGGGGGACGCACTAACTTTT-3'
	R2	5'-CCTTCTGCCTGAAGAACTCC-3'
	R3	5'-ACGGAGCTCCTACAGCAAAA-3'
	R4	5'-TCCCACATCATCATTACATGC-3'
	R5	5'-ATGGCAGCATTGTCTCTCAA-3'

RACE

<i>Liz</i> 5'exon1	5'-CTGACCGGATGGACGACTGAAGCACTA-3'
<i>Liz</i> 5'exon3	5'-CCATCTTGCTTCAGCTGGGGACAATC-3'
<i>Liz</i> 3'exon1	5'-AGCATGCAGAGCGCTGAACCTAAGGAC-3'
<i>Liz</i> 3'exon3	5'-GTCCCCAGCTGAAGCAAGATGGCTGTA-3'

Northern probe

<i>Liz</i>	F	5'-CCAAGCCCTTCAAACCTCAAT-3'
	R	5'-GGGAAACCAGCACTGAATCT-3'
<i>Liz/Zdbf2</i>	F1	5'-CAGAGCAGACGTCGGACAGC-3'
	F2	5'-GCAGCTATTGCCGTGTGCAG-3'
	F3	5'-GAGTGTTTCGAGGCCAGTAGC-3'
	R1	5'-CGAAAGAGCCGCTAGGCAAA-3'
	R2	5'-CTGCTGTCCCGTTCCAGTT-3'
	R3	5'-CAACTTCTGCGTGTTCTGA-3'

Pyrosequencing

<i>Liz</i>	F	5'-CTCATGCTACGGTGAACATCC-3'
	R	5'-biotinAAGTAGTTCGGTCCCTTAGGTTTCAG-3'
	Seq	5'-GAACATCCCATGCCT-3'
<i>Liz/Zdbf2</i>	F	5'-AACTAGTCACCGTCCCCCACA-3'
	R	5'-biotinCAACCAAACGAAGGGGCACTA-3'
	Seq	5'-CCGTCCCCCACAGCC-3'
<i>Liz</i> promoter	F	5'-TGAGTGGCTGGGATGAGTC-3'
	R	5'-GCTAGGGAGCCATGCTTTATGT-3'
	Seq	5'-TCTCCAGAAAGGCCG-3'
Paternal sDMR3 A	F	5'-CTGGCACAGCCTTTTCGTTCT-3'
	R	5'-GACCTAAAAATGCCCAAACCTG -3'
	Seq	5'-CCCAAACCTGAAACA-3'
Paternal sDMR3 B	F	5'-ATTGTCGGATGGATATTCTCAGAG-3'
	R	5'-GGGTGCACTTTGACTTCGGTTAT -3'
	Seq	5'-CTACATGCGGAGTGG-3'
<i>Gpr1</i> Cast	F	5'-CCATCCTTTACGTCCTAATAAGCA-3'
	R	5'-biotinACGCTTTAGTACCTCAGCAACAG-3'
	Seq	5'-ACGTCCTAATAAGCAAGA-3'
<i>Gpr1</i> JF1	F	5'-TGCTAGGGCATATAAGAACAGGG-3'
	R	5'-biotinTTACTCCCAGGAGTCTGACCC-3'
	Seq	5'-GAACGAGTCCCAGGTA-3'
<i>Adam23</i>	F	5'-GTGTTCTTGTTGATCCGCTGTACC-3'
	R	5'-biotinCAGACAAGCAAAATGGCATATTGA

	Seq	5'-AATGGGAGCTGGAAC-3'
<i>Eef1b2</i>	F	5'-biotinGATGTGGCTGCTTTTAACAAGA-3'
	R	5'-GAGCCAAGGGTTATCTTTCAAG-3'
	Seq	5'-TTTTATTTAAGTGCCTGAC-3'

ChIP-qPCR

<i>Liz</i> Promoter	F	5'- CAGACAAAGGCCACAGTCAA-3'
	R	5'- GTTTATCGTGAAGGCCGAAA-3'
Intergenic1	F	5'- TGGCATGAGTCTAGCCAGTG-3'
	R	5'- CAGCAGTTTCCACCATTCAA-3'
Intergenic2	F	5'- GCTCACTCTCACTGCTGCTG-3'
	R	5'- TCCCATCTCAGGGGTTACTG-3'
Intergenic CGI	F	5'- TTGGGGATCGGTGTTTAGAG-3'
	R	5'- CGTCTTCGCTCTCCGATAG-3'
sDMR2	F	5'- AACTTTAGAGCCCCAGAAGG-3'
	R	5'- CCATCCGACAATTCAAAATC-3'
sDMR3	F	5'- ACCCTGGTCCGACCTTATGT-3'
	R	5'- GGGAACCTCCATCCTTAGC-3'
<i>Zdbf2</i> Promoter	F	5'- CTTTAAGAGCGGGAGGAAGG-3'
	R	5'- AACACTGGGCCCTTCTG-3'
IAP	F	5'- CTCCATGTGCTCTGCCTTCC-3'
	R	5'- CCCCGTCCCTTTTATAGAGA-3'
<i>Tbx15</i>	F	5'- TCCCCCTTCTCTGTGTGTCAG-3'
	R	5'- CGGAAGCAAGTCTCAGATCC-3'
<i>Pax5</i>	F	5'-ATGGGAGTTTGTTCCTGTGT-3'
	R	5'-AGTGATGTTTGGCCTAATCCTG-3'

Human Bisulfite

<i>LIZ</i> promoter	OF	5'- GGGTTGGAGATTGTATTAATT-3'
	OR	5'- ATAATAACACATACCTAAATCC-3'
	IF	5'- TTAATGGTTAGTGGTTTAATTAAT-3'
	IR	5'- TCCCACTACTTAAAAACTAA-3'
Intergenic region1	OF	5'- TTATTTTGAATTATTTGGAAGAT-3'
	OR	5'- TCTACCAATACTTTCCCTCTC-3'
	IF	5'- AGTATCATAGTTTTAAATATTTATG-3'
	IR	5'- ATTCTACTTTATCACTTCCCT-3'
Intergenic region2	OF	5'- TGTTGGTTAGGTTGGTTTTGAA-3'
	OR	5'- CCAAACCTAAAATCACTAAATTAA-3'
	IF	5'- GTTTTGTAGTTAGATTGGAAAATA-3'
	IR	5'- AAAATAATAATTACCTAAAAATAAAAC-3'
Intergenic region3	OF	5'- TGGGAATAGGGGTTTGTGTA-3'
	OR	5'- CCTCCACCTCCTAAATTCAAA-3'
	IF	5'- GTTATAGGGAGTTGAGGGGTAT-3'
	IR	5'- TTCAAACAATTCTCCTACCTCAA-3'
Intergenic region4	F	5'- TAATTGAATATATTTAGTGGGAAAT-3'
	OR	5'- AAACACATTTAAACAATATCCCTA-3'
	IR	5'- ATTACCCATATAATTTCAATCAA-3'

Human RT-qPCR

<i>LIZ</i>	F	5'- GAGGGCGTCTGAGGGTACTG-3'
	F	5'- CAGCACCGCTTCAAGAAGTCT-3'
<i>ZDBF2</i>	F	5'- GCCGAAGTCGCCTATTTCT-3'
	R	5'- TCCCTTCTTATTCCGGTTCC-3'
<i>LIZ/ZDBF2</i>	F	5'- GGGTCATCGTCTGAAGTGGT-3'
	R	5'- TCGGTAGCATCCTCATCCTC-3'

<i>GPR1</i>	F	5'- ACTGGGCAGTGAAGGAATTG-3'
	R	5'- CTAGCCATTGCGGATTCAT-3'
<i>TBP</i>	F	5'- CAGGAGCCAAGAGTGAAGAACA-3'
	R	5'- CCAGGAAATAACTCTGGCTCATAAC-3'

Human RT-PCR

Short <i>LIZ</i>	F	5'- CCCTTCCCACAGACCTTGT-3'
	R	5'- AGAGGCTGAGGCAGGAGAAT-3'
Long <i>LIZ</i>	F	5'- CCCTTCCCACAGACCTTGT-3'
	R	5'- TCCAAATGCACCACTTCAGA-3'

Human Northern probe

<i>LIZ</i>	F	5'- CCCTTCCCACAGACCTTGT-3'
	R	5'- GCACCGCTTCAAGAAGTCTC-3'

Supplemental Table S2 : Antisera list.

	Furnisher	Reference
H3	Abcam	ab1791
H3K4me2	Millipore	07-030
H3K4me3	Millipore	07-473
H3K9me3	Millipore	07-442
H3K27me3	Abcam	ab6002

Supplemental Table S3 : shRNA sequences*Ezh2_sh1*

5'-CCGCAGAAGAAGTGAAGAAATTTCTTTCAGTTCTTCTGCGG-3'

Ezh2_sh2

5'-GCTAGGCTAATTGGGACCAAAATTTGGTCCCAATTAGCCTAGC-3'

Scramble Addgene plasmid 1864

5'-CCTAAGGTAAAGTCGCCCTCGCTCGAGCGAGGGCGACTTAACCTTAGG-3'

Supplemental Experimental Procedures

Mouse ES cells

Male ES cell lines Cg1 and gC6 were derived from reciprocal CAST/Ei x129 Sv and 129 Sv x CAST/Ei mice. The *Dnmt-TKO* ES cells were provided by Masaki Okano (RIKEN, Kobe) and were described in (Tsumura et al. 2006). The male *Ezh2*-conditional KO ES cell line (Ike16-3) was derived from mice described in (Su et al. 2003). Unless otherwise stated, derivation and culture was done in chemically defined medium (Ying et al. 2008). Cells were plated on 10 μ g/ μ L Hu plasma fibronectin (Millipore) cultured in N2B27 medium supplemented with MEK inhibitor PD0325901 (PD) at 1.5 μ M, the GSK3 inhibitor CHIR99021 (CH) at 4.5 μ M and mouse LIF (Millipore) at 20 μ g/ml. Accutase (Sigma) was used to dissociate cells every 2-3 days for passage. For differentiation, cells were plated in the presence of inhibitors and LIF and the following day, culture medium without LIF and inhibitors, and traces of retinoic acid (RA) were added. RA was added the following day, and N2B27 medium was replaced the two following days for a total of four days of differentiation. For *Ezh2* deletion, 4-Hydroxytamoxifen (Sigma) was administered for four days at a final concentration of 1 μ M and experiments were conducted at least ten days from the first day of 4-Hydroxytamoxifen addition. J1 and *Dnmt-TKO* ES cells were cultured according to published procedures in serum-based medium which consists of Glasgow medium (Sigma) supplemented with 15% FCS (Gibco), 0.1mM MEM nonessential amino acids (Gibco), 2mM L-glutamine (Life Technologies), 1mM sodium pyruvate (Gibco), 0.1mM 2-mercaptoethanol (Sigma) and mouse LIF 20 μ g/ml (Millipore) (Tsumura et al. 2006). Serum cultured cells were plated on gelatin and dissociated for passage every two-three days using Trypsin Replacement (Gibco).

Human samples and cells

Human adult tissue samples were commercially obtained (Ambion). Human placentas were collected from caesarean sections of normal pregnancies, according to protocols approved by the local ethics committee (authorization DC-2012-1645) and including signed informed consent from all patients. After removal of the maternal membranes, small villosity pieces were collected as previously described (Barbaux et al. 2012). Semen samples were obtained with informed consent from men with normal semen parameters according to the World Health Organization (Cooper et al. 2010). Spermatozoa were purified from other semen cell types using a Percoll gradient (PureSperm, JCD); purity was checked by inverted light microscopy and confirmed by DNA methylation analysis of the *H19* DMR (Boissonnas et al.

2010). The use of human embryonic stem cells (HUES1, H1 and H9) RNA and DNA was approved by the Agence de la Biomédecine. The different lines were cultured in DMEMF-12 medium supplemented with 20% KnockOut serum, 0.1mM NEAA, 1mM glutamine, 0.4μM β-mercaptoethanol and 4ng/ml β-FGF (all Invitrogen) on mouse feeder cells.

RNA extraction and expression analyses

Total RNA was isolated using Trizol (Invitrogen), except oocytes and early embryos for which the RNA Picopure kit (Applied Biosystems) was used as described in manufacturers' instructions. Nuclear and Cytoplasmic RNA fractions were isolated using the Cytoplasmic and Nuclear RNA purification kit (Norgen Biotek), and polyadenylated RNA was purified using the mRNA Direct kit (Dynabeads). DNase treatment (Turbo DNase Ambion or Qiagen DNase set) was carried out for all samples involved in PCR amplification except RACE. For PCR and qPCR-based applications, reverse transcription was performed on total or cellular-fractionated RNA using random hexamers and SuperscriptIII (Invitrogen). Real-time quantitative amplification of cDNA was performed using the SYBR Green Master Mix on the Vii7 thermal cycling system (Applied Biosystems). Relative expression levels were normalized to *Rrm2*, *Beta-Actin*, *Arp0*, *Ppia* or *MapK1* according to the $2^{-\Delta Ct}$ method. Quantification of allelic ratios of cDNAs was performed by pyrosequencing on samples of hybrid crosses, using the PyroMark Q24 Pyrosequencer (Qiagen). Absence of amplification bias was systematically determined using genomic DNA. Rapid amplification of 5' and 3' cDNA ends (RACE) was performed using total and poly-adenylated RNA, using the SMARTer RACE cDNA amplification kit and Advantage 2 PCR kit (Clontech). Amplified bands were gel extracted (Nucleotrap gel extraction kit, Clontech) and cloned using Topo-TA vector (Invitrogen). Northern blot was conducted on total RNA according to standard procedures. The probe specific to *Liz* maps to *Liz* exons 1 and 2 and was amplified from ES cell cDNA. Three pooled *Zdbf2* probes were amplified using pituitary cDNA and map to exons 3-6 of *Zdbf2*. RNA integrity and loading were confirmed by 18/28S ribosomal RNA integrity. All primers are described in Supplementary Table S1.

Chromatin Immunoprecipitation

Chromatin immunoprecipitation was performed on formaldehyde-fixed chromatin as described in (Proudhon et al. 2012), except that prot G magnetic beads were used (Dynabeads). Antibodies are listed in Supplemental Table S2. Quantitative analysis of immunoprecipitated DNA was determined in qPCR assays using SYBR Green technology (Applied Biosystems)(Primers in Supplemental Table S1). The fold enrichment of each target

sequence was calculated as $2^{-\Delta Ct}$ between Input and antibody bound fractions (% Input). Bisulfite treatment was performed on immunoprecipitated chromatin embedded in agarose beads, as described above.

ShRNA assay

Ezh2-shRNA targeting was performed on immortalized mouse embryonic fibroblasts (MEFs) of CAST/Ei x C57Bl6/J background. Immortalization was carried out by infection with the retroviral vector pBABE-hygro p53 DD plasmid (Addgene), which consists of a dominant negative P53 mutant carrying a 28AA deletion, according to published procedures (Hahn et al. 2002). To produce the hairpins, synthetic oligos (Supplemental Table S3) were hybridized and cloned into the lentiviral vector pLKO.1 (TRC RNAi Consortium) using *EcoRI* and *AgeI* (New England Biolabs). Lentiviruses were prepared in HEK293T cells with 4ug plasmid construct, 3ug psPAX2 packaging plasmid, 1ug mMd2.G envelope plasmid and Polyethyleneimine. Supernatant was collected two and three days after transfection and concentrated using 100K centrifuge tubes (Millipore). Concentrated virus was snap-frozen and stored at -80°C until infection. Puromycin selection was carried out 48h after infection, and cells were allowed to grow for ten days before processed for molecular analyses.

Supplemental References

- Barbaux S, Gascoin-Lachambre G, Buffat C, Monnier P, Mondon F, Tonanny MB, Pinard A, Auer J, Bessieres B, Barlier A et al. 2012. A genome-wide approach reveals novel imprinted genes expressed in the human placenta. *Epigenetics* **7**: 1079-1090.
- Boissonnas CC, Abdalaoui HE, Haelewyn V, Fauque P, Dupont JM, Gut I, Vaiman D, Jouannet P, Tost J, Jammes H. 2010. Specific epigenetic alterations of IGF2-H19 locus in spermatozoa from infertile men. *Eur J Hum Genet* **18**: 73-80.
- Bourc'his D, Bestor TH. 2004. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**: 96-99.
- Cooper TG, Noonan E, von Eckardstein S, Auger J, Baker HW, Behre HM, Haugen TB, Kruger T, Wang C, Mbizvo MT et al. 2010. World Health Organization reference values for human semen characteristics. *Hum Reprod Update* **16**: 231-245.
- Hahn WC, Dessain SK, Brooks MW, King JE, Elenbaas B, Sabatini DM, DeCaprio JA, Weinberg RA. 2002. Enumeration of the simian virus 40 early region elements necessary for human cell transformation. *Mol Cell Biol* **22**: 2111-2123.
- Mikkelsen TS, Hanna J, Zhang X, Ku M, Wernig M, Schorderet P, Bernstein BE, Jaenisch R, Lander ES, Meissner A. 2008. Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**: 49-55.
- Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**: 553-560.
- Proudhon C, Duffie R, Ajjan S, Cowley M, Iranzo J, Carbajosa G, Saadeh H, Holland ML, Oakey RJ, Rakyan VK et al. 2012. Protection against de novo methylation is instrumental in maintaining parent-of-origin methylation inherited from the gametes. *Mol Cell* **47**: 909-920.
- Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, Baglivo I, Pedone PV, 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.
- Su IH, Basavaraj A, Krutchinsky AN, Hobert O, Ullrich A, Chait BT, Tarakhovsky A. 2003. Ezh2 controls B cell development through histone H3 methylation and Igh rearrangement. *Nat Immunol* **4**: 124-131.
- Tsumura A, Hayakawa T, Kumaki Y, Takebayashi S, Sakaue M, Matsuoka C, Shimotohno K, Ishikawa F, Li E, Ueda HR et al. 2006. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**: 805-814.
- Ying QL, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, Cohen P, Smith A. 2008. The ground state of embryonic stem cell self-renewal. *Nature* **453**: 519-523.

3- Caractérisation d'une nouvelle DMR maternelle, tissu-spécifique et lignée-spécifique

3.1- Classification des DMR candidates en régions d'empreinte transitoire ou persistante

Parmi les 28 régions candidates révélées par le criblage MeDIP-Seq sur les embryons *Dnmt3L*^{-/+}, quatre nouvelles gDMR maternelles et potentiellement ICR ont donc été identifiées. Je me suis ensuite intéressée aux 24 DMR candidates restantes. Afin de choisir les plus susceptibles d'être réellement soumises à empreinte, j'ai classé ces DMR en fonction de leurs caractéristiques nucléotidiques dont la présence 1) de motifs de liaison à la protéine Zfp57, qui est un indicateur fort d'empreinte, et 2) d'îlots CpG, qui sont associés à toutes les ICR maternelles connues. Parmi les DMR candidates restantes, 6 remplissent ces deux critères : les DMR associées aux transcrits *Socs5*, *Plec1*, *Dvl3*, *Chrna2*, *Tmem132e* et *Fcmd*. Quatre contiennent des motifs Zfp57 sans être classés comme îlot CpG (*Hdac5*, *Camta1*, *Gper*, et une région intergénique). Enfin, j'ai de plus intégré quatre DMR candidates qui n'étaient pas présentes dans notre lot de 24 candidates mais qui avaient été découvertes parallèlement dans d'autres criblages (*Smox*, *Usp4*, *Neurog3* et *Pvt1*) (Kobayashi et al., 2012 ; Xie et al., 2012) (Table1).

Je me suis ensuite appliquée à catégoriser ces 14 DRM candidates en DMR transitoires ou DMR persistantes tout au long de la vie (qu'elles soient ubiquitaires ou tissu-spécifiques). Les DMR maternelles transitoires regagnant de la méthylation sur l'allèle paternel après l'implantation, j'ai analysé le statut de méthylation des DMR candidates les plus « intéressantes » dans des embryons à 9,5dpc (days post-coïtum) issus de croisements entre souris de fonds génétiques différents, C57Bl6/J (B) et CAST/Ei (C) (FigureSup1A). Ces embryons présentent des polymorphismes nucléotidiques (SNP) qui permettent de différencier l'allèle maternel de l'allèle paternel. Parmi les 14 DMR candidates que j'ai analysées 1) dix présentent près de 100% de méthylation ou bien une méthylation très hétérogène mais équivalente sur leurs deux allèles parentaux, 2) l'une est dépourvue de presque toute méthylation sur les deux allèles, 3) l'une a été impossible à amplifier de part la nature de sa séquence et 4) deux présentent une méthylation préférentielle de l'allèle maternel : *Neurog3* avec 52% de

méthylation maternelle versus 16,1% sur l'allèle paternel et *Socs5* avec 67,5% de méthylation maternelle versus 2,6% sur l'allèle paternel (Figure1 et Figure3B). Douze de ces DMR candidates seraient donc soit des DMR transitoires, soient des faux positifs du crible. Afin de pouvoir distinguer entre ces deux possibilités, une analyse des profils de méthylation dans des blastocystes hybrides, avant implantation, est nécessaire. Raquel Perez, chercheuse post-doctorante dans le laboratoire, explore actuellement le statut de méthylation de ces DMR à ce stade du développement.

La DMR de *Socs5* a donc attiré mon attention de par le maintien d'une méthylation parentale différentielle très nette au stade 9,5dpc de développement *in utero* et la possibilité qu'il s'agisse d'une nouvelle DMR persistante. La caractérisation de cette DMR, associée au gène *Socs5* (*Suppressor of Cytokine Signaling 5*) a occupé la fin de ma thèse. J'ai ainsi analysé **1) sa dynamique de méthylation au cours du développement, 2) les marques d'histones associées et 3) son rôle sur la régulation transcriptionnelle de son gène hôte**. A ce propos, il est important de souligner que cette nouvelle DMR candidate est en position intragénique. Au moins trois effets sur la transcription sont possibles pour ce type de gDMR/ICR. Tout d'abord, l'ICR peut agir comme un promoteur alternatif intragénique. L'absence de méthylation sur l'allèle paternel permet l'initiation d'un transcrit alternatif au niveau de l'ICR seulement à partir de cet allèle, comme nous l'avons précédemment décrit pour l'ICR tissu-spécifique de *Cdh15* (Proudhon *et al.*, 2012). Deuxièmement, l'ICR peut contenir un signal d'arrêt de la transcription. Dans ce cas, l'absence de méthylation sur l'allèle paternel permet la liaison de certains facteurs sur l'ICR qui induiront un arrêt précoce de la transcription à partir de ce même allèle, comme il a été reporté au locus *Mcts2* (Wood *et al.*, 2008). Enfin, l'ICR pourrait agir comme un « enhancer » où l'absence de méthylation de l'allèle paternel permettrait la liaison de certains facteurs pour la régulation à distance de la transcription des gènes avoisinants. Ce dernier scénario n'a pas encore été mis en évidence dans le cas d'ICR maternelle. J'ai envisagé ces trois hypothèses mécanistiques dans mon étude du rôle de la DMR intragénique de *Socs5* sur la transcription à ce locus (Figure2).

3.2- Résultats

3.2.1- Identification de la DMR de *Socs5* en tant que DMR maternelle non-transitoire

La DMR de *Socs5*, d'une taille estimée de 1kb (kilo base) par notre criblage MeDIP-seq, est localisée en position intragénique dans l'intron 2 du gène *Socs5* (17qE4). Ce gène code pour une protéine de la famille des Suppressor of Cytokine Signaling, qui sont des inhibiteurs de la voie de signalisation JAK/STAT. Par le criblage MeDIP-seq, nous savions que cette DMR présentait un défaut de méthylation dans les embryons mutants *Dnmt3L*^{-/+} par rapport aux embryons WT et qu'elle était dépourvue de méthylation dans le sperme. En terme de caractéristiques génomiques, elle co-localise avec un îlot CpG et contient 4 motifs de liaison à la protéine Zfp57 dans le génome murin de référence (C57Bl6/J) (Figure3A). Toutes ces caractéristiques génomiques et épigénétiques sont communes à toutes les ICR maternelles connues.

Afin de vérifier la spécificité allélique de la méthylation de la DMR, j'ai utilisé la technique de conversion au bisulfite suivie de clonage/séquençage sur des embryons à 9,5dpc issus de croisements entre souris de fonds génétiques différents, C57Bl6/J (B) et CAST/Ei (C) (FigureSup1A). Par cette technique, j'ai pu confirmer que les embryons BxC WT présentent une méthylation majoritaire de l'allèle maternel (67,5% de méthylation versus 2,6% sur l'allèle paternel) et que c'est bien cette méthylation maternelle qui est défaillante dans les embryons BxC *Dnmt3L* mutants, avec une méthylation résiduelle de 3,1%. De plus, l'allèle paternel est bien protégé par la vague de méthylation *de novo* après l'implantation et reste globalement dépourvu de méthylation à mi-gestation (Figure3B). La DMR de *Socs5* est donc une DMR maternelle non-transitoire, qui persiste au delà de l'implantation. J'ai ensuite poursuivi mon analyse pour déterminer s'il s'agissait d'une DMR ubiquitaire ou tissu-spécifique.

3.2.2- Variation de la méthylation de la DMR de *Socs5* en fonction des tissus et des lignées murines

Avant d'analyser les profils de méthylation de la DMR de *Socs5* dans différents tissus, j'ai d'abord tiré profit des données publiques de séquençage génomique après traitement au bisulfite disponibles sur UCSC (University of California Santa Cruz) (Kobayashi *et al.*, 2012 ; Hon *et al.*, 2013) (Figure3C). Ces données confirmaient la présence de méthylation dans l'ovocyte et non dans le sperme, avec un maintien autour de 50% de méthylation dans le blastocyste. Cependant, en vie adulte, la méthylation de

la DMR semble variable en fonction des tissus. En effet, le foie présente 100% de méthylation alors que le poumon et le cortex présentent des variations aux environs de 50% (Figure3C). On peut aussi noter que le promoteur du gène *Socs5* qui coïncide avec un îlot CpG est dépourvu de méthylation dans tous les tissus. Les variations de méthylation observées au niveau de la DMR de *Socs5* semblent indiquer que celle-ci serait tissu-spécifique.

J'ai donc tout d'abord confirmé par bisulfite clonage/séquençage les profils de méthylation variables observés dans le poumon, le cortex et le foie à 3dpp (days post-partum), en ajoutant une information allélique essentielle par l'utilisation de tissus issus de croisements hybrides réciproques BxC et CxB (Figure4A). Dans le foie, où l'on observait 100% de méthylation dans les données publiques, la DMR présente une méthylation de 85,8% sur l'allèle maternel versus 39,8% sur l'allèle paternel dans le sens BxC. L'empreinte parentale est donc largement nivelée par gain de méthylation sur l'allèle paternel dans ce tissu. Ceci est encore plus frappant dans le sens réciproque CxB, où l'on observe 72,7% sur l'allèle maternel et 68,5% sur l'allèle paternel. Dans le poumon où l'on observait environ 50% de méthylation, l'allèle maternel apparaît méthylé majoritairement (71,2% vs 16%) dans le sens de croisement BxC, mais cette spécificité allélique est perdue dans le croisement réciproque CxB, à nouveau par gain de méthylation sur l'allèle paternel (86,5% vs 89,4%). La DMR de *Socs5* est donc non seulement tissu-spécifique mais semble aussi être sensible au fond génétique : elle serait donc lignée-spécifique. On peut aussi observer des variations en fonction des sens de croisement dans le cortex : dans le sens BxC l'allèle maternel est méthylé spécifiquement (89% vs 13,5%) alors que dans le sens CxB on observe conjointement une perte de méthylation sur les allèles maternels et un gain de méthylation sur les allèles paternels (37,9% vs 49,7%). Pourtant, quand on regarde le profil de méthylation de la DMR dans ces mêmes tissus dans les deux fonds génétiques purs Bl6 et Cast, on observe environ 50% de méthylation, avec des allèles globalement méthylés et d'autres globalement déméthylés (48% pour Bl6 et 58,8% pour Cast dans le cortex ; 53,5% pour Bl6 et 49,9% pour Cast dans le poumon) (Figure4B). La DMR serait donc prône à des variations de méthylation uniquement dans les croisements hybrides entre ces lignées et dans un seul sens de croisement. Enfin, on remarque aussi une forte hétérogénéité allélique, également pour une même origine parentale (foie BxC et cortex BxC par exemple), ce qui sous-tend une forte hétérogénéité cellulaire de méthylation au sein d'un même tissu.

Afin de confirmer le statut tissu-spécifique et lignée-spécifique de la DMR de *Socs5*, j'ai appliqué une méthode quantitative de mesure de méthylation par pyroséquençage sur les lignées pures C57Bl6/J et CAST/Ei et les croisements réciproques entre ces lignées (4 animaux/fond génétique). Dans le foie, le niveau de méthylation reste à peu près constant et aux environs de 70% pour les 2 lignées pures (Figure4C). J'ai confirmé aussi le niveau de méthylation différent dans les croisements réciproques BxC (64%) et CxB (78%) observé par bisulfite clonage/séquençage, ce qui montre bien que ces deux techniques sont équivalentes pour déterminer le pourcentage de méthylation présent au niveau de la DMR. Dans le poumon, on observe une variation forte du pourcentage de méthylation de la DMR en fonction du sens du croisement : 48% de méthylation en BxC et 74% en CxB. Dans le cortex, alors qu'un même niveau de méthylation de 60% est retrouvé pour les deux lignées pures, on observe aussi un gain de méthylation dans le sens CxB par rapport au sens BxC (69% vs 47%). L'inclusion d'autres tissus (rein, cœur, bulbe olfactif) m'a permis de généraliser le taux de méthylation supérieur systématique des croisements CxB par rapport aux croisements BxC (Figure4C). Il s'agirait d'une caractéristique spécifique de la DMR de *Socs5* : en comparaison, on observe très peu de variation autour des 50% attendus à l'ICR maternelle intragénique du locus *Mcts2/H13*, quelque soit le tissu, la lignée murine ou le sens de croisement (FigureSup2).

J'ai ensuite cherché à dater développementalement la sensibilité accrue de l'allèle paternel Bl6 à un regain de méthylation dans le croisement CxB. La méthylation de l'allèle Bl6 paternel pourrait être déjà présente dans le sperme, ou bien être acquise après l'implantation par défaut de protection lors de la vague de méthylation *de novo* du génome. Bien que les données de MeDIP-seq et de séquençage génomique au bisulfite semblaient indiquer une hypométhylation de l'ADN de sperme Bl6, j'ai tout de même comparé le taux de de méthylation de spermatozoïdes en fonds C57Bl6/J et CAST/Ei (Figure4D). Les deux lignées présentent peu de méthylation dans le sperme (15,9% pour Cast et 12,8% pour Bl6) ce qui exclut la possibilité que la méthylation spécifique de l'allèle paternel Bl6 soit héritée du spermatozoïde.

En revanche, à l'issue de la vague génomique de méthylation que subit l'embryon post-implantatoire, l'allèle paternel Bl6 a déjà regagné de la méthylation à 8,5dpc dans le sens de croisement CxB, (94,3%), alors que l'allèle paternel Cast dans le sens BxC en est largement dépourvu (10,2%) (Figure4D). Il semble donc que l'allèle Bl6 ne soit pas

protégé de la vague de méthylation *de novo* du génome et c'est pour cette raison que l'on observe un gain de méthylation sur l'allèle paternel Bl6 dans les croisements CxB et donc une perte de l'empreinte maternelle dans les tissus post-nataux. Même s'il n'existe vraisemblablement pas de mouvement de méthylation *de novo* avant l'implantation, il me reste tout de même à vérifier que l'empreinte maternelle est bien conservée dans les sens BxC et CxB au stade blastocyste. Cette expérience est en cours.

3.2.3- Complexité du caractère lignée-spécifique de l'empreinte de la DMR de *Socs5*

Les croisements réciproques BxC et CxB apportant beaucoup de complexité à la compréhension du locus, j'ai donc choisi d'étendre la caractérisation de la DMR de *Socs5* à d'autres croisements, en utilisant des souches de souris moins divergentes. J'ai choisi la lignée 129S2/SvPasCrl, qui fait partie de la même sous-espèce *Mus musculus domesticus* que les souris Bl6, alors que les souris Cast sont des *Mus musculus castaneus* (FigureSup1B). Pour information, il existe environ 4,458,000 SNPs entre les souris Bl6 et 129 contre 17,673,700 SNPs entre les souris Bl6 et Cast, ces deux dernières sous-espèces étant séparées de 500,000 à un million d'années (Keane et al., 2011).

J'ai tout d'abord analysé le taux de méthylation de la DMR de *Socs5* par pyroséquençage dans le poumon à 3dpp, parmi tous les croisements réciproques entre les 3 lignées Bl6, Cast et 129, et recherché lesquels présentaient 1) des pourcentages de méthylation les plus proches de 50%, et 2) avec le moins de variation inter-croisement possible (Figure5A). Les croisements entre les lignées Bl6 et 129 sont ceux qui remplissent le mieux ces deux critères : 48% de méthylation en Bx129 et 41% en 129xB. En analysant l'allélisme de méthylation dans le poumon pour tous les croisements par bisulfite suivi de clonage/séquençage, j'ai confirmé une bonne conservation de l'empreinte maternelle à la fois dans les croisements Bx129 et 129xB (Figure5B). En effet, l'allèle maternel présente 98,3% et 66,5% de méthylation dans les croisements Bx129 et 129xB respectivement, et l'allèle paternel, 13% et 12,3%. Pour les autres croisements, on observe des pertes sur l'allèle maternel simultanément à des gains de méthylation sur l'allèle paternel, reflétant encore une fois une instabilité de l'empreinte et une hétérogénéité cellulaire de la DMR de *Socs5*. De fait de leur homogénéité, j'ai choisi de continuer la caractérisation de l'empreinte au locus *Socs5* à partir des croisements réciproques Bx129 et 129xB.

Incidemment, la comparaison des croisements CxB et 129xB montre que l'allèle Bl6 paternel n'est pas intrinsèquement susceptible à un regain de méthylation, mais cette susceptibilité dépend de l'interaction avec un/des facteurs *trans* présents dans le génome Cast mais pas 129. De même, la comparaison des croisements CxB et Cx129 montre que seul l'allèle Bl6 mais pas le 129 est susceptible à ce facteur *trans* Cast. J'ai donc analysé de plus près les séquences nucléotidiques de la DMR des lignées Bl6, Cast et 129 à la recherche de variations de séquence et potentiellement de motifs pour ce facteur *trans*. Après alignement des 3 séquences, j'ai observé que la séquence Cast présente une délétion d'une répétition de 23pb et une duplication d'une répétition de 34pb comportant un motif de liaison supplémentaire au facteur Zfp57 par rapport aux séquences Bl6 et 129 (Figure6). Cependant, Zfp57 protège les allèles méthylés des gDMR/ICR de la vague de déméthylation du génome avant l'implantation, mais n'intervient pas *a priori* dans la protection des allèles non-méthylés lors de la méthylation *de novo* embryonnaire (Quenneville *et al.*, 2011). Par contre, on observe un SNP dans la séquence de la DMR présent uniquement sur l'allèle Bl6 : présence d'une Thymine sur l'allèle Bl6 et d'une Cytosine sur les allèles Cast et 129 (Figure6). Il est possible que ce polymorphisme empêche la liaison d'un facteur protecteur *trans* Cast sur l'allèle Bl6 exclusivement et par conséquent celui-ci ne serait pas protégé de la vague de méthylation *de novo* dans les croisements BxC. Je reviendrai sur cette partie dans la discussion section 2.1.

3.2.4 Dynamique développementale de la méthylation de la DMR de Socs5 dans des croisements hybrides homogènes

L'analyse par pyroséquençage des profils de méthylation de la DMR dans d'autres tissus que le poumon (foie, rein, cœur, cortex, bulbe olfactif) et dans les deux sens Bx129 et 129xB révèle, d'une manière générale, que 1) le taux de méthylation est peu différent dans les deux sens, 2) les tissus issus du sens Bx129 présentent un pourcentage de méthylation légèrement supérieur (10% environ) par rapport au sens 129xB et 3) la méthylation dans le sens Bx129 est constante pour tous les tissus (45%), alors que les taux sont plus hétérogènes entre les tissus 129xB (Figure5C). Le foie présente toujours un pourcentage de méthylation supérieur par rapport aux autres tissus, comme je l'avais précédemment observé pour les lignées Bl6 et Cast et les croisements BxC et CxB, en relation avec une perte d'empreinte tissu-spécifique.

Le poumon étant un des tissus qui présente une méthylation maternelle prononcée à la naissance (3dpp) (FigureSup3), j'ai analysé la stabilité de l'empreinte dans ce tissu au cours de la vie (3 et 18 mois), dans les deux croisements Bx129 et 129xB. Il s'est avéré que la DMR de *Socs5* conserve bien une empreinte maternelle jusqu'au moins 3 mois (Figure5E). On observe des variations avec un gain de méthylation sur l'allèle paternel dans le sens Bx129 : 13% de méthylation à 3dpp ; 29,5% à 3mois et 42,2% à 18 mois. Par contre, dans le sens 129xB la méthylation paternelle reste stable entre 3dpp et 3 mois aux environs de 12/14% mais à 18 mois l'allèle maternel perd un peu de méthylation et l'allèle paternel en regagne (57,7% versus 41,2%). Etant donné qu'on observe de la méthylation sur l'allèle paternel dès 3dpp, j'ai voulu vérifier si cette méthylation était déjà présente à des stades plus précoces ou si elle était simplement le résultat d'une hétérogénéité cellulaire dans le poumon. J'ai donc analysé l'allélisme de méthylation de la DMR de *Socs5* dès le début du développement dans ces deux sens de croisements (Figure5D). J'ai alors pu observer que la DMR présentait une méthylation préférentielle de l'allèle maternel dans les embryons à 8,5dpc. On observe 71,3% de méthylation sur l'allèle maternel et 14,9% sur l'allèle paternel dans le sens Bx129 et 65,2% versus 10,4% dans le sens 129xB. Dans les croisements réciproques Bx129 et 129xB, les allèles paternels sont protégés de la vague de méthylation *de novo* du génome après l'implantation et les spécificités parentales de méthylation de la DMR sont globalement gardées tout au long de la vie, à l'exception du foie.

3.2.5- Marques d'histones associées à la DMR

Sachant que l'asymétrie de méthylation parentale des DMR/ICR est associée à des marques différentielles d'histones sur la chromatine (McEwen *et al.*, 2010), je me suis intéressée à la présence de marques permissives (H3K4me2/3) et répressives (H3K9me2/3 et H3K27me3) qui reflèteraient un état transcriptionnel différent des deux allèles parentaux, ainsi qu'à la présence de marques reflétant une activité potentielle d'« enhancer » de la DMR de *Socs5* (H3K4me1 et H3K27ac). Pour ce faire, j'ai utilisé la technique de ChIP (Immunoprécipitation de la chromatine). Le plus simple étant de travailler à partir de cellules en culture pour cette technique, j'ai d'abord regardé si l'empreinte maternelle de la DMR de *Socs5* était conservée dans différents modèles cellulaires hybrides disponibles au Laboratoire, des cellules souches Cx129 et 129xC

ainsi que des MEF BxC et CxB (FigureSup4). Malheureusement, les cellules ES montrent une perte d'empreinte parentale, étant dépourvues de méthylation dans les deux sens de croisement, avec moins de 10% de méthylation sur l'allèle maternel. Pour les MEF, alors que l'empreinte maternelle est globalement maintenue en BxC, les CxB regagnent de la méthylation sur l'allèle paternel, comme je l'avais déjà observé dans les tissus analysés du même sens de croisement (Figure4A).

Ne pouvant utiliser ces deux modèles cellulaires pour réaliser mon expérience de ChIP-qPCR, j'ai poursuivi directement sur un tissu conservant l'empreinte, le poumon, issu de croisements Bx129 et 129xB (Figure7A). En terme de marques répressives, j'ai observé un faible enrichissement en H3K9me3 mais pas en H3K9me2 au niveau de la DMR par rapport aux régions avoisinantes. Les marques Polycomb H3K27me3 ne sont également pas enrichies à la DMR et autour. En terme de marques activatrices ou permissives, H3K4me1 est retrouvée significativement plus fortement à la DMR de *Socs5* par rapport à la région promotrice/5' UTR du gène. On retrouve également un léger enrichissement en H3K4me2 par rapport aux régions avoisinantes mais en moindre quantité que sur la 5'UTR, avec une même tendance mais encore plus diminuée en H3K4me3. En rapport avec une potentielle fonction d'enhancer de la DMR que laisserait sous-tendre la présence de H3K4me1, on ne trouve cependant pas d'enrichissement en H3K27ac. Au vu des modifications d'histones associées à la DMR de *Socs5* dans le poumon à 3dpp, il est difficile d'en présager avec certitude le rôle sur la régulation du locus.

Cependant, je me suis tout de même intéressée à l'allélisme des différentes modifications d'histones présentes au niveau de la DMR, par une analyse par pyroséquençage après ChIP (Figure7B). J'ai d'abord vérifié que la quantité d'input et des histones H3 était égale sur les deux allèles parentaux : 55% versus 45% d'enrichissement sur l'allèle maternel et paternel respectivement pour l'input et 56% versus 44% pour H3. On peut voir que la marque permissive H3K4me2 est enrichie à 75% sur l'allèle paternel et que la marque répressive H3K9me3 est quant à elle enrichie à 75% mais sur l'allèle maternel. L'association de ces marques avec l'asymétrie de méthylation présente au niveau de la DMR irait dans le sens d'une permissivité transcriptionnelle à partir de l'allèle paternel déméthylé et une répression de l'allèle maternel méthylé.

3.2.6- Activité transcriptionnelle au niveau du locus *Socs5*

Afin de comprendre plus en avant le rôle de cette DMR, je me suis penchée sur sa relation avec la transcription du locus *Socs5*. Celle-ci se situant en position intronique, elle ne devrait pas être trouvée dans l'ARN messager mature du gène *Socs5*, à moins qu'elle soit incorporée dans un transcrit alternatif du gène qui n'a pas encore été décrit. J'ai collecté différents tissus (2 animaux/tissu) issus de souris C57Bl6/J à 3dpp et réalisé une RT-PCR quantitative afin de mesurer le niveau d'expression de la DMR et des différents transcrits caractérisés de *Socs5*, en 5'UTR et 3'UTR (Figure8A). Lorsque la mesure est prise en 5'UTR, le gène *Socs5* montre une forte expression dans le poumon et le cortex mais aucune expression dans le foie et la rate. On observe 10 fois plus d'expression dans le poumon et 18 fois plus dans le cortex comparé au foie ou à la rate. En 3'UTR, on retrouve des résultats tissulaires concordants, avec 6 fois plus d'expression dans le poumon et 10 fois plus dans le cortex par rapport au foie. On note aussi que pour un même tissu, la mesure de l'expression de *Socs5* en 3'UTR est plus importante qu'en 5'UTR (70 fois supérieure pour le poumon par exemple). Bien qu'il soit difficile de comparer l'efficacité relative de différents couples de primers, ceci pourrait suggérer la présence de plusieurs transcrits incorporant la région 3'UTR sans être initiés nécessairement en 5'UTR.

En relation avec la présence de méthylation en H3K4, une possibilité attrayante serait que la transcription de certaines formes alternatives de *Socs5* soit initiée dans la DMR intragénique. Par RT-qPCR, j'ai en effet pu mesurer une forte quantité de transcrit au niveau de la DMR, avec la même spécificité tissulaire que la forme canonique de *Socs5* (poumon et cortex, mais pas foie et rate). Dans le poumon en particulier, la DMR est entre 4 à 5 fois plus abondante que la 5'UTR mais la 3'UTR est 15 fois plus exprimée que la DMR. Il existerait différents transcrits alternatifs, certains débutant en 5'UTR et incorporant la DMR, et d'autres débutant après la 5'UTR et incorporant ou non la DMR.

Les modèles de prédiction de type AceView suggèrent l'existence de différentes formes transcrites de *Socs5*, avec notamment l'incorporation d'un exon alternatif situé environ 4 kb en amont de la DMR (Figure8B). J'ai donc cherché à caractériser les différents transcrits alternatifs, prédits et nouveaux, au niveau du locus de *Socs5* par RT-PCR suivi de séquençage à l'aide de différents couples de primers sur ARN total de poumon Bl6 à 3dpp (Figure7B). En plus de confirmer l'existence de différents transcrits alternatifs déjà décrits, j'ai mis en évidence de nouvelles isoformes incorporant la DMR

grâce à des primers ancrés dans la partie 5' de cette région. Ces transcrits ne semblent pas être soumis à un contrôle d'épissage strict puisqu'ils contiennent une part variable de la DMR, ainsi qu'une partie plus ou moins grande de l'intron 2 entre la DMR et l'exon 2. Les primers utilisés pour les détecter étant situés au début de la DMR pour le forward et au début de l'exon 2 pour le reverse, il m'est impossible de savoir si ces transcrits alternatifs s'étendent au delà en amont de la DMR et/ou en aval de l'exon 2. Toujours dans le poumon, par RT-PCR à l'aide d'amorces brin-spécifique, il semblerait que la transcription soit bidirectionnelle au niveau de la DMR, mais ces résultats restent à confirmer (données non présentées ici).

Pour déterminer la taille approximative de ces transcrits, j'ai réalisé un Northern blot sur de l'ARN total de foie, poumon et cortex à 3dpp, en utilisant 3 sondes (une au niveau du 5'UTR, une dans la DMR et une au 3'UTR du gène *Socs5*) (Figure8C). Malheureusement, il est difficile de prédire la structure de ces transcrits par les tailles observées sur la membrane. On peut quand même observer, avec la sonde reconnaissant la DMR, que l'on détecte spécifiquement une bande de 1,25kb présente uniquement dans le poumon et le cortex où l'empreinte parentale est conservée, alors que dans le foie où l'allèle paternel a regagné de la méthylation cette bande est absente. Pour définitivement identifier les sites d'initiation et de terminaison de ces nouveaux transcrits, j'ai voulu réaliser des expériences de RACE-PCR sur de l'ARN total de poumon à 3dpp ; malheureusement, le niveau d'expression ou de transcrits stables s'est avéré trop faible pour cette technique.

La partie codante (ORF) de *Socs5* est contenue dans le dernier exon, l'exon 2. Si les transcrits initiés dans la DMR contiennent l'exon 2 en entier, alors toutes les parties incorporées en amont pourraient correspondre à une 5'UTR et ces transcrits auraient la même séquence codante que la forme canonique de *Socs5*. C'est la prédiction que donne le logiciel CPC (Coding Potential Calculator). Afin d'en savoir plus sur le potentiel codant de ces transcrits contenant la DMR, j'ai répété mes RT-PCR après isolation des ARN messagers (fraction PolyA+), toujours à partir de poumon. Cependant, l'absence d'amplification en comparaison de l'ARN total indique que les transcrits incorporant la DMR ne sont vraisemblablement pas maturés en ARN messagers et n'ont donc pas de potentiel codant (Figure8D).

Enfin, j'ai tout de même regardé quel était l'allélisme d'expression de ces différents transcrits par RT-PCR suivi de pyroséquençage dans le poumon à 3dpp dans

les croisements Bx129 et 129xB toujours à l'aide de SNP exprimés entre ces deux lignées (Figure6E). Au niveau de la DMR au moins, on observe une représentation égale des deux allèles parentaux : 55% d'allèle maternel et 45% d'allèle paternel dans le sens Bx129 et 50% à partir des deux allèles dans le sens 129xB. Il semble donc que la méthylation différentielle au niveau de la DMR n'ait pas d'impact sur son incorporation dans des transcrits. Tout ceci pose donc la question du rôle de la DMR de *Socs5* et de ces nouveaux transcrits alternatifs au sein de ce locus. Cependant, l'enrichissement en marques H3K4me1, la présence de transcription et probablement bidirectionnelle, l'absence de régulation post-transcriptionnelle de type épissage et polyadénylation favorisent l'hypothèse que la DMR de *Socs5* pourrait agir comme un enhancer.

3.2.7- Conservation et évolution du locus *Socs5* dans d'autres espèces

Pour gagner en visibilité sur le potentiel rôle de la DMR de *Socs5* chez la souris, j'ai regardé si celle-ci était conservée chez d'autres mammifères. Bien que les gènes soumis à empreinte soient généralement conservés entre les différentes espèces de mammifères, il existe aussi des cas d'évolution espèce-spécifique. J'ai tout d'abord interrogé la conservation nucléotidique du gène *Socs5* chez différentes espèces mammifères par les outils PhyloP et GERP (Figure9A). Alors qu'on observe un fort degré de conservation pour les parties codantes et même certaines parties introniques, la séquence de la DMR en elle-même est très faiblement conservée (rat, homme, chien, cheval...). Une analyse plus poussée montre que la DMR contient une insertion de 181nt, riche en CG et qui contient les 4 motifs de liaison à la protéine Zfp57. Cette séquence particulière, qui est le siège de la méthylation différentielle, est spécifique du génome de la souris et n'est même pas présente chez d'autres rongeurs proches tel que le rat. Les DMR espèce-spécifique sont souvent le résultat d'une insertion d'un rétrogène, qui possède une séquence homologue ailleurs dans le génome (Wood *et al.*, 2007). Or, pour la DMR de *Socs5*, il n'existe aucune séquence homologue ailleurs dans le génome de la souris, ce qui pose la question de l'origine de cette séquence de 181nt (Figure9B). La DMR de *Socs5* est donc une acquisition évolutive récente, unique à la souris.

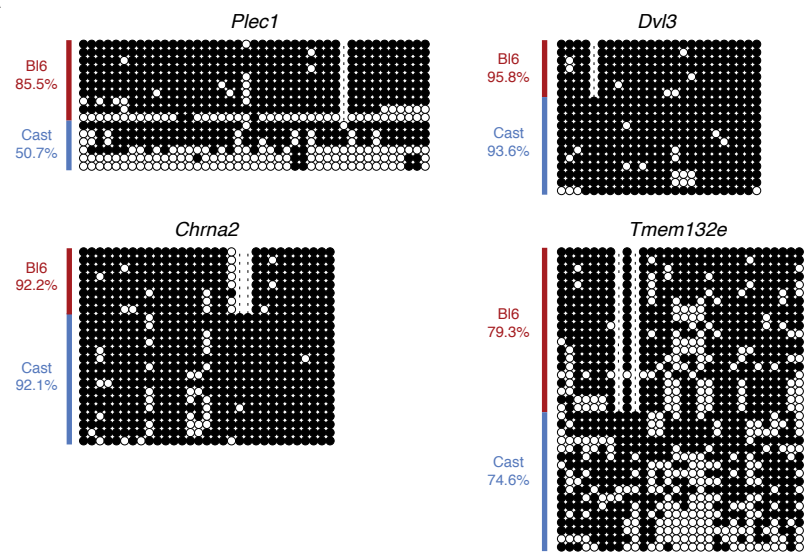
3.2.8- Délétion génétique de la DMR de *Socsc5* *in vivo* par le système CRISPR/Cas9

Afin de comprendre l'intérêt fonctionnel de la DMR de *Socsc5*, j'ai finalement réalisé une délétion de cette région *in vivo* chez la souris grâce à la technique de génie génétique CRISPR (clustered regularly interspersed short palindromic repeats)/Cas9 (Cong *et al.*, 2013 ; Sander *et al.*, 2014). Cette technique consiste en un système composé d'une endonucléase (la protéine Cas9) qui va cliver l'ADN au niveau d'une séquence spécifique ciblée par une molécule d'ARN simple brin appelée ARN guide (single guide RNA ou sgRNA). La protéine Cas9, de part son activité de nucléase, va générer des cassures double-brins au niveau de la séquence d'ADN qui aura été ciblée par le sgRNA. Ces cassures double-brins vont alors être réparées par les systèmes de recombinaison non-homologues des extrémités des cassures (non-homologous end joining ou NHEJ) ou par recombinaison homologue des extrémités (homology-directed repair ou HDR) quand une matrice de réparation est fournie. Pour la délétion de la DMR de *Socsc5*, j'ai utilisé deux sgRNA pour créer deux cassures : une en amont et une en aval de la séquence de la DMR, pour une délétion attendue de 1,8kb (Figure10A). Les sgRNA ont été générés à l'aide du logiciel CRISPR Design (<http://crispr.mit.edu>). Les possibles « off-targets » de ces sgRNA sont listées dans la table2.

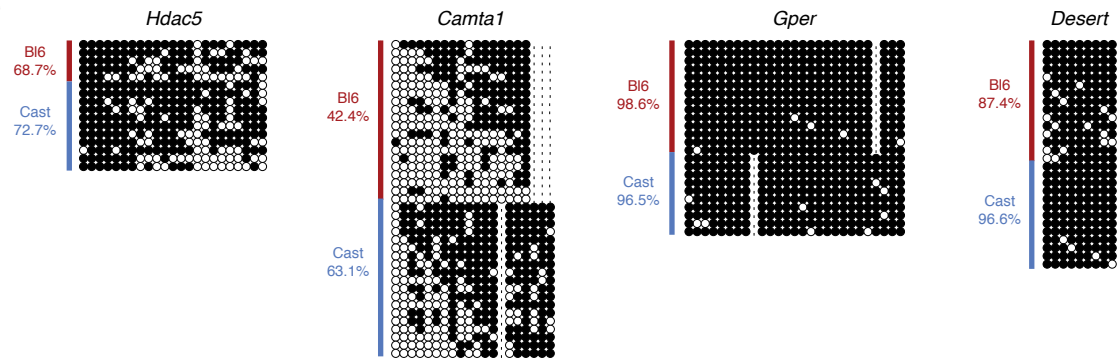
Afin de générer une délétion *in vivo* chez la souris, le système CRISPR/Cas9 nécessite la production de l'ARN messenger de Cas9 et des deux sgRNA et leur injection dans des embryons au stade zygote. En pratique, j'ai produit ces ARN par transcription *in vitro* selon les protocoles publiés (Wang *et al.* 2013) et les injections ont été réalisées par Fatima El Marjou et Colin Jouhanneau de la plateforme de transgénèse de l'Institut Curie dans des embryons fécondés C57Bl6/N collectés à 0.5dpc (Figure10B). Les ARN ont été injectés directement dans le cytoplasme des embryons à des concentrations de 100ng/ml pour l'ARNm de Cas9 et 50ng/ml pour les sgRNA. Après quelques heures en culture, les embryons injectés ont été transférés (stade zygote) dans des femelles pseudo-gestantes CD1 (25 à 30 embryons/femelle). En terme de rendement, 4 femelles ont donné naissance à 32 petits, parmi lesquels 3 présentaient une délétion au niveau de la DMR de *Socsc5*. L'analyse par séquençage d'ADN de queue a révélé que seul un animal présentait la délétion complète attendue (1,8kb) à l'état hétérozygote, les deux autres ayant des délétions de taille non conforme (Figure10C). Pour information, ce rendement est assez faible par rapport à d'autres loci que notre laboratoire a pu cibler *in vivo* par cette méthode.

L'animal mutant obtenu est de sexe mâle, viable et ne montre pas d'anomalies phénotypiques évidentes, mais il est impossible de savoir sur quel allèle la délétion a eu lieu. Ce fondateur a donc été croisé avec des femelles C57Bl6/J pour amplifier la colonie, et s'est ainsi avéré fertile. L'obtention d'animaux hétérozygotes F1 démontre de plus que la délétion était bien présente au niveau de la lignée germinale du fondateur. Les animaux F1 hétérozygotes mâles et femelles issus de ce back-cross, viables donc également, sont actuellement en cours de croisement, pour suivre les effets d'une transmission paternelle, maternelle mais également homozygote de la délétion. Je vais par la suite regarder dans le poumon et d'autres tissus à 3dpp chez les animaux F2 si la transmission de la délétion a un impact sur la transcription au niveau du locus, notamment sur les différents transcrits alternatifs où la DMR est incorporée mais également du transcrit canonique de *Socs5*. Ceci devrait nous permettre de tester s'il s'agit bien d'une séquence enhancer et si la méthylation parentale joue ou non un rôle sur sa fonction.

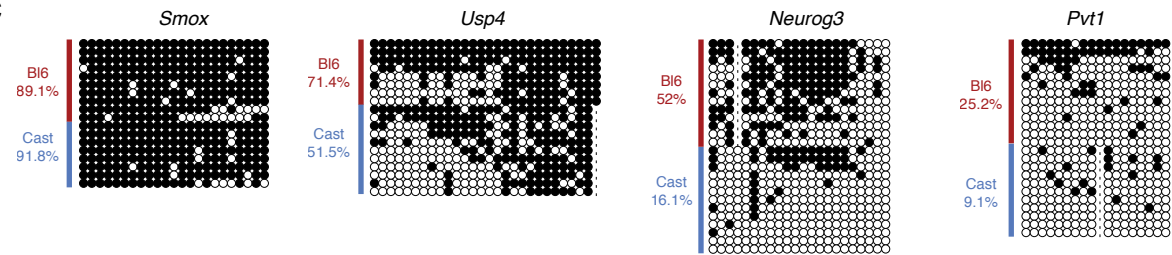
A



B



C



3.3- Figures et légendes

Figure1 : Methylation profil of 12 gDMR maternal candidates in BxC 9.5dpc embryos

A. Maternal gDMR candidates identified in our MeDIP-Seq screen, with a sliding window of 500bp, which co-localize with a CpG island and contain Zfp57 binding motif in their sequences.

B. Maternal gDMR candidates identified in our MeDIP-Seq screen, with a sliding window of 1000bp, which co-localize with contain Zfp57 binding motif in their sequences.

C. Maternal gDMR candidates identified in other screens (Kobayashi et al., 2012; Xie et al., 2012).

Methylation profile of the maternal candidates gDMR in hybrid WT embryos at 9.5dpc by bisulfite cloning/sequencing. Red and blue lines delineate maternal (Bl6 origin) and paternal (Cast origin) alleles respectively. Black circles : methylated CpG, white circles : unmethylated CpG, dash : absent CpG corresponding to strain-specific SNPs. Mouse strains : Bl6=C57Bl6/J, Cast=CAST/Ei.

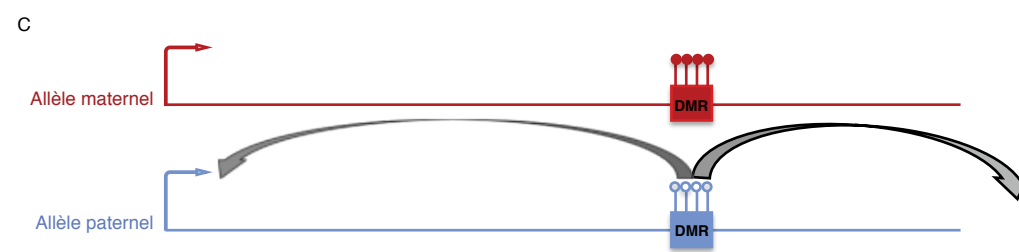
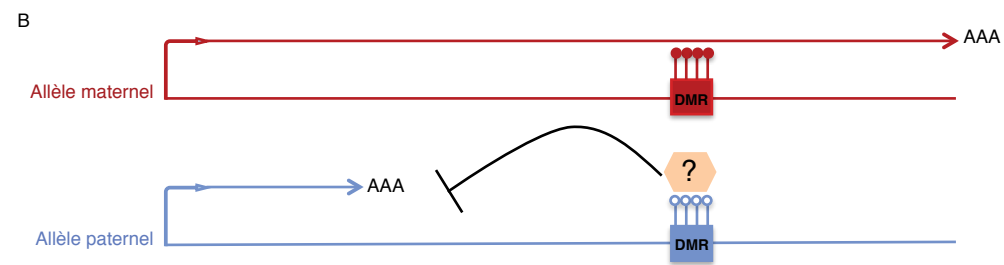
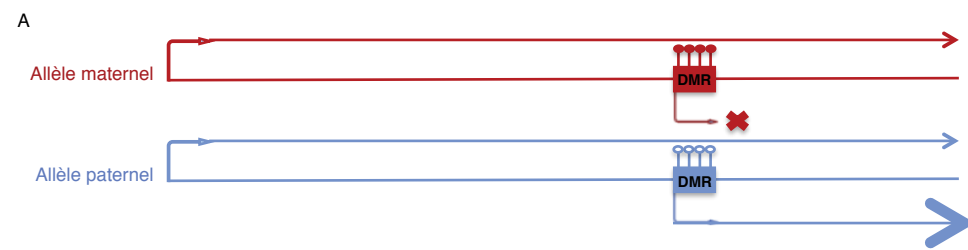


Figure2 : Three possible mechanisms for an imprinted intragenic gDMR

- A. The intragenic *Cdh15* gDMR acts as an alternatic promoter (Proudhon et al., 2012).
- B. The intragenic gDMR of *Mcts2/H13* led to a polyA signal arrest in the sequence of the canonical transcript (Wood et al., 2008).
- C. The intragenic gDMR acts as an enhancer.

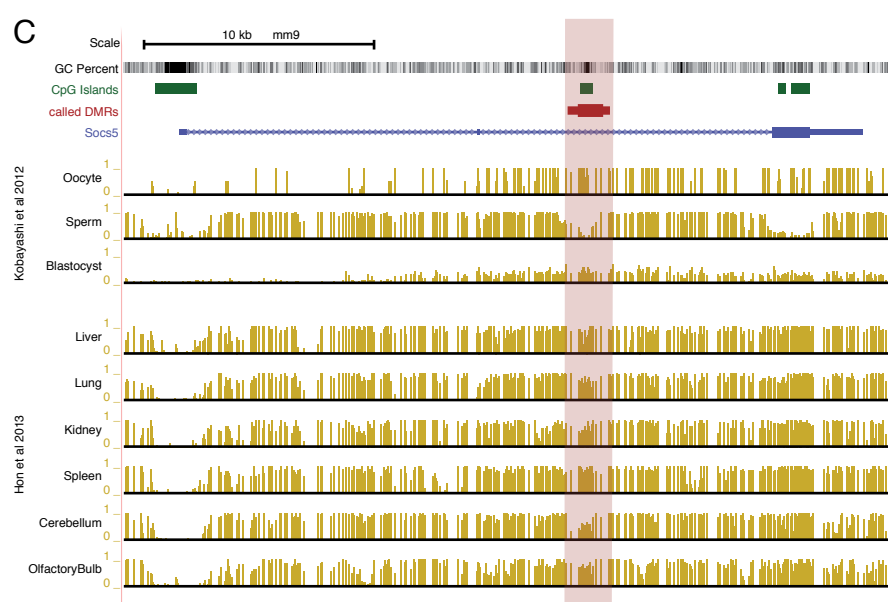
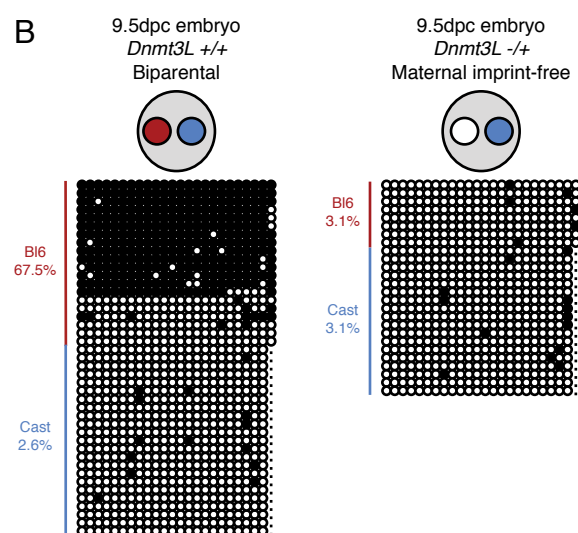
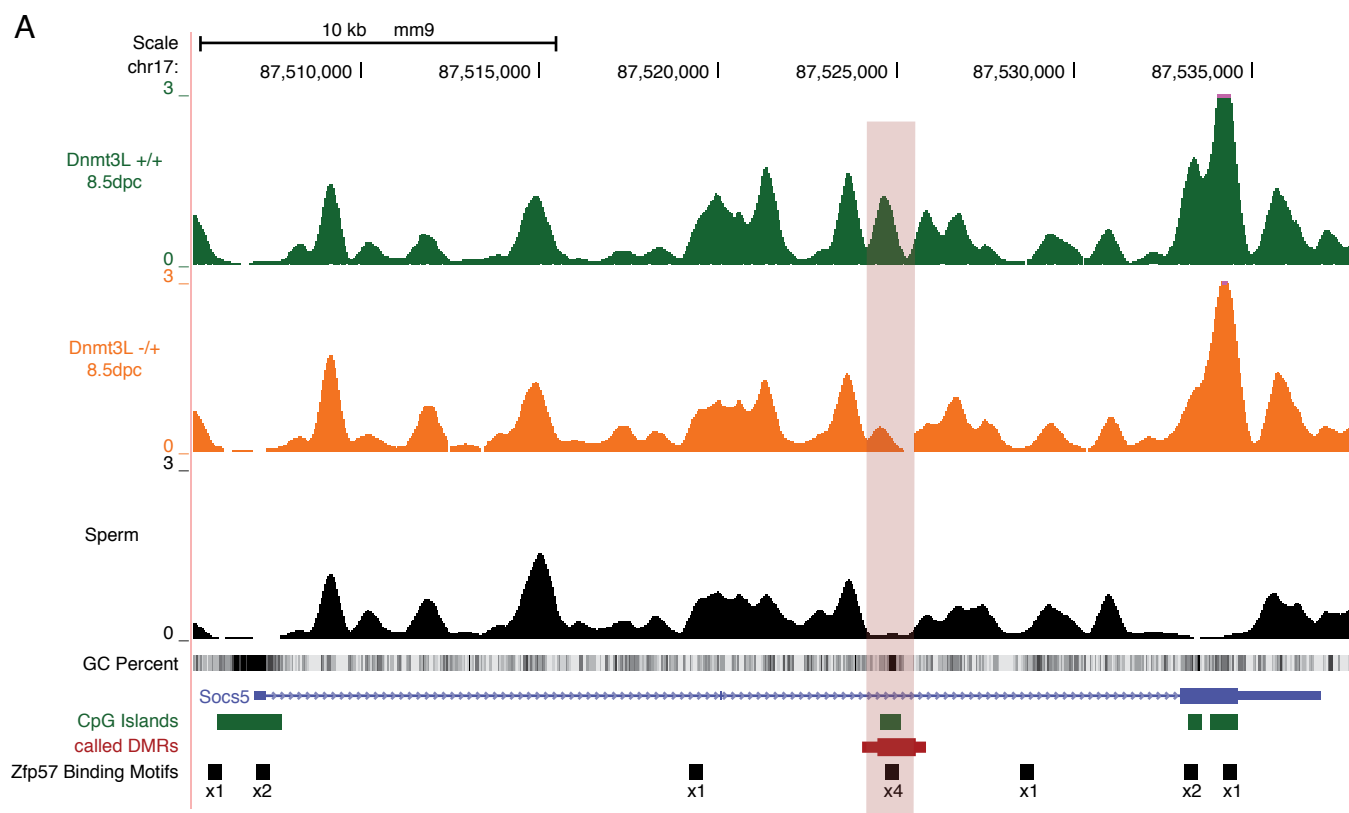


Figure3 : Characterization of the *Socs5* DMR as a new life-long maternal DMR

A. MeDIP-Seq profile of the *Socs5* locus. The gene is oriented 5' to 3' in blue. The MeDIP-Seq tracks show an intragenic CGI with hypomethylation in 8.5dpc *Dnmt3L*^{-/+} embryos (orange) and sperm (black) compared to WT embryos (in green). This intragenic CGI contain four Zfp57 binding motif. CpG islands are in green and the DMR is highlighted in red.

B. Confirmation of the methylation profile of the *Socs5* DMR in hybrid WT and mutant embryos at 9.5dpc by bisulfite cloning/sequencing. Red and blue lines delineate maternal (Bl6 origin) and paternal (Cast origin) alleles respectively. Black circles : methylated CpG, white circles : unmethylated CpG, dash : absent CpG corresponding to strain-specific SNPs. Mouse strains : Bl6=C57Bl6/J, Cast=CAST/Ei.

C. Published Whole Genome Bisulfite-Sequencing datasets over the *Socs5* gene (UCSC genome browser tracks). 0 to 1 values correspond to 0% and 100% of methylation. Reference publications for these data are indicated on the left.

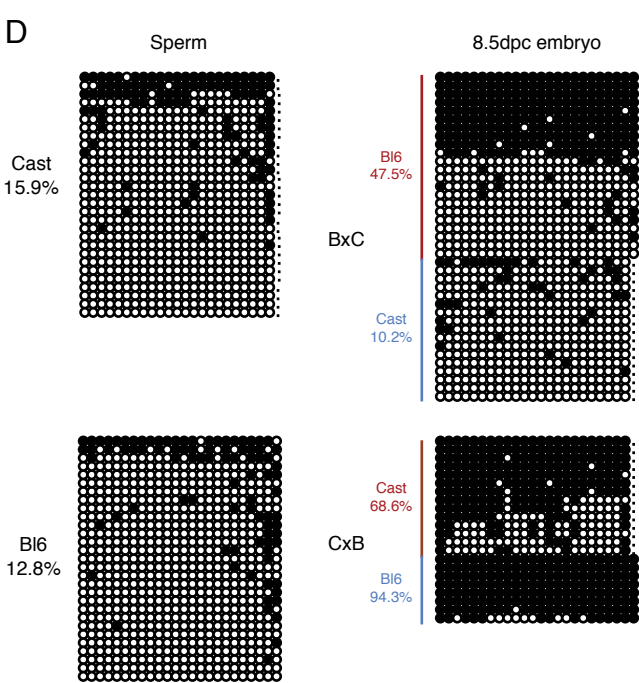
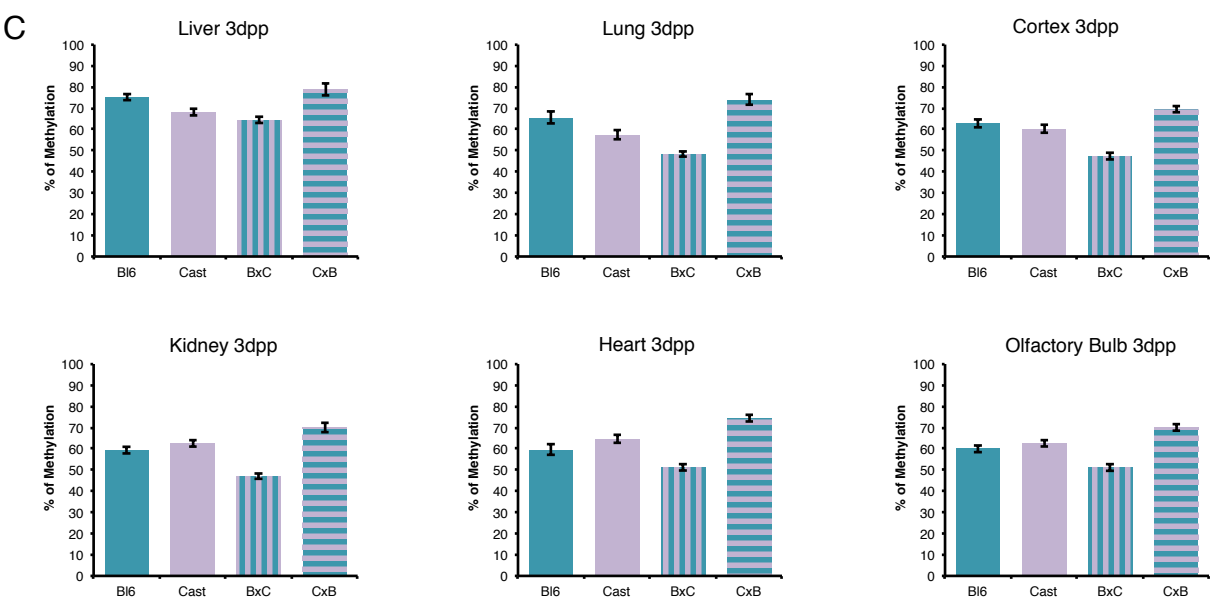
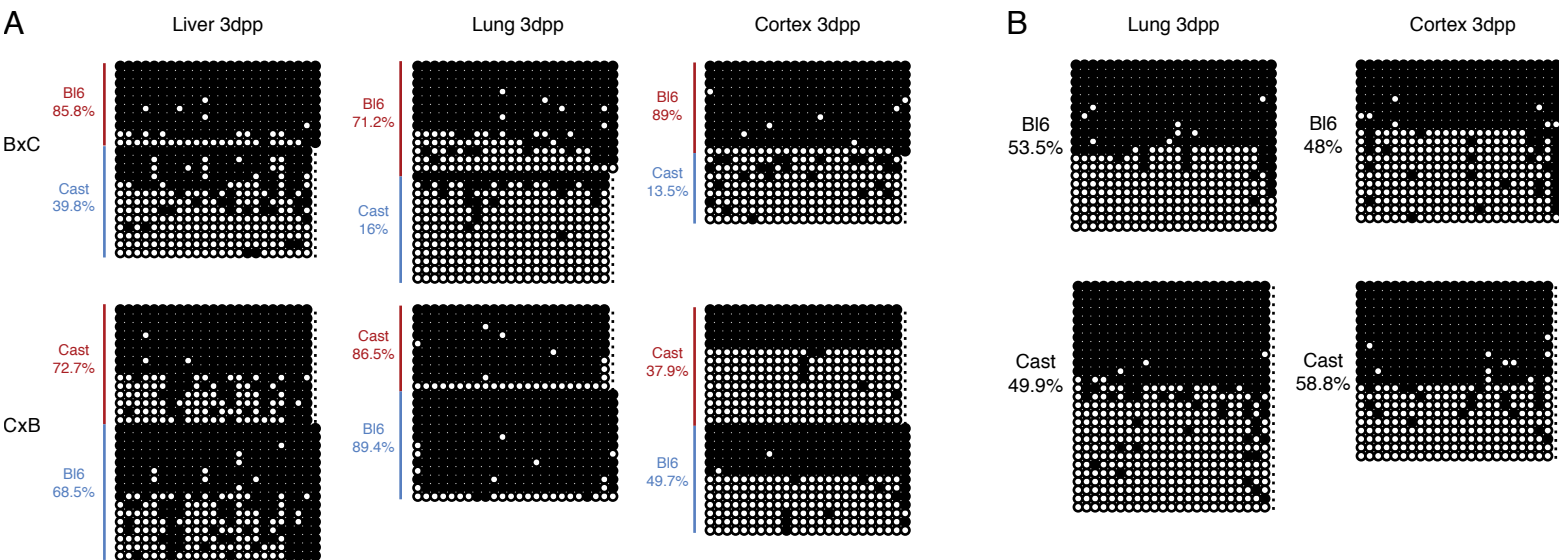


Figure4 : Variation of *Socs5* DMR methylation in different mouse tissues and strains

A. *Socs5* DMR methylation by bisulfite cloning/sequencing in three tissues (liver, lung and cortex) at 3dpp in BxC and CxB reciprocal crosses. The DMR shows methylation in liver on both parental alleles in both reciprocal crosses ; in lung, maternal-specific methylation in BxC but gain of methylation on the paternal allele in CxB crosses ; in cortex, maternal-specific methylation in BxC but gain of methylation on the paternal allele and loss on the maternal one in CxB crosses.

B. *Socs5* DMR methylation by bisulfite cloning/sequencing in lung and cortex of pure Bl6 and Cast strains. The methylation is around 50% for the two tissues in the two strains.

C. Percentage of methylation of the *Socs5* DMR by bisulfite pyrosequencing on Bl6 and Cast strains and reciprocal BxC and CxB crosses in different tissues at 3dpp. Error bars show the standard deviation from four biological replicates. Except for the liver where the methylation stays around 70%, CxB crosses show always more methylation than the BxC crosses for a same tissue.

D. Developmental analysis of *Socs5* DMR methylation by bisulfite cloning/sequencing. No methylation of the DMR is observed in Bl6 and Cast sperm. Maternal-specific methylation is observed in BxC embryos at 8.5dpc but parental specificity is lost in CxB embryos by methylation gain on the paternal allele.

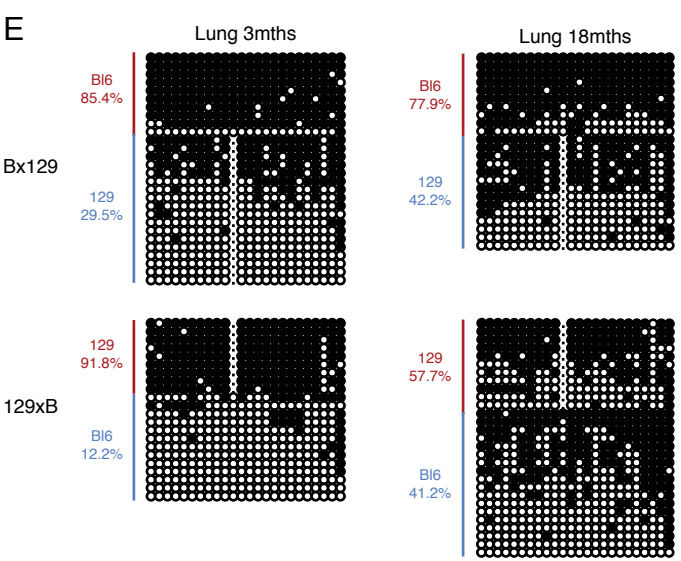
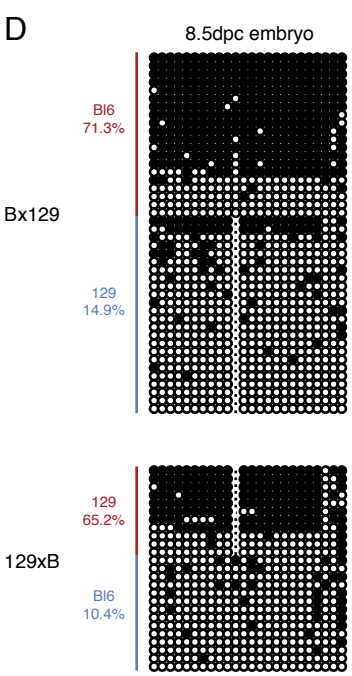
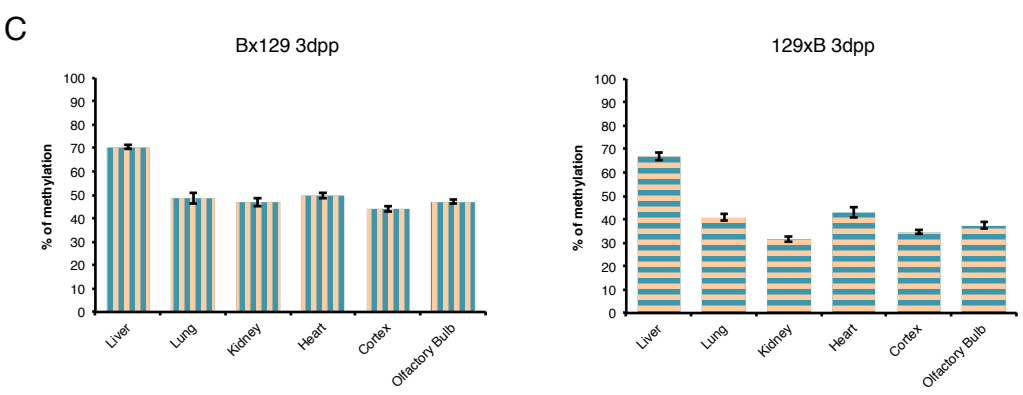
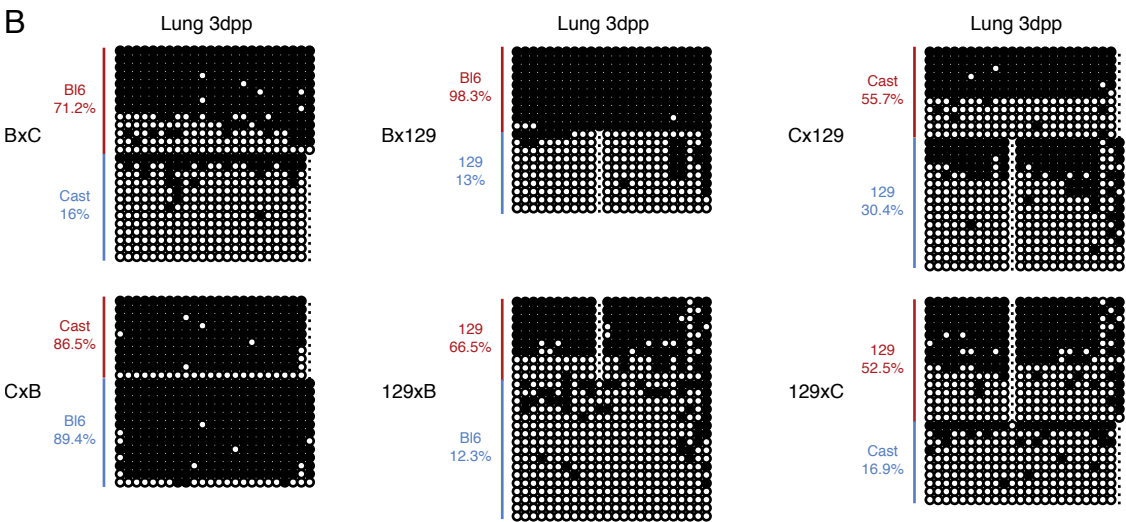
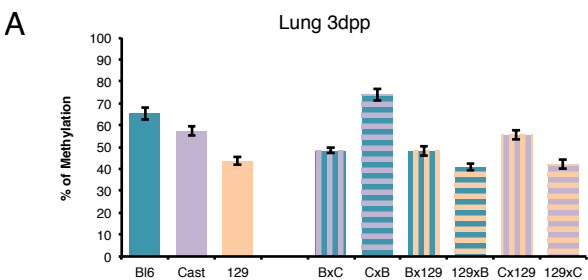


Figure5 : Tissue-specific and strain-specific methylation of the *Socs5* DMR

A. Percentage of methylation of the *Socs5* DMR by bisulfite pyrosequencing on all reciprocal crosses with Bl6, Cast and 129 strains, in 3dpp lung. The reciprocal Bx129 and 129xB crosses show 50% methylation with very little variation between the two crosses compared to the other strain combinations. Error bars show the standard deviation from four biological replicates.

B. Allelic methylation profile of the *Socs5* DMR in all Bl6 and 129 reciprocal hybrid crosses assessed in 3dpp lung by bisulfite cloning/sequencing. Bx129 and 129xB crosses show maternal-specific methylation, whereas Cx129 and 129xC crosses show some loss on the maternal allele.

C. Percentage of methylation of the *Socs5* DMR by bisulfite pyrosequencing in different tissues of 3dpp Bx129 and 129xB animals. Bx129 methylation is constant and around 50% for all tissues (except the liver) but 129xB methylation shows variations between tissues.

D. Methylation profile of the *Socs5* DMR by bisulfite cloning/sequencing in 8.5dpc embryos from Bx129 and 129xB crosses. The two crosses show maternal specific methylation, 71.3% of methylation on the Bl6 maternal allele and 65.2% of methylation on the Cast maternal allele. The Two paternal allele have around 10/14% of methylation in these crosses.

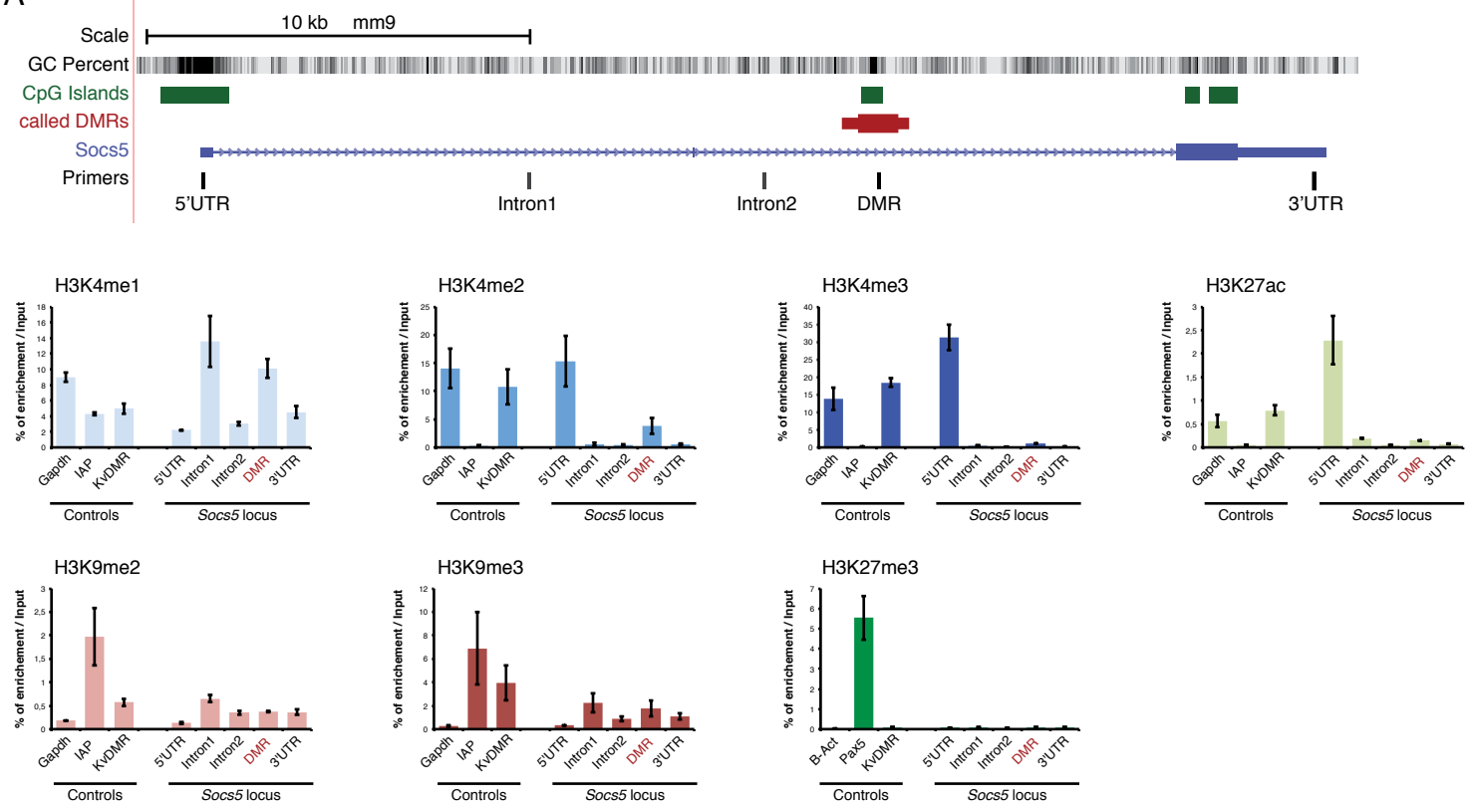
E. Methylation profile of the *Socs5* DMR by bisulfite cloning/sequencing in Bx129 and 129xB lung during life and aging. The maternal-specific methylation of the *Socs5* DMR is kept through ages until 3 months after birth but loses at 18 months by a gain of methylation on the paternal allele in both crosses. The Bx129 crosses show 85.4% and 77.9% of methylation on their maternal allele ; and 29.5% and 42.2% on their paternal allele at 3mths and 18mths respectively. The 129xB crosses show 91.8% and 57.7% of methylation on their maternal allele ; and 12.2% and 41.2% on their paternal allele at 3mths and 18mths respectively.

[illegible]

Figure6 : Sequence variations between mouse strains

Sequences alignment of the 3 mouse strains Bl6, Cast and 129. CpG are in red and the SNPs between mouse strains are highlighted in yellow. The Zfp57 binding sites (TGCCGC) are highlighted in green and the TATA box in blue. The deleted and duplicated repeats unique to the Cast strain are in orange. There is a duplicated repeat of 34bp unique to *Socs5* on the Cast strain that contains one more Zfp57 binding motif compared to the Bl6 and 129 strains but no transcription factor binding sites. There is a SNP unique to the Bl6 strain compared to the Cast and 129 strains (in red highlighted in yellow) : T on Bl6 and C on Cast and 129.

A



B

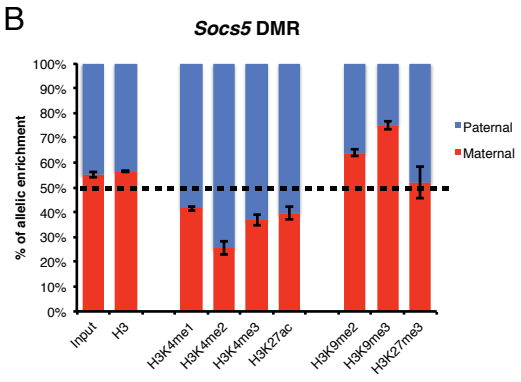


Figure7 : Post-translational histone modifications at the *Socs5* DMR

A. ChIP followed by RT-qPCR accross the *Socs5* locus on 3dpp lungs from Bx129 and 129xB crosses. The upper panel indicates the position of the qPCR primers over the *Socs5* gene. The percentage of enrichment of each specified histone mark was normalized to the input. No specific enrichment of repressive histone marks is seen at the DMR, compared to surrounding intronic sequences. An enrichment in H3K4me1 and H3K4me2 permissive modification is present at the DMR. On the left of the histograms appear some control loci: *Gapdh* as an active promoter, IAP retrotransposons as repressed regions and *KvDMR* as a canonical imprinted maternal ICR (associated with the *Kcnq1ot1* locus). For each cross, a pool of four biological tissues were used ; error bars show the standard deviation between the two crosses.

B. ChIP-pyrosequencing on 3dpp lungs from reciprocal Bx129 and 129xB. Genomic DNA (Input) and H3 IP were used to exclude assay-specific biases. Enrichment on the maternal allele is in red and on the paternal allele in blue. The permissive H3K4me2 mark and the repressive H3K9me3 mark show the strongest opposite allelic enrichment : H3K4me2 is enriched on the paternal allele (75%) whereas H3K9me3 is enriched on the maternal allele (75%). Error bars show the standard deviation between the two crosses.

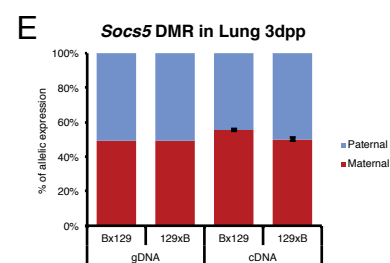
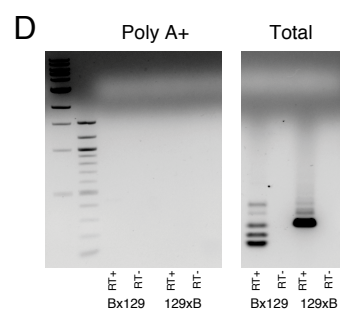
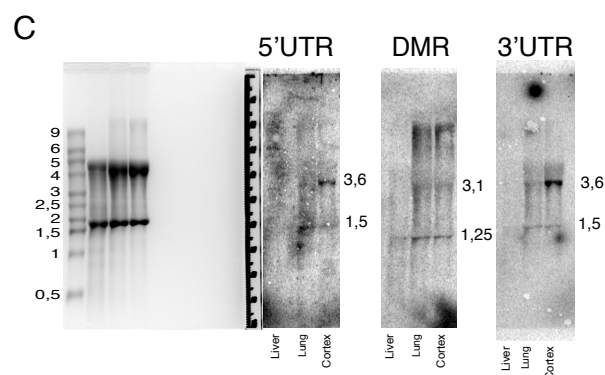
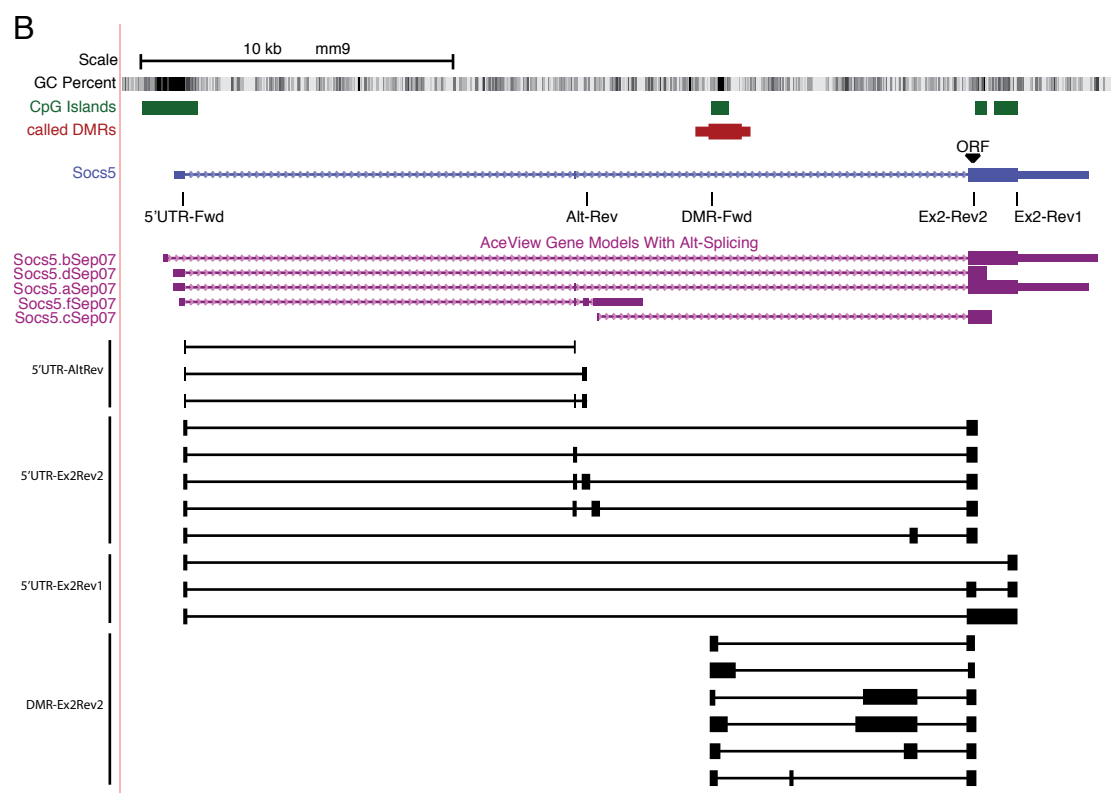
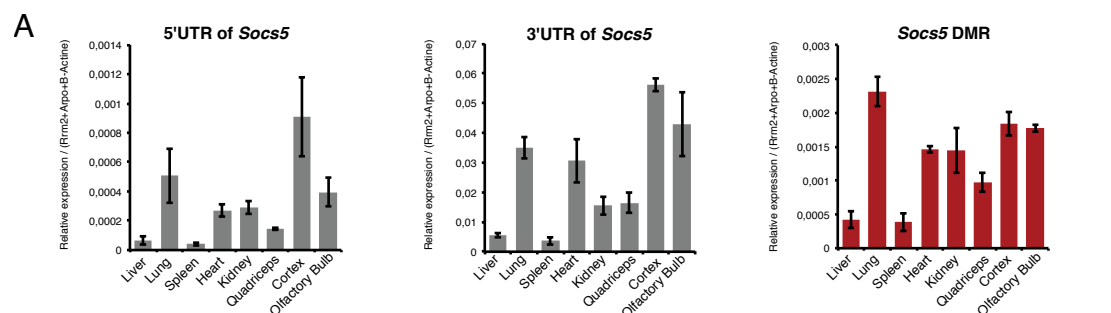


Figure8 : Transcription activity at the *Socs5* locus

A. RT-qPCR assays showing expression of the *Socs5* gene when measured at the 5'UTR and 3'UTR of *Socs5* in different tissues (in particular lung and cortex) at 3pp, but no expression in liver and spleen. Measurement at the DMR of *Socs5* shows the same tissue-specificity, in levels intermediate between the 5' and the 3'UTR. Error bars show the standard deviation from two biological replicates. Relative expression was normalized to the average between *Rrm2*, *Arpo* and β -*Actin* gene expressions.

B. RT-PCR assays followed by Sanger sequencing amplify different transcript isoforms of *Socs5* in total lung RNA at 3dpp. The upper panel indicates the position of the Open Reading Frame of *Socs5* (ORF- starting in exon 2), predicted transcripts by AceView and the position of the PCR primers. With primers mapping in the DMR and the exon 2, transcripts including the DMR were amplified. They also incorporate some parts of the intron 2, with various lengths.

C. Northern blot of total RNA at 3dpp from liver, lung and cortex. On the left, there is a picture of the ethidium bromide gel of total RNA used for the transfer on the membrane. The membrane has been hybridized with three probes : one in the 5'UTR, one in the DMR and one in the 3'UTR. With the 5' and 3'UTR probes we catch two different bands at 3.6 and 1.5kb only in lung and cortex. With the DMR probe, we catch two bands of smaller size at 3.1 and 1.25kb also in the lung and cortex.

D. Ethidium bromide gel of PolyA+ isolated fraction of RNA and total RNA after RT-PCR with the primers DMR-Forward and Ex2-Reverse2 in lung 3dpp. RT+ corresponds to the converted RNA in cDNA and RT- to non-converted RNA. No amplification is seen in the PolyA+ fraction, while different amplification products are seen in total RNA.

E. Allelic expression measured at the *Socs5* DMR by RT-PCR followed by pyrosequencing in total RNA from 3dpp lung in Bx129 and 129xB crosses. Genomic DNA is used to exclude assay-specific biases. Error bars show the standard deviation between two biological replicates for each cross orientation.

Figure9 : Evolution of the *Socs5* DMR among mammals

A. Conservation of the *Socs5* locus among mammalian species. GERP and PyloP show a low rate of conservation between the mouse and other mammalian species at the intragenic CpG island coinciding with the DMR. This lack of conservation reflects the insertion of a 181bp sequence specifically in the mouse genome, and even not present in the rat genome. This sequence is CG rich (CG motifs in red) and contains four TGCCGC Zfp57 binding sites (highlighted in green).

B. Evolutionary tree of a few imprinted genes in placental mammals (adapted from Wood *et al.*, 2007). Based on a multi-species comparative sequence analysis, the approximate points in the mammalian radiation at which each of the four imprinted retrogenes originated are indicated by an arrow. Three of them (*Nap1l5*, *U2af1-rs1*, and *Inpp5f_v2*) have been described that originated by retrotransposition and overlap CpG islands which undergo methylation during oogenesis. *Mcts2* has been identified in this study as a novel candidate imprinted retrogene on mouse. *Socs5* has been added on this tree based on the probability that it is also a retrogene.

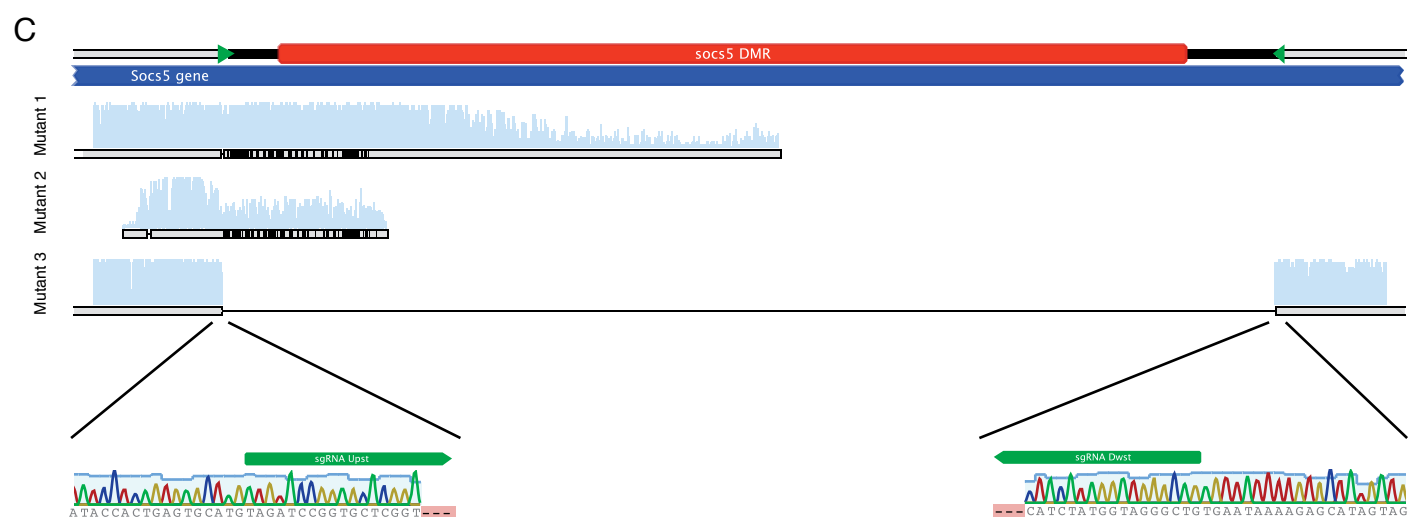
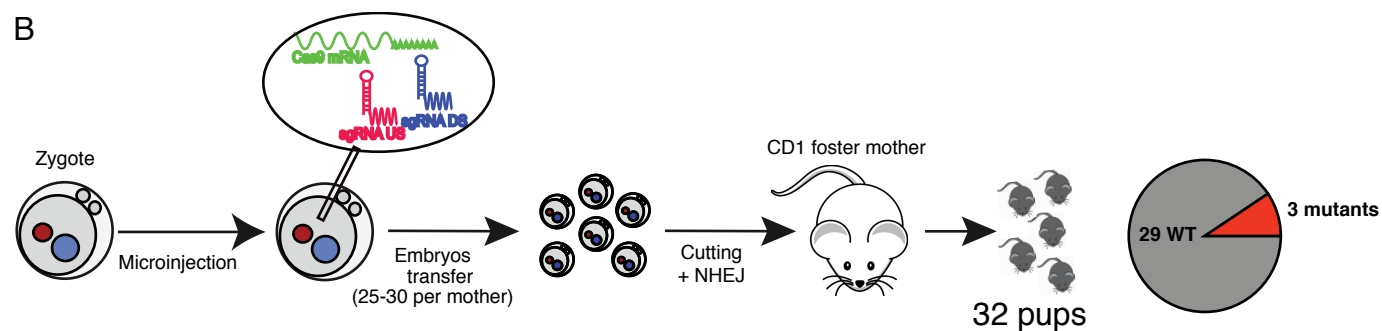
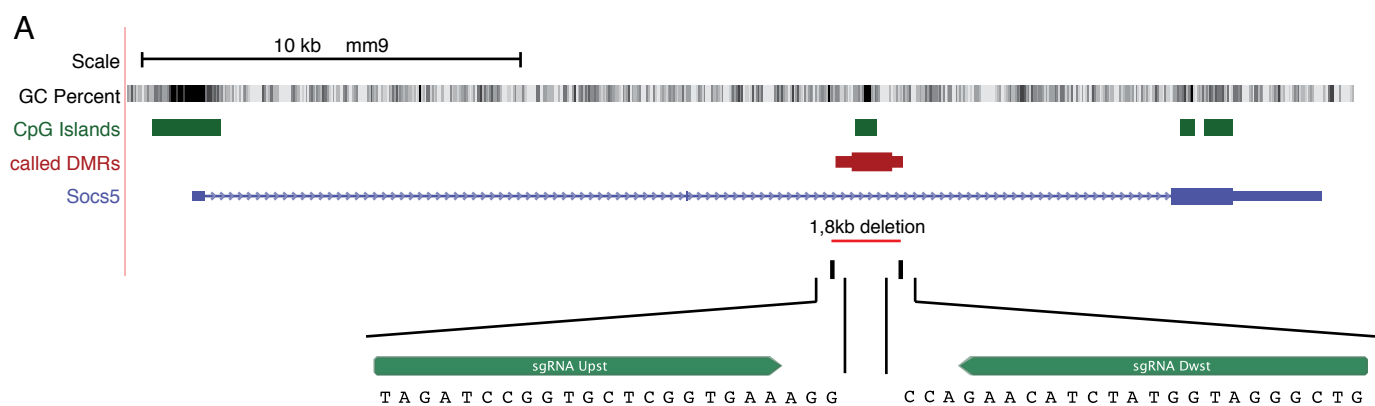


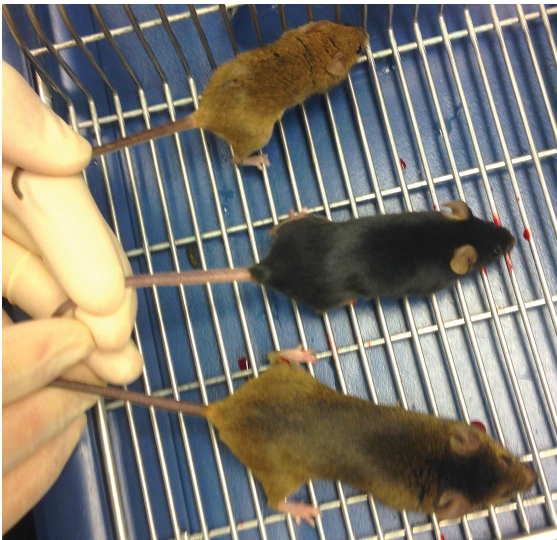
Figure10 : CRISPR/Cas9-targeted deletion of the *Socs5* DMR *in vivo*

A. Overview of the mouse *Socs5* locus, with a focus on the DMR and the CRISPR cutting sites (sgRNA Upst and sgRNA Dwst). The *Socs5* gene is represented in blue and the DMR in red. The green arrows correspond to the 20bp sgRNAs around the DMR. The PAM sequence is located at the 3' of the sgRNA. The expected deletion spanning the DMR is 1.8kb long.

B. Scheme of the procedure for generating the *Socs5* DMR deletion *in vivo* in mouse using the CRISPR/Cas9 system. The pie chart represents the number of mutant animals obtained (red) over all the 32 live pups.

C. Sanger sequencing of the three *Socs5* DMR mutants obtained by CRISPR/Cas9 engineering. The sequencing was performed on tail DNA. In red is represented the position of the DMR, the green arrows are the sgRNAs. Only the mutant 3 shows the expected deletion. The sequencing tracks highlight the position of the cutting sites, which map at 3rd nucleotide upstream of the PAM sequences (NGG). It has been reported that this position could be the site where the Cas9 cuts (Nishimasu et al., 2014 ; Gasiunas et al., 2012).

A



CAST/Ei

C57Bl/6

129SvPas



C57Bl/6NJ
9844 SNPs

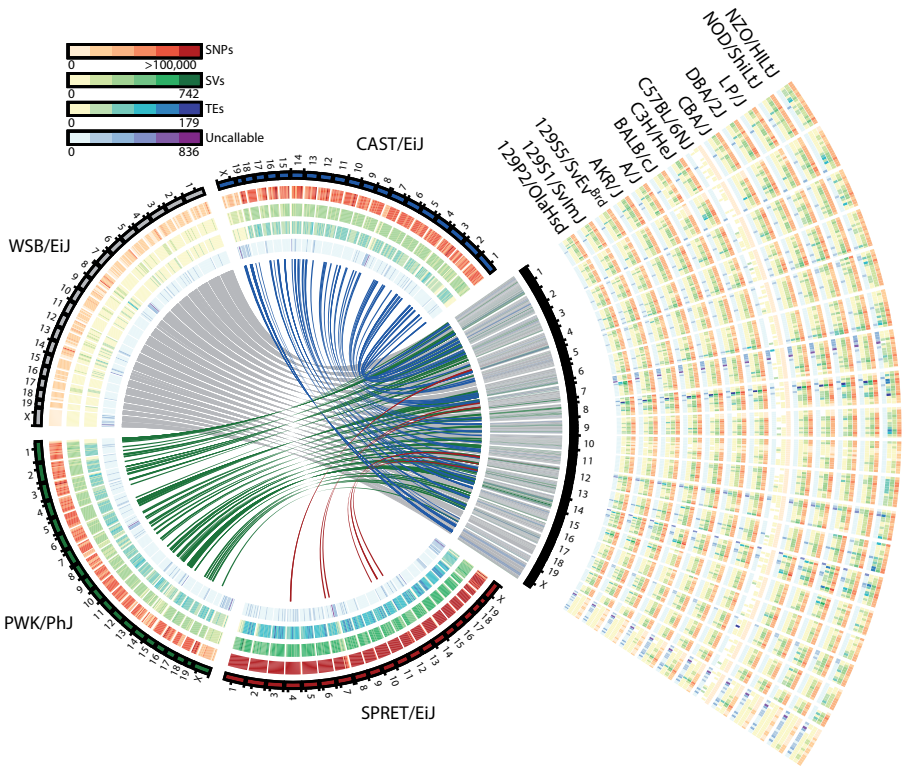


CAST/Ei
17,673,726 SNPs
entre Bl6 et Cast



129SvPas
4,458,004 SNPs
entre Bl6 et 129

B

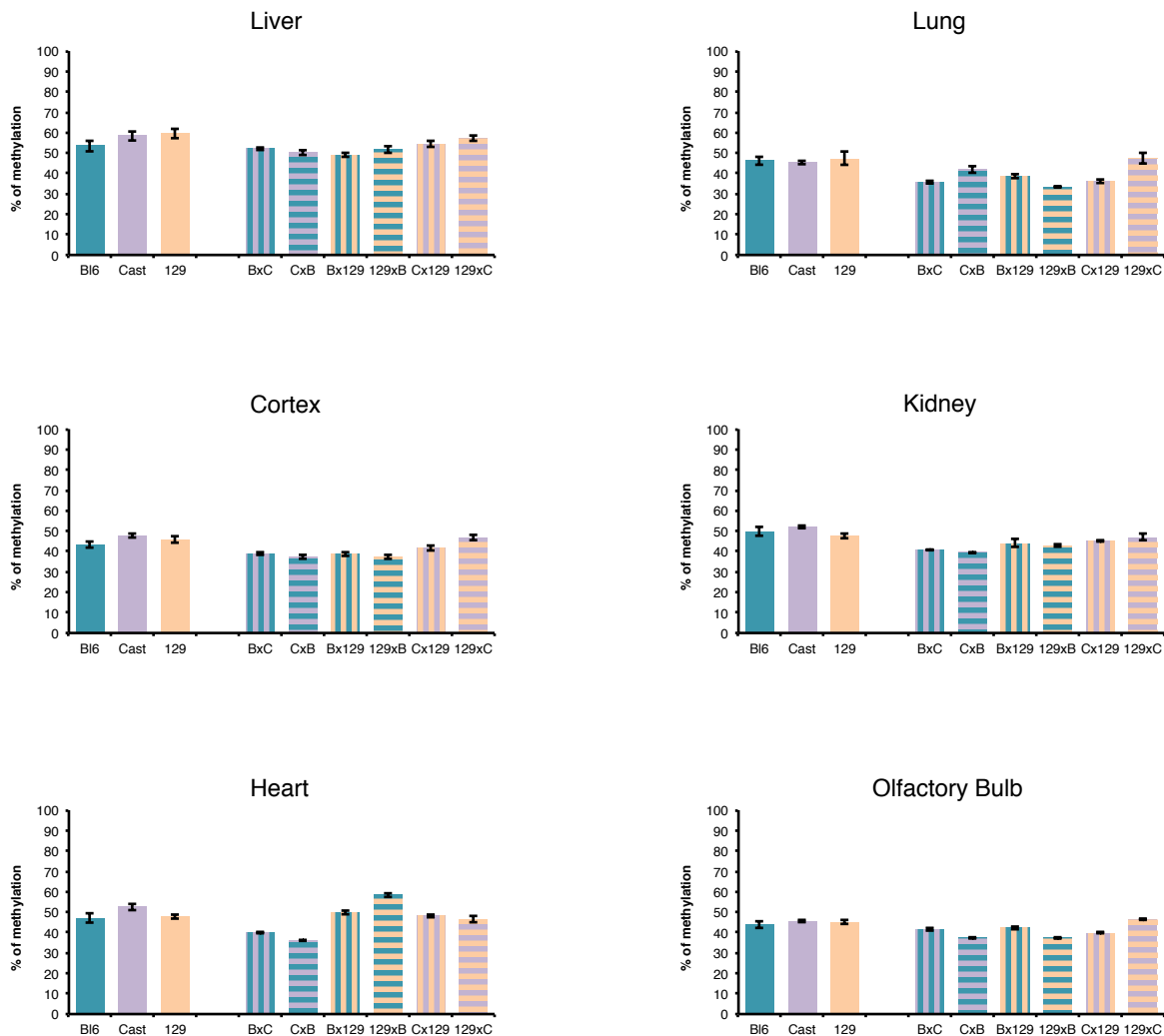


SupplementaryFigure1 : Mouse strains used for the characterization of the *Socs5* DMR

A. Pictures of the three mouse strains used in the study: 129SvPas, C57Bl/6J and CAST/Ei and the number of SNPs between Bl6 and the other strains.

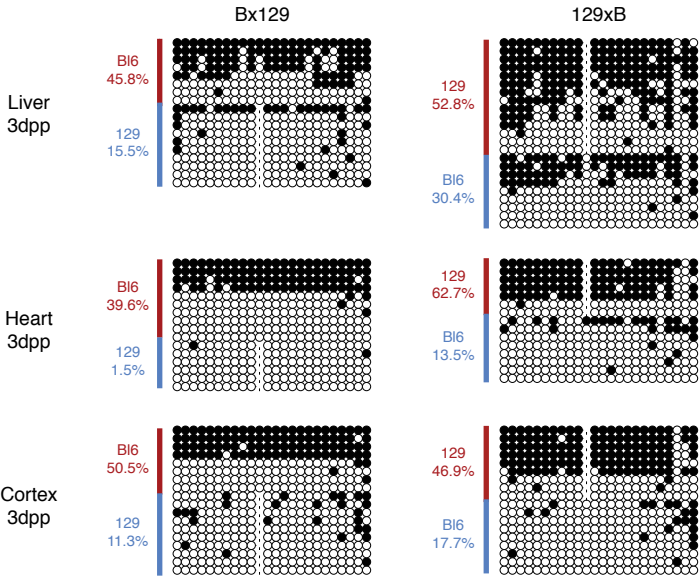
B. An overview variants called from 17 mouse genomes relative to the reference, adapted from Keane *et al.*, 2011. The four wild-derived strains (CAST/EiJ, WSB/EiJ, PWK/PhJ and SPRET/EiJ) are representative of the *Mus musculus castaneus*, *Mus musculus musculus*, *Mus musculus domesticus* and *Mus spretus* taxa and include the progenitors from which the classical laboratory strains were derived. These genomes are shown in a circle with tracks indicating the relative density of SNPs, structural variants (SVs) and uncallable regions (binned into 10-Mb regions). Transposable element (TE) insertions, which are a subset of the structural variant calls, are shown as a separate track. Corresponding tracks are shown for each of the 13 classical laboratory strains to the right of the circle. Links crossing the circle indicate regions on the reference where the wild-derived strain is closest to the reference (375-kb bins).

Mcts2/H13 maternal DMR in different tissues at 3dpp



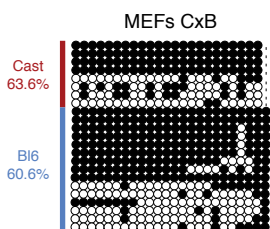
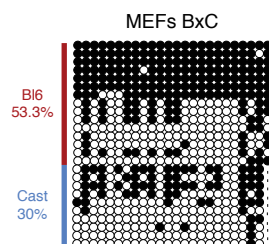
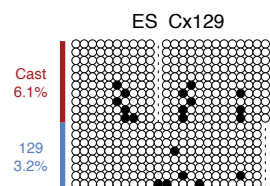
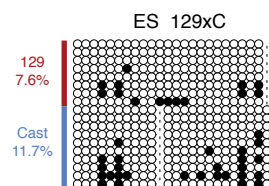
SupplementaryFigure2: Methylation status of the known *Mcts2/H13* intragenic maternal ICR

Percentage of methylation of the maternal *Mcts2/H13* ICR by bisulfite-pyrosequencing on 3dpp tissues of Bl6, Cast, 129 strains and their reciprocal crosses. For all tissues, the ICR has around 50% of methylation, as expected for a life-long, ubiquitous ICR. There is no differences of methylation among the strains and their reciprocal crosses within a same tissue, and in particular for the BxC and CxB crosses, which show the strongest variation at the *Socs5* DMR. Hybrid-specific variation can be noticed for the heart though.



SupplementaryFigure3 : Methylation status of the *Socs5* DMR among different tissues in Bx129 and 129xB crosses

Allelic methylation assessed by bisulfite cloning/sequencing in three tissues at 3dpp : liver, heart and cortex. All tissues show maternal-specific methylation, with instances of allelic losses. The paternal allele shows some gain of methylation only in the liver and in the 129xB cross orientation. Red and blue lines delineate maternal and paternal alleles respectively. Black circles : methylated CpG, white circles : unmethylated CpG, dash : absent CpG corresponding to strain-specific SNPs.



SupplementaryFigure4 : Loss of imprinting of the *Socs5* DMR in cultured cells

Methylation of the *Socs5* DMR assessed by bisulfite cloning/sequencing in two different hybrid systems of cellular models: ES cells and MEFs. The DMR shows an hypomethylated status on both parental alleles in ES cells in both reciprocal crosses 129xC and Cx129. Maternal methylation in both BxC and CxB MEFs is present but with some allelic losses ; paternal alleles regained methylation in the CxB cross.

A

Chromosome	Début	Fin	Transcript le plus proche	Promoteur	3'UTR	CGI	Blauffre-Seq (Smallwood et al., 2011 ; Xie et al., 2012 ; Hon et al., 2013)				Niveau de méthylation de l'lot CpG (Kobayashi et al., 2012)				
							Oocyte WT methylation	Sperme methylation	Blastocyste methylation	Tissus Somatics methylation	Motifs Zip57 (TGGCGC)	Wild-type ovocyte	Dmrt3L ^{4c} ovocyte	Sperme	Blastocyste
Chr8	125387861	125390344	Cdh15	Cdh15			+	-	+/-	+/-	2	100,0%	0,0%	6,7%	20,3%
Chr6	47974007	47975979	Zip777		Zip777		+	-	+/-	+	2	97,9%	0,0%	21,0%	44,4%
Chr13	54209856	54211139	Skn1				-	-	-	-	0	1,3%	0,0%	3,4%	1,2%
Chr17	87524084	87525819	Socs5				+	-	+/-	+	4	100,0%	3,9%	5,3%	45,2%
Chr7	86519538	86521097	Rbp1		Rbp1		-	+	-	+/-	0				
Chr7	6083480	6084890	Zip787		Zip787		+	-	+/-	+	3	84,7%	4,7%	3,6%	31,2%
Chr7	24992450	24993377	Zip111		Zip111		+	-	+/-	-	0	100,0%	0,0%	4,8%	14,5%
Chr15	76010966	76012080	Plec1				+	+	+/-	-	4	100,0%	23,3%	100,0%	33,5%
Chr11	102057005	102057998	Hdac5		Hdac5		+	+	+/-	+/-	1				
Chr13	60557950	60559042					-	+	-	+	0	100,0%	0,7%	25,0%	27,7%
Chr16	20530221	20531293	Dvl3				-	+	+/-	+	2	100,0%	0,0%	98,0%	19,6%
Chr4	150993001	150994022	Camta1				-	-	-	+/-	1	1,8%	3,2%	1,5%	14,4%
Chr4	53727006	53728024	Fcmd		Fcmd		-	-	-	-	1	4,2%	1,4%	6,8%	1,5%
Chr15	102047274	102048271	Ilgb7		Ilgb7		-	+	+/-	+	0	89,5%	0,0%	8,3%	45,1%
Chr5	106629403	106630408	nenese				-	+	-	+	0				
Chr10	122303419	122304463	Ppm1h				+	+	+/-	+	0				
Chr7	148034494	148035458	Odr3				+	+	+/-	+	0				
Chr7	147267611	147268611	Drd1p				-	-	-	-	0	5,8%	1,3%	1,6%	2,0%
Chr15	11250512	11251415	Adamts12				+	+	-	+/-	0				
Chr13	66815007	66816022	2410141K09Rlk				+	-	+/-	+	0				
Chr14	122056331	122057326	Dock9		Dock9		+	+	+/-	+/-	0				
Chr8	12262778	12263827					+	+	-	+	0				
Chr12	118489501	118490705	Pipm2				+	+	-	+	0	86,7%	6,3%	0,0%	21,7%
Chr19	45385459	45386455	sneefar	glrdoy, sneefar			-	-	-	-	0				
Chr6	125660888	125661895	Tmem16b	Tmem16b			-	+	+/-	+	0	82,1%		97,2%	37,1%
Chr11	115748842	115749776	Myo15b				+	+	-	+	0				
Chr10	74869979	74870980	Upb1	Upb1			-	+	-	+	0				
Chr13	96588183	96589194	Ilgap2				+	+	+/-	+	0	2,1%		5,1%	1,9%

B

chr6	47974479	47975613	Zip777		Zip777		+	-	+/-	+	3	97,9%	0,0%	21,0%	44,4%
chr8	125388035	125398665	Cdh15				+	-	+/-	+/-	3	100,0%	0,0%	6,7%	20,3%
chr7	6083787	6084472	Zip787				+	-	+/-	+	1	84,7%	4,7%	3,6%	31,2%
chr17	87524516	87525036	Socs5				+	-	+/-	+	4	100,0%	3,9%	5,3%	45,2%
chr15	76011065	76011621	Plec1				-	+	+/-	+	3	100,0%	23,3%	100,0%	33,5%
chr14	66767735	66768264	Chma2				-	-	-	+	1	1,9%	1,7%	39,2%	4,9%
chr5	139902103	139902615	311008217Rlk, Gper		Gper		+	+	+/-	+	1				
chr13	47106262	47106763	Tmem132e				+	-	+/-	+/-	4	94,2%	3,3%	15,4%	36,9%
chr11	82258076	82258615					?	+	-	+	3	78,6%	0,0%	33,9%	59,0%
chr10	22643161	22643660					+	+	-	+	1				
chr19	25152039	25152586	Dock8		Dock8		+	+	+/-	?	1	100,0%	0,0%	18,6%	25,0%
chr6	145135900	145156913	Casc1				+	-	+/-	+/-	0	97,1%	7,5%	7,8%	36,5%

C

chr2	131345997	131346497	Snox		Snox		+	-	+/-	+	0	100,0%	0,0%	42,2%	27,6%
chr9	108295183	108295684	Usp4				+	-	+/-	+	0	99,4%	1,9%	3,8%	20,3%
chr10	61590345	61591128	Neurog3				+	-	+/-	+/-	0	84,8%	2,2%	4,7%	25,1%
chr15	61868491	61868952	Pv11				+	-	+/-	+/-	1	100,0%	0,0%	3,3%	37,7%
chr7	13843983	13864718	633040a02Rlk				+	-	+/-	+		95,8%	0,4%	20,4%	33,4%
chr15	25631658	25649777	Myo10				+	-	+/-	+/-		100,0%	0,0%	29,0%	24,2%

Table1 : Maternal gDMR candidates identified in our screen and others screen

A. Maternal gDMR candidates from our screen with a sliding window of 500pb.

B. Maternal gDMR candidates from our screen with a sliding window of 1000pb that contain Zfp57 binding motif.

C. Maternal gDMR candidates from the screens made by Kobayashi et al., 2012 et Xie et al., 2012.

The maternal gDMR that we confirmed are in red and the ones that I have invalidated as not maintaining their parental differences of methylation after implantation are in grey. The *Socs5* gDMR is highlighted in yellow. The +, - et +/- signs represent 100%, 0% and 50% of methylation respectively in the bisulfite-seq tracks taken from the publications listed on top.

sgRNA Ups				
sequence	score	mismatches	UCSC gene	locus
GAGATGGGGTGCTCGGTGATCAG	0.3	4MMs [1:6:7:20]	NM_019521	chr8:+13468131
TGGATCCGGGGATCTGTGAATAG	0.2	4MMs [2:10:12:15]	NM_001039176	chr4:-118103305
AAGATCCAGTGCTCAGGGAATAG	0.2	4MMs [1:8:15:17]	NM_134471	chr4:-116845032
TAGTCCGGTGCCCGGCGAGGGG	0.1	4MMs [4:13:17:20]	NM_021310	chr13:+94269262
CAGACCCGGTGCTTGAGAAGGG	0.1	4MMs [1:5:14:17]	NM_029017	chr3:-32629020
CAGATCCGGTGCGCAGTGAGCAG	0.1	4MMs [1:13:15:20]	NR_027860	chr14:-73425614
CAGATCCGGTGCGCAGTGAGCGG	0.1	4MMs [1:13:15:20]	NM_026572	chr8:+119517198
TAGATCCGGTGCTAAGTCGAGAG	0.0	4MMs [14:15:18:19]	NM_001102607	chr9:+105591782

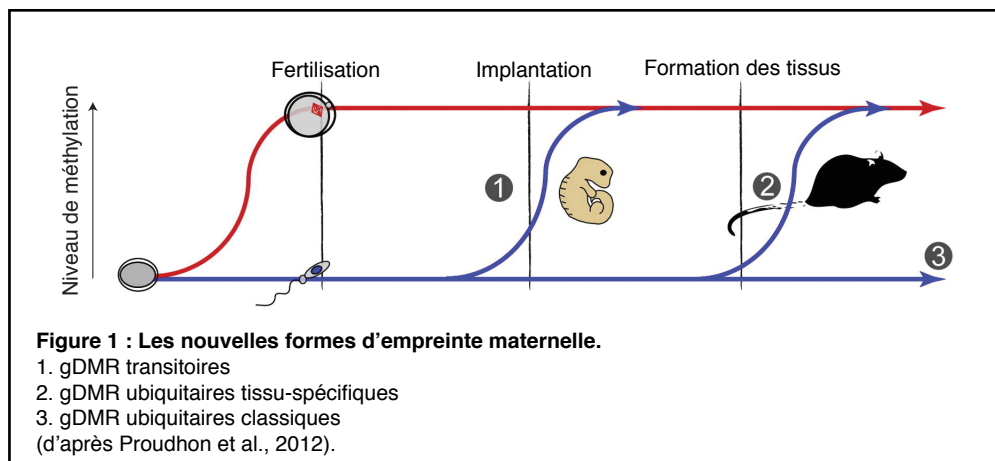
sgRNA Dwst				
sequence	score	mismatches	UCSC gene	locus
CAGTGATCCCATAGATGTTCCAG	0.8	4MMs [4:5:6:8]	NM_173048	chr11:+115445706
GAGCTGAACCATAGATGTTCCAG	0.6	4MMs [1:5:6:7]	NM_177755	chr15:+58146775
TAGGCATAGCATAGATGTTCCAG	0.5	4MMs [1:4:6:9]	NM_027324	chr13:+54202218
GGGCCCCACCATAGATGTCCTGG	0.4	4MMs [1:2:7:19]	NM_172990	chr4:-154345040
TAGGCCTACCCTAGATTTTCAGG	0.3	4MMs [1:4:11:17]	NM_198105	chrX:+147904365
CAGCCCTGCGAGAGATGTTTGAG	0.3	4MMs [8:10:12:20]	NM_001146022	chr14:-33911152
CAGCGGCACCATAGATGTGCAAG	0.2	4MMs [5:6:7:19]	NM_007566	chr17:-75076567
CAGGTGTACCATAGATGATCGGG	0.2	4MMs [4:5:6:18]	NM_177906	chr9:+28482714
AAGCCCCACCAAGAGATGCTCAGG	0.1	4MMs [1:7:12:18]	NM_001033228	chr13:+115891637
CAGCCATACAAGAGCTGTTCCAG	0.1	4MMs [6:10:12:15]	NM_001093778	chr12:+30468730
CAGCCCTGCCAAAGGTGTTGCGG	0.1	4MMs [8:12:15:20]	NM_001195271	chrX:+40946784
CAGCCCTAACATAGATTTAATGG	0.0	4MMs [9:17:19:20]	NM_025624	chr5:+148684339
CAGCCCTGCCCTAGCTGCTCCAG	0.0	4MMs [8:11:15:18]	NM_001085355	chr17:+5341618
CAGCCTCACCATAGATGCGCTGG	0.0	4MMs [6:7:18:19]	NM_178646	chr15:-75740794
CAGCCTTACCTTCGAGGTTCCAG	0.0	4MMs [6:11:13:16]	NM_183276	chr9:+110547123

Table2 : List of the possible off-targets for the *Socs5* gDMR CRISPR/Cas9 deletion in the mouse

Technique	Name of the primers	Primers
Bisulfite cloning/sequencing	Socs5-Out-Fwd Socs5-Out-Rev Socs5-Inner-Fwd Socs5-Inner-Rev	TGTTAGAGTTTAGTAGTGAGGT AACTCCAAACCCACCTCATAT ATGGTGTGGTTTGGGAGAAG AACCAACACTTTACTTATTAATTC
Northern blot	Socs5-5'UTR-Fwd Socs5-5'UTR-Rev Socs5-3'UTR-Fwd1 Socs5-3'UTR-Rev1 Socs5-3'UTR-Fwd2 Socs5-3'UTR-Rev2 Socs5-DMR-Fwd Socs5-DMR-Rev	CTCGCCCTGTCGCTGACT CGAGTAAGCTCCCACTACGC GGGGCATTTGCTATGAAGAA ATGCACACTGGGGCTTTTAC GTAAAAGCCCCAGTGTGCAT ACAGAACCGGGAACACAGTC CTAAGGAACGCGCTGAACTC AGGTGGATGAGACCAGCAAC
Bisulfite Pyrosequencing	Socs5-DMR-Fwd Socs5-DMR-Rev Socs5-DMR-Seq Mcts2-DMR-Fwd Mcts2-DMR-Rev Mcts2-DMR-Seq	GGATTATGTTGTTGGATTTGTTIAGG (Biot)CAATATCTTCCCATATATAATACCTACC GAGTTATATTTGAATTGTTAAGGAA TTGAAGAAGAATTAGTGGGGTAA (Biot)ACAATTAAACACACTTTCCTCTC ATGTAGGTGTTATTTTTTTGTAG
Pyrosequencing	Socs5-DMR-Fwd Socs5-DMR-Rev Socs5-DMR-BxC-Seq Socs5-DMR-Bx129/Cx129-Seq	TTGTCCAGGTCTGCACAGATTAGC (Biot)GGCCACACCATAAACGTCAATATC GCTTGAATGGGTGATG CTATACTTGAAGTCTAAGG
ChIP RT-qPCR	Socs5-5'UTR-Fwd Socs5-5'UTR-Rev Socs5-Intron1-Fwd Socs5-Intron1-Rev Socs5-Intron2-Fwd Socs5-Intron2-Rev Socs5-DMR-Fwd Socs5-DMR-Rev Socs5-3'UTR-Fwd Socs5-3'UTR-Rev KvDMR-Fwd KvDMR-Rev Gapdh-Fwd Gapdh-Rev IAP-Fwd IAP-Rev B-Actin-Fwd B-Actin-Rev Pax5-Fwd Pax5-Rev	TCACCTACCCGCCTCTGG CTCTCTGAAGGGCAGCCTAC TGGAATGGAATGACACCTG GTGGAACCTCTCTGCCAAGC GTGGGAGAGAGAACGACTGG CAGTCCTCCACACCATCTT TGTCCAGGTCTGCACAGATT TTCCCAGGAAAGACTGTTGC AGGGTTCGCTGGTTAGAACGA CTAGCAGCTGACTGCCTCCT AACTCCGAATAAGCAGCCTTC TGGAGTACGTGTTGTTATGTGG TCCATGACAACTTTGGCATTG CAGTCTTCTGGGTGGCAGTGA CTCCATGTGCTCTGCCTTCC CCCCGTCCCTTTTTTAGGAGA AAGTGACGTTGACATCCG GATCCACATCTGCTGGAAGG ATGGGAGTTTGTTTTCTGTGT AGTGATGTTTGGCCTAATCCTG
RT-qPCR	Socs5-5'UTR-Fwd Rev5'UTRSoc Socs5-DMR-Fwd Socs5-DMR-Rev Socs5-3'UTR-Fwd Socs5-3'UTR-Rev Rrm2-Fwd Rrm2-Rev Arpo-Fwd Arpo-Rev B-Actin-Fwd B-Actin-Rev	GCACCTCGGGCGCAGGCGGCA CGGCGGCACCGTGCGCAG TGTCCAGGTCTGCACAGATT TTCCCAGGAAAGACTGTTGC AGGGTTCGCTGGTTAGAACGA CTAGCAGCTGACTGCCTCCT CCGAGCTGGAAGTAAAGCG ATGGGAAAGACAACGAAGCG TCCAGAGGCACCATTGAAATT TCGCTGGCTCCACCTT AAGTGACGTTGACATCCG GATCCACATCTGCTGGAAGG
RT-PCR	Socs5-5'UTR-Fwd Socs5-AltExon1-Rev Socs5-DMR-Fwd Socs5-Exon2-Rev1 Socs5-Exon2-Rev2	GCGTAGTGGGAGCTTACTCG TTCCATCCCTCGACTTTCAG AACGCGCTGAACTCGCAACAGTCTT CTCGTTCTAACCAGCGAACC CTCAGTCCCAGCTGTAAGGC
CRISPR	Socs5-sgRNA-Ups Socs5-sgRNA-Dwst Socs5-geno-Fwd1 Socs5-geno-Rev1 Socs5-geno-Fwd2 Socs5-geno-Rev2 Socs5-geno-Fwd2 Socs5-geno-Rev2	TAGATCCGGTGCTCGGTGAA CAGCCCTACCATAGATGTTG CTCTTACGGTGGGAGAGCAG TCCAGGTCCAGCAAGAGACT CTAAGGAACGCGCTGAACTC CTGCCAGACTGAGATGTCCA GACTCAAGATACTCACCAGGT GTAATGAACTGACGATCTGATA

Table3 : Primers used in this study

DISCUSSION



1- Importance de la protection des allèles non-méthylés contre la méthylation.

Au cours de ma thèse, j'ai eu en charge la recherche de nouvelles gDMR maternelles chez la souris. Toutes les gDMR/ICR maternelles connues jusqu'alors suivaient une dynamique bien définie de méthylation de l'ADN. Par convention, la méthylation de ces gDMR est mise en place dans l'ovocyte spécifiquement. Puis, dans l'embryon précoce, l'allèle maternel méthylé doit être maintenu lors de la vague de déméthylation du génome post-fécondation, alors que l'allèle paternel non-méthylé de ces gDMR doit quant à lui être protégé de la vague de re-méthylation *de novo* du génome après l'implantation. Après cette période d'intenses changements, les profils de méthylation alléliques parentaux sont ensuite maintenus de façon ubiquitaire, à tous les stades du développement pré-nataux et post-nataux, dans tous les tissus et de manière homogène chez tous les individus d'une même espèce.

Lors de ma thèse, j'ai eu l'occasion de travailler sur des nouvelles formes d'empreinte maternelle, dites non-canoniques, mises en évidence grâce à notre crible MeDIP-seq : 1) les gDMR transitoires, qui regagnent de la méthylation sur l'allèle paternel juste après l'implantation (cas du locus *Gpr1/Zdbf2*), 2) les gDMR tissu-spécifiques, qui regagnent de la méthylation sur l'allèle paternel dans certains tissus (cas du locus *Cdh15*), et 3) les gDMR lignée-spécifiques, qui regagnent de la méthylation dans certains fonds génétiques de souris (cas du locus *Socs5*) (Figure1). Ceci montre que la protection de l'allèle paternel non-méthylé n'est sans doute pas un critère obligatoire de définition des gDMR soumises à empreinte, en tous cas dans le cas des gDMR maternelles. Finalement, l'acquisition d'une méthylation ovocytaire et la protection des allèles maternels méthylés après fécondation par la présence de motifs Zfp57 pourraient être les deux critères d'inclusion suffisants des gDMR/ICR maternelles. Ceci semble d'autant plus vrai que même dans le cas extrême des gDMR transitoires, qui n'existent que quelques jours au début de la vie, les différences brèves de méthylation parentale peuvent suffire à réguler l'expression allélique de certains gènes et à impacter sur la régulation du locus de manière indélébile (cas de *Gpr1/Zdbf2*) (Duffié et al., 2014).

Ces 3 nouvelles formes de gDMR maternelles sont soumises à une non-protection de leurs allèles paternels contre les vagues de méthylation que ce soit au moment de

l'implantation, au moment de la différenciation tissulaire, ou en fonction du fond génétique. Il est intéressant de noter que toutes ces gDMR labiles ont la particularité d'être en position intra-génique, non seulement dans l'ovocyte mais également dans les tissus embryonnaires et somatiques. La transcription traversante dans le corps des gènes pourrait donc avoir un rôle dans le regain de méthylation observé au niveau de ces gDMR non-canoniques (Smallwood et al., 2011). En effet, l'élongation de la transcription permet le dépôt de marques H3K36me3 et le recrutement des méthyltransférases *de novo* (Dhayalan et al., 2010 ; Hahn et al., 2011 ; Baubec et al., 2015). Cela expliquerait la perte d'empreinte aux gDMR intragéniques au moment de l'implantation, dans certains tissus ou chez certains individus, en relation avec le profil de transcription du gène hôte. Cependant, il existe des ICR maternelles canoniques en position intra-génique et celles-ci sont réfractaires à la méthylation: c'est le cas par exemple de la gDMR du locus *Mcts2/H13* qui est ubiquitaire ou encore de la gDMR du locus d'*Igf2r* (Wood et al., 2007 ; Zwart et al., 2001). Des contextes nucléotidiques et/ou chromatinien différents pourraient différencier les gDMR intra-géniques ubiquitaires des non-ubiquitaires. On peut imaginer que ces contextes variables permettraient d'attirer les Dnmt et/ou d'empêcher la liaison de facteurs de transcription protecteurs de la méthylation comme CTCF, Sp1 ou STAT1 (Stadler et al., 2011 ; Straussman et al., 2009) dans le cas des gDMR non ubiquitaires, ou au contraire d'exclure les Dnmt et de lier les facteurs de transcription dans le cas des gDMR ubiquitaires. De plus, le contexte tissulaire ou individuel pourrait jouer sur la liaison de ces facteurs protecteurs et expliquerait les spécificités de tissus et de lignées.

2- La gDMR de *Socs5*: le dernier locus soumis à empreinte persistante ?

Parmi les gDMR candidates restantes issues de notre criblage, je me suis focalisée particulièrement sur celles qui possédaient des sites de liaison au facteur Zfp57. En effet, du fait de la présence systématique d'au moins deux motifs Zfp57 dans toutes les ICR connues, la liaison du facteur Zfp57 au niveau des allèles méthylés des gDMR semble être un pré-requis obligatoire pour protéger ces allèles de la déméthylation globale du génome et ainsi maintenir un état différentiel jusqu'à l'implantation. La persistance des gDMR au delà de l'implantation implique de plus la capacité de maintenir l'allèle non-méthylé contre la vague de re-méthylation *de novo* du génome embryonnaire. Parmi les gDMR candidates, seules les gDMR de *Socs5* et de *Neurog3* se sont avérées maintenir un statut de méthylation différentielle après l'implantation, mesuré à 9,5dpc (Résultats Figures 1&3B). Cependant, la gDMR de *Socs5* présentait des différences alléliques parentales beaucoup plus prononcées que celles de la gDMR de *Neurog3*. Il semblerait donc que la gDMR de *Socs5* soit une des toutes dernières- si ce n'est la dernière- gDMR persistante maternelle à caractériser chez la souris. Ce caractère ultime m'a poussé à déployer de nombreux efforts pour comprendre sa cinétique développementale, sa régulation et son rôle fonctionnel.

La découverte de la gDMR de *Socs5* suscite quatre grandes questions : 1) pourquoi le locus *Socs5* est-il soumis à empreinte, 2) comment expliquer le caractère lignée-spécifique de la gDMR de *Socs5*, 3) quelle est la fonction du gène *Socs5*, et 4) comment fonctionne la gDMR de *Socs5* ?

2.1- Pourquoi le locus *Socs5* est-il soumis à empreinte ?

Sachant que l'empreinte parentale est finalement définie par deux phases : 1) l'acquisition de la méthylation sexe-spécifique dans les gamètes parentaux et 2) la capacité à maintenir l'allèle parental méthylé après fécondation, il est assez facile de décrypter pourquoi la gDMR de *Socs5* est soumise à empreinte maternelle chez la souris.

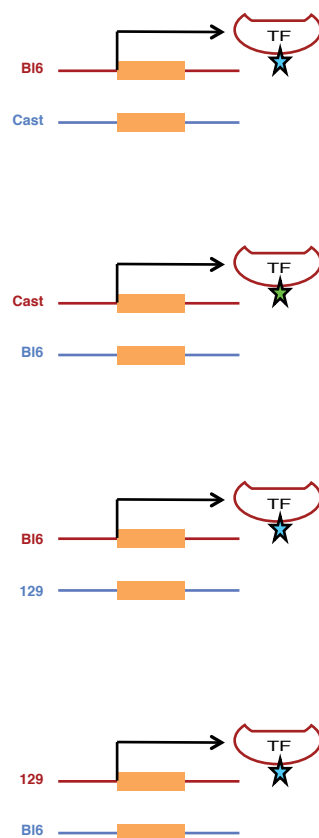
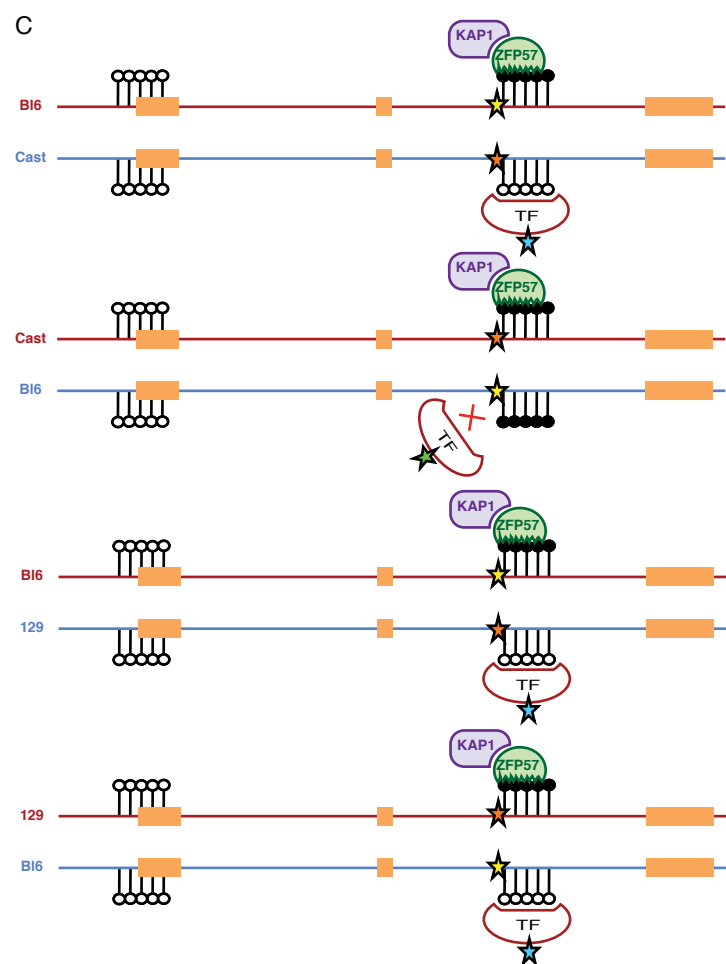
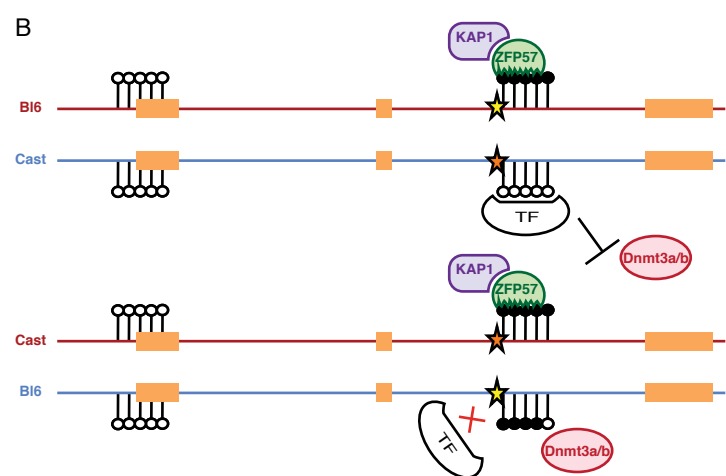
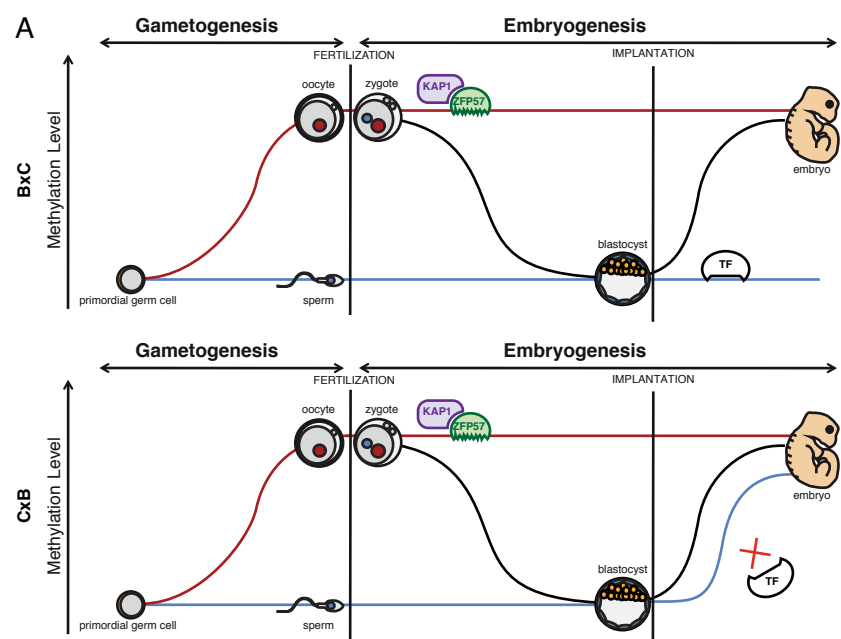
La phase d'acquisition sexe-spécifique peut s'expliquer par la position intra-génique de cette gDMR, dans l'intron 2 du gène *Socs5*. En effet, ce type de séquence intra-génique est méthylée *de novo* dans l'ovocyte mais rarement dans le spermatozoïde.

Il y a donc ici une opportunité d'acquérir de la méthylation différentielle dans les deux lignées germinales parentales. Ce modèle, actuellement reconnu, d'acquisition de la méthylation ovocytaire implique un mécanisme dépendant de la transcription (Smallwood et al., 2011 ; Kobayashi et al., 2012). L'élongation de la transcription par l'ARN Pol II est suivie du dépôt de la marque H3K36me3 qui elle même attire la machinerie de méthylation *de novo* (Dhayalan et al., 2010 ; Hahn et al., 2011 ; Baubec et al., 2015). Incidemment, ceci implique que le gène *Socs5* doit être exprimé dans l'ovocyte à partir du promoteur canonique en 5' pour stimuler la méthylation de la gDMR située dans le corps du gène. En accord avec ceci, les bases de données d'expression montrent un profil d'expression tissulaire plutôt large de *Socs5*, dont l'ovaire.

La capacité à maintenir l'allèle maternel méthylé après fécondation est liée à la présence de motifs Zfp57 et la gDMR de *Socs5* en contient quatre. Par une étude phylogénétique simple de la conservation nucléotidique de la gDMR de *Socs5* chez les mammifères, j'ai observé que ces 4 motifs sont contenus dans une séquence de 181nt riche en CG, qui est une acquisition spécifique du génome de la souris. Cette séquence n'est même pas présente chez le rat, alors que la séquence génomique autour de cette insertion est conservée (Résultats Figure9). L'origine de cette séquence est inconnue, je n'ai pas détecté de séquence homologue ailleurs dans le génome. C'est très certainement cette innovation murine qui confère la capacité de garder la méthylation héritée de l'ovocyte au niveau de la gDMR de *Socs5* chez la souris. Nous avons donc ici l'opportunité de démontrer le rôle essentiel de l'acquisition de motifs nucléotidiques TGCCGC dans l'évolution de l'empreinte. Pour cela, je vais prochainement analyser chez le rat le statut de méthylation de la séquence conservée autour de l'insertion, dans le poumon en particulier, en raison de son empreinte assez stable au locus *Socs5* chez la souris. On s'attend à une méthylation allélique uniforme dans ce cas, sans distinction parentale.

2.2- Comment expliquer le caractère lignée-spécifique de la gDMR de *Socs5* ?

J'ai constaté que la gDMR de *Socs5* pouvait être définie comme une gDMR persistante tissu-spécifique ou comme une gDMR transitoire, suivant les croisements de souris utilisés et donc le fond génétique. Dans les croisements réciproques entre les lignées Bl6 et Cast, la gDMR est transitoire dans le sens CxB, car elle regagne de la



méthylation sur l'allèle paternel au moment de l'implantation, mais elle est persistante et tissu-spécifique dans le sens BxC, avec au moins un maintien dans le poumon. Dans les croisements réciproques entre les lignées Bl6 et 129, la gDMR est persistante et tissu-spécifique, avec au moins un maintien dans le poumon à la fois dans le sens Bx129 et le sens 129xB. Dans les croisements réciproques entre les lignées 129 et Cast, l'empreinte est plus labile, mais l'allèle maternel reste globalement plus méthylé que l'allèle paternel dans le poumon. La gDMR de *Socs5* possède donc des spécificités de stades, de tissus et de lignées.

Ainsi, la comparaison des différents croisements montre que seul l'allèle Bl6 non méthylé en position paternelle peut être susceptible à la re-méthylation *de novo* au moment de l'implantation. Une hypothèse pour expliquer ce gain de méthylation spécifique serait que la gDMR en fond Bl6 porterait des caractéristiques de séquence *cis* qui ne lui permettent pas d'être protégée (Figure2A). Dans ce sens, la fixation de certains facteurs de transcription sur leur motif a été montrée comme capable d'empêcher localement la méthylation *de novo*, probablement en limitant l'accès à l'ADN des Dnmt (Stadler et al., 2011). J'ai donc analysé de plus près les séquences génomiques de la DMR dans les lignées C57Bl6/J, CAST/Ei et 129SvPas. Après alignement, j'ai pu identifier un SNP spécifique de l'allèle Bl6 en début de la gDMR : présence d'un T sur l'allèle Bl6 et d'un C sur les allèles Cast et 129 (Résultats Figure6). En accord avec l'hypothèse que ce SNP empêcherait la liaison d'un facteur de transcription sur l'allèle Bl6 exclusivement (Figure2A et 2B), j'ai recherché la présence de motifs de liaison pour des facteurs de transcription autour de ce SNP dans différentes bases de données (UCSC ; Genious ; Jaspard) : une étude plus poussée semble nécessaire, car je n'ai pas trouvé de motifs bien définis.

Un autre point important à considérer est que l'allèle Bl6 paternel n'est pas susceptible dans tous les contextes à la méthylation *de novo* : bien qu'il se re-méthyle en CxB, il reste globalement hypométhylé en 129xB et sans doute en fond pur Bl6 (BxB), où l'on observe 50% de méthylation globale. En plus d'une composante *cis* spécifique au fond Bl6, il existerait également une composante *trans* : le génome Cast, qui a divergé du génome Bl6 depuis environ 1 million d'années, apporterait un facteur de protection qui ne peut se lier à l'allèle Bl6. En revanche, le même facteur de protection apporté par le génome 129, plus proche du génome Bl6, n'a pas de problème à se lier à l'allèle Bl6. Cependant, dans le croisement CxB, si la version Bl6 du facteur *trans* était présente, elle

devrait pouvoir protéger l'allèle Bl6 en *cis*. Ceci suppose que le facteur Bl6 bien que présent génomiquement n'est pas exprimé. Une hypothèse possible serait donc que ce facteur *trans* de protection de l'empreinte soit lui même soumis à empreinte : dans le sens CxB, l'allèle paternel Bl6 de ce facteur *trans* serait réprimé et seul l'allèle Cast maternel serait produit, avec son incapacité à protéger l'allèle Bl6 de la gDMR en raison de divergence de séquence (Figure2C).

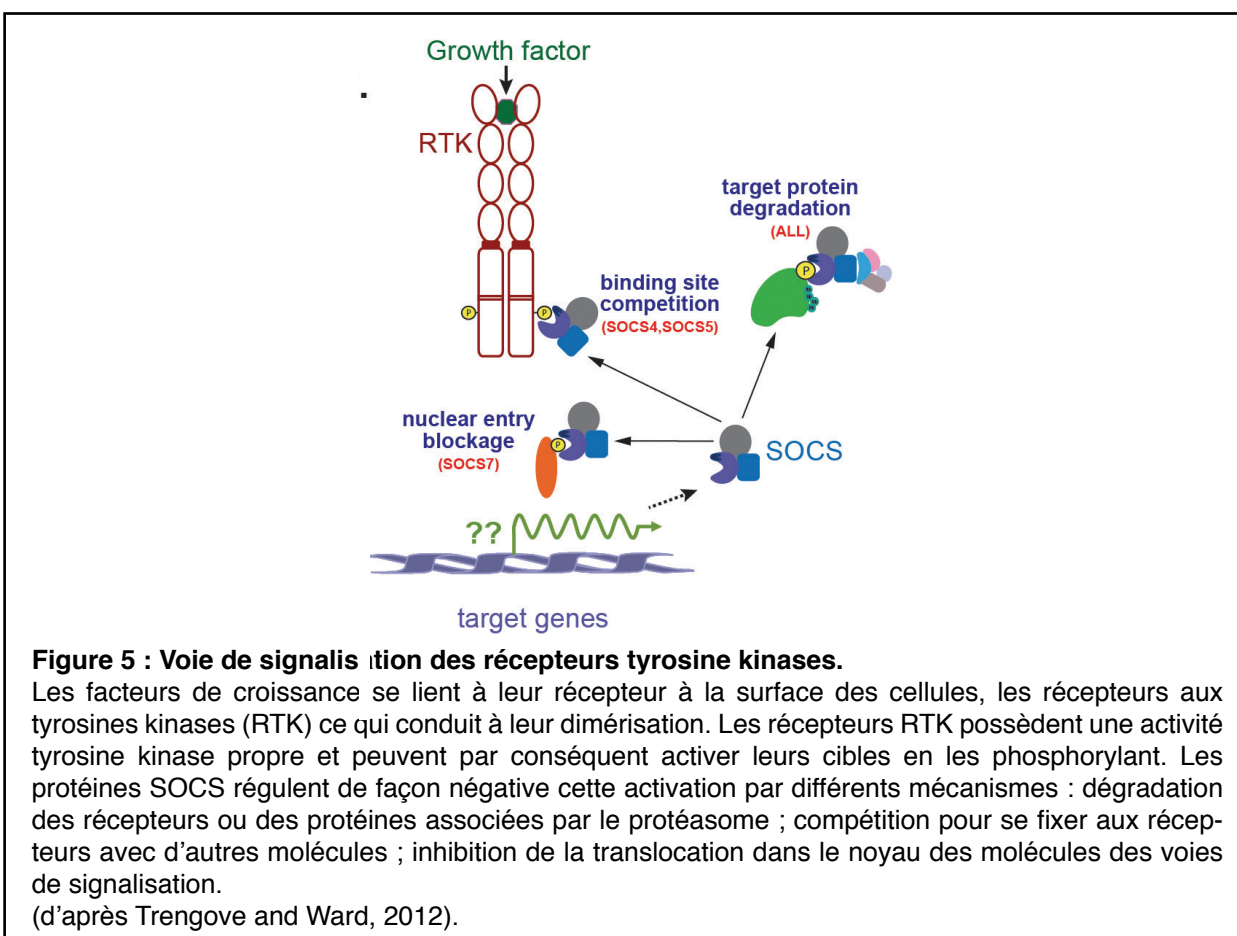
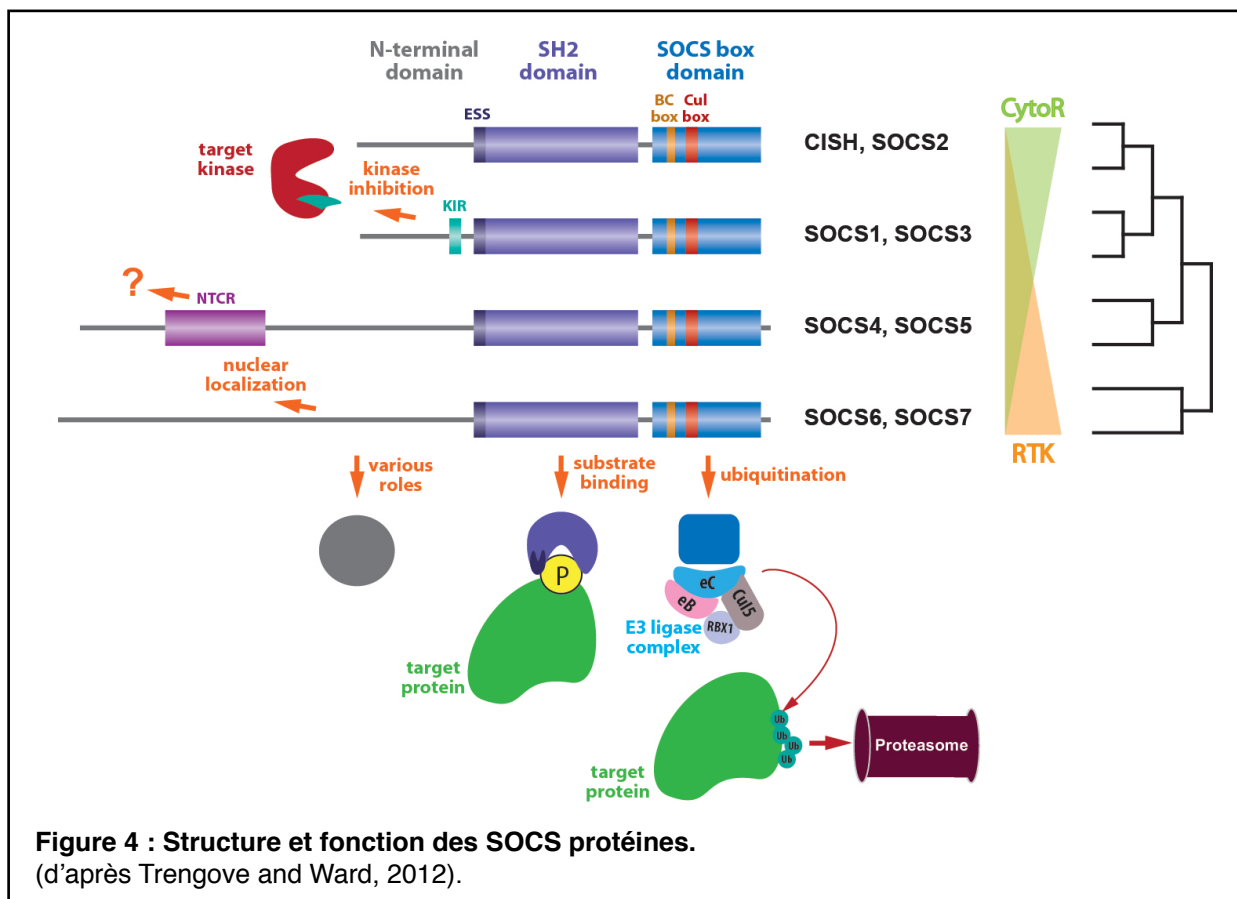
Ce genre de modèle de « l'empreinteur empreinté » a été discuté d'un point de vue théorique par Wolf et al., 2014. L'identification d'un tel gène modificateur impliqué dans la variabilité épigénétique de la gDMR de *Socs5* serait *a priori* possible par des expériences de cartographie génétique dans des croisements back-cross à partir de CxB. Ce type d'approche reste cependant très compliqué.

2.3- Quelle est la fonction du gène *Socs5* ?

La gDMR de *Socs5* n'est présente que chez la souris, mais le gène *Socs5* en lui-même a une fonction conservée chez les mammifères. Comprendre la fonction biologique de SOCS5 pourrait indiquer la pression de sélection exercée sur ce locus pour acquérir une gDMR soumise à empreinte et potentiellement, un mode de régulation de dose d'expression dépendant de la méthylation de l'ADN.

Le gène *Socs5* code pour une protéine inhibitrice de la voie de signalisation des récepteurs aux facteurs de croissance. Il existe 8 membres dans la famille des protéines SOCS : les SOCS1 à 7 et la Cytokine-inducible SH2-containing protein ou CISH. Du fait de leur similarité, un fort niveau de redondance fonctionnelle est attendu pour ces protéines. Les protéines CISH et SOCS1-3 sont le plus souvent associées à la régulation des récepteurs aux cytokines via la voie de signalisation JAK-STAT ; les protéines SOCS4-7 réguleraient plus spécifiquement la voie de signalisation des récepteurs aux facteurs de croissance (Kario et al., 2005 ; Krebs et al., 2002 ; Banks et al., 2005). Toutes les protéines SOCS fonctionnent par paire, et une paire de protéines a les mêmes structures et fonctions.

Chaque protéine est composée de 3 domaines distincts : un domaine N-terminal peu conservé entre elles, un domaine central Src-homology 2 (SH2) conservé et un domaine C-terminal très conservé appelé la boîte SOCS (Figure3). Le domaine SH2 a la



propriété d'interagir avec des résidus tyrosine phosphorylés présents à la surface de leurs protéines cibles, notamment les récepteurs de la surface cellulaire (O'Sullivan et al., 2007). La boîte SOCS est quant à elle subdivisée en deux sous-domaines : une boîte BC qui recrute les Elongin B et C et une boîte Cul qui se lie à la Cullin 5. Le complexe qui en résulte est capable de se lier à RBX1 ou RBX2 et d'autres ubiquitine ligase E3 afin de cibler l'ubiquitination des protéines et leur dégradation par le protéasome (Zhang et al., 1999 ; Kamura et al., 2004). De plus, SOCS5 tout comme SOCS4 possèdent une région très conservée entre elles dans leur domaine N-terminal appelée NTCR (N-terminal conserved region) mais son rôle est inconnu (Feng et al., 2012).

Les protéines SOCS régulent négativement les voies de signalisation contrôlées par les cytokines ou les facteurs de croissance. La régulation de ces voies est importante pour le contrôle de la prolifération cellulaire, du cycle cellulaire, de la différenciation, ou de la migration. Dans le cas de la signalisation par des facteurs de croissance, dans laquelle serait impliquée SOCS5, le signal est transmis par les récepteurs tyrosine kinase (RTK) mais on ne sait pas quels sont les facteurs activés par cette voie qui stimulent la production des protéines SOCS4-7 (Figure4). Toutes les protéines SOCS ciblent la dégradation d'autres protéines dans la cellule ; les protéines SOCS4 et SOCS5 entrent en plus en compétition avec certaines d'entre elles en se liant aux récepteurs RTK (Martens et al., 2005).

La fonction de chaque protéine SOCS a été étudiée par différentes approches, révélant leur rôle dans le développement et leur implication dans diverses maladies. La protéine SOCS5 est sans doute celle qui a été le moins bien caractérisée, mais on sait qu'elle est exprimée dans de nombreux tissus adultes et particulièrement dans les cellules lymphocytes primaires B et T (Brender et al., 2004). En rapport avec cette expression, SOCS5 a été impliquée dans la différenciation cellulaire des cellules T helper et notamment dans la balance entre le nombre de cellules Th1 et Th2 produites à partir de ces cellules T. Des souris transgéniques sur-exprimant le gène *Socs5* présentent une importante réduction du nombre de cellules Th2 : ceci passerait par la capacité de SOCS5 à se lier au récepteur IL-4R et à inhiber l'activation de STAT6 qui normalement stimule la différenciation des cellules T vers les cellules Th2 (Seki et al., 2002). Cependant, les souris dont le gène *Socs5* a été délété (*Socs5*^{-/-}) sont parfaitement normales, fertiles et aucune anomalie n'a été détectée quant au nombre de cellules Th1 et Th2 (Brender et al., 2004). La protéine SOCS5 serait aussi capable de réguler de façon

néglige la voie de signalisation EGF (Epidermal Growth Factor) *in vitro*, en stimulant la dégradation par le protéasome des protéines cibles de cette voie qui est notamment surexprimée dans les cancers du poumon (Nicholson et al., 2005). Elle pourrait ainsi avoir un rôle de suppresseur de tumeur : son expression est d'ailleurs réduite dans les cancers de la thyroïde (Yoon et al., 2012). Enfin, la protéine SOCS5 est associée indirectement dans une grande variété de maladies comme les conjonctivites allergiques (Ozaki et al., 2005), l'asthme (Ohshima et al., 2007) et les uvéites (Egwuagu et al., 2005).

En conclusion, bien que de nombreuses corrélations indirectes aient été établies quant à la fonction de SOCS5, son rôle précis et spécifique n'est finalement pas très bien défini. Une des causes possibles de ces incertitudes vient très certainement de la taille de la famille des SOCS (8 membres) et de la redondance possible d'activité entre les différentes protéines SOCS, qui pourraient masquer le rôle de SOCS5 dans des tests classiques d'invalidation fonctionnelle.

2.4- Comment fonctionne la gDMR de Socs5 ?

La gDMR de *Socs5* est très singulière : non seulement la séquence de cette gDMR est une acquisition évolutive récente chez la souris, mais son statut de méthylation allélique est plus ou moins stable au moment de la vague de méthylation *de novo* de l'embryon en fonction du fond génétique de la souris. D'un point de vue fonctionnel, les études chromatiniques et d'expression que j'ai réalisées ne permettent pas avec certitude d'anticiper la fonction régulatrice de la gDMR de *Socs5*. J'ai démontré que l'empreinte maternelle persistait dans les poumons au cours du développement et qu'il existait différents transcrits alternatifs, incorporant la DMR, et exprimés plus fortement que le transcrit canonique initié en 5' dans ce tissu. La DMR de *Socs5* est en position intra-génique : au moins trois effets sur la transcription sont possibles pour ce type de gDMR/ICR (Résultats Figure2).

Le premier mécanisme serait que la DMR agisse comme un promoteur alternatif intra-génique, comme nous l'avons précédemment décrit pour *Cdh15* (Proudhon et al., 2012). Malheureusement, même si j'ai pu détecter la présence de transcrits alternatifs incorporant la gDMR par RT-PCR dans le poumon à 3dpp, j'ai été dans l'incapacité de déterminer l'origine de ces transcrits par RACE-PCR, leurs niveaux d'expression étant

trop faible pour cette technique. On peut voir effectivement que le niveau d'expression est 100 fois plus faible que des normaliseurs tels que *Rrm2*, *Arpo* et *β -Actine*. J'ai tout de même regardé quel était l'allélisme d'expression de ces différents transcrits dans le poumon, mais j'ai observé une expression bi-allélique, même dans les fonds Bx129 et 129xB. Ces transcrits ne semblent pas sensibles aux marques de méthylation opposées sur les deux allèles. Par contre, j'ai pu observer des marques d'histones opposées sur les deux allèles en relation avec la méthylation différentielle de l'ADN : H3K4me2 sur l'allèle paternel et H3K9me3 sur l'allèle maternel. Cependant, je n'ai pas observé d'enrichissement en H3K4me3. Il est donc peu probable que la DMR de *Socs5* soit un promoteur alternatif, du moins dans le poumon.

Un second mécanisme possible pour une gDMR intra-génique est que celle-ci contiendrait un signal d'arrêt de la transcription. L'absence de méthylation sur l'allèle paternel permettrait la liaison de facteurs d'arrêt de la transcription à partir de ce même allèle, comme il a été reporté au locus *Mcts2* (Wood et al., 2008). Encore une fois ici, les expériences de RACE n'ayant pas fonctionné, je suis dans l'incapacité de déterminer si les transcrits identifiés continuent en amont de la DMR.

Le dernier mécanisme possible pour une gDMR intra-génique est qu'elle pourrait agir comme un « enhancer » où l'absence de méthylation de l'allèle paternel permettrait la liaison de certains facteurs afin de réguler à distance la transcription des gènes avoisinants. Il a été montré que plusieurs régions différentiellement méthylées dans certains tissus sont en fait des enhancers tissu-spécifiques (Hon et al., 2013). Les expériences de ChIP réalisées au niveau de la DMR révèlent un certain enrichissement d'une marque d'histones classiquement associée aux enhancers, H3K4me1. De plus, je n'ai pu observer aucun mécanisme de contrôle post-transcriptionnel au niveau des transcrits alternatifs, pas de polyAdénylation et pas d'épissage strict, et il y aurait peut être une transcription bidirectionnelle au niveau de la DMR, ce qui est une autre caractéristique des enhancers (Andersson et al., 2014 ; Pnueli et al., 2015 ; Kim et al., 2015). Par contre, je n'ai pas observé d'enrichissement pour la marque H3K27ac, classiquement associée aux enhancers. Ceci est peut être lié à un problème technique, bien que mes contrôles soient correctement enrichis ou déplétés et que la 5'UTR du gène arbore bien cette marque. Cependant, toutes les expériences chromatiniques et transcriptionnelles ont été réalisées à 3dpp ; il reste possible que la gDMR joue un rôle d'enhancer à des stades plus précoces du développement, en vie embryonnaire. Il est

connu que de nombreux enhancers sont non-fonctionnels dans les tissus post-nataux, mais actifs dans les tissus embryonnaires (Hon et al., 2013).

Face à ces incertitudes, la délétion de la gDMR chez la souris par le système CRISPR/Cas9 va me permettre d'en savoir plus quant à son rôle, biologique et moléculaire. Compte-tenu de l'absence de phénotype des souris *Socs5*^{-/-} (Brender et al., 2004) et de la redondance attendue de fonction des différents membres de la famille des SOCS, nous ne nous attendons *a priori* pas à un phénotype drastique dans nos modèles mutants de la gDMR de *Socs5*. Par contre, cette expérience sera essentielle pour en décrypter le rôle sur l'activité transcriptionnelle du locus. Cela me permettra notamment de voir si la gDMR de *Socs5* contrôle la transcription des nouveaux transcrits alternatifs dans le poumon- la gDMR serait alors un promoteur alternatif tissu-spécifique- ou bien si la gDMR joue un rôle sur l'expression du transcrit canonique *Socs5*- la gDMR serait alors bien un enhancer tissu- ou stade-spécifique. Les animaux F1 hétérozygotes mâles et femelles que j'ai déjà obtenus sont actuellement en cours de croisement, pour suivre les effets d'une transmission paternelle, maternelle et homozygote de la délétion. Si la gDMR agit bien comme un enhancer, la possibilité la plus simple serait qu'elle agisse à courte distance sur l'activité du promoteur canonique de *Socs5*. Je vais donc regarder, dans différents tissus et à différents stades de développement, si la transcription du gène *Socs5* ainsi que des transcrits alternatifs est altérée dans mes mutants.

3- Combien de gènes soumis à empreinte reste-il à identifier ?

Lorsque j'ai commencé ma thèse en 2011, plus d'une centaine de gènes soumis à empreinte étaient connus chez la souris et conservés pour la majorité chez l'humain. Chez la souris, on connaissait 20 gDMR/ICR contrôlant ces GSE. La liste de gDMR candidates que nous avons établie à la suite de notre criblage a mis en évidence un nombre très restreint de régions supplémentaires qui seraient potentiellement soumises à empreinte : 28 exactement, dont 5 ont été confirmées (*Cdh15*, *AK008011*, *Zfp777*, *Zfp787* et *Socs5*). De façon très intéressante, ce crible nous a permis de mettre en évidence de nouvelles formes d'empreinte qui diffèrent de la définition classique d'un gène soumis à empreinte.

Parmi les 23 régions restantes, il semblerait que la majorité d'entre elles soit des gDMR transitoires qui regagneraient de la méthylation sur les allèles paternels après l'implantation ou dans le pire des cas, des faux positifs du criblage. Seule la gDMR de *Socs5* a montré un maintien de son statut différenciellement méthylé après la naissance, ce qui nous fait penser que c'est sûrement la dernière gDMR soumise à empreinte maternelle de façon persistante, avec des spécificités de tissus et de lignées. De manière intéressante et rassurante, la gDMR du gène *Socs5* a été retrouvée comme candidate potentielle dans de nombreux cribles d'identification de nouveaux GSE ou ICR (Kobayashi et al., 2012 ; Xie et al., 2012 ; Hon et al., 2013). Elle n'avait cependant jamais été analysée plus en avant, la complexité de sa régulation, telle que nous l'avons mise en évidence, ayant sûrement découragé les auteurs à poursuivre sa caractérisation. Ces autres cribles ont permis d'identifier d'autres régions potentiellement soumises à empreinte : cependant, la majorité d'entre elles a ensuite été invalidée par des expériences complémentaires, renforçant l'idée qu'il ne resterait que très peu de nouvelles gDMR à découvrir. Il reste quand même une gDMR candidate au niveau du gène *Neurog3* qui pourrait être une gDMR persistante au delà de l'implantation.

Il semble donc aujourd'hui que la recherche de régions soumises à empreinte qui persistent en vie adulte soit arrivée à saturation. Il s'agit maintenant de comprendre leur rôle et leur mode de régulation au cours du développement. En revanche il pourrait rester un bon nombre de gDMR transitoires à découvrir, leur nombre estimé étant de quelques centaines, sur la base d'un maintien d'environ 50% de méthylation au stade blastocyste (Smallwood et al., 2011 ; Smith et al., 2012 ; Kobayashi et al., 2012). D'une

manière générale, les gènes soumis à empreinte étant, de part leur haploïdie fonctionnelle, susceptibles d'être impliqués dans des maladies et des cancers, une meilleure compréhension de leur régulation nous permettra de mieux comprendre ces situations pathologiques chez l'humain.

4- Conclusion

La gDMR de *Socs5* représente la première région de contrôle parentale qui ait été aussi bien caractérisée en matière de variation inter-individuelle. De par la sensibilité de son statut empreinté en fonction du fond génétique, le cas de la gDMR de *Socs5* ouvre la perspective de l'existence de polymorphismes d'empreinte entre individus, y compris chez l'homme. Ce type de régulation pourrait sous-tendre des susceptibilités différentes aux maladies : une combinaison d'empreinte génomique parentale (effet parental d'expression de certains gènes) et de polymorphisme génétique pourrait avoir un effet délétère chez certains individus seulement. Chez l'homme, certaines régions héritées seulement de la mère ou du père pourraient alors avoir un effet sur la transcription délétère et causer des maladies chez certains individus. Ces régions, différentes en terme de séquences, arboreraient une méthylation différente en fonction du polymorphisme présent dans la séquence et cette méthylation jouerait un rôle sur l'expression des gènes adjacents (Kerkel et al., 2008). Aujourd'hui, un exemple connu de ce type de régulation est la méthylation variable apposée aux QTL (methylation Quantitative Trait Locus, meQTL) qui confèrerait à ces régions une susceptibilité à la maladie (Plongthongkum et al., 2014 ; Almli et al., 2015). On peut étendre ce type de régulation à un concept de meQTL soumis à empreinte comme il a été montré au locus *GNAS/NESPAS* où un polymorphisme présent dans l'ICR du locus joue sur la méthylation de celle-ci (Renteria et al., 2013). Identifier les régions portant des polymorphismes génétiques corrélées à des différences épigénétiques peut être un moyen de comprendre l'étiologie de certaines maladies.

Figure2 : Dynamic model of the methylation profil of the *Socs5* DMR.

A. Dynamic model of the methylation profile of the *Socs5* DMR during development in the reciprocal BxC and CxB crosses. The maternal allele is in red and the paternal one in blue, the black line represents the global methylation of the genome. The KAP1/Zfp57 complex protects the maternal allele against the wave of de-methylation of the genome after fertilization. Transcription factors protect the paternal allele from the wave of *de novo* methylation of the genome after implantation except in the case where the paternal allele is a Bl6 allele.

B & C. Hypothesis modele for the gain of the methylation on the paternal Bl6 allele after implantation. Maternal allele is in red and paternal allele in blue. The CpG island correspond to the white circles (unmethylated) and black circles (methylated). The stars represent the SNPs : on the *Socs5* DMR sequence, yellow = T and orange = C, and on the *trans* transcription factor, blue for the Bl6 and 129 factors and green for the Cast factor.

REFERENCES

BIBLIOGRAPHIQUES

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

Abramowitz, L.K., and Bartolomei, M.S. (2012). Genomic imprinting: recognition and marking of imprinted loci. *Current opinion in genetics & development* 22, 72-78.

Allis, C.D., Berger, S.L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., Shiekhatar, R., *et al.* (2007). New nomenclature for chromatin-modifying enzymes. *Cell* 131, 633-636.

Almli, L.M., Stevens, J.S., Smith, A.K., Kilaru, V., Meng, Q., Flory, J., Abu-Amara, D., Hammamieh, R., Yang, R., Mercer, K.B., *et al.* (2015). A genome-wide identified risk variant for PTSD is a methylation quantitative trait locus and confers decreased cortical activation to fearful faces. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., *et al.* (2014). An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455-461.

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 genetics* 8, e1002750.

Aravin, A.A., and Bourc'his, D. (2008). Small RNA guides for de novo DNA methylation in mammalian germ cells. *Genes & development* 22, 970-975.

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. *Molecular cell* 31, 785-799.

Arnaud, P. (2010). Genomic imprinting in germ cells: imprints are under control. *Reproduction* 140, 411-423.

Arney, K.L., Bao, S., Bannister, A.J., Kouzarides, T., and Surani, M.A. (2002). Histone methylation defines epigenetic asymmetry in the mouse zygote. *The International journal of developmental biology* 46, 317-320.

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 biology* 15, 545.

Babak, T., Deveale, B., Armour, C., Raymond, C., Cleary, M.A., van der Kooy, D., Johnson, J.M., and Lim, L.P. (2008). Global survey of genomic imprinting by transcriptome sequencing. *Current biology : CB* 18, 1735-1741.

Babak, T., DeVeale, B., Tsang, E.K., Zhou, Y., Li, X., Smith, K.S., Kukurba, K.R., Zhang, R., Li, J.B., van der Kooy, D., *et al.* (2015). Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. *Nature genetics* 47, 544-549.

Ball, M.P., Li, J.B., Gao, Y., Lee, J.H., LeProust, E.M., Park, I.H., Xie, B., Daley, G.Q., and Church, G.M. (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nature biotechnology* 27, 361-368.

Banks, A.S., Li, J., McKeag, L., Hribal, M.L., Kashiwada, M., Accili, D., and Rothman, P.B. (2005). Deletion of SOCS7 leads to enhanced insulin action and enlarged islets of Langerhans. *The Journal of clinical investigation* 115, 2462-2471.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 410, 120-124.

Barlow, D.P., and Bartolomei, M.S. (2014). Genomic imprinting in mammals. *Cold Spring Harbor perspectives in biology* 6.

Barlow, D.P., Stoger, R., Herrmann, B.G., Saito, K., and Schweifer, N. (1991). The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature* 349, 84-87.

Bartolomei, M.S., Webber, A.L., Brunkow, M.E., and Tilghman, S.M. (1993). Epigenetic mechanisms underlying the imprinting of the mouse H19 gene. *Genes & development* 7, 1663-1673.

Bartolomei, M.S., Zemel, S., and Tilghman, S.M. (1991). Parental imprinting of the mouse H19 gene. *Nature* 351, 153-155.

Barton, S.C., Surani, M.A., and Norris, M.L. (1984). Role of paternal and maternal genomes in mouse development. *Nature* 311, 374-376.

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schubeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* 520, 243-247.

Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* 405, 482-485.

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. *Journal of molecular biology* 203, 971-983.

Bestor, T.H. (2000). The DNA methyltransferases of mammals. *Human molecular genetics* 9, 2395-2402.

Bestor, T.H., Gundersen, G., Kolsto, A.B., and Prydz, H. (1992). CpG islands in mammalian gene promoters are inherently resistant to de novo methylation. *Genetic analysis, techniques and applications* 9, 48-53.

Bestor, T.H., and Ingram, V.M. (1983). Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proceedings of the National Academy of Sciences of the United States of America* 80, 5559-5563.

Bird, A. (1997). Does DNA methylation control transposition of selfish elements in the germline? *Trends in genetics : TIG* 13, 469-472.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes & development* 16, 6-21.

Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. *Nature* 321, 209-213.

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

Bostick, M., Kim, J.K., Esteve, 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. *Nature genetics* 47, 479-485.

Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 431, 96-99.

Bourc'his, D., and Proudhon, C. (2008). Sexual dimorphism in parental imprint ontogeny and contribution to embryonic development. *Molecular and cellular endocrinology* 282, 87-94.

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.

Braidotti, G., Baubec, T., Pauler, F., Seidl, C., Smrzka, O., Stricker, S., Yotova, I., and Barlow, D.P. (2004). The Air noncoding RNA: an imprinted cis-silencing transcript. *Cold Spring Harbor symposia on quantitative biology* 69, 55-66.

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A., and Cedar, H. (1994). Sp1 elements protect a CpG island from de novo methylation. *Nature* 371, 435-438.

Brandeis, M., Kafri, T., Ariel, M., Chaillet, J.R., McCarrey, J., Razin, A., and Cedar, H. (1993). The ontogeny of allele-specific methylation associated with imprinted genes in the mouse. *The EMBO journal* 12, 3669-3677.

Brender, C., Columbus, R., Metcalf, D., Handman, E., Starr, R., Huntington, N., Tarlinton, D., Odum, N., Nicholson, S.E., Nicola, N.A., *et al.* (2004). SOCS5 is expressed in primary B and T lymphoid cells but is dispensable for lymphocyte production and function. *Molecular and cellular biology* 24, 6094-6103.

Brinkman, A.B., Gu, H., Bartels, S.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 research* 22, 1128-1138.

Carlson, L.L., Page, A.W., and Bestor, T.H. (1992). Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting. *Genes & development* 6, 2536-2541.

Carone, B.R., Fauquier, L., Habib, N., Shea, J.M., Hart, C.E., Li, R., Bock, C., Li, C., Gu, H., Zamore, P.D., *et al.* (2010). Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell* 143, 1084-1096.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P., *et al.* (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123, 581-592.

Cattanach, B.M., and Kirk, M. (1985). Differential activity of maternally and paternally derived chromosome regions in mice. *Nature* 315, 496-498.

Chapman, V., Forrester, L., Sanford, J., Hastie, N., and Rossant, J. (1984). Cell lineage-specific undermethylation of mouse repetitive DNA. *Nature* 307, 284-286.

Chedin, F. (2011). The DNMT3 family of mammalian de novo DNA methyltransferases. *Progress in molecular biology and translational science* 101, 255-285.

Chedin, F., Lieber, M.R., and Hsieh, C.L. (2002). The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proceedings of the National Academy of Sciences of the United States of America* 99, 16916-16921.

Chen, T. (2011). Mechanistic and functional links between histone methylation and DNA methylation. *Progress in molecular biology and translational science* 101, 335-348.

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

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. *The Journal of biological chemistry* 277, 38746-38754.

Cheng, X., and Blumenthal, R.M. (2008). Mammalian DNA methyltransferases: a structural perspective. *Structure* 16, 341-350.

Choufani, S., Shapiro, J.S., Susiarjo, M., Butcher, D.T., Grafodatskaya, D., Lou, Y., Ferreira, J.C., Pinto, D., Scherer, S.W., Shaffer, L.G., *et al.* (2011). A novel approach identifies new differentially methylated regions (DMRs) associated with imprinted genes. *Genome research* 21, 465-476.

Ciccone, D.N., Su, H., Hevi, S., Gay, F., Lei, H., Bajko, J., Xu, G., Li, E., and Chen, T. (2009). KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature* 461, 415-418.

Clouaire, T., Webb, S., Skene, P., Illingworth, R., Kerr, A., Andrews, R., Lee, J.H., Skalnik, D., and Bird, A. (2012). Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes & development* 26, 1714-1728.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452, 215-219.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., *et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819-823.

Constancia, M., Dean, W., Lopes, S., Moore, T., Kelsey, G., and Reik, W. (2000). Deletion of a silencer element in Igf2 results in loss of imprinting independent of H19. *Nature genetics* 26, 203-206.

Coombes, C., Arnaud, P., Gordon, E., Dean, W., Coar, E.A., Williamson, C.M., Feil, R., Peters, J., and Kelsey, G. (2003). Epigenetic properties and identification of an imprint mark in

the Nesp-Gnasxl domain of the mouse Gnas imprinted locus. *Molecular and cellular biology* 23, 5475-5488.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., *et al.* (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* 146, 67-79.

Cowley, M., Garfield, A.S., Madon-Simon, M., Charalambous, M., Clarkson, R.W., Smalley, M.J., Kendrick, H., Isles, A.R., Parry, A.J., Carney, S., *et al.* (2014). Developmental programming mediated by complementary roles of imprinted Grb10 in mother and pup. *PLoS biology* 12, e1001799.

Curley, J.P., Barton, S., Surani, A., and Keverne, E.B. (2004). Coadaptation in mother and infant regulated by a paternally expressed imprinted gene. *Proceedings Biological sciences / The Royal Society* 271, 1303-1309.

da Rocha, S.T., Edwards, C.A., Ito, M., Ogata, T., and Ferguson-Smith, A.C. (2008). Genomic imprinting at the mammalian Dlk1-Dio3 domain. *Trends in genetics : TIG* 24, 306-316.

Das, R., Lee, Y.K., Strogantsev, R., Jin, S., Lim, Y.C., Ng, P.Y., Lin, X.M., Chng, K., Yeo, G., Ferguson-Smith, A.C., *et al.* (2013). DNMT1 and AIM1 Imprinting in human placenta revealed through a genome-wide screen for allele-specific DNA methylation. *BMC genomics* 14, 685.

Davis, E., Caiment, F., Tordoir, X., Cavaille, J., Ferguson-Smith, A., Cockett, N., Georges, M., and Charlier, C. (2005). RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. *Current biology : CB* 15, 743-749.

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

DeChiara, T.M., Robertson, E.J., and Efstratiadis, A. (1991). Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* 64, 849-859.

DeVeale, B., van der Kooy, D., and Babak, T. (2012). Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. *PLoS genetics* 8, e1002600.

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. *The Journal of biological chemistry* 285, 26114-26120.

Duffie, R., Ajjan, S., Greenberg, M.V., Zamudio, N., Escamilla del Arenal, M., Iranzo, J., Okamoto, I., Barbaux, S., Fauque, P., and Bourc'his, D. (2014). The Gpr1/Zdbf2 locus provides new paradigms for transient and dynamic genomic imprinting in mammals. *Genes & development* 28, 463-478.

Duffie, R., and Bourc'his, D. (2013). Parental epigenetic asymmetry in mammals. *Current topics in developmental biology* 104, 293-328.

Edmunds, J.W., Mahadevan, L.C., and Clayton, A.L. (2008). Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *The EMBO journal* 27, 406-420.

Edwards, C.A., and Ferguson-Smith, A.C. (2007). Mechanisms regulating imprinted genes in clusters. *Current opinion in cell biology* 19, 281-289.

Edwards, J.R., O'Donnell, A.H., Rollins, R.A., Peckham, H.E., Lee, C., Milekic, M.H., Chanrion, B., Fu, Y., Su, T., Hibshoosh, H., *et al.* (2010). Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. *Genome research* 20, 972-980.

Egwuagu, C.E., Yu, C.R., Li, Z., and Nussenblatt, R.B. (2005). SOCS5 mRNA levels in peripheral blood mononuclear cells (PBMC): a potential bio-marker for monitoring response of uveitis patients to Daclizumab therapy. *Journal of autoimmunity* 24, 39-46.

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 genetics* 4, e1000116.

Feil, R., and Berger, F. (2007). Convergent evolution of genomic imprinting in plants and mammals. *Trends in genetics : TIG* 23, 192-199.

Feinberg, A.P., Cui, H., and Ohlsson, R. (2002). DNA methylation and genomic imprinting: insights from cancer into epigenetic mechanisms. *Seminars in cancer biology* 12, 389-398.

Feng, Z.P., Chandrashekar, I.R., Low, A., Speed, T.P., Nicholson, S.E., and Norton, R.S. (2012). The N-terminal domains of SOCS proteins: a conserved region in the disordered N-termini of SOCS4 and 5. *Proteins* 80, 946-957.

Ferguson-Smith, A.C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. *Nature reviews Genetics* 12, 565-575.

Ferguson-Smith, A.C., Sasaki, H., Cattanach, B.M., and Surani, M.A. (1993). Parental-origin-specific epigenetic modification of the mouse H19 gene. *Nature* 362, 751-755.

Filippova, G.N. (2008). Genetics and epigenetics of the multifunctional protein CTCF. *Current topics in developmental biology* 80, 337-360.

Fitzpatrick, G.V., Soloway, P.D., and Higgins, M.J. (2002). Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. *Nature genetics* 32, 426-431.

Fournier, C., Goto, Y., Ballestar, E., Delaval, K., Hever, A.M., Esteller, M., and Feil, R. (2002). Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes. *The EMBO journal* 21, 6560-6570.

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., and Paul, C.L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America* 89, 1827-1831.

Fuks, F., Hurd, P.J., Deplus, R., and Kouzarides, T. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic acids research* 31, 2305-2312.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 109, E2579-2586.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. *Annual review of biochemistry* 74, 481-514.

Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G.P., Haig, D., and Dulac, C. (2010). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* 329, 643-648.

Gu, T.P., Guo, F., Yang, H., Wu, H.P., Xu, G.F., Liu, W., Xie, Z.G., Shi, L., He, X., Jin, S.G., *et al.* (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* 477, 606-610.

Guenatri, M., Duffie, 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., and Weber, M. (2013). Functions of DNA methylation and hydroxymethylation in mammalian development. *Current topics in developmental biology* 104, 47-83.

Hackett, J.A., Sengupta, R., Zylitz, 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.

Hahn, M.A., Wu, X., Li, A.X., Hahn, T., and Pfeifer, G.P. (2011). Relationship between gene body DNA methylation and intragenic H3K9me3 and H3K36me3 chromatin marks. *PloS one* 6, e18844.

Haig, D., and Graham, C. (1991). Genomic imprinting and the strange case of the insulin-like growth factor II receptor. *Cell* 64, 1045-1046.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M.A. (2002). Epigenetic reprogramming in mouse primordial germ cells. *Mechanisms of development* 117, 15-23.

Hajkova, P., Jeffries, S.J., Lee, C., Miller, N., Jackson, S.P., and Surani, M.A. (2010). Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* 329, 78-82.

Hata, K., Okano, M., Lei, H., and Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* 129, 1983-1993.

Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., *et al.* (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell stem cell* 6, 479-491.

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.

Henckel, A., Chebli, K., Kota, S.K., Arnaud, P., and Feil, R. (2012). Transcription and histone methylation changes correlate with imprint acquisition in male germ cells. *The EMBO journal* 31, 606-615.

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 & development* 22, 1607-1616.

Hirasawa, R., and Feil, R. (2010). Genomic imprinting and human disease. *Essays in biochemistry* 48, 187-200.

Hirasawa, R., and Sasaki, H. (2009). Dynamic transition of Dnmt3b expression in mouse pre- and early post-implantation embryos. *Gene expression patterns : GEP* 9, 27-30.

Hiura, H., Sugawara, A., Ogawa, H., John, R.M., Miyauchi, N., Miyanari, Y., Horiike, T., Li, Y., Yaegashi, N., Sasaki, H., *et al.* (2010). A tripartite paternally methylated region within the Gpr1-Zdbf2 imprinted domain on mouse chromosome 1 identified by meDIP-on-chip. *Nucleic acids research* 38, 4929-4945.

Hodges, E., Smith, A.D., Kendall, J., Xuan, Z., Ravi, K., Rooks, M., Zhang, M.Q., Ye, K., Bhattacharjee, A., Brizuela, L., *et al.* (2009). High definition profiling of mammalian DNA

methylation by array capture and single molecule bisulfite sequencing. *Genome research* 19, 1593-1605.

Holliday, R., and Pugh, J.E. (1975). DNA modification mechanisms and gene activity during development. *Science* 187, 226-232.

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. *Nature genetics* 45, 1198-1206.

Hore, T.A., Rapkins, R.W., and Graves, J.A. (2007). Construction and evolution of imprinted loci in mammals. *Trends in genetics : TIG* 23, 440-448.

Horsthemke, B., and Buiting, K. (2008). Genomic imprinting and imprinting defects in humans. *Advances in genetics* 61, 225-246.

Hotchkiss, R.D. (1948). The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *The Journal of biological chemistry* 175, 315-332.

Howard, G., Eigens, R., Gaudet, F., Jaenisch, R., and Eden, A. (2008). Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. *Oncogene* 27, 404-408.

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.

Ichihanagi, T., Ichihanagi, K., Miyake, M., and Sasaki, H. (2013). Accumulation and loss of asymmetric non-CpG methylation during male germ-cell development. *Nucleic acids research* 41, 738-745.

Ideraabduallah, F.Y., Vigneau, S., and Bartolomei, M.S. (2008). Genomic imprinting mechanisms in mammals. *Mutation research* 647, 77-85.

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 research* 21, 1670-1676.

Inoue, A., and Zhang, Y. (2011). Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* 334, 194.

Iqbal, K., Jin, S.G., Pfeifer, G.P., and Szabo, P.E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proceedings of the National Academy of Sciences of the United States of America* 108, 3642-3647.

Isles, A.R., Davies, W., and Wilkinson, L.S. (2006). Genomic imprinting and the social brain. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 361, 2229-2237.

Iyengar, S., and Farnham, P.J. (2011). KAP1 protein: an enigmatic master regulator of the genome. *The Journal of biological chemistry* 286, 26267-26276.

Iyengar, S., Ivanov, A.V., Jin, V.X., Rauscher, F.J., 3rd, and Farnham, P.J. (2011). Functional analysis of KAP1 genomic recruitment. *Molecular and cellular biology* 31, 1833-1847.

Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* 416, 556-560.

Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature genetics* 33 *Suppl*, 245-254.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. *Science* 293, 1074-1080.

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. (1999). The DNA methylation paradox. *Trends in genetics : TIG* 15, 34-37.

Kacem, S., and Feil, R. (2009). Chromatin mechanisms in genomic imprinting. *Mammalian genome : official journal of the International Mammalian Genome Society* 20, 544-556.

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. *The EMBO journal* 32, 340-353.

Kamura, T., Maenaka, K., Kotoshiba, S., Matsumoto, M., Kohda, D., Conaway, R.C., Conaway, J.W., and Nakayama, K.I. (2004). VHL-box and SOCS-box domains determine binding specificity for Cul2-Rbx1 and Cul5-Rbx2 modules of ubiquitin ligases. *Genes & development* 18, 3055-3065.

Kaneda, M., Hirasawa, R., Chiba, H., Okano, M., Li, E., and Sasaki, H. (2010). Genetic evidence for Dnmt3a-dependent imprinting during oocyte growth obtained by conditional knockout with Zp3-Cre and complete exclusion of Dnmt3b by chimera formation. *Genes to cells : devoted to molecular & cellular mechanisms* 15, 169-179.

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.

Kaneko, S., Bonasio, R., Saldana-Meyer, R., Yoshida, T., Son, J., Nishino, K., Umezawa, A., and Reinberg, D. (2014). Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Molecular cell* 53, 290-300.

Kario, E., Marmor, M.D., Adamsky, K., Citri, A., Amit, I., Amariglio, N., Rechavi, G., and Yarden, Y. (2005). Suppressors of cytokine signaling 4 and 5 regulate epidermal growth factor receptor signaling. *The Journal of biological chemistry* 280, 7038-7048.

Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., Okano, M., Li, E., Nozaki, M., and Sasaki, H. (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Human molecular genetics* 16, 2272-2280.

Keane, T.M., Goodstadt, L., Danecek, P., White, M.A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., *et al.* (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 477, 289-294.

Kebede, A.F., Schneider, R., and Daujat, S. (2015). Novel types and sites of histone modifications emerge as players in the transcriptional regulation contest. *The FEBS journal* 282, 1658-1674.

Kim, T.K., Hemberg, M., and Gray, J.M. (2015). Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers. *Cold Spring Harbor perspectives in biology* 7, a018622.

Kobayashi, H., Sakurai, T., Imai, M., Takahashi, N., Fukuda, A., Yayoi, O., Sato, S., Nakabayashi, K., Hata, K., Sotomaru, Y., *et al.* (2012). Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS genetics* 8, e1002440.

Kobayashi, H., Yamada, K., Morita, S., Hiura, H., Fukuda, A., Kagami, M., Ogata, T., Hata, K., Sotomaru, Y., and Kono, T. (2009). Identification of the mouse paternally expressed imprinted gene *Zdbf2* on chromosome 1 and its imprinted human homolog *ZDBF2* on chromosome 2. *Genomics* 93, 461-472.

Kobayashi, S., Isotani, A., Mise, N., Yamamoto, M., Fujihara, Y., Kaseda, K., Nakanishi, T., Ikawa, M., Hamada, H., Abe, K., *et al.* (2006). Comparison of gene expression in male and female mouse blastocysts revealed imprinting of the X-linked gene, *Rhox5/Pem*, at preimplantation stages. *Current biology : CB* 16, 166-172.

Kota, S.K., Lleres, D., Bouschet, T., Hirasawa, R., Marchand, A., Begon-Pescia, C., Sanli, I., Arnaud, P., Journot, L., Girardot, M., *et al.* (2014). ICR noncoding RNA expression controls imprinting and DNA replication at the *Dlk1-Dio3* domain. *Developmental cell* 31, 19-33.

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693-705.

Krebs, D.L., Uren, R.T., Metcalf, D., Rakar, S., Zhang, J.G., Starr, R., De Souza, D.P., Hanzinikolas, K., Eyles, J., Connolly, L.M., *et al.* (2002). SOCS-6 binds to insulin receptor substrate 4, and mice lacking the SOCS-6 gene exhibit mild growth retardation. *Molecular and cellular biology* 22, 4567-4578.

Kurukuti, S., Tiwari, V.K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenkov, V., Reik, W., and Ohlsson, R. (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proceedings of the National Academy of Sciences of the United States of America* 103, 10684-10689.

La Salle, S., Mertineit, C., Taketo, T., Moens, P.B., Bestor, T.H., and Trasler, J.M. (2004). Windows for sex-specific methylation marked by DNA methyltransferase expression profiles in mouse germ cells. *Developmental biology* 268, 403-415.

La Salle, S., and Trasler, J.M. (2006). Dynamic expression of DNMT3a and DNMT3b isoforms during male germ cell development in the mouse. *Developmental biology* 296, 71-82.

Ladopoulos, V., Hofemeister, H., Hoogenkamp, M., Riggs, A.D., Stewart, A.F., and Bonifer, C. (2013). The histone methyltransferase KMT2B is required for RNA polymerase II association and protection from DNA methylation at the MagohB CpG island promoter. *Molecular and cellular biology* 33, 1383-1393.

Latos, P.A., Pauler, F.M., Koerner, M.V., Senergin, H.B., Hudson, Q.J., Stocsits, R.R., Allhoff, W., Stricker, S.H., Klement, R.M., Warczok, K.E., *et al.* (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* 338, 1469-1472.

Lehnertz, B., Ueda, Y., Derijck, A.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Current biology : CB* 13, 1192-1200.

Leonhardt, H., Page, A.W., Weier, H.U., and Bestor, T.H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* 71, 865-873.

Lewis, A., Mitsuya, K., Umlauf, D., Smith, P., Dean, W., Walter, J., Higgins, M., Feil, R., and Reik, W. (2004). Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. *Nature genetics* 36, 1291-1295.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. *Nature* 366, 362-365.

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, H., Rauch, T., Chen, Z.X., Szabo, P.E., Riggs, A.D., and Pfeifer, G.P. (2006). The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. *The Journal of biological chemistry* 281, 19489-19500.

Li, T., Vu, T.H., Ulaner, G.A., Yang, Y., Hu, J.F., and Hoffman, A.R. (2004). Activating and silencing histone modifications form independent allelic switch regions in the imprinted *Gnas* gene. *Human molecular genetics* 13, 741-750.

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. *Developmental cell* 15, 547-557.

Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F., and Schubeler, D. (2011). Identification of genetic elements that autonomously determine DNA methylation states. *Nature genetics* 43, 1091-1097.

Lin, S., Ferguson-Smith, A.C., Schultz, R.M., and Bartolomei, M.S. (2011). Nonallelic transcriptional roles of CTCF and cohesins at imprinted loci. *Molecular and cellular biology* 31, 3094-3104.

Lin, S.P., Youngson, N., Takada, S., Seitz, H., Reik, W., Paulsen, M., Cavaille, J., and Ferguson-Smith, A.C. (2003). Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the *Dlk1-Gtl2* imprinted cluster on mouse chromosome 12. *Nature genetics* 35, 97-102.

Lindroth, A.M., Park, Y.J., McLean, C.M., Dokshin, G.A., Persson, J.M., Herman, H., Pasini, D., Miro, X., Donohoe, M.E., Lee, J.T., *et al.* (2008). Antagonism between DNA and H3K27 methylation at the imprinted *Rasgrf1* locus. *PLoS genetics* 4, e1000145.

Lister, R., Pelizzola, M., Downen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., *et al.* (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462, 315-322.

Liu, J., Chen, M., Deng, C., Bourc'his, D., Nealon, J.G., Erlichman, B., Bestor, T.H., and Weinstein, L.S. (2005). Identification of the control region for tissue-specific imprinting of the stimulatory G protein alpha-subunit. *Proceedings of the National Academy of Sciences of the United States of America* 102, 5513-5518.

Liu, X., Gao, Q., Li, P., Zhao, Q., Zhang, J., Li, J., Koseki, H., and Wong, J. (2013). UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nature communications* 4, 1563.

Liu, Y., Toh, H., Sasaki, H., Zhang, X., and Cheng, X. (2012). An atomic model of *Zfp57* recognition of CpG methylation within a specific DNA sequence. *Genes & development* 26, 2374-2379.

Lucifero, D., Mann, M.R., Bartolomei, M.S., and Trasler, J.M. (2004). Gene-specific timing and epigenetic memory in oocyte imprinting. *Human molecular genetics* 13, 839-849.

Lucifero, D., Mertineit, C., Clarke, H.J., Bestor, T.H., and Trasler, J.M. (2002). Methylation dynamics of imprinted genes in mouse germ cells. *Genomics* 79, 530-538.

Luedi, P.P., Dietrich, F.S., Weidman, J.R., Bosko, J.M., Jirtle, R.L., and Hartemink, A.J. (2007). Computational and experimental identification of novel human imprinted genes. *Genome research* 17, 1723-1730.

Luedi, P.P., Hartemink, A.J., and Jirtle, R.L. (2005). Genome-wide prediction of imprinted murine genes. *Genome research* 15, 875-884.

Macleod, D., Charlton, J., Mullins, J., and Bird, A.P. (1994). Sp1 sites in the mouse *aprt* gene promoter are required to prevent methylation of the CpG island. *Genes & development* 8, 2282-2292.

Malagnac, F., Bartee, L., and Bender, J. (2002). An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation. *The EMBO journal* 21, 6842-6852.

Mancini-Dinardo, D., Steele, S.J., Levorse, J.M., Ingram, R.S., and Tilghman, S.M. (2006). Elongation of the *Kcnq1ot1* transcript is required for genomic imprinting of neighboring genes. *Genes & development* 20, 1268-1282.

Marks, H., Kalkan, T., Menafrá, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A., *et al.* (2012). The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* 149, 590-604.

Martens, N., Uzan, G., Wery, M., Hooghe, R., Hooghe-Peters, E.L., and Gertler, A. (2005). Suppressor of cytokine signaling 7 inhibits prolactin, growth hormone, and leptin signaling by interacting with STAT5 or STAT3 and attenuating their nuclear translocation. *The Journal of biological chemistry* 280, 13817-13823.

Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., *et al.* (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 466, 253-257.

McCole, R.B., Loughran, N.B., Chahal, M., Fernandes, L.P., Roberts, R.G., Fraternali, F., O'Connell, M.J., and Oakey, R.J. (2011). A case-by-case evolutionary analysis of four imprinted retrogenes. *Evolution; international journal of organic evolution* 65, 1413-1427.

McEwen, K.R., and Ferguson-Smith, A.C. (2010). Distinguishing epigenetic marks of developmental and imprinting regulation. *Epigenetics & chromatin* 3, 2.

McGrath, J., and Solter, D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* 37, 179-183.

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.

Mertineit, C., Yoder, J.A., Taketo, T., Laird, D.W., Trasler, J.M., and Bestor, T.H. (1998). Sex-specific exons control DNA methyltransferase in mammalian germ cells. *Development* 125, 889-897.

Messerschmidt, D.M., de Vries, W., 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.

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.

Moarefi, A.H., and Chedin, F. (2011). ICF syndrome mutations cause a broad spectrum of biochemical defects in DNMT3B-mediated de novo DNA methylation. *Journal of molecular biology* 409, 758-772.

Monk, M., Boubelik, M., and Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* 99, 371-382.

Morselli, M., Pastor, W.A., Montanini, B., Nee, K., Ferrari, R., Fu, K., Bonora, G., Rubbi, L., Clark, A.T., Ottonello, S., *et al.* (2015). In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. *eLife* 4.

Mouse, E.C., Stamatoyannopoulos, J.A., Snyder, M., Hardison, R., Ren, B., Gingeras, T., Gilbert, D.M., Groudine, M., Bender, M., Kaul, R., *et al.* (2012). An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome biology* 13, 418.

Murrell, A., Heeson, S., Bowden, L., Constancia, M., Dean, W., Kelsey, G., and Reik, W. (2001). An intragenic methylated region in the imprinted Igf2 gene augments transcription. *EMBO reports* 2, 1101-1106.

Murrell, A., Heeson, S., and Reik, W. (2004). Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nature genetics* 36, 889-893.

Nabel, C.S., Manning, S.A., and Kohli, R.M. (2012). The curious chemical biology of cytosine: deamination, methylation, and oxidation as modulators of genomic potential. *ACS chemical biology* 7, 20-30.

Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 322, 1717-1720.

Nakabayashi, K., Trujillo, A.M., Tayama, C., Camprubi, C., Yoshida, W., Lapunzina, P., Sanchez, A., Soejima, H., Aburatani, H., Nagae, G., *et al.* (2011). Methylation screening of reciprocal genome-wide UPDs identifies novel human-specific imprinted genes. *Human molecular genetics* 20, 3188-3197.

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. *Nature cell biology* 9, 64-71.

Nakamura, T., Liu, Y.J., Nakashima, H., Umehara, H., Inoue, K., Matoba, S., Tachibana, M., Ogura, A., Shinkai, Y., and Nakano, T. (2012). PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* 486, 415-419.

Nicholson, S.E., Metcalf, D., Sprigg, N.S., Columbus, R., Walker, F., Silva, A., Cary, D., Willson, T.A., Zhang, J.G., Hilton, D.J., *et al.* (2005). Suppressor of cytokine signaling (SOCS)-5 is a potential negative regulator of epidermal growth factor signaling. *Proceedings of the National Academy of Sciences of the United States of America* 102, 2328-2333.

Nikaido, I., Saito, C., Mizuno, Y., Meguro, M., Bono, H., Kadomura, M., Kono, T., Morris, G.A., Lyons, P.A., Oshimura, M., *et al.* (2003). Discovery of imprinted transcripts in the mouse transcriptome using large-scale expression profiling. *Genome research* 13, 1402-1409.

Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156, 935-949.

O'Doherty, A.M., Rutledge, C.E., Sato, S., Thakur, A., Lees-Murdock, D.J., Hata, K., and Walsh, C.P. (2011). DNA methylation plays an important role in promoter choice and protein production at the mouse Dnmt3L locus. *Developmental biology* 356, 411-420.

O'Sullivan, L.A., Liongue, C., Lewis, R.S., Stephenson, S.E., and Ward, A.C. (2007). Cytokine receptor signaling through the Jak-Stat-Socs pathway in disease. *Molecular immunology* 44, 2497-2506.

Ohno, R., Nakayama, M., Naruse, C., Okashita, N., Takano, O., Tachibana, M., Asano, M., Saitou, M., and Seki, Y. (2013). A replication-dependent passive mechanism modulates DNA demethylation in mouse primordial germ cells. *Development* 140, 2892-2903.

Ohshima, M., Yokoyama, A., Ohnishi, H., Hamada, H., Kohno, N., Higaki, J., and Naka, T. (2007). Overexpression of suppressor of cytokine signalling-5 augments eosinophilic

airway inflammation in mice. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* 37, 735-742.

Okabe, H., Hiura, H., Nishida, Y., Funayama, R., Tanaka, S., Chiba, H., Yaegashi, N., Nakayama, K., Sasaki, H., and Arima, T. (2012). Re-investigation and RNA sequencing-based identification of genes with placenta-specific imprinted expression. *Human molecular genetics* 21, 548-558.

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.

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

Ooi, S.K., 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.

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

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

Ozaki, A., Seki, Y., Fukushima, A., and Kubo, M. (2005). The control of allergic conjunctivitis by suppressor of cytokine signaling (SOCS)3 and SOCS5 in a murine model. *Journal of immunology* 175, 5489-5497.

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Molecular cell* 32, 232-246.

Pant, V., Kurukuti, S., Pugacheva, E., Shamsuddin, S., Mariano, P., Renkawitz, R., Klenova, E., Lobanenko, V., and Ohlsson, R. (2004). Mutation of a single CTCF target site within the H19 imprinting control region leads to loss of Igf2 imprinting and complex patterns of de novo methylation upon maternal inheritance. *Molecular and cellular biology* 24, 3497-3504.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., *et al.* (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132, 422-433.

Pauler, F.M., Barlow, D.P., and Hudson, Q.J. (2012). Mechanisms of long range silencing by imprinted macro non-coding RNAs. *Current opinion in genetics & development* 22, 283-289.

Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A., *et al.* (2001). Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* 107, 323-337.

Pnueli, L., Rudnizky, S., Yosefzon, Y., and Melamed, P. (2015). RNA transcribed from a distal enhancer is required for activating the chromatin at the promoter of the gonadotropin alpha-subunit gene. *Proceedings of the National Academy of Sciences of the United States of America* 112, 4369-4374.

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.

Proudhon, C., and Bourc'his, D. (2010a). [Evolution of genomic imprinting in mammals: what a zoo!]. *Medecine sciences : M/S* 26, 497-503.

Proudhon, C., and Bourc'his, D. (2010b). Identification and resolution of artifacts in the interpretation of imprinted gene expression. *Briefings in functional genomics* 9, 374-384.

Proudhon, C., Duffie, R., Ajjan, S., Cowley, M., Iranzo, J., Carbajosa, G., Saadeh, H., Holland, M.L., Oakey, R.J., Rakyan, V.K., *et al.* (2012). Protection against de novo methylation is instrumental in maintaining parent-of-origin methylation inherited from the gametes. *Molecular cell* 47, 909-920.

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. *Molecular cell* 44, 361-372.

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proceedings of the National Academy of Sciences of the United States of America* 97, 5237-5242.

Rand, E., Ben-Porath, I., Keshet, I., and Cedar, H. (2004). CTCF elements direct allele-specific undermethylation at the imprinted H19 locus. *Current biology : CB* 14, 1007-1012.

Rapkins, R.W., Hore, T., Smithwick, M., Ager, E., Pask, A.J., Renfree, M.B., Kohn, M., Hameister, H., Nicholls, R.D., Deakin, J.E., *et al.* (2006). Recent assembly of an imprinted domain from non-imprinted components. *PLoS genetics* 2, e182.

Razin, A., and Cedar, H. (1994). DNA methylation and genomic imprinting. *Cell* 77, 473-476.

Reddington, J.P., Perricone, S.M., Nestor, C.E., Reichmann, J., Youngson, N.A., Suzuki, M., Reinhardt, D., Dunican, D.S., Prendergast, J.G., Mjoseng, H., *et al.* (2013). Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. *Genome biology* 14, R25.

Regha, K., Sloane, M.A., Huang, R., Pauler, F.M., Warczok, K.E., Melikant, B., Radolf, M., Martens, J.H., Schotta, G., Jenuwein, T., *et al.* (2007). Active and repressive chromatin are interspersed without spreading in an imprinted gene cluster in the mammalian genome. *Molecular cell* 27, 353-366.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 447, 425-432.

Reik, W., Collick, A., Norris, M.L., Barton, S.C., and Surani, M.A. (1987). Genomic imprinting determines methylation of parental alleles in transgenic mice. *Nature* 328, 248-251.

Reik, W., Dean, W., and Walter, J. (2001). Epigenetic reprogramming in mammalian development. *Science* 293, 1089-1093.

Reik, W., and Walter, J. (2001). Genomic imprinting: parental influence on the genome. *Nature reviews Genetics* 2, 21-32.

Renteria, M.E., Coolen, M.W., Statham, A.L., Choi, R.S., Qu, W., Campbell, M.J., Smith, S., Henders, A.K., Montgomery, G.W., Clark, S.J., *et al.* (2013). GWAS of DNA methylation variation within imprinting control regions suggests parent-of-origin association. *Twin research and human genetics : the official journal of the International Society for Twin Studies* 16, 767-781.

Riggs, A.D. (1975). X inactivation, differentiation, and DNA methylation. *Cytogenetics and cell genetics* 14, 9-25.

Rollins, R.A., Haghighi, F., Edwards, J.R., Das, R., Zhang, M.Q., Ju, J., and Bestor, T.H. (2006). Large-scale structure of genomic methylation patterns. *Genome research* 16, 157-163.

Rougier, N., Bourc'his, D., Gomes, D.M., Niveleau, A., Plachot, M., Paldi, A., and Viegas-Pequignot, E. (1998). Chromosome methylation patterns during mammalian preimplantation development. *Genes & development* 12, 2108-2113.

Ruchirawat, M., Noshari, J., and Lapeyre, J.N. (1987). Kinetic mechanisms and interaction of rat liver DNA methyltransferase with defined DNA substrates. *Molecular and cellular biochemistry* 76, 45-54.

Ruf, N., Bähring, S., Galetzka, D., Pliushch, G., Luft, F.C., Nurnberg, P., Haaf, T., Kelsey, G., and Zechner, U. (2007). Sequence-based bioinformatic prediction and QUASEP identify genomic imprinting of the KCNK9 potassium channel gene in mouse and human. *Human molecular genetics* 16, 2591-2599.

Ruf, N., Dunzinger, U., Brinckmann, A., Haaf, T., Nurnberg, P., and Zechner, U. (2006). Expression profiling of uniparental mouse embryos is inefficient in identifying novel imprinted genes. *Genomics* 87, 509-519.

Saito, Y., Liang, G., Egger, G., Friedman, J.M., Chuang, J.C., Coetzee, G.A., and Jones, P.A. (2006). Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. *Cancer cell* 9, 435-443.

Sakai, Y., Suetake, I., Shinozaki, F., Yamashina, S., and Tajima, S. (2004). Co-expression of de novo DNA methyltransferases Dnmt3a2 and Dnmt3L in gonocytes of mouse embryos. *Gene expression patterns : GEP* 5, 231-237.

Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology* 32, 347-355.

Sandovici, I., Smith, N.H., Nitert, M.D., Ackers-Johnson, M., Uribe-Lewis, S., Ito, Y., Jones, R.H., Marquez, V.E., Cairns, W., Tadayyon, M., *et al.* (2011). Maternal diet and aging alter the epigenetic control of a promoter-enhancer interaction at the Hnf4a gene in rat pancreatic islets. *Proceedings of the National Academy of Sciences of the United States of America* 108, 5449-5454.

Sanli, I., and Feil, R. (2015). Chromatin mechanisms in the developmental control of imprinted gene expression. *The international journal of biochemistry & cell biology*.

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

Santos, F., Peters, A.H., Otte, A.P., Reik, W., and Dean, W. (2005). Dynamic chromatin modifications characterise the first cell cycle in mouse embryos. *Developmental biology* 280, 225-236.

Schaefer, C.B., Ooi, S.K., Bestor, T.H., and Bourc'his, D. (2007). Epigenetic decisions in mammalian germ cells. *Science* 316, 398-399.

Schmidl, C., Klug, M., Boeld, T.J., Andreesen, R., Hoffmann, P., Edinger, M., and Rehli, M. (2009). Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome research* 19, 1165-1174.

Schoenherr, C.J., Levorse, J.M., and Tilghman, S.M. (2003). CTCF maintains differential methylation at the Igf2/H19 locus. *Nature genetics* 33, 66-69.

Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., and Rauscher, F.J., 3rd (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes & development* 16, 919-932.

Schultz, D.C., Friedman, J.R., and Rauscher, F.J., 3rd (2001). Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. *Genes & development* 15, 428-443.

Schulz, R., Proudhon, C., Bestor, T.H., Woodfine, K., Lin, C.S., Lin, S.P., Prissette, M., Oakey, R.J., and Bourc'his, D. (2010). The parental non-equivalence of imprinting control regions during mammalian development and evolution. *PLoS genetics* 6, e1001214.

Schuster-Gossler, K., Bilinski, P., Sado, T., Ferguson-Smith, A., and Gossler, A. (1998). The mouse Gtl2 gene is differentially expressed during embryonic development, encodes multiple alternatively spliced transcripts, and may act as an RNA. *Developmental dynamics : an official publication of the American Association of Anatomists* 212, 214-228.

Seitz, H., Royo, H., Bortolin, M.L., Lin, S.P., Ferguson-Smith, A.C., and Cavaille, J. (2004). A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. *Genome research* 14, 1741-1748.

Seitz, H., Youngson, N., Lin, S.P., Dalbert, S., Paulsen, M., Bachellerie, J.P., Ferguson-Smith, A.C., and Cavaille, J. (2003). Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nature genetics* 34, 261-262.

Seki, Y., Hayashi, K., Matsumoto, A., Seki, N., Tsukada, J., Ransom, J., Naka, T., Kishimoto, T., Yoshimura, A., and Kubo, M. (2002). Expression of the suppressor of cytokine signaling-5 (SOCS5) negatively regulates IL-4-dependent STAT6 activation and Th2 differentiation. *Proceedings of the National Academy of Sciences of the United States of America* 99, 13003-13008.

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.

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 genetics* 9, e1003439.

Shovlin, T.C., Bourc'his, D., La Salle, S., O'Doherty, A., Trasler, J.M., Bestor, T.H., and Walsh, C.P. (2007). Sex-specific promoters regulate Dnmt3L expression in mouse germ cells. *Human reproduction* 22, 457-467.

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., Esteve, P.O., Pradhan, S., and Carey, M. (2007). Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes & development* 21, 1169-1178.

Smallwood, S.A., Tomizawa, S., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S.R., and Kelsey, G. (2011). Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nature genetics* 43, 811-814.

Smilnich, N.J., Day, C.D., Fitzpatrick, G.V., Caldwell, G.M., Lossie, A.C., Cooper, P.R., Smallwood, A.C., Joyce, J.A., Schofield, P.N., Reik, W., *et al.* (1999). A maternally methylated CpG island in KvLQT1 is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. *Proceedings of the National Academy of Sciences of the United States of America* 96, 8064-8069.

Smith, F.M., Garfield, A.S., and Ward, A. (2006). Regulation of growth and metabolism by imprinted genes. *Cytogenetic and genome research* 113, 279-291.

Smith, R.J., Dean, W., Konfortova, G., and Kelsey, G. (2003). Identification of novel imprinted genes in a genome-wide screen for maternal methylation. *Genome research* 13, 558-569.

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.

Smith, Z.D., Gu, H., Bock, C., Gnirke, A., and Meissner, A. (2009). High-throughput bisulfite sequencing in mammalian genomes. *Methods* 48, 226-232.

Smits, G., Mungall, A.J., Griffiths-Jones, S., Smith, P., Beury, D., Matthews, L., Rogers, J., Pask, A.J., Shaw, G., VandeBerg, J.L., *et al.* (2008). Conservation of the H19 noncoding RNA and H19-IGF2 imprinting mechanism in therians. *Nature genetics* 40, 971-976.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, 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.

Stec, I., Nagl, S.B., van Ommen, G.J., and den Dunnen, J.T. (2000). The PWWP domain: a potential protein-protein interaction domain in nuclear proteins influencing differentiation? *FEBS letters* 473, 1-5.

Stein, R., Gruenbaum, Y., Pollack, Y., Razin, A., and Cedar, H. (1982). Clonal inheritance of the pattern of DNA methylation in mouse cells. *Proceedings of the National Academy of Sciences of the United States of America* 79, 61-65.

Stoger, R., Kubicka, P., Liu, C.G., Kafri, T., Razin, A., Cedar, H., and Barlow, D.P. (1993). Maternal-specific methylation of the imprinted mouse *Igf2r* locus identifies the expressed locus as carrying the imprinting signal. *Cell* 73, 61-71.

Straussman, R., Nejman, D., Roberts, D., Steinfeld, I., Blum, B., Benvenisty, N., Simon, I., Yakhini, Z., and Cedar, H. (2009). Developmental programming of CpG island methylation profiles in the human genome. *Nature structural & molecular biology* 16, 564-571.

Strichman-Almashanu, L.Z., Lee, R.S., Onyango, P.O., Perlman, E., Flam, F., Frieman, M.B., and Feinberg, A.P. (2002). A genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes. *Genome research* 12, 543-554.

Stricker, S.H., Steenpass, L., Pauler, F.M., Santoro, F., Latos, P.A., Huang, R., Koerner, M.V., Sloane, M.A., Warczok, K.E., and Barlow, D.P. (2008). Silencing and transcriptional properties of the imprinted *Air* ncRNA are independent of the endogenous promoter. *The EMBO journal* 27, 3116-3128.

Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., and Tajima, S. (2004). DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *The Journal of biological chemistry* 279, 27816-27823.

Sun, X.J., Wei, J., Wu, X.Y., Hu, M., Wang, L., Wang, H.H., Zhang, Q.H., Chen, S.J., Huang, Q.H., and Chen, Z. (2005). Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. *The Journal of biological chemistry* 280, 35261-35271.

Surani, M.A., Barton, S.C., and Norris, M.L. (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* 308, 548-550.

Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nature reviews Genetics* 9, 465-476.

Swain, J.L., Stewart, T.A., and Leder, P. (1987). Parental legacy determines methylation and expression of an autosomal transgene: a molecular mechanism for parental imprinting. *Cell* 50, 719-727.

Szabo, P.E., and Pfeifer, G.P. (2012). H3K9me2 attracts PGC7 in the zygote to prevent Tet3-mediated oxidation of 5-methylcytosine. *Journal of molecular cell biology* 4, 427-429.

Takeshima, H., Yamashita, S., Shimazu, T., Niwa, T., and Ushijima, T. (2009). The presence of RNA polymerase II, active or stalled, predicts epigenetic fate of promoter CpG islands. *Genome research* 19, 1974-1982.

Talbert, P.B., and Henikoff, S. (2010). Histone variants--ancient wrap artists of the epigenome. *Nature reviews Molecular cell biology* 11, 264-275.

Tamaru, H., and Selker, E.U. (2001). A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. *Nature* 414, 277-283.

Tanay, A., O'Donnell, A.H., Damelin, M., and Bestor, T.H. (2007). Hyperconserved CpG domains underlie Polycomb-binding sites. *Proceedings of the National Academy of Sciences of the United States of America* 104, 5521-5526.

Terranova, R., Yokobayashi, S., Stadler, M.B., Otte, A.P., van Lohuizen, M., Orkin, S.H., and Peters, A.H. (2008). Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Developmental cell* 15, 668-679.

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

Trasler, J.M. (2006). Gamete imprinting: setting epigenetic patterns for the next generation. *Reproduction, fertility, and development* 18, 63-69.

Tucker, K.L., Beard, C., Dausmann, J., Jackson-Grusby, L., Laird, P.W., Lei, H., Li, E., and Jaenisch, R. (1996). Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. *Genes & development* 10, 1008-1020.

Umlauf, D., Goto, Y., Cao, R., Cerqueira, F., Wagschal, A., Zhang, Y., and Feil, R. (2004). Imprinting along the *Kcnq1* domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. *Nature genetics* 36, 1296-1300.

Vavouri, T., and Lehner, B. (2012). Human genes with CpG island promoters have a distinct transcription-associated chromatin organization. *Genome biology* 13, R110.

Wagschal, A., Sutherland, H.G., Woodfine, K., Henckel, A., Chebli, K., Schulz, R., Oakey, R.J., Bickmore, W.A., and Feil, R. (2008). G9a histone methyltransferase contributes to imprinting in the mouse placenta. *Molecular and cellular biology* 28, 1104-1113.

Wan, J., Oliver, V.F., Zhu, H., Zack, D.J., Qian, J., and Merbs, S.L. (2013). Integrative analysis of tissue-specific methylation and alternative splicing identifies conserved transcription factor binding motifs. *Nucleic acids research* 41, 8503-8514.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153, 910-918.

Wang, X., Soloway, P.D., and Clark, A.G. (2011). A survey for novel imprinted genes in the mouse placenta by mRNA-seq. *Genetics* 189, 109-122.

Watanabe, T., Tomizawa, S., Mitsuya, K., Totoki, Y., Yamamoto, Y., Kuramochi-Miyagawa, S., Iida, N., Hoki, Y., Murphy, P.J., Toyoda, A., *et al.* (2011). Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse *Rasgrf1* locus. *Science* 332, 848-852.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature genetics* 39, 457-466.

Weber, M., and Schubeler, D. (2007). Genomic patterns of DNA methylation: targets and function of an epigenetic mark. *Current opinion in cell biology* 19, 273-280.

Webster, K.E., O'Bryan, M.K., Fletcher, S., Crewther, P.E., Aapola, U., Craig, J., Harrison, D.K., Aung, H., Phutikanit, N., Lyle, R., *et al.* (2005). Meiotic and epigenetic defects in *Dnmt3L*-knockout mouse spermatogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 102, 4068-4073.

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., *et al.* (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451, 796-801.

Wiench, M., John, S., Baek, S., Johnson, T.A., Sung, M.H., Escobar, T., Simmons, C.A., Pearce, K.H., Biddie, S.C., Sabo, P.J., *et al.* (2011). DNA methylation status predicts cell type-specific enhancer activity. *The EMBO journal* 30, 3028-3039.

Wigler, M., Levy, D., and Perucho, M. (1981). The somatic replication of DNA methylation. *Cell* 24, 33-40.

Williamson, C.M., Turner, M.D., Ball, S.T., Nottingham, W.T., Glenister, P., Fray, M., Tymowska-Lalanne, Z., Plagge, A., Powles-Glover, N., Kelsey, G., *et al.* (2006). Identification of an imprinting control region affecting the expression of all transcripts in the *Gnas* cluster. *Nature genetics* 38, 350-355.

Wolf, J.B., Oakey, R.J., and Feil, R. (2014). Imprinted gene expression in hybrids: perturbed mechanisms and evolutionary implications. *Heredity* 113, 167-175.

Wood, A.J., Roberts, R.G., Monk, D., Moore, G.E., Schulz, R., and Oakey, R.J. (2007). A screen for retrotransposed imprinted genes reveals an association between X chromosome homology and maternal germ-line methylation. *PLoS genetics* 3, e20.

Wood, A.J., Schulz, R., Woodfine, K., Koltowska, K., Beechey, C.V., Peters, J., Bourc'his, D., and Oakey, R.J. (2008). Regulation of alternative polyadenylation by genomic imprinting. *Genes & development* 22, 1141-1146.

Wossidlo, M., Arand, J., Sebastiano, V., Lepikhov, K., Boiani, M., Reinhardt, R., Scholer, H., and Walter, J. (2010). Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes. *The EMBO journal* 29, 1877-1888.

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. *Nature communications* 2, 241.

Xie, W., Barr, C.L., Kim, A., Yue, F., Lee, A.Y., Eubanks, J., Dempster, E.L., and Ren, B. (2012). Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* 148, 816-831.

Yamaguchi, S., Hong, K., Liu, R., Inoue, A., Shen, L., Zhang, K., and Zhang, Y. (2013). Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. *Cell research* 23, 329-339.

Yang, X., Han, H., De Carvalho, D.D., Lay, F.D., Jones, P.A., and Liang, G. (2014). Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer cell* 26, 577-590.

Yang, Y., Li, T., Vu, T.H., Ulaner, G.A., Hu, J.F., and Hoffman, A.R. (2003). The histone code regulating expression of the imprinted mouse *Igf2r* gene. *Endocrinology* 144, 5658-5670.

Yearim, A., Gelfman, S., Shayevitch, R., Melcer, S., Glaich, O., Mallm, J.P., Nissim-Rafinia, M., Cohen, A.H., Rippe, K., Meshorer, E., *et al.* (2015). HP1 is involved in regulating the global impact of DNA methylation on alternative splicing. *Cell reports* 10, 1122-1134.

Yen, Z.C., Meyer, I.M., Karalic, S., and Brown, C.J. (2007). A cross-species comparison of X-chromosome inactivation in Eutheria. *Genomics* 90, 453-463.

Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends in genetics : TIG* 13, 335-340.

Yoh, S.M., Lucas, J.S., and Jones, K.A. (2008). The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes & development* 22, 3422-3434.

Yokomine, T., Hata, K., Tsudzuki, M., and Sasaki, H. (2006). Evolution of the vertebrate DNMT3 gene family: a possible link between existence of DNMT3L and genomic imprinting. *Cytogenetic and genome research* 113, 75-80.

Yoon, B., Herman, H., Hu, B., Park, Y.J., Lindroth, A., Bell, A., West, A.G., Chang, Y., Stablewski, A., Piel, J.C., *et al.* (2005). Rasgrf1 imprinting is regulated by a CTCF-dependent methylation-sensitive enhancer blocker. *Molecular and cellular biology* 25, 11184-11190.

Yoon, S., Yi, Y.S., Kim, S.S., Kim, J.H., Park, W.S., and Nam, S.W. (2012). SOCS5 and SOCS6 have similar expression patterns in normal and cancer tissues. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* 33, 215-221.

Zhang, J.G., Farley, A., Nicholson, S.E., Willson, T.A., Zugaro, L.M., Simpson, R.J., Moritz, R.L., Cary, D., Richardson, R., Hausmann, G., *et al.* (1999). The conserved SOCS box motif in suppressors of cytokine signaling binds to elongins B and C and may couple bound proteins to proteasomal degradation. *Proceedings of the National Academy of Sciences of the United States of America* 96, 2071-2076.

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 research* 38, 4246-4253.

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. *Molecular cell* 40, 939-953.

Zuo, X., Sheng, J., Lau, H.T., McDonald, C.M., Andrade, M., Cullen, D.E., Bell, F.T., Iacovino, M., Kyba, M., Xu, G., *et al.* (2012). Zinc finger protein ZFP57 requires its co-factor to recruit DNA methyltransferases and maintains DNA methylation imprint in embryonic stem cells via its transcriptional repression domain. *The Journal of biological chemistry* 287, 2107-2118.

Zwart, R., Sleutels, F., Wutz, A., Schinkel, A.H., and Barlow, D.P. (2001). Bidirectional action of the Igf2r imprint control element on upstream and downstream imprinted genes. *Genes & development* 15, 2361-2366.

