



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par :
Annabelle CONGRAS

le vendredi 12 décembre 2014

Titre :

Analyse de la méthylation de l'ADN spermatique et développement de cellules pluripotentes induites chez des verrats infertiles porteurs ou non de remaniements chromosomiques.

École doctorale et discipline ou spécialité :

ED BSB : Biologie cellulaire

Unité de recherche :

UMR1388 Génétique, Physiologie et Systèmes d'Elevage (GenPhySE)

Directeur(s) de Thèse :

Martine YERLE-BOUISSOU
Hervé ACLOQUE

Jury :

Kerstin BYSTRICKY - Présidente du jury - Professeur
Hélène JAMMES - Rapporteur - Directrice de recherche
Pierre SAVATIER - Rapporteur - Directeur de recherche
Michael WEBER - Examineur - Chargé de recherche
Franck PELLESTOR - Examineur - Maître de conférence des universités Praticien hospitalier



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par :

Annabelle CONGRAS

le vendredi 12 décembre 2014

Titre :

Analyse de la méthylation de l'ADN spermatique et développement de cellules pluripotentes induites chez des verrats infertiles porteurs ou non de remaniements chromosomiques.

École doctorale et discipline ou spécialité :

ED BSB : Biologie cellulaire

Unité de recherche :

UMR1388 Génétique, Physiologie et Systèmes d'Élevage (GenPhySE)

Directeur(s) de Thèse :

Martine YERLE-BOUISSOU

Hervé ACLOQUE

Jury :

Kerstin BYSTRICKY - Présidente du jury - Professeur

Hélène JAMMES - Rapporteur - Directrice de recherche

Pierre SAVATIER - Rapporteur - Directeur de recherche

Michael WEBER - Examineur - Chargé de recherche

Franck PELLESTOR - Examineur - Maître de conférence des universités Praticien hospitalier

REMERCIEMENTS

Tout d'abord, je remercie chaleureusement mes deux directeurs de thèse de m'avoir donné l'opportunité de travailler avec eux sur ce projet depuis plus de 3 ans : Martine, merci pour tes conseils, ton expertise scientifique et ton attention du détail, et Hervé, pour ta disponibilité et ta confiance, ton flot de nouvelles idées, et ta capacité à me pousser vers l'avant à chaque obstacle rencontré sur la route.

Je souhaiterais également remercier chaleureusement les membres du jury, Kerstin Bystricky, Hélène Jammes, Pierre Savatier, Michael Weber et Franck Pellestor d'avoir accepté de participer au grand final de cette aventure, ainsi que les membres de mon comité de thèse, Daniel Vaiman, Bertrand Pain et Louis Casteilla, pour leurs conseils avisés sur le déroulement de ce projet.

Cette thèse a été financée par le Département de Génétique Animale de l'INRA ainsi que par la Direction Générale de l'Enseignement et de la Recherche : je remercie donc Denis Milan, Alain Milon, Christine Citti et Alain Ducos pour la confiance qu'ils ont accordée à ce projet de recherche.

Je voudrais remercier les membres des différentes plateformes techniques qui ont contribué à ce travail (Contrôle chromosomique, Get-PlaGE, imagerie-TRI, GeT-Biopuces), ainsi que Marielle Afanassieff, Yann Tapponier et Pierre Osteil du SBRI, Stéphane Ferchaud de l'UE1372 GeneESI, et Olivier Féraud et Ali Thurán de l'unité U935 pour leur collaboration.

Ce travail n'aurait pas pu voir le jour sans la contribution précieuse de nombreuses personnes du laboratoire : Chantal pour son travail de fourmi et ses cachettes d'écureuil, Florence V pour sa réactivité et son efficacité, Harmonie pour les centaines de métaphases, Florence M et Alain P pour leurs conseils techniques, David et Yann pour avoir essayé de m'inculquer quelques bases de biostatistiques, Kamila pour son engagement rapide sur ce projet dès son arrivée dans l'équipe, Manu pour la pêche aux morceaux de cochon dans l'azote, et Sylvain pour ses envolées lyriques sur tableau blanc.

Sur un plan plus personnel, je remercie l'ensemble des membres du laboratoire que j'ai croisé depuis 2011 pour leur accueil, leur soutien et leurs sourires, et en particulier l'équipe des GOs qui font du bâtiment (LG)C un petit paradis pour thésards. A Claire, Laure F, Laure G, Manu, Maria, Maguy, Sophie, Pitou, Emmanuelle P, Manuëla, Sarah, Maria-Inès, Valentin et tous les autres, merci!

Enfin, un grand merci à mes parents et aux membres de ma famille et de ma belle-famille qui m'ont soutenue, avec une pensée pour les piles de magazines « La Recherche » de mon père qui m'a transmis sa passion du monde scientifique. Merci également à mes ami(e)s pour leurs encouragements, les déjeuners mensuels, les soirées filles etc... et pour finir, à Clément, mon magicien du quotidien, de m'avoir supporté (dans tous les sens du terme !) et d'être notre meilleure moitié.

TABLE DES MATIERES

REMERCIEMENTS.....	3
TABLE DES MATIERES.....	5
TABLE DES FIGURES	15
TABLE DES TABLEAUX.....	21
RESUME	25
ABSTRACT	27
INTRODUCTION.....	29
DONNEES BIBLIOGRAPHIQUES.....	35
I. L'infertilité : rôles des altérations structurales et épigénétiques au cours du développement des cellules germinales	37
1. Généralités.....	37
2. Infertilité et remaniements chromosomiques	39
a. Effets des remaniements chromosomiques sur la méiose.....	39
b. Cas d'infertilité en présence de remaniements chromosomiques chez le porc.....	43
3. Infertilité et mécanismes épigénétiques.....	45
a. Méthylation et déméthylation de l'ADN chez les mammifères	45
b. L'empreinte parentale, un processus régulé par la méthylation de l'ADN	49
c. Variations épigénétiques au cours de l'embryogénèse.....	51
i. Reprogrammation épigénétique post-fertilisation	53
ii. Reprogrammation de l'épigénome dans les cellules germinales primordiales.....	53
iii. Modifications épigénétiques au cours de la gamétogénèse mâle	57
iv. Comportement spécifique des gènes soumis à empreinte	57
d. Cas d'infertilité en présence de défauts épigénétiques	59
i. Variations de la méthylation de l'ADN dans les gènes soumis à empreinte	61
ii. Des variations sur l'ensemble du génome	63
iii. Modifications chromatinienne impliquées dans l'altération de la gamétogénèse	65
4. Effets des remaniements chromosomiques sur l'épigénome	65
5. Techniques d'études de la méthylation de l'ADN.....	67
II. La reprogrammation cellulaire.....	71
1. La pluripotence dans l'embryon : dérivation de lignées ESCs et EpiSCs	71
a. Caractéristiques des cellules souches embryonnaires	71
i. Etablissement de lignées de cellules souches embryonnaires.....	71
ii. Caractéristiques morphologiques et moléculaires des cellules pluripotentes	71
iii. Capacités fonctionnelles des mESCs	73
b. Deux états de pluripotence définis : amorcé et naïf	75

i.	Caractéristiques des lignées mEpiSCs et comparaison avec les mESC	75
ii.	Plasticité entre les deux états de pluripotence	75
iii.	Caractéristiques des cellules souches embryonnaires humaines	77
2.	Développement des cellules pluripotentes induites par reprogrammation cellulaire	77
a.	Les différentes méthodes de reprogrammation	77
i.	Du transfert de noyau à la reprogrammation cellulaire	77
ii.	Evolution des techniques de reprogrammation cellulaire	79
b.	Le mécanisme de la reprogrammation cellulaire	83
i.	Un processus stochastique ou déterministe ?	83
ii.	Les étapes fondamentales de la reprogrammation cellulaire	85
c.	Un outil à fort potentiel	89
i.	De nombreuses applications thérapeutiques envisageables	89
ii.	Un exemple d'application : la différenciation germinale mâle	89
d.	Un outil à améliorer	91
i.	Instabilité génomique des lignées pluripotentes	91
ii.	Reprogrammation partielle des iPSCs	93
3.	Le challenge des lignées pluripotentes porcines	95
a.	La pluripotence chez le porc	95
i.	Faible succès de l'établissement de lignées ES porcines	95
ii.	Un développement embryonnaire précoce particulier	97
b.	Dérivation de lignées iPSCs porcines	101

MATERIELS ET METHODES 105

I.	Composition des solutions utilisées	107
1.	Solutions	107
2.	Milieux de cultures	107
II.	Etude de la méthylation de l'ADN spermatique porcine	109
1.	Choix des animaux	109
2.	Extraction d'ADN à partir de semence	109
3.	Mesure du niveau global de méthylation de l'ADN : LUMA	109
4.	Immunoprécipitation d'ADN méthylé (MeDIP) et hydroxyméthylé (hMeDIP)	111
5.	Séquençage haut-débit de l'ADN méthylé et hydroxyméthylé d'animaux contrôles	111
a.	Préparation des bibliothèques et séquençage haut débit	111
b.	Analyse des résultats	113
i.	Alignement des lectures sur le génome porcine	113
ii.	Détection des pics de méthylation	113
6.	Analyse locale de la méthylation des animaux à qualité de semence dégradée	113
a.	PCR quantitative sur les fractions d'ADN méthylé et témoin	113
i.	Choix des gènes et PCR quantitative	113
ii.	Exploitation des résultats	115
b.	Conversion bisulfite et pyroséquençage	115
i.	Dessin des amorces	115
ii.	Conversion bisulfite de l'ADN et amplification par PCR	115
iii.	Pyroséquençage	117
c.	Recherche d'empreinte parentale chez le porc	117
d.	Expression des transcrits du locus GNAS dans le testicule	119

III. Dérivation et caractérisation de lignées pluripotentes induites porcines	119
1. Choix des animaux	119
2. Reprogrammation cellulaire	119
a. Culture de fibroblastes	119
b. Systèmes de reprogrammation cellulaire.....	119
i. Infection par le lentivirus PL-SIN-EOS-C(3+)-EIP.....	119
ii. Infection par les virus pMXs	121
iii. Système de reprogrammation non intégratif	121
3. Culture des lignées de cellules pluripotentes	121
4. Validation des caractères de pluripotence	123
a. Expression de la phosphatase alcaline	123
b. Immunocytochimie.....	123
c. Analyse du cycle cellulaire	125
d. Expression des gènes de pluripotence	125
i. Extraction d'ARN et transcription inverse.....	125
ii. PCR conventionnelle	125
iii. PCR quantitative.....	125
5. Différenciation des piPSCs	127
a. Différenciation en corps embryoïdes	127
b. Formation de tératomes.....	127
6. Analyse transcriptomique	129
a. Hybridation sur la puce porcine Agilent 60K customisée	129
b. Exploitation des résultats	129
i. Analyse statistique	129
ii. Analyse fonctionnelle.....	129
7. Etude de l'instabilité génomique	131
a. Caryotypes en bandes G	131
b. Hybridation génomique comparative (CGH)	131
c. Hybridation in situ en fluorescence (FISH)	133

ETUDE DE LA METHYLATION DE L'ADN SPERMATIQUE PORCIN 137

I. Présentation des résultats	139
1. Un ADN spermatique fortement méthylé et hydroxyméthylé	139
2. Profil de méthylation dans le génome des cellules spermatiques.....	141
3. Focus sur les gènes soumis à empreinte parentale	143
a. Choix des gènes et recherche d'empreinte parentale chez le porc	143
b. Profil de méthylation des gènes soumis à empreinte conservé chez le porc.....	145
c. Etude locale ou globale : des résultats comparables	147
4. Etude de la méthylation de l'ADN spermatique chez des verrats à qualité de semence dégradée	149
a. Un niveau de méthylation global conservé chez les verrats infertiles	149
b. Variations spécifiques dans le locus GNAS	149
II. Conservation du profil de méthylation spermatique entre mammifères	151
1. Un ADN spermatique fortement méthylé à l'exception des régions promotrices	151
2. Profils de méthylation spécifiques des gènes soumis à empreinte	153
a. Evaluation du statut d'empreinte de 5 gènes chez le porc	155
b. Profil de méthylation des gènes soumis à empreinte conservé entre les espèces	155
i. Régions faiblement méthylées.....	157
ii. Régions fortement méthylées.....	157

3.	Différentes approches d'analyse de la méthylation de l'ADN	159
III.	Identification d'une région différentiellement méthylée dans le sperme de qualité dégradée	159
1.	Augmentation du niveau de méthylation dans le locus GNAS	159
a.	Description du locus complexe GNAS.....	159
b.	Une région liée à des maladies du développement et à la croissance	163
c.	Un profil de méthylation modifié chez certains animaux à spermogramme anormal.....	163
2.	Stabilité du niveau global de méthylation	165

DERIVATION DE LIGNEES IPS PORCINES A PARTIR D'ANIMAUX INFERTILES PORTEURS DE REMANIEMENTS CHROMOSOMIQUES169

I.	Bilan des différentes expériences de reprogrammation cellulaire	171
II.	Caractéristiques de pluripotence des clones transloqués I3 et I4 et de la lignée contrôle PB20	175
1.	Morphologie et cycle cellulaire ES-like	175
2.	Expression des gènes de la pluripotence	177
a.	Activation des gènes du réseau central de pluripotence	177
b.	Etude d'un ensemble de gènes de la pluripotence naïve et amorcée	179
3.	Des capacités de différenciation limitées	181
III.	Reprogrammation cellulaire par un système non intégratif	183
1.	Expression de nombreux gènes de pluripotence.....	183
2.	Formation de structures « blastocyst-like »	185
IV.	Instabilité génomique des piPSCs	185
1.	Accumulation d'anomalies chromosomiques au fil des passages	185
2.	Cas du système non intégratif	187
3.	Caractérisation des remaniements des lignées I3 et I4	187
a.	Apparitions de duplications dans le génome	187
b.	Caractérisation de la zone dupliquée du chromosome 4	189
c.	Expression des gènes de pluripotence entre passages précoces et tardifs.....	191
V.	Etude transcriptomique	191
1.	Comparaison des lignées reprogrammées	193
2.	Comparaison des passages précoces et tardifs	195
3.	Profil d'expression des corps embryoides	197
VI.	Discussion et perspectives	197
1.	Bilan moléculaire et fonctionnel	197
2.	Bilan caryotypique	205
3.	Perspectives	209

CONCLUSION GENERALE213

REFERENCES BIBLIOGRAPHIQUES217

ANNEXES243

ANNEXE 1 : Article publié dans <i>Biology of Reproduction</i>	245
ANNEXE 2 : Liste des réactifs et matériels utilisés	297
ANNEXE 3 : Tableau bilan des clones reprogrammés analysés	299

TABLE DES FIGURES

Figure 1 : Déroulement de la méiose mâle et femelle chez les mammifères.....	38
Figure 2 : Effets des mésappariements sur la méiose mâle.	40
Figure 3 : Ségrégation chromosomique des quadrivalents.....	40
Figure 4 : Evolution au cours du temps du nombre de porcs caryotypés et du nombre de remaniements identifiés par la plateforme de contrôle chromosomique de l'unité GenPhySE, UMR 1388.....	42
Figure 5 : Caryotype de l'individu porteur de la translocation réciproque t(13;17).....	42
Figure 6 : Caractéristiques de l'individu porteur de la translocation réciproque t(Y;14).....	44
Figure 7 : Mécanisme de maintenance de la méthylation de l'ADN chez les mammifères.	46
Figure 8 : Mécanismes de déméthylation active et passive.....	48
Figure 9 : Représentation du locus IGF2/H19.....	50
Figure 10 : Reprogrammations épigénétiques au cours du développement embryonnaire de la souris.	52
Figure 11 : Transition histone-protamine au cours de la spermiogénèse humaine.....	56
Figure 12 : Comportement des DMRs gamétiques stables et transitoires dans le remodelage épigénétique pré-implantatoire.....	56
Figure 13 : Héritabilité épigénétique transgénérationnelle.	58
Figure 14 : Analyse du profil de méthylation de H19 et MEST dans le sperme de patients normozoospermiques (NZ) ou sévèrement oligospermiques (OZ) par conversion bisulfite, clonage et séquençage.....	60
Figure 15 : A. Immunoprécipitation de l'ADN méthylé (MeDIP) par un anticorps spécifique des cytosines méthylées. B. Conversion des cytosines non méthylées en uraciles par sulphonation bisulfite HSO_3^- , déamination et désulphonation. C. Pyroséquençage.	66
Figure 16 : Evolution des marques bivalentes au cours de la différenciation des ESCs.	72
Figure 17 : Capacité de contribution au développement embryonnaire des cellules pluripotentes embryonnaires ou induites.	72
Figure 18 : Principales voies de signalisation régulant le maintien de la pluripotence naïve et amorcée (Bieberich and Wang, 2013).	74
Figure 19 : Les différents états de pluripotence dans les cellules souches embryonnaires murines.....	76
Figure 20 : Construction PL-SIN-EOS-C(3+)-EiP permettant la sélection des cellules reprogrammées par l'expression de l'eGFP et résistance à la puromycine.	80
Figure 21 : Le modèle stochastique et les différentes étapes de l'établissement de la pluripotence stable lors de la reprogrammation cellulaire.	84
Figure 22 : Incidence des modifications génomiques dans les cellules souches pluripotentes humaines au cours des passages.....	92
Figure 23 : Développement embryonnaire précoce chez le porc.	96

Figure 24 : A. Niveau de méthylation global de l'ADN spermatique dans 5 animaux à qualité de semence normale. B. Profil de méthylation de la séquence répétée LINE1 dans les verrats contrôles.....	138
Figure 25 : Pourcentage de couverture des lectures et des pics de méthylation sur chaque chromosome.	138
Figure 26 : Corrélation entre les réplicats techniques (répétitions du MeDIP et hMeDIP) et biologiques (animaux N1 et N2) établie par DiffBind.	140
Figure 27 : Pourcentage d'éléments annotés sur le génome porcin ayant une intersection avec les pics de méthylation.....	140
Figure 28 : Pourcentage des gènes impliqués dans la biologie des cellules germinales et du développement embryonnaire précoce ayant une intersection avec les pics de méthylation.....	142
Figure 29 : La méthylation locale de l'ADN spermatique est majoritairement conservée entre espèces et varie peu en fonction du statut de fertilité des animaux.	144
Figure 30 : Profil de méthylation de l'ADN spermatique porcin autour des gènes MEST, DAZL, IMPACT, RASGRF1, MEG3 et GNAS obtenu par MeDIP-Seq révélant la faible méthylation spécifique des TSS et des îlots CpGs (barres vertes) correspondants aux régions promotrices ainsi que la méthylation sexe-spécifique des DMRs (trait rouge) des gènes soumis à empreinte.	146
Figure 31 : Le niveau de méthylation global de l'ADN spermatique ne dépend pas de la qualité de la semence chez le porc.....	148
Figure 32 : Profil de méthylation du locus GNAS par conversion bisulfite et pyroséquençage.	150
Figure 33 : Niveau d'expression des transcrits du locus GNAS dans les testicules d'animaux à bonne ou mauvaise qualité de semence.....	150
Figure 34 : A. Représentation schématique du locus GNAS chez la souris. B. Représentation des différents isoformes de GSA et GNASXL	160
Figure 35 : A. Le profil obtenu en cytométrie de flux (FACS) et l'observation au microscope à fluorescence confirment l'expression de la GFP dans les lignées I3 et I4. B. Les courbes de croissance exponentielles des lignées I3 et I4 reflètent leur forte prolifération cellulaire. C. Morphologie iPS-like des lignées I3, I4 et PB20 en milieu piPS et en milieu LIF+2i.....	174
Figure 36 : Analyse du cycle cellulaire par FACS.....	174
Figure 37 : Expression des facteurs-clés de pluripotence dans les lignées I3 et I4.....	176
Figure 38 : Expression de nombreux gènes de pluripotence dans les lignées I3, I4 et PB20.....	178
Figure 39 : Faible capacité de différenciation des lignées I3 et I4.....	180
Figure 40 : Caractères de pluripotence des lignées NI.....	182
Figure 41 : Heatmap représentant les Δ Ct entre gènes cibles et gènes de références, de blanc, expression équivalente au gène de référence, à noir, gène non détecté : de nombreux gènes de pluripotence sont plus fortement exprimés dans les lignées issues du système non intégratif.	182
Figure 42 : A. La morphologie des EBs issus des lignées NI est similaire à celle des blastocystes porcins et ces structures expriment des marqueurs des territoires embryonnaires (SOX2, rouge) et extraembryonnaires (CDX2, vert) B et C. Evolution de l'expression d'un sous-ensemble de gènes lors de la formation de corps embryoides à partir des lignées NI13 et NI20... ..	184

Figure 43 : Evolution du caryotype des lignées reprogrammées au cours du temps.	186
Figure 44 : Caractérisation des remaniements majoritaires dans les lignées I3 et I4.	188
Figure 45 : Variation du niveau d'expression de gènes de pluripotence entre passages précoce et tardif.	188
Figure 46 : A. Analyse en composante principale montrant le fort regroupement des données par réplicats techniques (marqueurs) et l'absence d'effet de la lame d'hybridation (couleurs) ainsi que la grande variabilité de la condition F (triangle inversé) représentant la lignée fibroblastique contre les autres conditions. B. Diagramme de Venn des comparaisons des lignées I3 (A), I4 (C) et NI12 (G) par rapport à la lignée de fibroblastes (F) C. Heatmap de l'expression (log) de 5136 gènes présentant au moins un différentiel d'expression ($p < 0.01$, $lfc > 2$) entre les 7 conditions. D. Heatmap des gènes différentiels ($p < 0.01$, $lfc > 2$) entre la lignée NI12 et les EBs.	192
Figure 47 : Les quatre lignées reprogrammées I3, I4, PB20 et NI12 sur-expriment par rapport aux fibroblastes un ensemble de gènes impliqués dans les fonctions régulant le cycle cellulaire. Les couleurs représentent la p-valeur pour chaque procédé biologique	196
Figure 48 : Procédés biologiques (BP) les plus significativement enrichis ($FDR < 0.01$ et nombre de gènes > 10) dans la liste de gènes communément sous-régulés dans les quatre lignées reprogrammées par rapport aux fibroblastes.	198
Figure 49 : Principales fonctions biologiques (Ingenuity Knowledge Database) représentées dans les gènes communément sur- ou sous-régulés dans les quatre lignées reprogrammées par rapport aux fibroblastes.	200
Figure 50 : Représentation en nuage des 30 voies de signalisation (KEGG pathways) les plus significativement enrichies dans les gènes différentiellement exprimés entre la lignée NI12 et les lignées I3 et I4.	202
Figure 51 : Activation (vert) et inactivation (rouge) d'acteurs des voies de signalisation régulant la pluripotence des hESCs (Ingenuity Pathways Analysis) dans la lignée NI12 par rapport aux lignées I3 et I4.	202
Figure 52 : Principales fonctions biologiques (Ingenuity Knowledge Database) représentées dans les gènes communément sur- ou sous-régulés dans les corps embryoïdes par rapport aux cellules NI12. La taille des carrés est fonction de la p-valeur pour chaque fonction et la couleur est fonction du z-score indiquant une tendance à l'activation (rouge) ou à l'inactivation (bleu)	204
Figure 53 : Dérégulation de la voie LIF/JAK/STAT3 dans les corps embryoïdes de la lignée NI12 et activation de la voie BMP menant à la répression de NANOG (Ingenuity Pathway Analysis)	206
Figure 54 : Phénomène de chromothripsis.	208

TABLE DES TABLEAUX

Tableau 1 : Paramètres limites de qualité de semence humaine définis par l’OMS.....	36
Tableau 2 : Régulateurs épigénétiques chromatinien et leurs effets sur la méiose.....	64
Tableau 3 : Principales techniques de mesure de la méthylation de l’ADN.....	66
Tableau 4 : Caractéristiques des différentes expérimentations de reprogrammation cellulaire chez le porc.....	100
Tableau 5 : Caractéristiques des animaux étudiés normaux et à mauvaise qualité de semence.....	108
Tableau 6 : Amorces utilisées pour la PCR quantitative suivant le MeDIP.....	112
Tableau 7 : Amorces utilisées pour le pyroséquençage.....	114
Tableau 8 : Amorces d’amplification (F,R) et de séquençage (S) utilisées pour la détection de l’empreinte parentale.....	116
Tableau 9 : Couples d’amorces utilisés pour l’analyse de l’expression des transcrits GNAS dans le testicule porcin.....	118
Tableau 10 : Caractéristiques des animaux choisis pour la reprogrammation cellulaire.....	118
Tableau 11 : Amorces et paramètres de PCR pour la détection des marqueurs de pluripotence.....	124
Tableau 12 : Amorces et paramètres de PCR pour la détection des transgènes.....	124
Tableau 13 : Lignées cellulaires analysées par PCR quantitative.....	124
Tableau 14 : Amorces utilisées pour la PCR quantitative.....	126
Tableau 15 : Lignées cellulaires dans différents états de culture utilisées pour l’analyse transcriptomique.....	128
Tableau 16 : Mélange pour la réaction de marquage des ADNs.....	130
Tableau 17 : Marquage des ADNs.....	130
Tableau 18 : Marquage par peintures chromosomiques.....	132
Tableau 19 : Marquage par BAC.....	132
Tableau 20 : Filtre des lectures alignées sur le génome après le séquençage haut-débit.....	138
Tableau 21 : Détection de l’expression biallélique ou monoallélique dans des tissus fœtaux.....	142
Tableau 22 : Bilan des expériences de reprogrammation cellulaire par système intégratif (bleu) ou non intégratif (orange).....	170
Tableau 23 : Comparaison de l’expression de plusieurs gènes dans les lignées obtenues par système non intégratif par rapport aux lignées I.....	182
Tableau 24 : Gènes contenus dans la région dupliquée 111296543-116660879 du chr1 humain (Ensembl, Assemblage GRCh38).....	190
Tableau 25 : Nombre de sondes sur et sous-réglées pour chaque comparaison entre deux conditions.....	192

Tableau 26 : Annotations fonctionnelles les plus significativement enrichies dans les gènes sur-exprimés par les lignées reprogrammées (GeneCodis).	194
--	-----

RESUME

Les problèmes de fertilité s'appliquent à la fois à l'homme, où ils touchent une part croissante de la population, et aux espèces d'élevage, dont le porc où ils freinent la diffusion des caractères d'intérêt agronomique et représentent une perte économique importante pour les éleveurs. Dans ce projet, je me suis intéressée à deux mécanismes liés à l'infertilité : la dérégulation du profil de méthylation de l'ADN spermatique et la présence de remaniements chromosomiques.

La première partie de ma thèse a été consacrée à l'étude du profil de méthylation de l'ADN spermatique de verrats fertiles et infertiles. J'ai ainsi constaté que le niveau de méthylation de l'ADN spermatique est relativement conservé entre trois espèces de mammifères, au niveau global ainsi qu'au niveau de certains gènes, dont des gènes soumis à empreinte. L'étude de la méthylation de l'ADN de la semence à qualité dégradée a révélé que le niveau global de méthylation du sperme ne varie pas chez les animaux infertiles, et que la méthylation de la majorité des gènes étudiés n'est pas affectée. Cependant, j'ai mis en évidence une augmentation de la méthylation dans le locus complexe GNAS ainsi qu'une dérégulation de son expression chez certains animaux infertiles. Notre étude est la première à mettre en évidence un lien entre hyperméthylation du locus GNAS et infertilité mâle chez les mammifères.

Pour étudier ensuite l'impact des remaniements chromosomiques sur la fertilité des animaux, particulièrement dans le cas d'animaux stériles, j'ai choisi de produire des lignées de cellules souches pluripotentes induites (iPSCs) par reprogrammation cellulaire de fibroblastes issus de ces animaux, afin de pouvoir par la suite étudier leur différenciation dans la voie germinale. La caractérisation des lignées obtenues pour un verroat azoospermique porteur de la translocation réciproque t(Y;14) a révélé qu'elles possèdent de nombreuses caractéristiques des cellules pluripotentes telles que l'expression de gènes spécifiques et un cycle cellulaire proche de celui des cellules souches embryonnaires (ESCs). Bien qu'elles ressemblent plus à des lignées iPS humaines, caractéristiques de l'état de pluripotence « amorcé », elles sont capables de se maintenir en culture dans le milieu spécifique de la pluripotence naïve des ES murines, laissant présager une capacité à évoluer vers l'état naïf de pluripotence. Toutefois, ces lignées possèdent une faible capacité de différenciation ainsi qu'une instabilité génomique augmentant avec les passages. J'ai ainsi mis en évidence que l'utilisation de stratégies intégratives pour produire des lignées iPS porcines augmente leur instabilité chromosomique, aussi bien dans des lignées issues d'animaux à caryotypes normaux que porteurs de remaniements chromosomiques. L'utilisation d'une autre technique de reprogrammation cellulaire (non intégrative) m'a permis de démontrer que les cellules iPS obtenues par cette méthode ne présentent pas cette instabilité. Leur caractérisation partielle laisse présager l'obtention de lignées plus stables et rassemblant plus de caractéristiques de pluripotence.

ABSTRACT

Fertility issues concern both humans, affecting a growing part of the population, and farm animals including pigs in which they slow down the diffusion of agronomical traits of interest and represent an economic loss for breeders. In this project, I focused on two mechanisms linked to infertility: alterations in DNA methylation of sperm cells and chromosomal rearrangements.

The first part of my thesis was dedicated to the study of DNA methylation in sperm cells of fertile or infertile boars. I observed that the methylation level of spermatid DNA is conserved between three mammalian species, both at the global level and at the local level of specific genomic regions, including some imprinted genes. The study of DNA methylation in low quality semen revealed that the global level of DNA methylation does not vary in infertile boars and that the majority of imprinted genes are not affected as well. Nevertheless I highlighted an increase in DNA methylation in the GNAS complex locus as well as a deregulation of its expression in some boars with low quality semen. This study is the first one to pinpoint a link between hypermethylation of the GNAS locus and male infertility in mammals.

To study the effect of chromosomal rearrangements on fertility, especially for sterile boars, I chose to produce induced pluripotent stem cell lines (iPSCs) derived from fibroblasts of these animals, as a tool for studying their differentiation towards the germ cell lineage. Characterisation of the cell lines derived from an azoospermic boar carrier of a reciprocal translocation $t(Y;14)$ revealed that these cell lines harbor several characteristics of pluripotency including the expression of specific genes and a cell cycle resembling the one of embryonic stem cells. Even though these cell lines look like human pluripotent stem cells which are characteristics of the primed state of pluripotency, they are able to proliferate in a naïve-specific culture medium, suggesting an ability to evolve into the naïve state. However these cell lines revealed a poor differentiation potential and a genomic instability increasing with passaging. Thus, I highlighted that the use of integrative strategies for cell reprogramming in pig iPSCs production increases their chromosomal instability in cell lines derived from both karyotypically normal and abnormal somatic cells. Nevertheless the use of a non-integrative technique allowed me to demonstrate that the cell lines obtained with this method did not harbor this instability. Their preliminary characterization may be predictive of production of more stable cell lines gathering more characters of pluripotency.

INTRODUCTION

La consommation de viande dans le monde a considérablement augmenté dans la seconde partie du XX^{ème} siècle, atteignant 42.2 kg/habitant/an en 2011. Représentant près de 37% de cette consommation, la viande de porc était la plus fortement consommée dans le monde en 2011, en particulier en Europe, en Chine et aux Etats-Unis. Si cette consommation tend à se stabiliser, tandis que la consommation de volaille est en pleine expansion, le porc reste un animal d'intérêt agronomique majeur, y compris en France où la production annuelle dépasse les 2 millions de tonnes. Accompagnant l'industrialisation de la production porcine dans les années 60, la sélection génétique a permis d'améliorer les performances des animaux en termes de croissance, prolificité, qualité de la viande, qualité de la carcasse, et ces dernières années, de robustesse. La diffusion de ces caractères d'intérêts agronomiques passant par la sélection et la diffusion de verrats reproducteurs, l'augmentation du nombre d'animaux infertiles parmi cette population est problématique. En effet, la mauvaise qualité de semence est une des causes principales de réforme des animaux dans les Centres d'Insémination Artificielle (33%).

Chez l'homme, l'infertilité est un problème de santé publique qui touche environ 15% des couples souhaitant fonder une famille. Dans ce contexte, l'étude de l'infertilité chez le porc permet de comparer les mécanismes menant à la formation de gamètes fonctionnels entre les deux espèces et de développer de nouvelles hypothèses quant à l'infertilité humaine. En effet, le porc est un animal utilisé comme modèle biomédical pour diverses pathologies (pathologies cardiaques, diabète, mélanome...), car il possède de nombreuses similitudes avec l'homme sur les plans anatomiques, physiologiques et génétiques, comme l'a confirmé le séquençage complet de son génome en 2012. Les problèmes de fertilité peuvent être d'origines diverses - génétique, hormonale, environnementale - mais à l'heure actuelle de nombreux cas restent inexplicables. Parmi les causes génétiques, la présence de remaniements chromosomiques chez l'homme ou le porc peut interférer avec le déroulement de la méiose et empêcher ainsi la production de gamètes de bonne qualité. Ces dernières années, plusieurs études ont également corrélé la qualité de la semence humaine avec des altérations du profil de méthylation de l'ADN spermatique, en particulier au niveau des gènes soumis à empreinte parentale. Le développement embryonnaire précoce et la biologie des cellules germinales étant soumis à de nombreuses variations de l'épigénome, qui permettent entre autres d'effacer l'empreinte parentale dans les cellules germinales primordiales puis de la rétablir dans les gamètes en fonction du sexe de l'embryon, une dérégulation de cette reprogrammation épigénétique est une hypothèse forte dans l'obtention de gamètes déficients.

L'étude de l'expression génique et de la méthylation de l'ADN dans les cellules germinales porcines à différents stades de différenciation, chez des animaux infertiles porteurs ou non de remaniements chromosomiques, permettrait de répondre à de nombreuses questions : quel est l'effet des remaniements chromosomiques sur le développement précoce des cellules germinales ? Des variations de la méthylation de l'ADN apparaissent-elles chez les animaux infertiles ? Les remaniements chromosomiques impactent-ils la reprogrammation épigénétique des cellules germinales ?

Le méthylome spermatique porcine étant à ce jour très peu décrit, l'établissement du profil de méthylation de l'ADN issu de la semence d'animaux fertiles et normaux a été la première étape de ce projet. Ensuite, une étude comparative du profil de méthylation de l'ADN d'animaux fertiles et d'animaux à qualité de semence dégradée a été menée en particulier au niveau des gènes soumis à empreinte, afin d'identifier des marqueurs épigénétiques caractéristiques des altérations de la qualité de semence.

Cette première étude se focalisant sur les spermatozoïdes issus d'animaux adultes, elle exclut les individus azoospermiques et ne permet pas d'évaluer la dynamique des modifications épigénétiques dans les cellules germinales. De plus, les animaux étant stériles et ayant été sacrifiés, l'accès aux ressources biologiques est restreinte. Pour dépasser ces limitations, la seconde partie du projet a porté sur l'établissement d'une librairie de cellules souches pluripotentes dérivées d'animaux infertiles et porteurs de remaniements, l'objectif étant de permettre un accès continu à du matériel biologique de ces animaux pour tout type d'étude, et en particulier pour le développement d'un système *in vitro* de différenciation des cellules souches en cellules germinales. Pour cela, le second objectif de ce projet de thèse a été la production et la caractérisation de lignées de cellules souches pluripotentes induites porcines (iPSCs) à partir d'animaux infertiles porteurs de remaniements chromosomiques. En effet, les cellules iPS, de par leur capacité à s'auto-renouveler et à se différencier dans les trois feuillets embryonnaires, possèdent un éventail d'applications potentielles très large et leur développement est un challenge pour les espèces d'élevage. Dans ce contexte, notre modèle a également permis d'étudier l'effet des remaniements chromosomiques sur la reprogrammation cellulaire chez le porc, et de mettre en évidence, par une analyse cytogénétique détaillée, leur instabilité chromosomique ou leur propension à accumuler des anomalies du génome.

Le présent manuscrit se découpe en 4 parties. La première présente une étude bibliographique sur le rôle des remaniements chromosomiques sur la fertilité, puis sur les variations épigénétiques survenant au cours du développement embryonnaire précoce et dans les cellules germinales, et sur les corrélations existantes entre profils de méthylation et qualité de la semence masculine. Elle se termine par une présentation de l'état de l'art concernant le développement de cellules souches pluripotentes, induites ou embryonnaires, chez la souris, l'homme et le porc. Les méthodes utilisées tout au long de ce projet sont regroupées dans la seconde partie. La troisième partie permet de décrire et discuter les

résultats obtenus par l'étude du méthylome spermatique porcin et par la comparaison du niveau de méthylation de gènes particuliers entre verrats fertiles et infertiles ayant mené à l'identification d'un locus différenciellement méthylé chez les animaux infertiles (article publié en Annexe 1). La quatrième partie porte sur la mise en place d'un protocole de reprogrammation cellulaire à partir de fibroblastes d'animaux porteurs de remaniements chromosomiques et infertiles, ayant permis l'obtention de lignées par deux systèmes de reprogrammation différents. Leur pluripotence a été analysée ainsi que leur stabilité génomique au cours du temps. Dans ces deux dernières parties thématiques, les résultats sont présentés puis discutés et les perspectives qui en découlent sont présentées. Enfin, les principaux résultats de ce projet de thèse sont récapitulés dans une conclusion générale clôturant ce mémoire.

DONNEES

BIBLIOGRAPHIQUES

Tableau 1 : Paramètres limites de qualité de semence humaine définis par l'OMS.

Paramètre	Valeur limite	Anomalie associée
Volume de semence	< 1.5 mL	Hypospermie
	> 6 mL	Hyperspermie
Nombre de spermatozoïdes par éjaculat	< $39 \cdot 10^6$	
Concentration	= 0 /mL	Azoospermie
	< $15 \cdot 10^6$ /mL	Oligospermie
	> $200 \cdot 10^6$ /mL	Polyspermie
Mobilité totale	< 40%	Asthénospermie
Mobilité progressive	< 32%	
Morphologie normale	< 4%	Tératospermie
Vitalité	< 58%	

Cette étude bibliographique se découpe en deux parties. Dans un premier temps, nous dresserons un état des lieux de l'impact des remaniements chromosomiques sur la biologie des cellules germinales, puis nous nous intéresserons aux fluctuations du profil de méthylation de l'ADN dans les cellules germinales, depuis la fertilisation jusqu'à l'obtention de gamètes fonctionnels, et aux altérations connues de ce profil chez des patients à spermogramme anormal. Dans un second temps, nous nous pencherons sur la biologie des cellules souches pluripotentes, avec un accent sur l'obtention de cellules pluripotentes induites par reprogrammation de cellules somatiques chez l'homme, la souris et le porc.

I. L'infertilité : rôles des altérations structurales et épigénétiques au cours du développement des cellules germinales

1. Généralités

L'infertilité est fréquemment définie chez l'homme comme l'incapacité à concevoir après un à deux ans de rapports non protégés, que ce soit par incapacité à déclencher une grossesse ou à la mener à terme. L'infertilité est un état difficile à quantifier, car sa mesure se base entre autres sur des critères subjectifs tels que le délai nécessaire à la conception et les difficultés ressenties par les couples (Spira et al., 2012). Néanmoins, on estime aujourd'hui à 1 sur 6 le nombre de couples rencontrant des difficultés à concevoir (Shah et al., 2003). Les causes de l'infertilité sont nombreuses et variées (génétiques, hormonales, immunologiques, psychologiques, liées à l'âge, au style de vie, à un traumatisme, une maladie ou une intervention chirurgicale ou associées à un défaut des gamètes) mais beaucoup de cas sont encore inexplicables.

Chez le porc, on constate que la principale cause de réforme des reproducteurs est la mauvaise qualité de semence (33%) (Ferchaud et al., 2009; Martinat-Botté et al., 2009). Ce chiffre élevé représente un enjeu économique pour les centres d'insémination et les éleveurs, ainsi qu'un frein pour la diffusion des caractères et l'adaptation des espèces agronomiques aux enjeux de l'agriculture mondiale. Chez l'homme, ce sont les normes de l'Organisation Mondiale de la Santé qui définissent les valeurs minimales ou maximales caractérisant une semence masculine de bonne qualité (WHO, 2010). Les valeurs de certains de ces paramètres sont résumées dans le Tableau 1. Chez le porc, la qualité de la semence est évaluée en mesurant le volume de l'éjaculat (en moyenne 200mL), en examinant sa couleur (présence de sang), en déterminant le nombre de spermatozoïdes anormaux, en évaluant la motilité des spermatozoïdes et en calculant la concentration en spermatozoïdes. En 1997, Althouse et collègues ont défini comme valeurs limites un nombre minimal de spermatozoïdes par éjaculat de 15.10^9 , un taux de motilité minimal de 70% (ou 3 sur l'échelle de Bishop) et un nombre maximum de spermatozoïdes anormaux de 20% (Althouse, 1997).

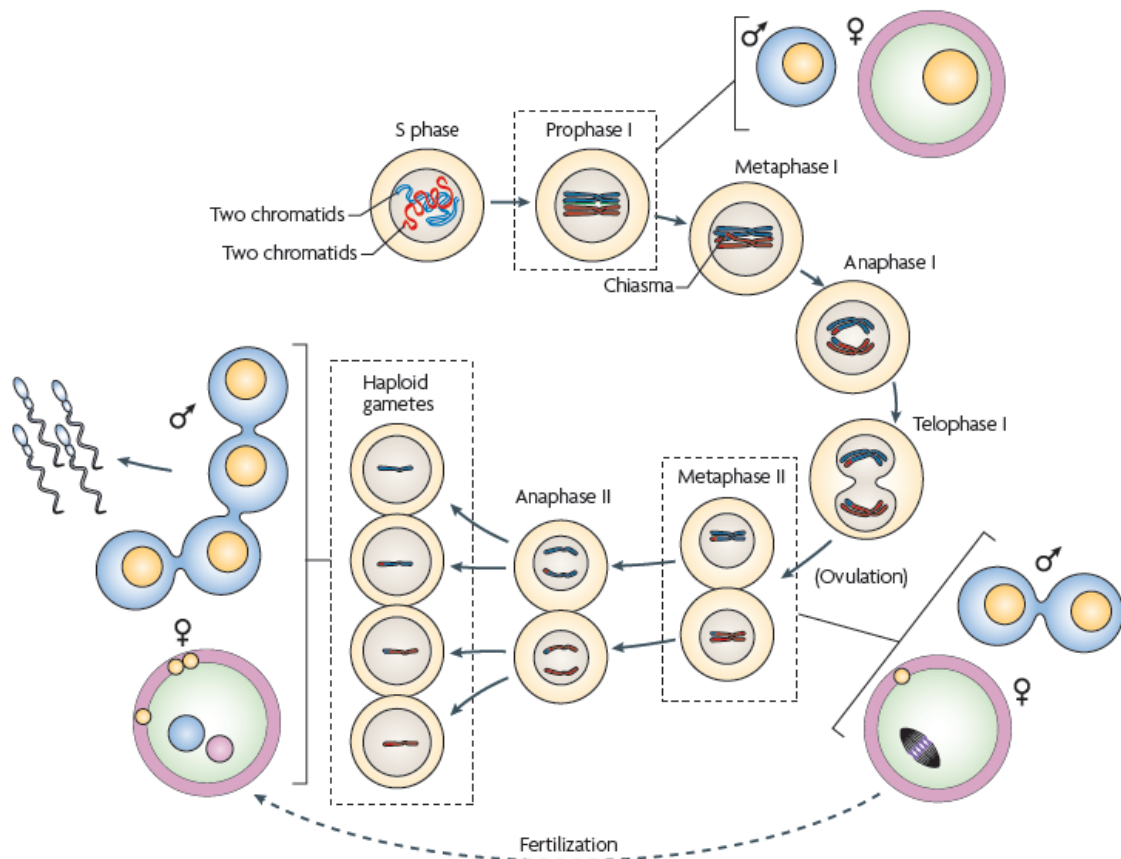


Figure 1 : Déroulement de la méiose mâle et femelle chez les mammifères (Handel and Schimenti, 2010). La première division méiotique, réductionnelle, permet le brassage génétique via la formation d'enjambements. La seconde division, équationnelle, permet la production de cellules haploïdes. Chez le mâle, la méiose se déroule dans le testicule à partir de la puberté. Chez la femelle, la première division de méiose commence dans l'embryon puis est bloquée en prophase I jusqu'à la puberté. Le processus redémarre à l'ovulation, ou une seule des quatre cellules filles haploïdes devient un ovocyte fonctionnel, et s'achève lors de la fécondation.

Sur le plan génétique, plusieurs causes liées à l'infertilité ou à la sous fertilité ont été décrites dans la littérature. Des centaines de gènes candidats, identifiés parfois sur la base de quelques cas seulement, ont été validés sur le modèle murin et classés en fonction du phénotype d'infertilité associé au gène muté (Matzuk and Lamb, 2008). On peut ainsi observer la présence de mutations dans des gènes impliqués dans la formation des gonades tels que *Sox9*, *Cftr*, *Dax1* ou *Wtn4* (Shah et al., 2003) ou dans les gènes participant à la méiose et aux mécanismes qui y sont rattachés comme la formation de cassures de l'ADN double-brin, la recombinaison homologue et la réparation de l'ADN, l'appariement des chromatides, la formation du complexe synaptonemal ou encore la ségrégation des chromosomes (*Spo11*, *Rad51c*, *Smc1b*, *Rec8*, *Stag3*, *Scyp1-2-3*, *Hspa2*, *Cdk2*, *Aurka*) (Handel and Schimenti, 2010). Un défaut lymphocytaire peut également jouer sur la qualité de la semence via l'augmentation du stress oxydant dans les spermatozoïdes. Enfin, une dérégulation du nombre de cellules germinales primordiales (PGCs) supportées par les cellules de Sertoli par apoptose sélective peut entraîner soit une sur-apoptose des cellules germinales, soit une sous-apoptose et la malformation des gamètes (Ioannou and Griffin, 2011).

2. Infertilité et remaniements chromosomiques

a. Effets des remaniements chromosomiques sur la méiose

Les remaniements chromosomiques sont également une cause possible de l'infertilité. Ils consistent en des échanges de matériel entre deux chromosomes (translocations) ou des délétions, duplications ou inversion de matériel génétique au sein d'un même chromosome. Une étude a montré que l'on observe de 4 à 10 fois plus de translocations autosomales chez les hommes infertiles que dans la population normale (O'Flynn O'Brien et al., 2010). Parmi les cas d'aneuploïdies constitutionnelles liées à l'infertilité, les plus connues sont le syndrome de Klinefelter (47,XXY), qui peut exister sous forme mosaïque chez des hommes fertiles, le syndrome de Turner (45,X) pour lequel des gènes nécessaires au bon développement des gonades sont manquants, et le syndrome de Down, aussi appelé trisomie 21, lié à une moindre prolifération des cellules germinales primordiales. Les microdélétions de l'Y sont aussi un phénotype courant. Elles peuvent apparaître dans trois zones du chromosome Y (AZFa, AZFb et AZFc) où elles touchent des gènes nécessaires au bon fonctionnement de la voie germinale, notamment la famille DAZ (Shah et al., 2003; Ioannou and Griffin, 2011).

Les remaniements chromosomiques peuvent avoir un effet sur le bon déroulement de la méiose. La méiose est un processus complexe (Figure 1) qui traverse plusieurs étapes clés qui permettent de réguler son bon fonctionnement. Les recombinaisons, qui garantissent le brassage génétique, sont assurées par la formation de cassures de l'ADN double brin (DSB). Ces cassures sont catalysées par la protéine SPO11, notamment au niveau de points chauds de recombinaison identifiables par la présence

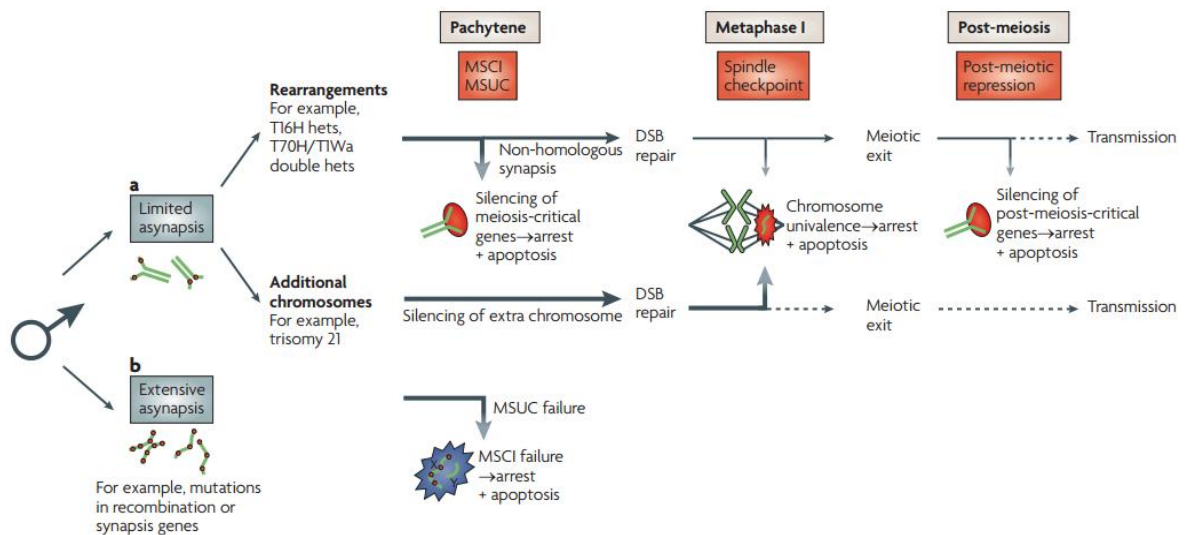


Figure 2 : Effets des mésappariements sur la méiose mâle (Burgoyne et al., 2009). **a.** Le mésappariement peut interférer avec les phénomènes de MSUC (Meiotic Silencing of Unsynapsed Chromatin) et MSCI (Meiotic Sex Chromosome Inactivation) en retenant les protéines impliquées au niveau de la chromatide inappariée, tout en risquant de réprimer des gènes essentiels sur les chromosomes remaniés ou d'empêcher la réparation des cassures à l'ADN, et mener à l'arrêt de la méiose au stade pachytène en fin de prophase. Un appariement non-homologue permet le franchissement de ce point de contrôle mais peut générer de l'univalence lors de l'assemblage des chromosomes sur le fuseau, entraînant l'apoptose des cellules. Lorsque toutes les barrières sont franchies, le remaniement est transmis à la génération suivante. **b.** Un mésappariement trop étendu retient les protéines impliquées dans le MSUC et le MSCI au niveau des chromatides inappariées et mène à l'apoptose des cellules au stade pachytène.

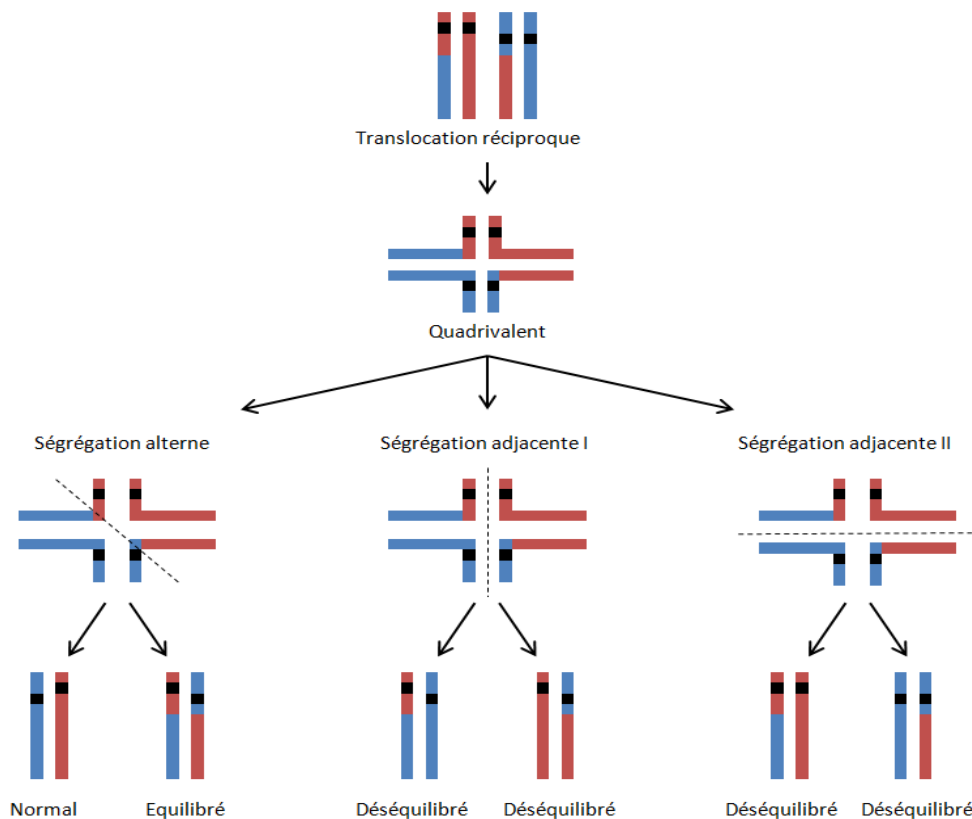


Figure 3 : Ségrégation chromosomique des quadrivalents. En présence d'une translocation réciproque, l'appariement des chromosomes peut mener à la formation de quadrivalents. Lors de la ségrégation des chromosomes, ces quadrivalents peuvent se séparer suivant plusieurs axes. La ségrégation alterne donne soit une paire normale de chromosomes, soit une paire anormale de chromosomes mais équilibrée, tandis que la ségrégation adjacente donne des paires de chromosomes déséquilibrées, présentant des duplications ou des délétions de matériel génétique.

de la marque épigénétique H3K4me3. Lors de l'appariement des chromatides sœurs, les cassures sont réparées par l'invasion de la chromatide homologue, en présence de RAD51 et DMC1. Au stade pachytène de la prophase I, un point de contrôle permet de s'assurer que les cassures de l'ADN ont bien été réparées avant de poursuivre le processus méiotique (Handel and Schimenti, 2010). La présence de remaniements chromosomiques peut affecter le déroulement de la méiose en générant des mésappariements ou des inappariements (Figure 2). Dans ces cas là, les protéines impliquées dans la détection et de la réparation des DSBs telles que γ H2AX et ATR sont retenues au niveau de la chromatide inappariée via le processus d'inactivation de la chromatine inappariée (MSUC) (Burgoyne et al., 2009), ce qui ralentit le processus de réparation. Lorsque le nombre de DSBs non réparés est trop grand, les cellules entrent en apoptose au niveau du point de contrôle du stade pachytène. Une autre étape fondamentale de la méiose est l'inactivation des chromosomes sexuels (MSCI) par les mêmes protéines γ H2AX et ATR. De la même manière que précédemment, si ces protéines sont déjà retenues sur une région inappariée, l'inactivation des chromosomes sexuels est défectueuse et les cellules entrent également en apoptose au stade pachytène. Outre la conservation de DSBs, l'inactivation de la chromatine inappariée générée par les remaniements chromosomiques pourrait inactiver des gènes essentiels pour la survie cellulaire ou le bon déroulement de la méiose et donc conduire les cellules en apoptose. L'appariement non-homologue permet de franchir malgré tout ce point de contrôle mais la méiose peut encore s'arrêter notamment au niveau du point de contrôle du fuseau en métaphase I (Figure 2).

Les mésappariements ne sont pas les seuls problèmes causés par les remaniements chromosomiques. Les translocations réciproques, c'est-à-dire les échanges de matériel génétique entre deux chromosomes, peuvent mener à la formation de quadrivalent lors de l'étape d'appariement des chromatides sœurs. Le délai imposé par le traitement de telles structures peut générer du retard dans le processus méiotique, entraînant alors une apoptose cellulaire. En cas de translocation entre autosome et chromosome sexuel, l'inactivation des chromosomes sexuels peut entraîner une extinction non désirée de gènes sur l'autosome impliqué dans le quadrivalent. Si les quadrivalents sont gérés normalement, ils peuvent également mener à la formation de gamètes déséquilibrés (manque ou surplus de matériel génétique) non viables ou inefficaces pour la fécondation, dépendant du type de ségrégation des chromosomes (Figure 3). En cas d'inversion péri ou para-centrique dans un chromosome, on observe la formation d'une boucle d'appariement qui peut mener à une réduction des crossing-overs ainsi qu'à l'apparition de gamètes déséquilibrés (Shah et al., 2003). Enfin, la présence d'un chromosome surnuméraire ou une aneuploïdie peuvent être résolues un temps par la recombinaison non homologue, mais l'univalence n'est pas acceptée par le point de contrôle du fuseau en métaphase I et les cellules contenant des chromosomes univalents entrent en apoptose (Figure 2).

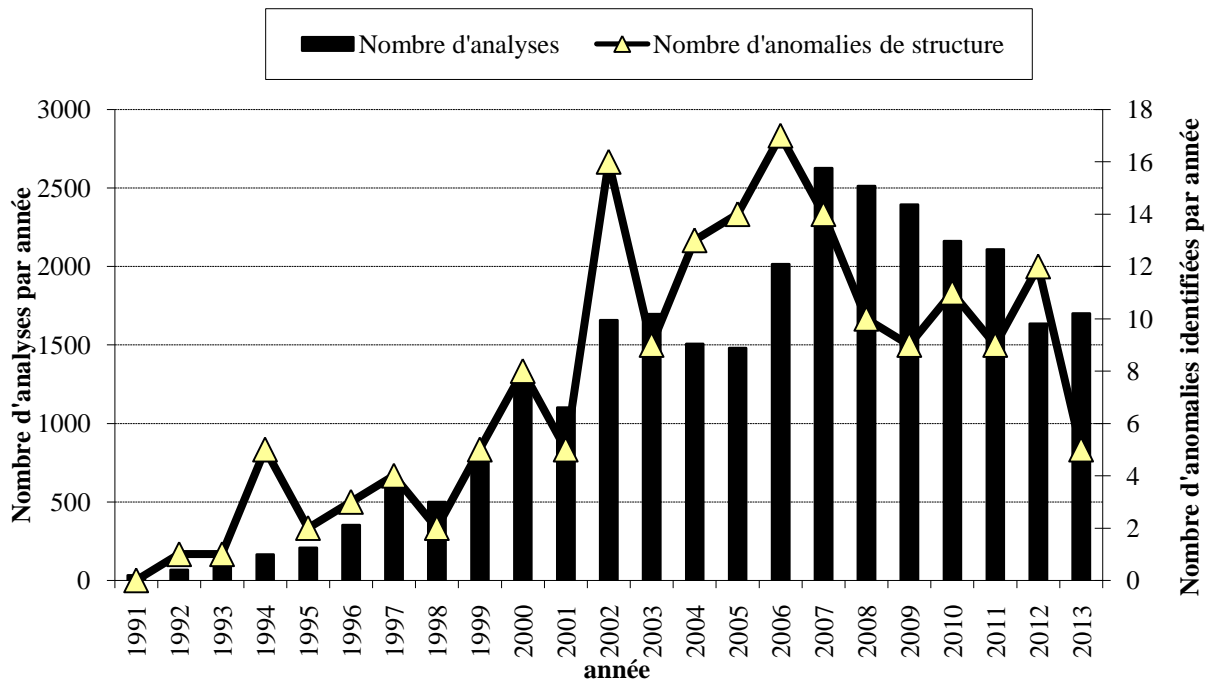


Figure 4 : Evolution au cours du temps du nombre de porcs caryotypés et du nombre de remaniements identifiés par la plateforme de contrôle chromosomique de l'unité GenPhySE, UMR 1388 ((Ducos et al., 2007) ; Alain Pinton, communication personnelle))

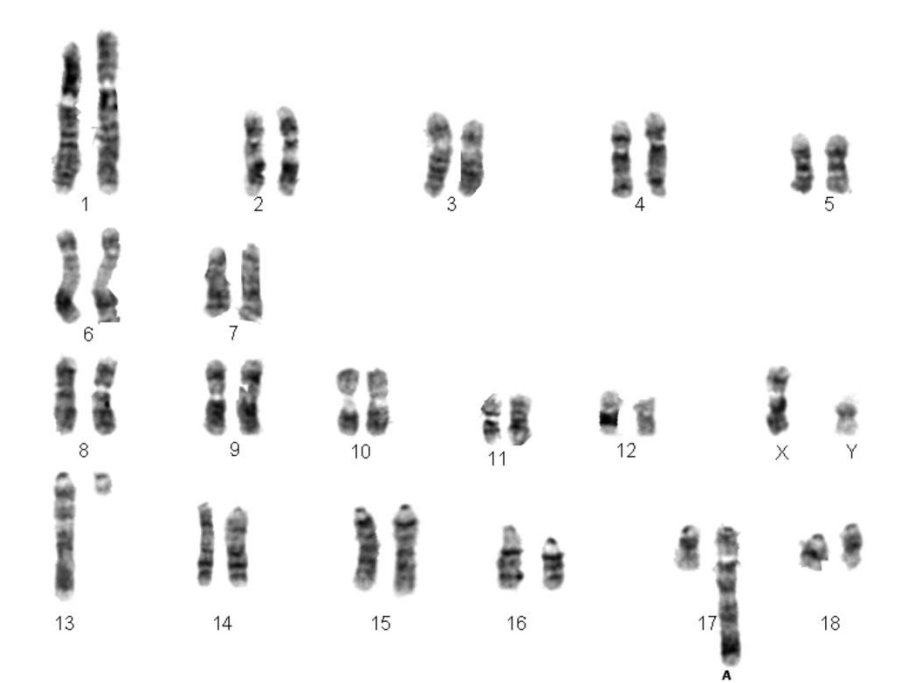


Figure 5 : Caryotype de l'individu porteur de la translocation réciproque $t(13;17)$ (Ducos et al., 1998b).

b. Cas d'infertilité en présence de remaniements chromosomiques chez le porc

Les premières études chromosomiques sur le porc domestique ont été réalisées au début du 20^{ème} siècle (Woodsdalek, 1913). Depuis, l'obtention du caryotype de référence dans les années 1980 (Ford et al., 1980; Gustavsson, 1988) et la mise en place du diagnostic des verrats par des plateformes de contrôle chromosomique ont permis d'identifier de nombreux cas de remaniements chromosomiques, principalement des translocations réciproques (Ducos et al., 1998a; Ducos et al., 1998b). Parmi ces remaniements, certains ont été associés à des phénotypes d'hypoprolificité (Popescu et al., 1986; Ducos et al., 1998a; Villagómez and Pinton, 2008). Des études plus poussées menées sur des animaux à performance reproductrice faible ont montré que 50% des animaux hypoprolifériques étaient porteurs de remaniements chromosomiques. Afin de limiter la diffusion de ces remaniements dans les populations animales d'élevage, un contrôle chromosomique systématique a été mis en place sur les verrats entrant en centre d'insémination artificielle. En France, la plateforme de cytogénétique animale, intégrée au laboratoire de Génétique, Physiologie et Système d'Élevage (GenPhySE, UMR 1388 INRA-ENVY-ENSAT) au sein de notre équipe de recherche Cytogene, réalise un grand nombre d'analyses sur les animaux français et européen, avec plus de 23000 individus analysés à ce jour (Figure 4).

Parmi les anomalies de structure identifiées, on retrouve majoritairement des translocations réciproques (86%) entre autosomes ou impliquant un chromosome sexuel, des inversions (10%), quelques translocation Robertsoniennes (ex : t(13;17)), et enfin quelques duplications. Des anomalies de nombres sont elles aussi détectées (ex : anomalie mosaïque 38,XY/39,XXY).

Parmi toutes ces anomalies, 40% sont liées à une baisse de la prolificité chez l'animal (ex : t(13;17), t(1;14), t(Y;14), mosaïc 38,XY/39,XXY). La translocation réciproque t(13;17)(q4.1;q1.1) (Figure 5) a par exemple été constatée sur un porc Large White x Piétrain hypoproliférique donnant des tailles de portée réduite de 8,1 porcelets par portée sur un total de 26 portées, soit une réduction de 34% par rapport à la prolificité moyenne (12.3 porcelets/portée).

La translocation réciproque t(Y;14) détectée chez un animal de race Duroc azoospermique (Ducos et al., 2007) a été étudiée plus en détails afin de mieux comprendre le phénotype de l'animal porteur (Figure 6a et 6b). Des analyses histologiques de coupes testiculaires ont montré l'arrêt complet de la spermatogénèse au stade des spermatocytes primaires, la présence de débris cellulaires dans la lumière des testicules et une hyperplasie des cellules de Leydig. L'analyse de l'appariement méiotique au stade pachytène a montré la présence d'un quadrivalent formé par les chromosomes X, 14, et les dérivés des chromosomes Y et 14 dans 84.7% des cas. L'immunolocalisation de γ H2AX a révélé l'accumulation de cette histone spécifique dans des fragments inappariés le long des chromosomes X et Y mais aussi sur les chromosomes 14 et dérivé 14 présents dans le quadrivalent, suggérant ainsi un phénomène de

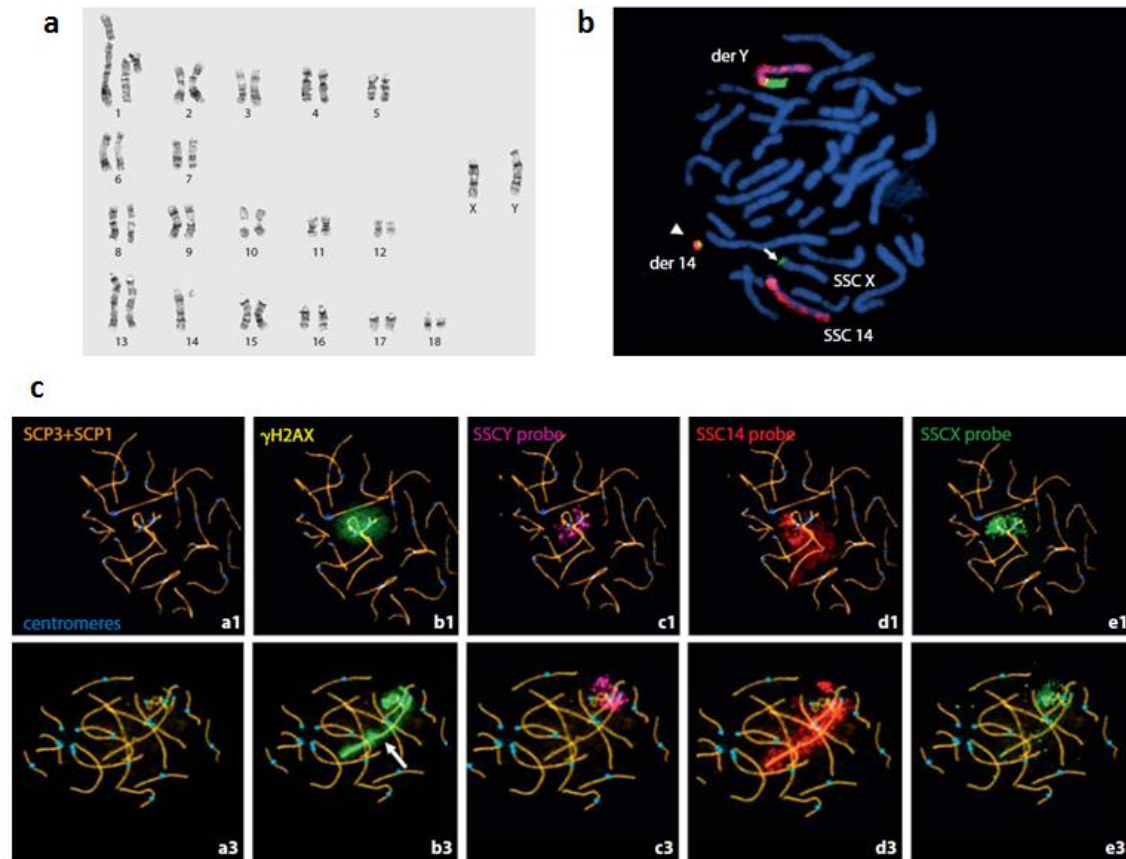


Figure 6 : Caractéristiques de l'individu porteur de la translocation réciproque $t(Y;14)$ adapté de (Pinton et al., 2008). **a.** Caryotype en bandes G présentant la translocation réciproque $t(Y;14)$. **b.** FISH permettant l'identification des chromosomes impliqués dans le remaniement et de leurs dérivés par peinture chromosomique des chromosomes 14 (rouge) et Y (vert). Flèche blanche : région pseudoautosomale du chromosome X marquée par la peinture du chromosome Y. **c.** Immunolocalisation de protéines du fuseau (SCP3 et SCP1) et des centromères couplée à l'immunolocalisation de γ H2AX. Les chromosomes Y, 14 et X sont visualisés après hybridation in situ (FISH) avec les sondes de peinture spécifiques de ces chromosomes. La combinaison entre immunolocalisation de protéines et FISH permet d'identifier les zones inactivées. Le chromosome 14 est ainsi localisé dans la zone d'inactivation des chromosomes sexuels, caractérisée par l'expression de γ H2AX (ligne 1), zone qui diffuse le long du bras du chromosome 14 dans certains cas (ligne 2).

répression transcriptionnelle inapproprié sur ces fragments (Figure 6c). Les gènes *CSK2* et *GAKIN* présents dans ces régions réprimées sont impliqués dans les processus méiotiques et pourraient expliquer l'arrêt de la spermatogénèse observé dans l'animal porteur de la translocation (Pinton et al., 2008).

Notre équipe a également mené en 2011 une analyse d'un porc porteur du remaniement mosaïque 38,XY/39,XXY (Pinton et al., 2011), mieux connu chez l'homme sous le nom de syndrome de Klinefelter (46,XY/47,XXY) et largement associé à l'infertilité masculine (Morris et al., 2008). Le mosaïcisme de cette anomalie a été confirmée bien que seulement 3 à 5% des cellules soient normales (38,XY) dans le sang ou les testicules. L'étude histologique des testicules a révélé une réduction du nombre de tubes séminifères ainsi qu'une spermatogénèse partielle, avec des spermatogonies et des spermatocytes I, et peu ou pas de spermatozoïdes, souvent mal formés (Pinton et al., 2011).

Enfin, l'analyse par FISH de la ségrégation méiotique chez des animaux porteurs d'inversions a permis d'estimer la proportion de gamètes déséquilibrés résultant de la méiose en présence d'inversion péri ou para-centrique. Cette proportion s'est avérée faible, allant de 0,62% à 4,12% selon les cas, mais significativement plus élevée que dans le cas d'animaux à caryotype normal (Massip et al., 2009). Contrairement aux résultats obtenus chez l'homme, la taille de l'inversion ne semble pas être corrélée au taux de gamètes déséquilibrés : par exemple, ce taux est faible à la fois chez le mâle (4%) et chez la femelle (3.7%) dans le cas d'une très grande inversion couvrant 76% du chromosome 4 (Massip et al., 2010).

Les remaniements chromosomiques représentent donc une source reconnue de dysfonctionnement au cours de la méiose, pouvant entraîner l'arrêt de la formation de gamètes ou l'obtention de gamètes de mauvaise qualité. D'autres phénomènes peuvent avoir un effet sur le développement précoce des cellules germinales puis sur la gamétogénèse. Dans la partie suivante, nous allons nous intéresser à l'établissement du profil épigénétique des cellules au cours du développement et à ses possibles effets sur l'obtention de gamètes fonctionnels.

3. Infertilité et mécanismes épigénétiques

a. Méthylation et déméthylation de l'ADN chez les mammifères

Les modifications épigénétiques permettent de moduler l'expression des gènes sans modification de la séquence d'ADN. Ces modifications sont caractérisées par la présence de groupements chimiques particuliers sur l'ADN ou sur les histones. Chez les mammifères, l'une des marques épigénétiques les plus étudiées est la méthylation de l'ADN au niveau du cinquième carbone des cytosines, notée 5mC.

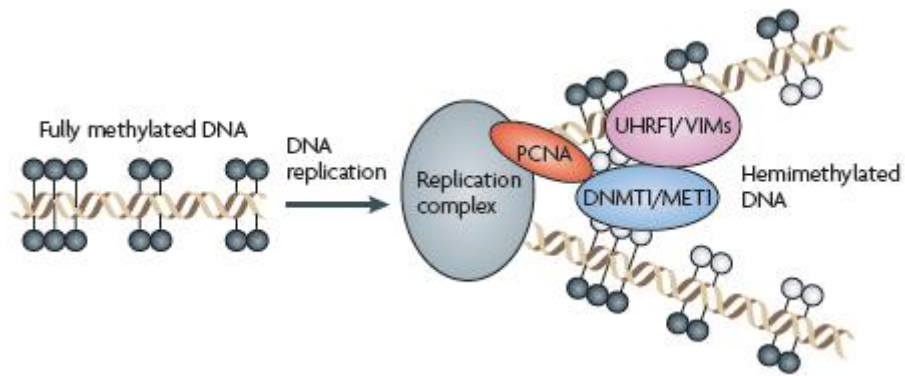


Figure 7 : Mécanisme de maintenance de la méthylation de l'ADN chez les mammifères (Law and Jacobsen, 2010). Au cours de la réplication de l'ADN, la fixation de UHRF1 sur l'ADN hemiméthylé permet le recrutement de DNMT1 et la méthylation du brin nouvellement synthétisé.

Elle est présente majoritairement lorsque cette cytosine est suivie d'une guanine c'est-à-dire dans un contexte nucléotidique NCGN ou N peut être n'importe quel nucléotide. Cette méthylation recouvre 70 à 80% des CpGs du génome des mammifères (Ehrlich et al., 1982), se situant dans les gènes, les transposons, les séquences répétées ou l'ADN intergénique (Suzuki and Bird, 2008). Les CpGs non méthylés sont regroupés dans des îlots à forte densité en CpG, que l'on appelle îlots CpG, et qui sont souvent associés aux promoteurs (Suzuki and Bird, 2008). La méthylation des promoteurs est souvent reliée à une répression du gène (Hammoud et al., 2014), et les marques de méthylation de l'ADN sont souvent associées aux marques d'histone H3K9me2 et H3K9me3 répressives pour la transcription (Saitou et al., 2012).

Le bon établissement du profil de méthylation du génome (ou méthylome) est primordial puisqu'un défaut de méthylation de l'ADN peut-être létal chez l'embryon (Law and Jacobsen, 2010). La méthylation de l'ADN est régulée chez les mammifères par différentes enzymes de la famille des méthyltransférases. DNMT3a, DNMT3b et DNMT3L sont responsables de la méthylation *de novo* de l'ADN tandis que DNMT1 permet la maintenance de la méthylation aux mêmes positions à chaque cycle cellulaire (Figure 7). DNMT3L interagit avec les queues d'histones H3K4 non méthylées et y recrute les enzymes DNMT3a et DNMT3b qui effectuent la méthylation de l'ADN. En effet, on observe une forte corrélation négative entre la méthylation des histones H3K4 et la méthylation de l'ADN (Law and Jacobsen, 2010). De plus, le bon fonctionnement de KDM1B, une lysine déméthylase qui supprime la méthylation sur les histones H3K4, est nécessaire pour l'établissement correct du profil de méthylation. DNMT1 quand à elle interagit avec UHFR1 qui s'associe à la chromatine pour le recrutement de la méthyltransférase aux foyers de réplication (Kagiwada et al., 2013).

Le processus de déméthylation peut être actif ou passif (Figure 8). Dans le cas d'une déméthylation passive, la perte de méthylation est progressive au fil des cycles cellulaires. Malgré l'expression continue de l'enzyme de maintenance DNMT1, la répression d'UHFR1 ne permet pas la bonne localisation de DNMT1 aux foyers de réplifications, et la méthylation de l'ADN n'est pas maintenue après la réplication de l'ADN (Kagiwada et al., 2013).

Il existe plusieurs hypothèses quand au processus de déméthylation active. La première est que le groupement méthyl serait retiré directement de la cytosine. La seconde hypothèse privilégie l'excision totale du nucléotide via l'action de glycosylases entraînant la déamination de la cytosine méthylée, créant ainsi un mésappariement T/G résolu par un système de réparation de l'ADN de type BER (Base Excision Repair). Enfin, la troisième hypothèse met en évidence le rôle des enzymes de la famille TET, qui transforment la cytosine méthylée en cytosine hydroxyméthylée par réaction oxydative. Cette cytosine hydroxyméthylée, notée 5hmC, peut à son tour se diluer au fil des réplifications car elle est moins spécifique de DNMT1. Elle peut également être convertie en 5hmU via le complexe

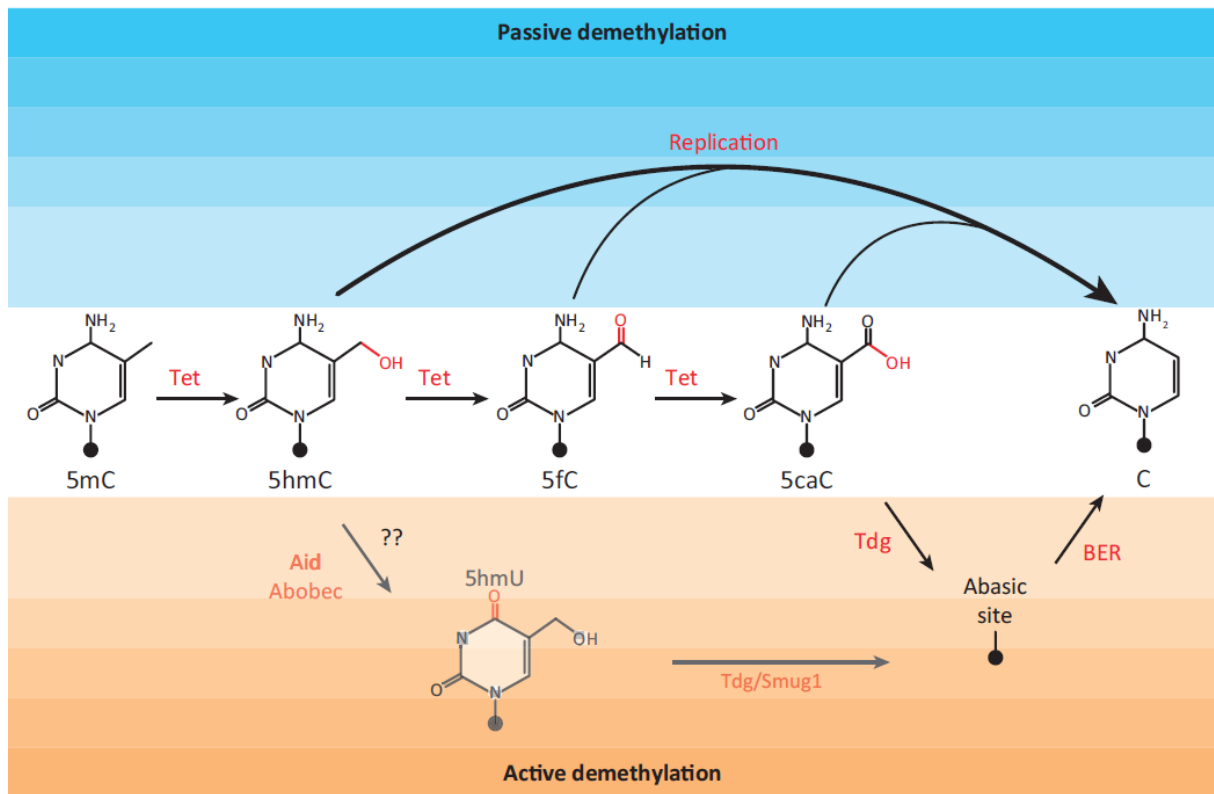


Figure 8 : Mécanismes de déméthylation active et passive (Piccolo and Fisher, 2014). Les réactions de conversion successives des cytosines méthylées, catalysées par les enzymes de la famille TET, génèrent des formes chimiques non reconnues par le système de maintenance de méthylation de l'ADN, qui peuvent se diluer au cours des réplifications de l'ADN. La forme finale, 5-carboxycytosine, et la 5-hydroxyméthyluracile obtenue par conversion de 5hmC par le complexe Aid/Apobec, servent de cibles pour les carboxylases comme TDG qui les excisent. La séquence est ensuite réparée via le système BER.

Aid/Apobec, constituant alors un bon candidat pour l'excision via les glycosylases. De plus, la réaction oxydative peut se poursuivre par les enzymes de la famille TET jusqu'à l'obtention de cytosines carboxylées 5caC qui sont aussi de bonnes cibles pour les glycosylases comme TDG (Piccolo and Fisher, 2014). L'hydroxyméthylation des 5mC est donc généralement considérée comme une étape intermédiaire du processus de déméthylation de l'ADN (Law and Jacobsen, 2010).

b. L'empreinte parentale, un processus régulé par la méthylation de l'ADN

L'expression monoallélique de certains gènes, dépendante de l'origine parentale des allèles, est appelé empreinte parentale (Reik and Walter, 2001). Ce phénomène a été découvert dès les années 1980s au cours d'études d'embryons uniparentaux ou porteurs de disomie uniparentale (McGrath and Solter, 1984; Surani et al., 1984; Cattanach and Kirk, 1985). On parle de gènes soumis à empreinte maternelle lorsque le gène est exprimé sur l'allèle maternel et de gène soumis à empreinte paternelle lorsque le gène est exprimé sur l'allèle paternel. Chez l'homme, plus de 100 gènes ont été identifiés comme soumis à empreinte, bien que des études récentes dans le cerveau indiquent que ce nombre pourrait être bien plus élevé (Gregg et al., 2010). Ces gènes soumis à empreinte sont généralement réunis en clusters autour d'une région chromosomique appelée Imprinted Control Region (ICR) qui contrôle leur expression. Une des principales caractéristiques communes à de nombreux gènes soumis à empreinte est le différentiel de méthylation observé dans des régions spécifiques appelées DMRs (Differentially Methylated Region) entre les deux allèles parentaux. Le niveau de méthylation de ces DMRs peut varier au cours du développement et en fonction des tissus, résultant en une empreinte parentale tissu-spécifique de certains gènes.

L'une des hypothèses principales sur le rôle de cette empreinte est celle du conflit parental : au cours du développement fœtal, le génome paternel activerait des gènes nécessaires au bon développement de l'embryon tandis que le génome maternel serait programmé pour activer des gènes protégeant les ressources de la mère (Moore and Haig, 1991). En effet, des gènes à expression paternelle encouragent la croissance fœtale tandis que des gènes à expression maternelle la répriment. De plus, la majorité des gènes soumis à empreinte sont présents dans le placenta, certains de manière exclusive (Court et al., 2014). Le rôle important de plusieurs de ces gènes dans le développement explique que de nombreuses maladies liées au développement présentent une altération de l'empreinte génomique. C'est le cas par exemple du syndrome de Prader-Willi, caractérisé par une hypotonie sévère, un retard mental, des problèmes comportementaux notamment alimentaires, ou du syndrome d'Angelman, caractérisé également par un retard mental, des rires inappropriés, et des crises de tremblements. Ces deux syndromes sont causés par une mutation dans le gène soumis à empreinte *UBE3A* (Knoll et al., 1989). Le syndrome de Beckwith-Wiedemann est quant à lui dû à une dérégulation de l'empreinte dans la région chromosomique 11p15.5 contenant les clusters *IGF2/H19* et *CDKN1C/KCNQ1OT1*, et est

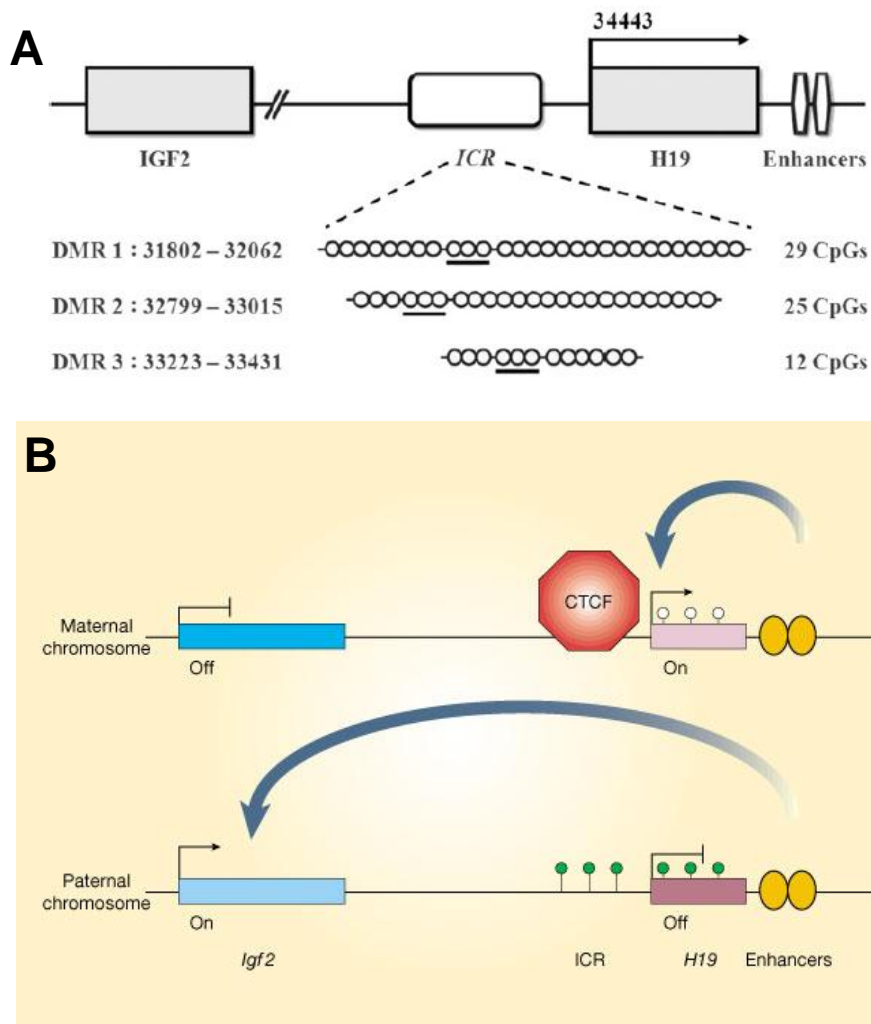


Figure 9 : Représentation du locus IGF2/H19. **A**. Schéma du locus IGF2/H19 chez le porc indiquant les coordonnées (numéro d'accèsion GenBank AY044827) des différents DMRs situés dans l'ICR de H19 (Park et al., 2009). Les ronds représentent les CpG et les traits situés en dessous indiquent les sites de fixation de CTCF. La flèche indique le site d'initiation de transcription de H19. **B**. Régulation de l'expression dans le locus IGF2/H19 (Reik and Murrell, 2000). Sur l'allèle maternel non méthylé (ronds blancs), la fixation de l'insulateur CTCF dans l'ICR de H19 empêche l'activation d'IGF2 par ses « enhancers » distaux : IGF2 est réprimé et H19 est activé. Sur l'allèle paternel, la méthylation de l'ADN (ronds verts) au niveau de l'ICR bloque la fixation de CTCF : IGF2 est activé par ses enhancers et H19 est réprimé. Les DMRs décrits dans IGF2 ne sont pas représentés sur ces deux schémas.

caractérisé par une croissance excessive, des malformations congénitales et une prédisposition tumorale (Weksberg et al., 2010; Soejima and Higashimoto, 2013).

La base de données d'empreinte génomique en ligne « Mouse Book of Imprinted Gene » (http://www.har.mrc.ac.uk/research/genomic_imprinting/) (Williamson et al., 2013) rassemble les régions soumises à empreinte chez la souris et leurs homologues humains. Le « Catalogue of Imprinted Gene » (www.otago.ac.nz/IGC) (Morison et al., 2001) compare les données obtenues pour l'homme, la souris, le rat et d'autres espèces telles que le porc, le mouton, la vache, le lapin et les marsupiaux. Un grand nombre de ces régions est conservé entre l'homme et la souris. Il est toutefois intéressant de noter que le nombre de gènes soumis à empreinte dans le placenta murin est plus élevé que dans le placenta humain (Monk et al., 2006).

De nombreuses études chez le porc, réalisées par croisement entre races génétiquement éloignées ou par l'utilisation du modèle parthénogénétique, ont confirmé l'empreinte parentale des clusters *DLK1/MEG3 (DLK1/GTL2)* et *IGF2/H19*, et des gènes *NDN*, *MAGEL2*, *MEST (PEG1)*, *CDKN1C*, *NAP1L5*, *PEG3*, *PEG10*, *SNRPN*, *NCD*, *NNAT*, *DIRAS3*, *PLAGL1*, *PRIM2A*, *SGCE*, *PPP1R*, *AMPD3*, *COPG2*, *DHCR7*, *IGF2R*, *OSBPL1A*, *PHLDA2*, *PON2*, *TFPI2*, *RASGRF1* et *NESP55* (Cheng et al., 2007; Zhang et al., 2007; Li et al., 2008b; Bischoff et al., 2009; Park et al., 2011; Li et al., 2012; Oczkowicz et al., 2012; Wang et al., 2012; Zhang et al., 2012; Ding et al., 2014) et ce avec parfois une empreinte tissue-spécifique (Bischoff et al., 2009; Li et al., 2012; Zhang et al., 2012; Ding et al., 2014). Les gènes *ASB4*, *ASCL2*, *CD81*, *COMMD1*, *DCN*, *DLX5*, *HMI3* et *UBE3A-AS* et *NAP1L4* ont eux été reportés d'expression biallélique chez le porc (Bischoff et al., 2009; Li et al., 2012). Des études contradictoires présentent *IGF2R* soit comme soumis à empreinte maternelle (Bischoff et al., 2009; Park et al., 2011) soit comme d'expression biallélique, malgré la présence de DMRs (Braunschweig, 2012). L'étude du cluster *IGF2/H19* (Figure 9) a permis d'identifier la présence de zones différenciellement méthylées entre les allèles paternels et maternels, avec une grande similarité de séquence dans la région entre le génome humain et le génome porcin (Amarger et al., 2002; Park et al., 2009; Braunschweig et al., 2011; Zhao et al., 2013). Il est donc généralement admis que le mécanisme d'empreinte parentale est majoritairement conservé entre plusieurs mammifères y compris chez le porc.

c. Variations épigénétiques au cours de l'embryogénèse

De la fécondation à la formation de gamètes, de nombreuses modulations des marques épigénétiques entrent en jeu. Elles sont résumées dans la Figure 10. Ces variations peuvent être découpées en trois étapes : reprogrammation épigénétique pré-implantatoire, modifications spécifiques des cellules germinales primordiales et enfin changements au cours de la gamétogénèse mâle.

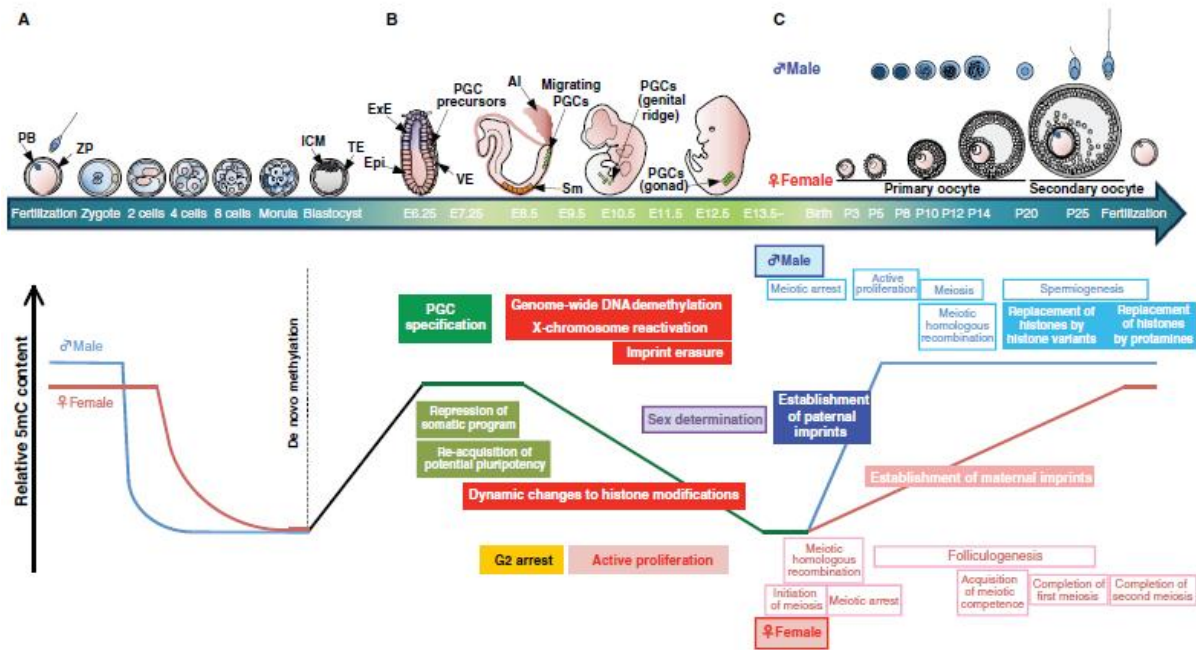


Figure 10 : Reprogrammations épigénétiques au cours du développement embryonnaire de la souris (Saitou et al., 2012). **A**. Après la fertilisation, le pronucléus mâle est rapidement et activement déméthylé tandis que la déméthylation du pronucléus maternel est plus lente. Au stade blastocyste, l'ADN est de nouveau méthylé. **B**. Dans les PGCs en migration, l'ADN est largement déméthylé pour garantir l'effacement de l'empreinte. **C**. L'empreinte parentale est réétablie rapidement dans les gamètes mâles et plus progressivement dans les gamètes femelles. AI, allantois; Epi, epiblast; ExE, extraembryonic ectoderm; ICM, inner cell mass; PB, polar body; PGCs, primordial germ cells; Sm, somite; TE, trophectoderm; VE, visceral endoderm; ZP, zona pelucida.

i. Reprogrammation épigénétique post-fertilisation

L'étude de la méthylation de l'ADN dans les gamètes montre que le sperme est fortement méthylé, entre 70% et 90%, tandis que l'ovocyte l'est moins chez l'homme, la souris (40%) et le porc (Petkov et al., 2009; Popp et al., 2010; Molaro et al., 2011; Kobayashi et al., 2012). Les zones qui échappent à la méthylation sont les régions très riches en CG, et notamment les promoteurs. On retrouve une corrélation négative entre expression des gènes et méthylation des promoteurs.

Le niveau global de méthylation de l'ADN cellulaire chute après la fertilisation (Monk 1987). Chez certains mammifères tels que la souris, l'homme, le porc et le bovin, le pronucléus paternel est activement et rapidement déméthylé après la fécondation (Dean et al., 2001; Beaujean et al., 2004; Fulka et al., 2006), tandis que la déméthylation du pronucléus maternel est délayé dans le temps et plus progressive (Fulka et al., 2006). Bien que ce procédé de déméthylation paternel soit conservé entre plusieurs espèces, il est plus lent chez l'homme et le bovin que chez le porc et la souris (Fulka et al., 2004) et n'a pas été détecté chez les ovins, le lapin, le zebrafish et *Xenopus* (Dean et al., 2001; Beaujean et al., 2004). La déméthylation du pronucléus paternel passe par l'action de TET3, en l'absence de laquelle l'embryon ne se développe pas (Gu et al., 2011). En effet, la perte de 5mC est corrélée avec un gain en 5hmC (Iqbal et al., 2011; Wossidlo et al., 2011). La désamination des 5hmCs en 5hmU par le complexe Aid/Apobec permet ensuite l'excision et la réparation du nucléotide par le système BER (Hajkova et al., 2010). De plus la déméthylation peut également se poursuivre passivement, puisque la marque 5hmC ne possède pas de système de maintenance qui garantit sa conservation après la réplication (Beaujean et al., 2004; Hackett et al., 2012). La faible vitesse de déméthylation du pronucléus maternel indique que les marques de méthylation sont diluées passivement lors de la réplication, en accord avec l'absence de 5hmC (Hackett et al., 2012; Piccolo and Fisher, 2014). Néanmoins, une étude récente a pu détecter la présence de ce dérivé dans le pronucléus maternel, suggérant une part de déméthylation active du génome maternel (Wang et al., 2014).

Il est intéressant de noter que chez la souris certaines parties du génome échappent à cette forte déméthylation. C'est le cas notamment des IAPs (Intracisternal A Particule), des régions répétées et des gènes soumis à empreinte (Lane et al., 2003; Daxinger and Whitelaw, 2012; Kobayashi and Kono, 2012; Teng and Zhou, 2013).

ii. Reprogrammation de l'épigénome dans les cellules germinales primordiales

La méthylation du génome augmente à nouveau au moment de l'implantation grâce à l'action des enzymes DNMT3a et DNMT3b (Okano et al., 1999). On observe ensuite une différence entre les cellules somatiques et les cellules germinales primordiales (PGCs) : tandis que les cellules somatiques conservent leur niveau de méthylation, les cellules germinales subissent une reprogrammation de leur

épigénome caractérisée par une vague de déméthylation et de reméthylation de l'ADN (Figure 10). L'effacement de toutes les marques épigénétiques dans les PGCs est important pour garantir la totipotence de ces cellules (Surani et al., 2007), éviter la transmission d'épimutations aux générations suivantes (Hackett et al., 2012), effacer l'empreinte parentale en fonction du sexe de l'embryon (Sasaki and Matsui, 2008) et garantir la bonne expression des gènes spécifiques de la lignée germinale (Borgel et al., 2010; Hammoud et al., 2014). Chez la souris, le processus de déméthylation des PGCs démarre à J8 du développement embryonnaire, lorsque les cellules sont migrantes, mais la plus forte déméthylation s'engage lorsque les cellules atteignent les crêtes génitales (Seki et al., 2005). Cette déméthylation touche en premier les promoteurs, les îlots CpG, puis les exons, les introns, les régions intergéniques et enfin les gènes soumis à empreinte (Kerjean et al., 2000; Teng and Zhou, 2013). Elle concerne entre 70% et 80% du génome (Popp et al., 2010). Le processus de déméthylation s'étend jusqu'à E13.5 chez la souris (Hajkova et al., 2002).

Chez le porc, les cellules germinales primordiales sont identifiables entre J18 et J20 dans le mésenchyme dorsal et colonisent les crêtes génitales autour de J23-J24 (Takagi et al., 1997). Bien que certains gènes tels qu'IGF2R se déméthylent dès J22, avant l'arrivée dans les crêtes génitales (Hyldig et al., 2011a), la majorité de la déméthylation des PGCs porcines est observable entre J24 et J28 lorsque les cellules se situent dans les crêtes génitales (Petkov et al., 2009). Les PGCs mâles sont reméthylées à J30-31 tandis que les cellules femelles restent déméthylées jusqu'après la naissance, comme chez la souris (Petkov et al., 2009; Hyldig et al., 2011a; Hyldig et al., 2011b).

Certaines régions du génome murin échappent à cette profonde déméthylation, comme les IAPs ou des éléments de régulations (Hajkova et al., 2002; Lane et al., 2003; Hackett et al., 2013; Kagiwada et al., 2013) tandis que d'autres y sont moins sensibles, avec une déméthylation moindre et plus longue à se mettre en place, comme pour les éléments répétés SINEs chez le porc (Hyldig et al., 2011a). La déméthylation des PGCs est médiée chez les mammifères par les enzymes TET1 et TET2 tandis que TET3 n'est pas exprimée (Hackett et al., 2013). Ce processus est également associé à la répression de la méthylation *de novo* (répression de DNMT3a et 3b) et de maintenance (répression d'UHRF1) de l'ADN (Hackett et al., 2013; Kagiwada et al., 2013). Ici encore, il est probable que la déméthylation passe à la fois par un processus actif, via les enzymes de la famille TET puis des glycosylases telles que TDG, ainsi que par un processus passif, avec la dilution par la réplication des marques 5mC et 5hmC.

La modification du niveau de méthylation des PGCs est précédée de changement de l'environnement chromatinien, qui évolue pour ressembler à celui des cellules souches embryonnaires (ES) (Surani 2007). Chez la souris comme chez le porc, le niveau de H3K9me2 diminue, suivi d'une augmentation de H3K27me3, tandis que les PGCs sont en arrêt en phase G2 du cycle cellulaire (Seki et al., 2005; Hajkova et al., 2008; Hyldig et al., 2011a).

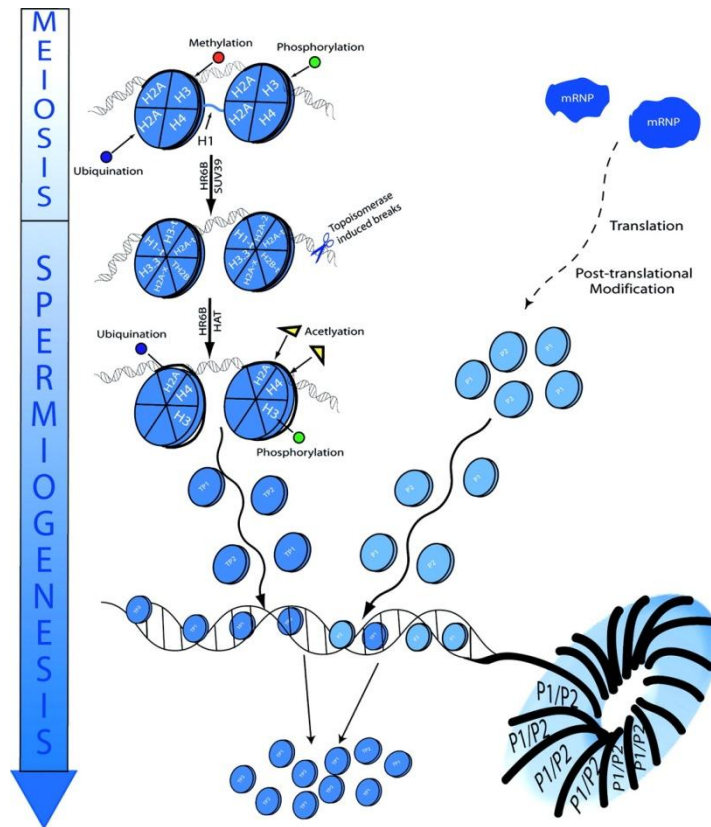


Figure 11 : Transition histone-protamine au cours de la spermiogénèse humaine (Carrell et al., 2007). Après remplacement des histones par des variants d'histones, l'hyperacétylation de H4 permet le relâchement de la structure de l'ADN et facilite le remplacement des variants par les protamines P1 et P2. HR6B, ubiquitin-conjugating enzyme E2B ; HAT, histone acetyltransferase ; Suv39, H3 Lys 9 histone methyltransferase ; mRNP, mature RiboNucleoProtein

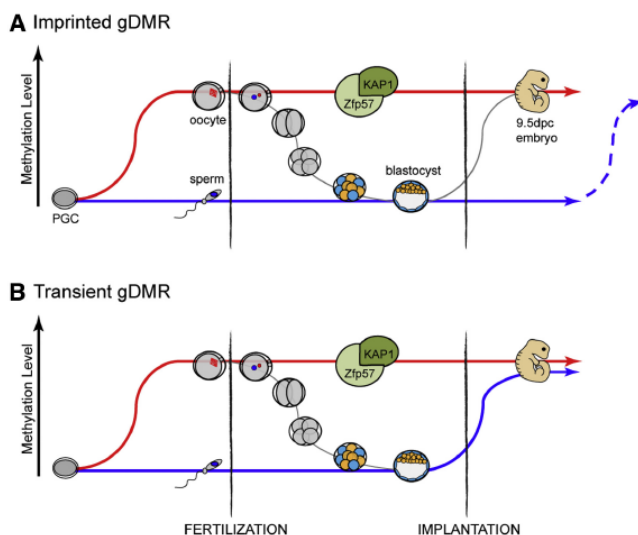


Figure 12 : Comportement des DMRs gamétiques stables et transitoires dans le remodelage épigénétique pré-implantatoire (Proudhon et al., 2012). **A.** Les gDMRs soumis à empreinte sont protégés de la déméthylation sur l'allèle maternel (rouge) par le complexe ZFP57/KAP1 et protégés de la reméthylation globale de l'embryon sur l'allèle paternel (bleu). **B.** Les gDMRs transitoires de l'allèle maternel n'échappent pas à la déméthylation des PGCs et ceux de l'allèle paternel sont reméthylés à l'implantation.

iii. Modifications épigénétiques au cours de la gamétogénèse mâle

Les marques de méthylation réapparaissent dans les cellules germinales mâles avant la méiose par méthylation *de novo* médiée par DNMT3L, DNMT3a et DNMT3b (Piccolo and Fisher, 2014). En l'absence de DNMT3L, l'ADN gamétique est hypométhylé et la méiose est altérée (Bourc'his and Bestor, 2004). Le contexte chromatinien est également important en cours de méiose, notamment la méthylation de H3K4 et de H3K9 (Carrell and Hammoud, 2010; Kota and Feil, 2010) mais aussi des acétylations et ubiquitinations de variants d'histones (Carrell and Hammoud, 2010). En fin de méiose, H3K9 est déméthylé. Enfin, la spermiogénèse est marquée par un bouleversement de la structure de l'ADN : afin d'assurer la compaction de l'ADN dans le noyau spermatique, les histones sont remplacées par des protamines (Figure 11) (Ward and Coffey, 1991). Cet échange est lui aussi accompagné de variations du profil des histones, notamment la déméthylation de H3K9 en fin de méiose et l'hyperacétylation de H4. Malgré cette transition histone-protamine intensive, on observe dans le sperme une rétention des histones et du nucléosome sur un faible pourcentage du génome (5-15%) (Carrell and Hammoud, 2010; Kota and Feil, 2010) autour de gènes essentiels pour le développement embryonnaire (Hammoud et al., 2009; Erkek et al., 2013).

iv. Comportement spécifique des gènes soumis à empreinte

Comme indiqué précédemment, les gènes soumis à empreinte possèdent un comportement particulier au cours des vagues de reprogrammation épigénétiques de l'embryogénèse. En effet, ces gènes échappent à la première vague de reprogrammation des pronuclei paternels et maternels. Les gDMRs (DMRs gamétiques) méthylés dans l'ovocyte sont protégés de la déméthylation grâce à la fixation du complexe ZFP57/KAP1 (Li et al., 2008a) et les gDMRs déméthylés dans les cellules germinales mâles sont protégés de la méthylation *de novo* se produisant après l'implantation, et restent donc déméthylés dans l'embryon (Figure 12). Ces gDMRs sont dits stables, ou à empreinte. Il existe une seconde catégorie de gDMRs, appelés gDMRs transitoires, qui ne sont pas conservés au cours de l'embryogénèse précoce, soit en n'étant pas protégé de la déméthylation pour les gDMRs maternels, soit en étant reméthylés en même temps que l'ensemble du génome pour les gDMRs paternels (Figure 12). La différence de méthylation entre les deux allèles est donc perdue dans les cellules somatiques. Il est à noter que les gDMRs paternels stables peuvent être reméthylés plus tard au cours du développement en fonction des tissus (Proudhon 2012).

Malgré tout, les gDMRs ne sont pas tout à fait insensibles à la vague de déméthylation de l'embryon préimplantatoire. Le DMR gamétique est souvent différent de celui que l'on trouve dans les cellules du blastocyste, ou dans les cellules somatiques. La région peut-être étendue, décalée, ou allongée d'un côté ou de l'autre dans les gamètes (Tomizawa et al., 2011). Une explication possible est qu'un DMR plus grand serait plus résistant à la déméthylation et permettrait d'assurer malgré tout la conservation

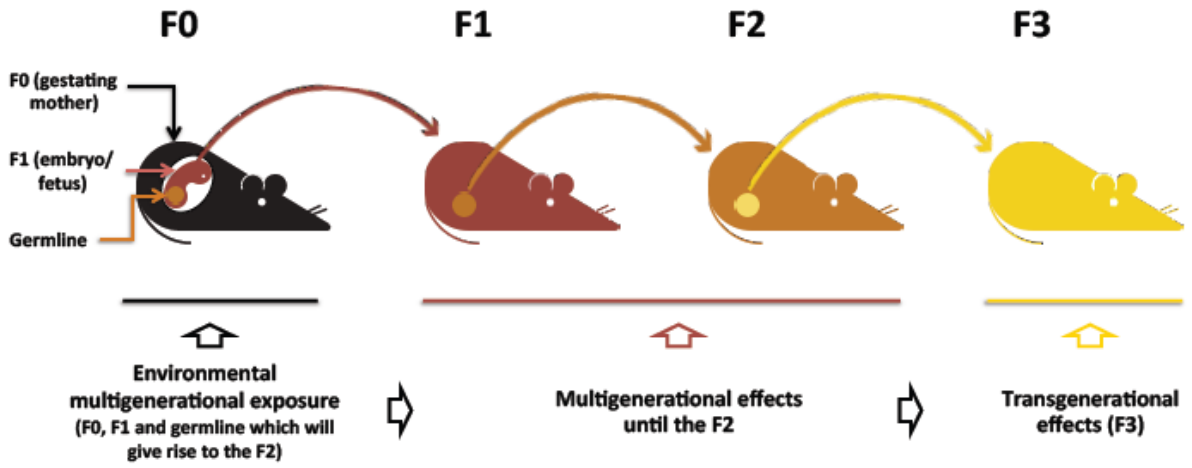


Figure 13 : Héritabilité épigénétique transgénérationnelle : les modifications épigénétiques et les phénotypes associés provoqués par une modification de l'environnement sur une femelle gestante de génération F0 se transmettent via la lignée germinale à la génération F3 qui n'a pas été directement exposée (Martínez et al., 2014)

du DMR au cours du développement précoce et ce jusqu'aux cellules somatiques (Proudhon et al., 2012). Au contraire, la vague de déméthylation ayant lieu dans les PGCs est plus intense, puisque les gènes soumis à empreinte n'y échappent pas. Cela permet de reprogrammer les marques épigénétiques en fonction du sexe de l'embryon. A l'issue de cette déméthylation, les DMRs devant rester méthylés échappent à la méthylation *de novo* du génome (Kerjean et al., 2000).

Cette différence de comportement entre gènes soumis à empreinte et gènes d'expression biallélique, que l'on retrouve également au niveau des transgènes, permet aujourd'hui d'envisager une héritabilité épigénétique via les gamètes (Figure 13). L'héritabilité épigénétique a été observée dans de nombreux cas, principalement sur des transgènes, (Daxinger and Whitelaw, 2012) mais il est difficile de différencier la transmission par les gamètes de celle pouvant avoir lieu par l'environnement placentaire maternel, ou par le lait. C'est toutefois plus facile dans le cas d'une transmission paternelle car le génome paternel contribue moins à l'environnement fœtal. L'exemple le plus connu de transmission intergénérationnelle d'un épiallèle endogène est celui du phénotype Agouti chez la souris (Morgan et al., 1999). La méthylation du promoteur se trouvant dans l'IAP inséré en amont du locus agouti sur l'allèle Avy est inversement corrélée à l'activité transcriptionnelle du gène. Les mères porteuses de l'allèle Avy donnent plus de filles jaunes, quelque soit le père, que les mères non porteuses de cet allèle, dont les portées présentent différents phénotypes de couleur allant de brun à jaune. Cette transmission maternelle est indépendante des effets maternels post-fertilisation, c'est bien une héritabilité épigénétique via les gamètes (Morgan et al., 1999). Des études plus récentes ont depuis démontré l'existence d'effets multigénérationnels (Figure 13) chez les rongeurs, avec par exemple une dérégulation du métabolisme des lipides chez les descendants de mâles ayant subi une restriction alimentaire in utero, associé à la transmission via les gamètes d'une altération du profil de méthylation d'un gène impliqué dans le métabolisme lipidique (Martínez et al., 2014). Chez l'homme, plusieurs études ont montré les effets sur les descendants de modifications de l'environnement, en particulier de la nutrition (Patti, 2013). La transmission via les gamètes est donc également possible, en cas de mauvais effacement ou rétablissement de l'empreinte, mais reste encore incertaine chez l'homme (Carrell and Hammoud, 2010). Malgré tout, des études ont montré l'héritabilité par la voie paternelle d'épimutations dans le locus *SNRPN-SNURF* (Buiting et al., 2003), ainsi que la conservation d'une hypométhylation de *H19* entre le sperme et un embryon obtenu par procréation médicalement assistée (Kobayashi et al., 2009).

d. Cas d'infertilité en présence de défauts épigénétiques

Les modifications épigénétiques jouent donc un rôle important au cours du développement embryonnaire et de la formation des cellules germinales. Un mauvais profil épigénétique peut alors refléter un mauvais fonctionnement des gamètes.

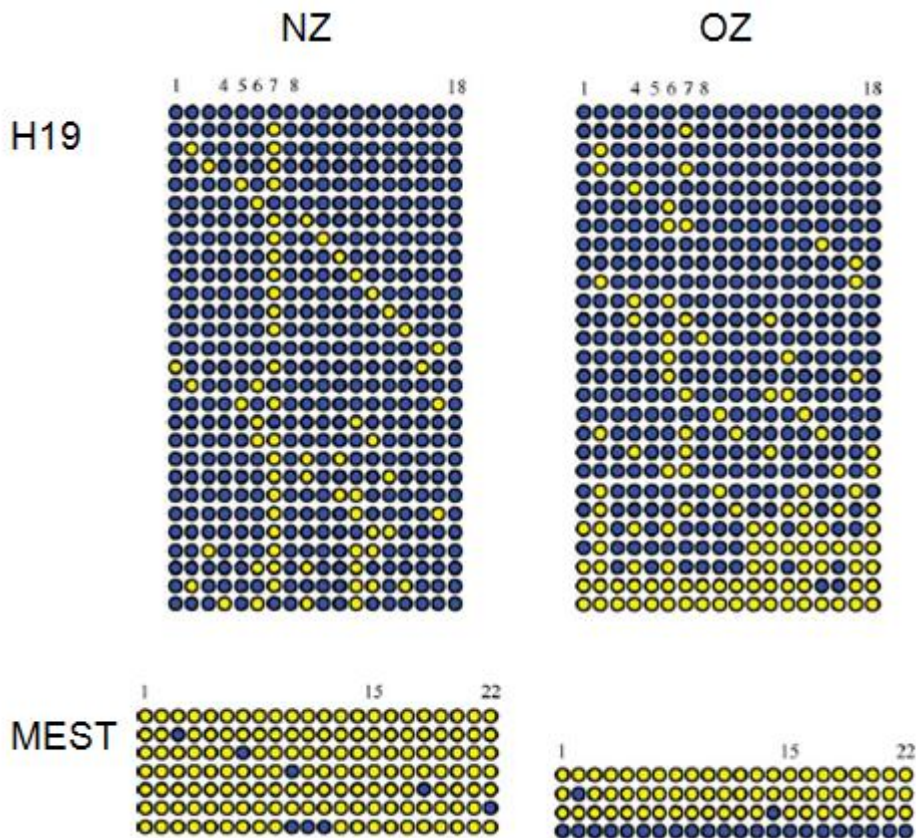


Figure 14 : Analyse du profil de méthylation de *H19* et *MEST* dans le sperme de patients normozoospermiques (NZ) ou sévèrement oligospermiques (OZ) par conversion bisulfite, clonage et séquençage. Chaque ligne représente un clone et chaque rond représente un CpG. CpG méthylé (bleu) et non méthylé (jaune), site de fixation CTCF dans le locus *H19*: CpG 4-8. Adapté de (Marques et al., 2008)

i. *Variations de la méthylation de l'ADN dans les gènes soumis à empreinte*

De nombreuses études ont corrélé modifications des marques épigénétiques des gènes soumis à empreinte et qualité de la semence. Chez l'homme, un défaut de méthylation au niveau du gène *H19*, exprimé maternellement et normalement méthylé dans le sperme, a été observé chez des patients oligospermiques (Marques et al., 2004; Kobayashi et al., 2007; Marques et al., 2008; Hammoud et al., 2010; Poplinski et al., 2010; Li et al., 2013) (Figure 14). L'hypométhylation sur le site de fixation de CTCF dans le DMR du gène *H19* se retrouve aussi chez des patients tératospermiques, où elle s'accompagne d'une perte de méthylation dans le second DMR d'*IGF2*, gène appartenant au même cluster et exprimé par l'allèle paternel. Chez les patients oligo-astheno-tératospermiques, la perte de méthylation est corrélée avec la concentration du sperme (Boissonnas et al., 2010), tout comme dans l'étude de Poplinski *et al* où la perte de méthylation sur le locus *IGF2/H19* est associée à un faible comptage des spermatozoïdes (Poplinski et al., 2010). L'étude de spermatozoïdes testiculaires de patients azospermiques révèle elle aussi une hypométhylation du site de fixation de CTCF dans le locus du DMR d'*H19* en cas d'hypospermatogénèse (Marques et al., 2010). Cette hypométhylation n'a pas été retrouvée chez des patients asthénospermiques (Li et al., 2013) ni dans des spermatogonies prélevées dans des tubules présentant un arrêt de la spermatogénèse au stade spermatogonie, ni dans des spermatocytes issues de tubules présentant un arrêt de la spermatogénèse au stade spermatocyte (Hartmann et al., 2006). Le défaut de méthylation sur *H19* a été retrouvé chez des patients infertiles à semence normale où il a été corrélé avec une hyperméthylation du promoteur du gène *MHTFR* codant pour une enzyme régulatrice nécessaire au maintien de la biodisponibilité des groupements méthyls endogènes. L'hyperméthylation de ce promoteur, que l'on retrouve également chez des patients oligospermiques, pourrait donc entraîner une répression de ce gène et donc l'indisponibilité des groupements méthyls, générant alors des défauts de méthylation sur le site de fixation de CTCF (Wu et al., 2010; Rotondo et al., 2013).

Chez le bovin, aucune différence notable n'a été remarquée sur les différents DMRs du locus *IGF2/H19* entre des groupes d'animaux à haute et faible fertilité. Seule une différence a été constatée sur le site de fixation de CTCF dans le DMR d'*H19* (Jena et al., 2014).

Ces variations des marques épigénétiques pourraient être responsables d'une altération de l'expression d'*IGF2* au cours de la spermatogénèse ou du développement embryonnaire précoce (Marques et al., 2004; Jena et al., 2014). *MEG3* et *ZDBF2*, deux autres gènes soumis à empreinte maternelle, ont également été reportés comme hypométhylés dans le sperme de patients oligospermiques (Kobayashi et al., 2007; Sato et al., 2011).

Plusieurs études ont aussi montré que des gènes à empreinte paternelle, normalement déméthylés dans le sperme, arborent une hyperméthylation dans le cas d'une mauvaise qualité de la semence (Houshdaran et al., 2007; Kobayashi et al., 2007; Marques et al., 2008; Hammoud et al., 2010;

Poplinski et al., 2010; Sato et al., 2011). C'est notamment le cas de *MEST* (ou *PEG1*) (Figure 14), *PAX8*, *LIT1*, *PEG3*, *ZAC*, *SNRPN*, *DIRAS3*, *PLAGL1*, *SFN*, *MT1A*, *NTF3* et *HRAS*. Il est toutefois intéressant de noter que la région la plus touchée dans le cas de patients oligospermiques est le DMR du gène *MEST*, situé dans son promoteur (Hammoud et al., 2010; Sato et al., 2011) et que son hyperméthylation semble liée à la sévérité du cas (Marques et al., 2004; Marques et al., 2008). L'hyperméthylation de *MEST* peut-être associée à plusieurs phénotypes : oligospermie, mais aussi asthénospermie et tératospermie (Poplinski et al., 2010). Enfin, les cas les plus sévères présentent à la fois le défaut de méthylation sur des gènes d'expression paternelle et la méthylation aberrante des gènes soumis à empreinte maternelle (Kobayashi et al., 2007; Marques et al., 2008; Sato et al., 2011).

D'une manière générale, une hypothèse forte est qu'une reprogrammation défectueuse de la méthylation de l'ADN dans les PGCs entraînerait des altérations du profil de méthylation dans les cellules germinales, particulièrement au niveau des gènes soumis à empreinte, et que le type d'altération pourrait être lié au phénotype observé (type d'anomalie, sévérité du cas). Cette hypothèse se confirme par le phénotype oligospermique de souris déficientes pour les gènes *Dnmt3a* et *Dnmt3l*, régulateurs de la méthylation *de novo* des PGCs.

ii. Des variations sur l'ensemble du génome

Des études mesurant le niveau de méthylation dans le sperme sur un ensemble d'éléments plus vaste que les gènes soumis à empreinte parentale, avec en particulier l'analyse des îlots CpGs et des éléments répétés, ont mis en évidence des variations de niveau de méthylation sur d'autres séquences dans le sperme de patients infertiles. Houshdaran et collaborateurs ont observé une augmentation du niveau global de méthylation et Pacheco et collaborateurs une altération généralisée du niveau de méthylation des CpGs dans des échantillons d'individus asthénospermiques (Houshdaran et al., 2007; Pacheco et al., 2011). Néanmoins ces résultats sont opposés aux résultats de Marques et collaborateurs qui n'observaient aucune variation de méthylation dans la séquence *LINE1*, généralement choisie comme séquence de référence pour mesurer le niveau global de méthylation du génome (Marques et al., 2008). Malgré ce, des études locales sur des gènes non soumis à empreinte jouant un rôle important dans la gamétogénèse ont montré des variations du niveau de méthylation en cas d'infertilité. L'identification par exemple d'un DMR tissu-spécifique contrôlant l'expression de *DDX4* (ou *VASA*) dans des biopsies testiculaires a permis de mettre en avant l'hyperméthylation de cette région pour 6 patients sur 17 atteints d'un arrêt de la maturation des gamètes, corrélée à la répression de ce gène (Sugimoto et al., 2009). La régulation de l'expression de *DDX4* par la méthylation de son promoteur, garantissant son inactivation dans les cellules somatiques et son activation par la déméthylation dans les cellules germinales spécifiquement, a également été confirmée dans la spermatogénèse bovine (Luo et al., 2013). De même, l'étude du promoteur du gène *DAZL* à quant à

Tableau 2 : Régulateurs épigénétiques chromatiniens et leurs effets sur la méiose (adapté de (Kota and Feil, 2010)).

Protéine	Substrat	Phénotype de la souris KO	Référence
KMT1A / KMT1B	H3K9	défauts de méiose, infertilité mâle et femelle	(Peters et al., 2001)
KMT1C (G9A)	H3K9	défauts de méiose, infertilité mâle et femelle	(Tachibana et al., 2007)
KMT2B (MLL2)	H3K4	infertilité mâle et femelle	(Glaser et al., 2009)
PRDM9 (MEISETZ)	H3K4	arrêt de la méiose, infertilité mâle et femelle	(Hayashi et al., 2005)
KDM1B (LSD2)	H3K4me	altérations dans l'établissement de l'empreinte maternelle	(Ciccone et al., 2009)
KDM3A (JHDM2A)	H3K9me	altérations de la condensation de la chromatine post-méiotique, infertilité mâle	(Okada et al., 2007)
UBR2	HR6B-H2A	défauts de méiose, infertilité mâle	(An et al., 2010)
RNF8	γ H2AX	altérations de la transition histone-protamine, infertilité mâle	(Lu et al., 2010)
HR6B (UBE2B)	H2A	infertilité mâle	(Roest et al., 1996)
LSH (HELLS)	chromatine	défauts de méiose, infertilité mâle	(De La Fuente et al., 2006)
BRDT (BRD6)	H4ac	sperme anormal, infertilité mâle	(Shang et al., 2007)

elle permis de mettre en évidence une augmentation de la méthylation de ce promoteur dans du sperme de mauvaise qualité (Navarro-Costa et al., 2010; Li et al., 2013).

iii. *Modifications chromatiniennes impliquées dans l'altération de la gamétogénèse*

Comme vu précédemment, la méthylation de l'ADN n'est pas la seule marque épigénétique jouant un rôle important dans la biologie des cellules germinales. Les modifications de la chromatine, et en particulier des histones, sont elles aussi primordiales. Des « knock-outs » réalisés chez la souris sur des gènes codant pour des protéines régulant ces marques d'histones, telles que les méthyltransférases KMT1A, KMT1B, KMT1C ou la déméthylase KDM1, affectent la méiose, comme indiqué dans le Tableau 2 (Kota and Feil, 2010). Dans de nombreux cas, c'est l'appariement des chromosomes qui est impacté. Par exemple, H3K4me3 est recrutée sur les points chauds de recombinaison où elle permet la fixation de l'enzyme SPO11 qui catalyse les cassures doubles-brins de l'ADN nécessaires à la formation de crossing-overs lors de l'assemblage des chromosomes (Hayashi et al., 2005). La qualité des gamètes peut enfin être altérée au cours de la spermiogénèse, étape durant laquelle la chromatine est fortement remodelée, notamment par la transition histone-protamine. Bien que cette transition ne concerne pas l'ensemble du génome, un taux résiduel d'histones anormalement élevé a été constaté dans du sperme à qualité dégradée d'hommes hypofertiles (Ramos et al., 2008). De plus, un défaut de protamination des spermatozoïdes, notamment un rapport entre les protamines P1 et P2 différent de 1, a aussi été corrélé avec des altérations du profil de méthylation, en particulier une hyperméthylation des gènes à empreinte maternelle *LIT1* et *SNRPN* (Hammoud et al., 2010).

4. Effets des remaniements chromosomiques sur l'épigénome

L'épigénome d'une cellule dépend donc en partie du profil de méthylation de son ADN et des diverses modifications chimiques des histones, qui sont souvent caractéristiques de l'état de la chromatine environnante. Or, les foyers d'euchromatine et d'hétérochromatine jouent un rôle dans la régulation de l'expression des gènes, l'hétérochromatine étant généralement enrichie en marques répressives telles que H3K27me3 et H3K9me3. Ainsi, le repositionnement d'un gène d'un foyer d'euchromatine vers un foyer d'hétérochromatine peut entraîner sa réactivation de façon mosaïque : c'est le phénomène de variéfaction (Henikoff, 1990). L'analyse de l'architecture 3D du noyau a révélé que les chromosomes sont organisés en territoires chromosomiques et que cette organisation est étroitement associée à une organisation supérieure de la chromatine qui suit des règles bien déterminées ; notamment les domaines pauvres en gènes sont plus fréquemment en proximité de l'enveloppe nucléaire (Cremer and Cremer, 2001; Scheuermann et al., 2004). Ainsi, la présence d'un remaniement chromosomique peut entraîner le changement de position d'une région dans l'espace nucléaire et donc son statut d'expression. C'est le cas de la translocation déséquilibrée t(1;2)(q12;p13) apparaissant dans plusieurs

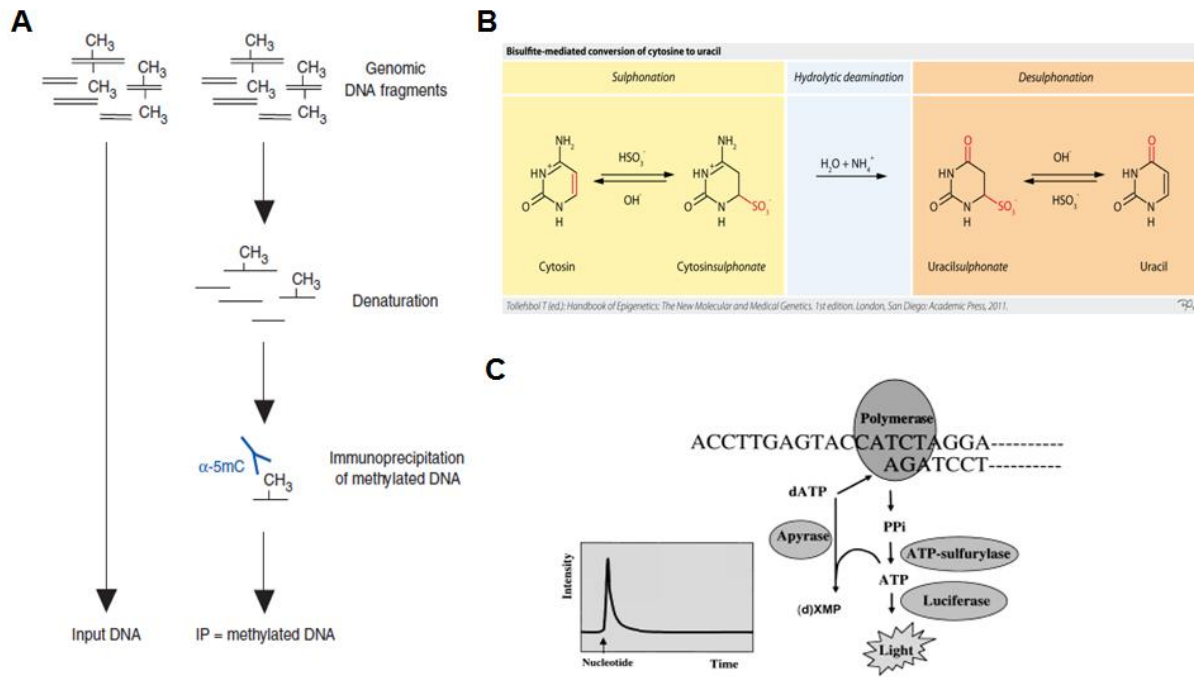


Figure 15 : *A. Immunoprécipitation de l'ADN méthylé (MeDIP) par un anticorps spécifique des cytosines méthylées (Weber et al., 2005). B. Conversion des cytosines non méthylées en uraciles par sulphonation bisulfite HSO_3^- , déamination et désulphonation (Tollefsbold, 2011). C. Pyroséquence (Ronaghi, 2001). Lorsque le nucléotide présent dans le milieu réactionnel est celui attendu par la polymérase, son incorporation déclenche une cascade enzymatique générant l'émission d'un signal lumineux via l'action de la luciférase.*

Tableau 3 : *Principales techniques de mesure de la méthylation de l'ADN (Laird, 2010)*

Pretreatment	Analytical step			
	Locus-specific analysis	Gel-based analysis	Array-based analysis	NGS-based analysis
Enzyme digestion	<ul style="list-style-type: none"> • <i>HpaII</i>-PCR 	<ul style="list-style-type: none"> • Southern blot • RLGS • MS-AP-PCR • AIMS 	<ul style="list-style-type: none"> • DMH • MCAM • HELP • MethylScope • CHARM • MMAPS 	<ul style="list-style-type: none"> • Methyl-seq • MCA-seq • HELP-seq • MSCC
Affinity enrichment	<ul style="list-style-type: none"> • MeDIP-PCR 		<ul style="list-style-type: none"> • MeDIP • mDIP • mCIP • MIRA 	<ul style="list-style-type: none"> • MeDIP-seq • MIRA-seq
Sodium bisulphite	<ul style="list-style-type: none"> • MethylLight • EpiTYPER • Pyrosequencing 	<ul style="list-style-type: none"> • Sanger BS • MSP • MS-SNuPE • COBRA 	<ul style="list-style-type: none"> • BiMP • GoldenGate • Infinium 	<ul style="list-style-type: none"> • RRBS • BC-seq • BSPP • WGSBS

AIMS, amplification of inter-methylated sites; BC-seq, bisulphite conversion followed by capture and sequencing; BiMP, bisulphite methylation profiling; BS, bisulphite sequencing; BSPP, bisulphite padlock probes; CHARM, comprehensive high-throughput arrays for relative methylation; COBRA, combined bisulphite restriction analysis; DMH, differential methylation hybridization; HELP, *HpaII* tiny fragment enrichment by ligation-mediated PCR; MCA, methylated CpG island amplification; MCAM, MCA with microarray hybridization; MeDIP, mDIP and mCIP, methylated DNA immunoprecipitation; MIRA, methylated CpG island recovery assay; MMAPS, microarray-based methylation assessment of single samples; MS-AP-PCR, methylation-sensitive arbitrarily primed PCR; MSCC, methylation-sensitive cut counting; MSP, methylation-specific PCR; MS-SNuPE, methylation-sensitive single nucleotide primer extension; NGS, next-generation sequencing; RLGS, restriction landmark genome scanning; RRBS, reduced representation bisulphite sequencing; -seq, followed by sequencing; WGSBS, whole-genome shotgun bisulphite sequencing.

types de tumeurs chez l'homme, où une région hétérochromatique proche du centromère du chromosome 1 est transloquée sur le chromosome 2, générant un foyer aberrant d'hétérochromatine (Fournier et al., 2010). Le chromosome 2 remanié est alors repositionné à la périphérie nucléaire, et le foyer aberrant d'hétérochromatine est enrichi en marques d'histone répressives qui se diffusent sur la partie du chromosome 2 non remaniée. Ce processus entraîne la dérégulation de l'expression de gènes situés sur la partie non remaniée du chromosome. Cette étude confirme que les remaniements chromosomiques peuvent avoir un effet sur l'expression des gènes via la modification de leur environnement chromatinien.

D'autre part, les remaniements chromosomiques, lorsqu'ils touchent une région soumise à empreinte, peuvent être la cause de pathologies résultantes d'altérations épigénétiques. Par exemple, le syndrome de Beckwith-Wiedemann peut être du à une disomie uniparentale paternelle, des duplications de la région chromosomique 11p15 contenant les clusters soumis à empreinte *IGF2/H19* et *CDKN1C/KCNQ1OT1*, ou bien à des translocations ou inversions incluant cette région (Soejima and Higashimoto, 2013).

5. Techniques d'études de la méthylation de l'ADN

Il existe aujourd'hui de nombreuses méthodes pour étudier la méthylation de l'ADN soit localement sur quelques centaines de bases ou bien sur l'ensemble du génome (« genome-wide »). Plusieurs de ces méthodes sont basées sur la digestion enzymatique, sur l'enrichissement de fractions d'ADN méthylé par un anticorps spécifique (MeDIP, Figure 15A) (Weber et al., 2005), ou sur la conversion de l'ADN par le bisulfite (HSO_3^-) représentée sur la Figure 15B qui permet de transformer les cytosines non méthylées en uracile (Frommer et al., 1992). Après un séquençage il est donc facile de différencier les cytosines méthylées (restées cytosines) des cytosines non méthylées (converties en uracile puis thymine par PCR).

Pour une analyse locale, une amplification par PCR spécifique de l'ADN converti au bisulfite suivie d'un clonage et du séquençage d'un nombre significatif de clones permet de définir le niveau de méthylation de l'ADN. Le pyroséquençage (Figure 15C) peut également être utilisé pour quantifier le niveau de méthylation de chaque CpG dans l'amplicon sans passer par une étape de clonage (Dupont et al., 2004). Le MeDIP peut quant à lui être suivi d'une PCR quantitative classique pour définir le niveau de méthylation relatif d'une région dans un échantillon.

Pour une analyse à l'échelle du génome, plusieurs méthodes peuvent être couplées soit au MeDIP soit à la conversion bisulfite de l'ADN telles que le séquençage haut-débit (Next Generation Sequencing, NGS) ou les puces ADN (Tableau 3). Bien que les résultats obtenus soient comparables en termes de niveau de méthylation, les différentes méthodes n'ont pas les mêmes avantages et ne souffrent pas des

mêmes biais (Beck, 2010; Bock et al., 2010; Harris et al., 2010; Laird, 2010; Pelizzola and Ecker, 2011). Les méthodes utilisant la conversion bisulfite sont plus résolutive que les méthodes d'enrichissement car le résultat par cytosine est indépendant de la séquence environnante. En effet, les méthodes d'enrichissement souffrent souvent d'un biais lié à la densité de CpG (Down et al., 2008) et leur résolution est limitée par la taille des fragments immunoprécipités (~300pb). Le RRBS (Reduced Representation Bisulfite Sequencing), qui allie enrichissement par digestion enzymatique et conversion bisulfite, est très résolutif mais la couverture du génome est globalement limitée aux îlots CpGs. Le WGSBS (Whole Genome Shotgun Bisulfite Sequencing) permet d'avoir de bons résultats en termes de couverture du génome et de résolution, mais il est économiquement désavantageux. Malgré ses biais, le MeDIP-Seq reste un bon compromis pour obtenir une couverture du génome satisfaisante à moindre coût. Cette technique est donc largement employée pour étudier la méthylation de l'ADN dans différents tissus et états ou pour identifier des zones différenciellement méthylées entre deux états (Kerkel et al., 2010; Ruike et al., 2010; Li et al., 2011; Sati et al., 2012; Bang et al., 2013; Hu et al., 2013; Zhang et al., 2013b; Su et al., 2014)

En conclusion, le développement de gamètes fonctionnels est un procédé complexe qui peut être altéré par de nombreux facteurs génétiques et environnementaux, parmi lesquels la présence d'anomalies chromosomiques. Ce processus est également soumis à des mécanismes épigénétiques dont les systèmes de régulation ne sont pas encore totalement éclaircis. Dans ce contexte, une meilleure compréhension du développement embryonnaire précoce ainsi que la possibilité de dériver des cellules germinales à partir de cellules souches *in vitro* font des cellules pluripotentes un outil remarquable pour étudier les mécanismes sous-tendant le développement des cellules germinales et les perturbations que peuvent induire les réarrangements chromosomiques sur ce développement. Dans la seconde partie, nous nous intéressons donc au développement de ces cellules pluripotentes, embryonnaires et induites, chez la souris, l'homme, et le porc.

II. La reprogrammation cellulaire

1. La pluripotence dans l'embryon : dérivation de lignées ESCs et EpiSCs

a. Caractéristiques des cellules souches embryonnaires

i. Etablissement de lignées de cellules souches embryonnaires

La pluripotence est définie comme la capacité d'une cellule à se différencier dans tous les types cellulaires issus des trois feuilletts embryonnaires : mésoderme, ectoderme et endoderme. Les cellules souches pluripotentes ne peuvent donc que contribuer à la formation de l'embryon, à la différence des cellules totipotentes qui peuvent contribuer aussi bien aux tissus embryonnaires qu'extra-embryonnaires. Mais les cellules pluripotentes ont davantage de potentiel que les cellules multipotentes qui sont capables de générer plusieurs types de cellules au sein d'une seule et même catégorie tissulaire (exemple : cellules souches hématopoïétiques, cellules souches musculaires, etc...). Chez l'homme et la souris, les cellules pluripotentes embryonnaires se trouvent dans la masse cellulaire interne du blastocyste, qui deviendra l'embryon. En prélevant des cellules directement de cette masse cellulaire interne, plusieurs équipes ont pu établir des lignées de cellules souches embryonnaires (ES) chez la souris (mESCs) (Evans and Kaufman, 1981) , l'homme (hESCs) (Thomson et al., 1998), le singe rhésus (rhESCs) (Thomson et al., 1995; Thomson et al., 1998; Thomson and Marshall, 1998) et avec plus de difficultés chez le rat (Brenin et al., 1997; Schulze et al., 2006; Buehr et al., 2008) et chez quelques espèces agronomiques comme le bovin et le lapin (Roach et al., 2006; Wang et al., 2007; Osteil et al., 2013). Les cellules ES ont de plus la capacité de s'auto-renouveler à l'infini.

ii. Caractéristiques morphologiques et moléculaires des cellules pluripotentes

La morphologie des mESCs est caractéristique : les cellules sont petites, sphériques, ont un fort ratio noyau/cytoplasme et forment des agrégats très denses qui poussent en dômes. Leur rapide prolifération s'explique par un cycle cellulaire particulier présentant une phase G1 raccourcie (Savatier et al., 1994) ne dépendant pas des mêmes voies de signalisation que les cellules somatiques : le rôle central de la protéine RB dans le cycle des cellules différenciées est modifié puisque, bien que la protéine soit présente, son hyperphosphorylation la rend inactive (Burdon et al., 2002) La durée de la phase G1 est en lien direct avec le niveau de pluripotence des cellules : elle est la phase la plus propice à la différenciation et sa durée s'allonge dès que les cellules entrent en différenciation (Coronado et al., 2013). De plus, les cellules souches embryonnaires ne possèdent pas de point de contrôle des dommages de l'ADN en G1 (Aladjem et al., 1998) et ont donc tendance à accumuler plus facilement des chromosomes surnuméraires (Mantel et al., 2007). Elles expriment de nombreux gènes permettant

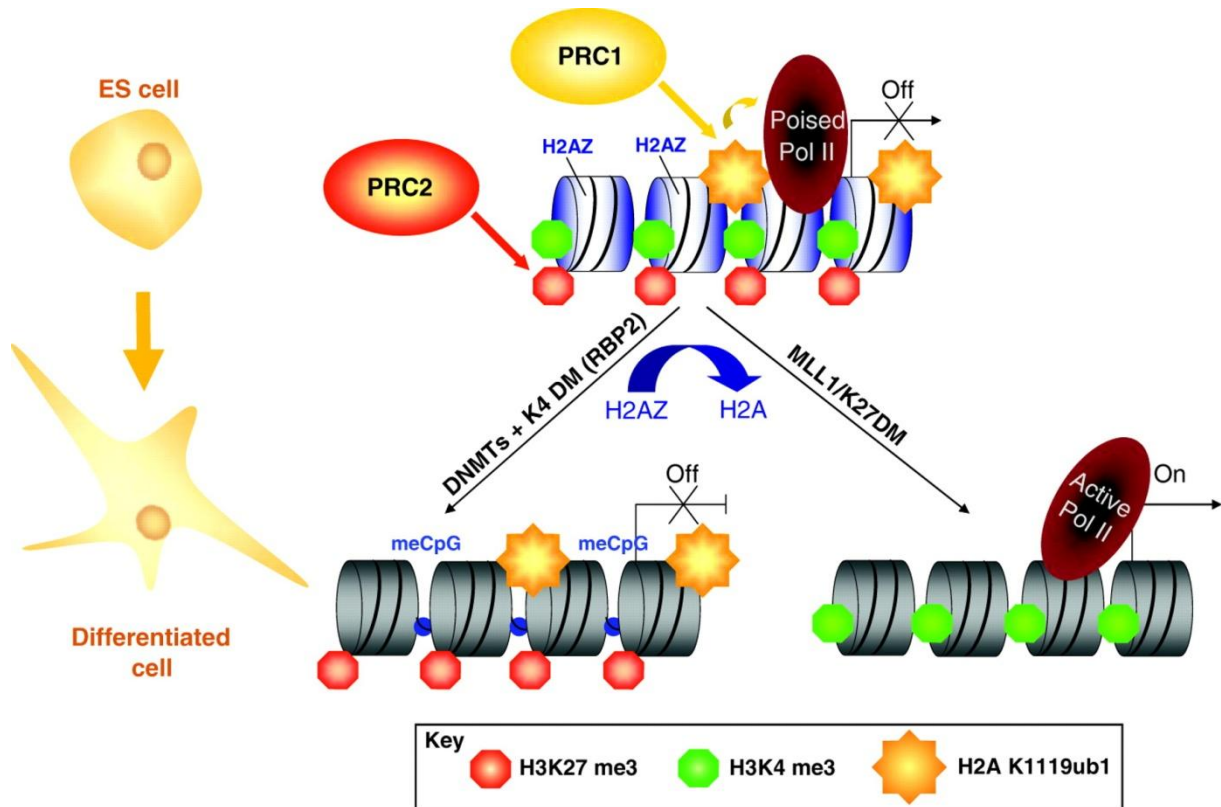


Figure 16 : Evolution des marques bivalentes au cours de la différenciation des ESCs. Les gènes régulateurs du développement sont marqués par la marque activatrice H3K4me3 qui permet le recrutement de la polymérase au niveau du promoteur et par la marque répressive H3K27me3 qui bloque la transcription du gène. Au cours de la différenciation, seule une des deux marques est conservée, entraînant la transcription ou non du gène. PRC1 et PRC2 : complexes Polycomb (Schuettengruber and Cavalli, 2009)

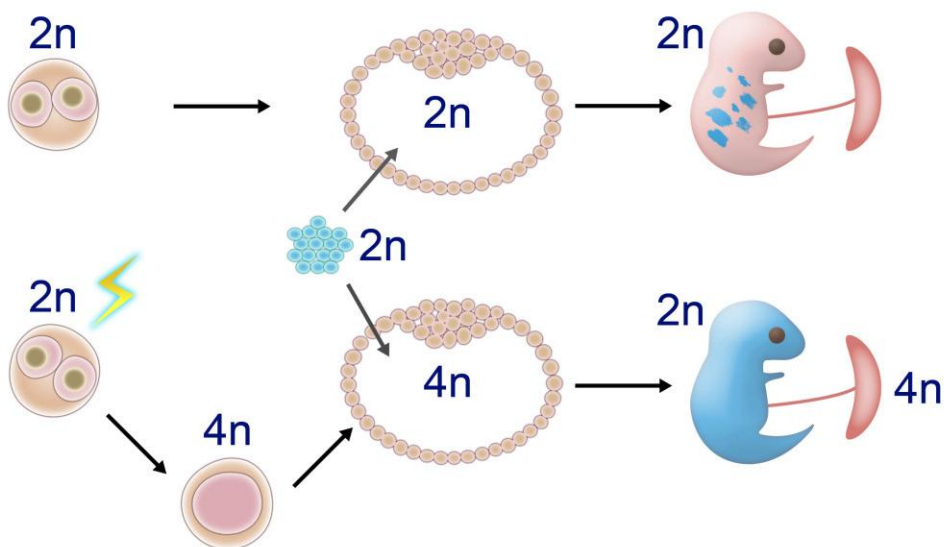


Figure 17 : Capacité de contribution au développement embryonnaire des cellules pluripotentes embryonnaires ou induites (en bleu) : formation de chimères par injection de cellules pluripotentes dans un blastocyste et formation de fœtus entièrement dérivé de cellules pluripotentes par complémentération tétraploïde (Kang and Gao, 2012)

le maintien de la pluripotence et l'auto-renouvellement tels que *Nanog*, *Oct4*, *Sox2*, *Lin28*, *Fbxo15* et *Rex1* (ou *Zfp42*). Les cellules souches embryonnaires de souris sont également positives pour l'expression de la phosphatase alcaline, bien que ce marqueur ne soit pas spécifique des cellules pluripotentes.

Le profil épigénétique des mESCs est spécifique. Il est caractérisé par une distribution bimodale des niveaux de méthylation des CpGs, qui sont soit largement non méthylés, comme dans les régions denses en CG, soit largement méthylés dans les zones peu denses. Mais certaines régions ne répondent pas à cette règle. En effet, le profil de méthylation de l'ADN est plutôt fortement corrélé au profil de méthylation des histones : la présence de H3K4me2 et/ou H3K4me3 est caractéristique d'une région non méthylée tandis que la présence de H3K27me3 (ou H3K9me3) est spécifique des régions fortement méthylées. Les promoteurs à forte densité de CpGs associés aux gènes impliqués dans le développement sont enrichis en H3K4me3, une marque active, mais aussi en H3K27me3 qui est une marque répressive (Figure 16). Ces promoteurs à marque bivalente sont fortement méthylés et les gènes associés sont généralement éteints dans les cellules ES. Les promoteurs à faible densité en CpG, associés à des gènes d'expression tissu-spécifique, sont généralement méthylés à l'exception de ceux qui sont enrichis en H3K4me3 ou H3K4me2 et présentent des niveaux réduits de méthylation. Le profil de méthylation de l'ADN, ainsi que celui des histones, évolue au cours de la différenciation cellulaire (Meissner et al., 2008).

iii. Capacités fonctionnelles des mESCs

Les mESCs sont capables de se différencier en tous les types cellulaires, ce qui est confirmé *in vitro* par la formation de corps embryoïdes exprimant des marqueurs du mésoderme, de l'ectoderme et de l'endoderme, ainsi que par leur différenciation dirigée vers des lignages spécifiques en présence de molécules activatrices (cardiomyocytes (Boheler et al., 2002), mélanocytes (Nissan et al., 2011), cellules neurales (Bibel et al., 2004), cellules germinales (Geijsen et al., 2004), cellules hématopoïétiques (Schmitt et al., 1991), cellules vasculaires (Yamashita et al., 2000), etc.). *In vivo*, l'injection de cellules pluripotentes dans un tissu mène à la formation de tératomes, tumeurs constituées de cellules étrangères à la région injectée, à différents niveaux de différenciation et pouvant appartenir aux différents feuillets embryonnaires. De plus, des mESCs injectées dans un blastocyste receveur sont capables de coloniser cet embryon au cours du développement et ainsi de former des chimères, certaines avec transmission germinale (Pease and Williams, 1990) (Figure 17). Enfin, ces cellules ont la capacité de générer des souris adultes dérivées d'ESCs par complémentation tétraploïde, c'est-à-dire en intégrant des cellules ES dans un embryon tétraploïde, obtenu par la fusion des deux cellules diploïdes du stade 2C du développement. Le fœtus est alors uniquement dérivé des cellules ES tandis que les tissus extra-embryonnaires sont dérivés de l'embryon receveur tétraploïde,

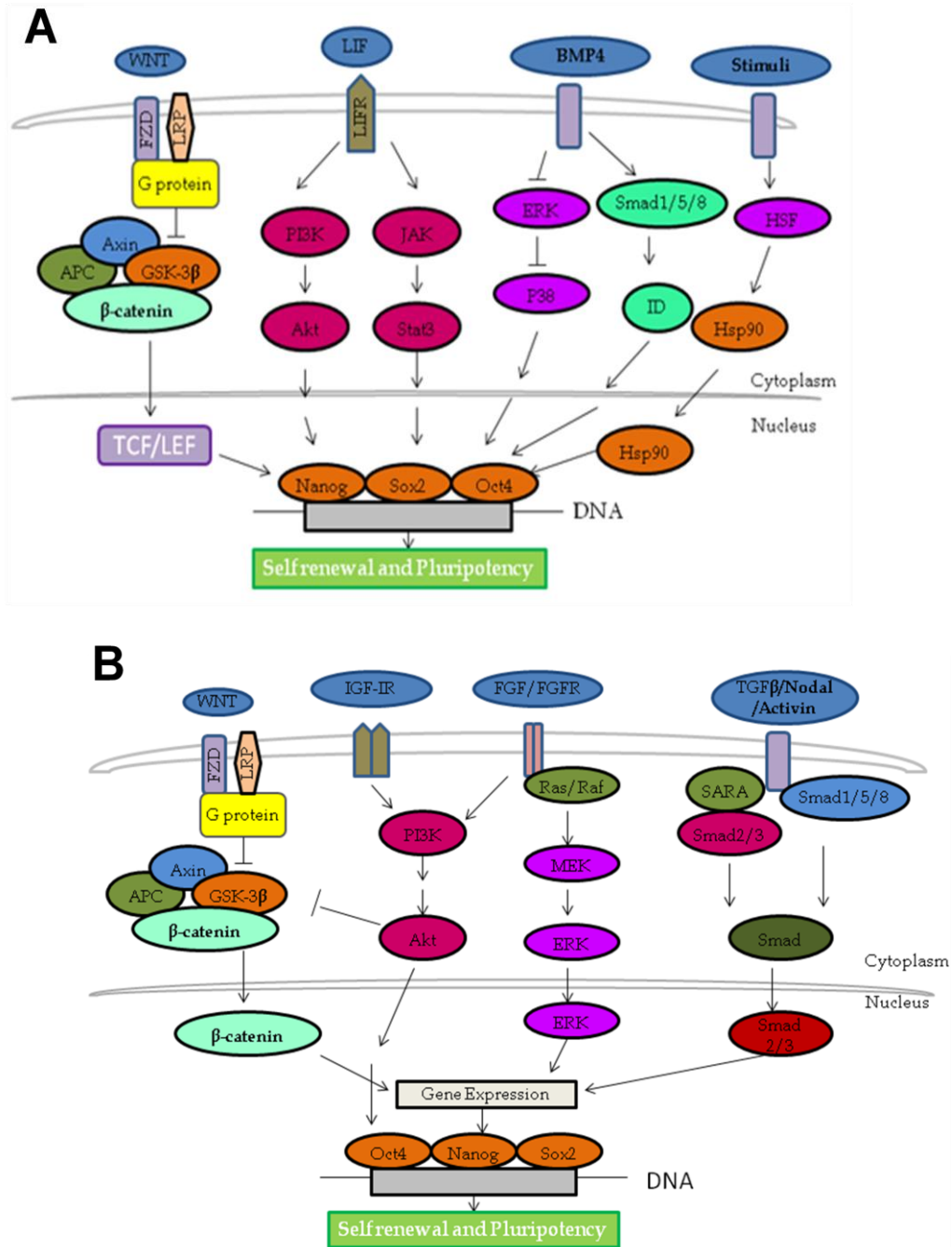


Figure 18 : Principales voies de signalisation régulant le maintien de la pluripotence naïve et amorcée (Bieberich and Wang, 2013). **A.** Maintien de la pluripotence naïve par l'action conjuguée des voies LIF/JAK/STAT, BMP4 et WNT **B.** Maintien de la pluripotence amorcée par les voies WNT, FGF/MEK/ERK et Activin/Nodal (TGFβ).

prouvant ainsi la pluripotence complète des cellules ES testées (Nagy et al., 1990; Eggan et al., 2001) (Figure 17).

b. Deux états de pluripotence définis : amorcé et naïf

i. Caractéristiques des lignées mEpiSCs et comparaison avec les mESCs

Chez la souris, des cellules souches pluripotentes peuvent également être prélevées plus tardivement au cours du développement embryonnaire, dans l'épiblaste à J5.5 (Brons et al., 2007; Tesar et al., 2007). Les cellules ainsi dérivées sont appelées cellules souches de l'épiblaste (mEpiSCS). Bien qu'elles expriment des marqueurs de pluripotence et soient capables de se différencier dans de multiples lignages, elles diffèrent légèrement des mESCs. On parle alors d'un stade amorcé de pluripotence, contre un stade dit naïf représenté par les mESCs (Nichols and Smith, 2009). Les mEpiSCs forment des colonies plus plates, sont incapables de former des chimères, ont déjà inactivé le chromosome X et expriment quelques marqueurs de différenciation précoce comme *Gata6* (Hanna, 2010). Les mESCs sont LIF-dépendantes, ce qui signifie que leur état pluripotent est maintenu en présence de LIF (Leukemia Inhibitory Factor) dans le milieu de culture via l'activation de la voie de signalisation de JAK/STAT3 et la phosphorylation de STAT3 sur la Tyr750 (Ying et al., 2008) (Figure 18A). En présence de bFGF (basic Fibroblast Growth Factor) un effecteur de la voie MEK/ERK/FGF, les cellules mES se différencient. A l'inverse, les mEpiSCs dépendent de la présence de bFGF pour maintenir leur état pluripotent (Figure 18B). Ces deux populations se démarquent également par l'expression de gènes caractéristiques de l'état naïf (*Rex1*, *Stella*, *Dazl*, *Dax1*, etc.) ou de l'état amorcé (*Gata6*, *Dkk1*, *Otx2*, *Lefty2*, etc.) (Tesar et al., 2007), par la durée de la phase G1 du cycle cellulaire qui est raccourcie à l'état naïf (Coronado et al., 2013) ainsi que par leur profil épigénétique, en particulier une forte déméthylation de l'ADN à l'état naïf (Bao et al., 2009).

ii. Plasticité entre les deux états de pluripotence

Il est possible de reprogrammer des cellules amorcées vers l'état naïf de pluripotence en modifiant les conditions de culture (Silva et al., 2008; Ying et al., 2008), comme l'indique la Figure 19. L'utilisation de milieu contenant du LIF et des inhibiteurs permet de moduler les voies de signalisation intracellulaires de manière à modifier le niveau de pluripotence des cellules :

- Le LIF active la voie JAK/STAT3 caractéristique de l'état naïf.
- Des molécules inhibent la voie bFGF/MEK/ERK comme l'inhibiteur de Mek (PD0325901) et l'inhibiteur de FGFR (PD161570) (Ying et al., 2008)
- L'inhibition de GSK3 β permet d'activer la voie WNT : Lorsque GSK3 β est inactivé, la β -caténine n'est plus phosphorylée ni dégradée par les protéasomes, elle s'accumule dans le noyau cellulaire et entraîne l'activation de la voie WNT canonique. Cette voie joue un rôle

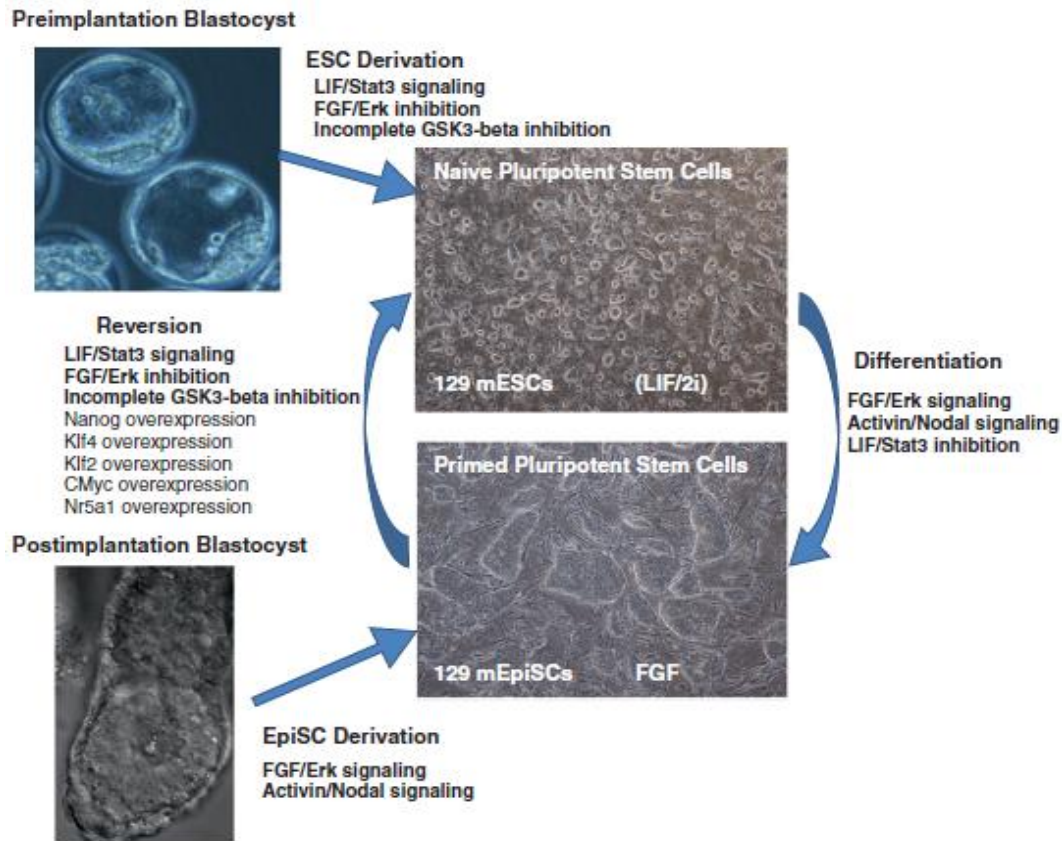


Figure 19 : Les différents états de pluripotence dans les cellules souches embryonnaires murines (De Los Angeles et al., 2012). La dérivation de cellules de l'ICM murin produit des cellules pluripotentes naïves LIF-dépendantes tandis que la dérivation de cellules de l'épiblaste produit des cellules pluripotentes amorcées bFGF-dépendantes. Ces deux états peuvent être interchangés en modifiant les conditions de culture pour activer/désactiver les voies de signalisations dont dépendent ces deux états.

dans la maintenance de la pluripotence en favorisant l'expression continue de marqueurs tels que *Rex1*, *Oct4*, *Sox2* et *Nanog* (Sato et al., 2004; Reya and Clevers, 2005). Il est intéressant de noter que la β -caténine active aussi c-Myc et pourrait donc jouer un rôle dans la reprogrammation cellulaire et le contrôle du cycle cellulaire (Zhang et al., 2014).

iii. *Caractéristiques des cellules souches embryonnaires humaines*

Chez l'homme, les cellules prélevées dans la masse cellulaire interne du blastocyste et appelées hES sont plus proches des cellules murines EpiSCs que des ES à la fois sur le plan morphologique, expressionnel, sur leur dépendance à la voie de signalisation bFGF/Activin et une propension à posséder un chromosome X inactif. De plus, l'utilisation d'un milieu de culture contenant du LIF, un inhibiteur de la voie bFGF/MEK/ERK et un inhibiteur de GSK3 β , noté LIF/2i, n'est pas suffisante pour l'obtention de lignées de cellules souches embryonnaires dans l'état naïf de pluripotence (De Los Angeles et al., 2012). Ce n'est qu'en développant un nouveau milieu bien plus complexe que Gafni et collaborateurs ont obtenus des lignées humaines répondant aux critères caractérisant l'état naïf de pluripotence (Gafni et al., 2013). Les lignées humaines seraient donc moins plastiques que les lignées murines, bien qu'elles soient sensibles aux mêmes voies de signalisations que les mEpiSCs pour leur maintenance (voie TGF β /Activin, cascade bFGF/MEK/ERK) ou leur différenciation (voie LIF/STAT3) (Pera and Tam, 2010).

2. Développement des cellules pluripotentes induites par reprogrammation cellulaire

a. Les différentes méthodes de reprogrammation

i. Du transfert de noyau à la reprogrammation cellulaire

Les premières avancées dans le domaine de la reprogrammation ont eu lieu dès les années 50, avec le développement de la technique de clonage par transfert de noyau (SCNT). Le développement d'un organisme entier suite à la transplantation du noyau d'une cellule somatique dans un ovocyte énucléé a mis en avant l'existence de la reprogrammation nucléaire : le noyau d'une cellule somatique contient toute l'information nécessaire au développement de l'embryon, et placés dans un contexte adéquat ces gènes sont réactivés (Gurdon, 1975). Le SCNT est donc une technique permettant de produire *in vitro* des cellules souches totipotentes et pluripotentes, mais elle présente de nombreux désavantages tels que l'apparition de remaniements chromosomiques ou de déficiences développementales sévères des organismes ainsi obtenus (Wilmut et al., 1997; Wakayama et al., 1998; Thuan et al., 2010).

La fusion cellulaire est une autre technique qui permet, en fusionnant une cellule somatique avec une cellule pluripotente, de transformer la cellule somatique en cellule pluripotente (Stadtfield and Hochedlinger, 2010; Yamanaka and Blau, 2010). En étudiant le profil d'expression des cellules ES et les mécanismes propres à la fusion cellulaire, qui indiquent une réactivation du gène Oct4 au cours de la fusion, Yamanaka et Takahashi (Takahashi and Yamanaka, 2006) ont fait l'hypothèse que les facteurs jouant un rôle dans le maintien de l'identité des cellules ES (comme *Oct4* et *Nanog*) pouvaient aussi prendre part à l'induction de la pluripotence. En infectant des cellules somatiques murines, puis humaines, avec un cocktail de quatre de ces gènes, *Oct4*, *Sox2*, *Klf4* et *c-Myc*, ils ont rendu possible la reprogrammation cellulaire et générés les premières cellules souches pluripotentes induites murines (miPSCs) et humaines (hiPSCs), possédant des caractéristiques morphologiques, moléculaires et fonctionnelles similaires à celles des cellules souches embryonnaires (Takahashi and Yamanaka, 2006; Takahashi et al., 2007). Il est également possible de faire passer une cellule d'un type cellulaire à un autre en utilisant d'autres facteurs que les 4 déterminés ici : ce phénomène est appelé la transdifférenciation.

ii. *Evolution des techniques de reprogrammation cellulaire*

La technique de reprogrammation par infection rétrovirale des quatre facteurs de transcription *Oct4*, *Sox2*, *Klf4* et *c-Myc* a été largement utilisée pour dériver des lignées de cellules iPS dans différentes espèces et types cellulaires (Aasen et al., 2008; Liu et al., 2008; Lowry et al., 2008; Park et al., 2008b; Qin et al., 2008; Li and Ding, 2010; Rajarajan et al., 2012; Tancos et al., 2012; Osteil et al., 2013). Dans la majorité des cas, un gène rapporteur fluorescent comme l'eGFP (Enhanced Green Fluorescent Protein) est ajouté à la suite des 4 facteurs pour pouvoir suivre en temps réel l'entrée des facteurs exogènes dans les cellules et pour sélectionner les cellules transduites. Mais la transduction par rétrovirus comporte des effets négatifs, car elle nécessite l'intégration de l'ADN viral dans le génome de la cellule hôte. Suivant les sites d'insertion, cela peut donc entraîner des mutations génétiques et augmenter les risques liés au développement de cancers, limitant ainsi les applications en santé humaine. De plus, il existe un risque de réactivation des gènes exogènes qui sont normalement éteints une fois la reprogrammation achevée.

Afin de contourner ces limitations, plusieurs méthodes ont été mises en œuvre pour optimiser la transduction virale comme l'utilisation de vecteurs lentiviraux polycistroniques (Carey et al., 2009; Sommer and Mostoslavsky, 2010), pouvant aussi être dotés d'un système d'induction temporaire des transgènes en présence de doxycycline (Brambrink et al., 2008; Stadtfield et al., 2008a). Un autre avantage est que l'efficacité de la reprogrammation est augmentée par l'utilisation des lentivirus. Le système inductible a été utilisé pour développer les cellules iPS dites « de seconde génération » : des iPSCs générées par une première reprogrammation cellulaire sont différenciées soit *in vitro* soit *in vivo* par réinjection dans le blastocyste, puis les cellules somatiques ainsi obtenues sont reprogrammées une

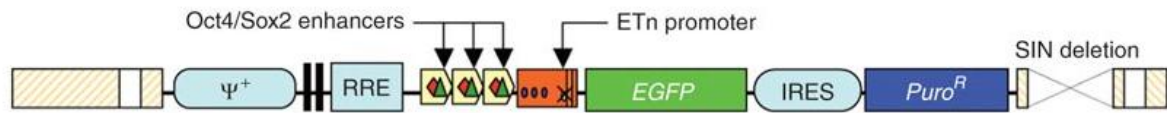


Figure 20 : Construction PL-SIN-EOS-C(3+)-EiP permettant la sélection des cellules reprogrammées par l'expression de l'eGFP et résistance à la puromycine (Hotta et al., 2009a). RRE = Rev Response Element ; ETn = Early Transposon ; IRES = Internal Ribosome Entry Site ; SIN = Self inactivating

seconde fois avec une plus grande efficacité en les cultivant dans un milieu contenant la doxycycline (Wernig et al., 2008a; Nagy, 2013).

Pour éviter l'intégration de transgènes dans le génome de la cellule, des systèmes non intégratifs ont été mis au point (Sommer and Mostoslavsky, 2010; Stadtfeld and Hochedlinger, 2010). Parmi eux, on peut citer les vecteurs plasmidiques contenant des sites loxP permettant l'excision des transgènes après expression de la Cre-recombinase (Kaji et al., 2009), les adénovirus non intégratifs (Stadtfeld et al., 2008b), les virus Sendai (Fusaki et al., 2009), les épisomes qui sont naturellement supprimés dans les cellules mais dont l'efficacité est limitée (Yu et al., 2009a), les transposons tels que le piggyBac qui nécessite l'expression de la transposase pour être excisé (Woltjen et al., 2009; Yusa et al., 2009) et les plasmides pouvant être porteurs d'unités polycistroniques (Okita et al., 2008; Gonzalez et al., 2009) et du système d'induction par la doxycycline (Ye et al., 2010). Toutes ces études ont permis de démontrer qu'une expression temporaire des facteurs de reprogrammation est suffisante pour induire la pluripotence, bien que l'efficacité de reprogrammation puisse être diminuée (Stadtfeld and Hochedlinger, 2010).

Des modifications de la composition du cocktail de facteurs de reprogrammation ont été testées. Des lignées ont été dérivées en modulant les niveaux des différents facteurs (Carey et al., 2011), ou en modifiant le cocktail de reprogrammation, notamment en supprimant l'oncogène *c-MYC* ou en utilisant *NANOG* et *LIN28* en remplacement ou en complément des marqueurs initialement décrits (Yu et al., 2007; Huangfu et al., 2008b; Nakagawa et al., 2008; Wernig et al., 2008b; Nakagawa et al., 2010). De petites molécules ont été utilisées pour augmenter l'efficacité de la reprogrammation cellulaire telles que la vitamine C qui agit sur la sénescence cellulaire (Esteban et al., 2010b), des modificateurs des marques épigénétiques comme l'acide valproïque, un inhibiteur des histones désacétylases qui accroît la reprogrammation d'un facteur 100 ou la 5-azacytidine qui inhibe les DNA méthyltransférases (Huangfu et al., 2008a; Mikkelsen et al., 2008).

L'efficacité de la reprogrammation cellulaire étant assez faible (0.1%-1% par les méthodes intégratives (Stadtfeld and Hochedlinger, 2010)), il est important de sélectionner avec précision les clones reprogrammés. Pour ce faire, Hotta et al. ont construit le vecteur lentiviral EOS, pour « Early Transposon promoter (ETn) and Oct4 and Sox2 promoters », qui marque les colonies iPS émergentes par la protéine fluorescente eGFP et permet leur sélection par sa résistance à la puromycine (Figure 20) (Hotta et al., 2009a; Hotta et al., 2009b). En effet, lorsque les cellules se reprogramment, le recrutement des activateurs de transcription SP1 et SP3 dans la région promotrice de ETn, qui est spécifique des cellules indifférenciées, accompagné de celui des facteurs de transcription OCT4 et SOX2 sur leurs enhanceurs respectifs, déclenche l'expression de la eGFP.

L'avantage de ce large éventail de techniques est qu'elles sont facilement combinables, comme le démontre l'expérience de Tsukiyama en 2011 qui utilise un transposon piggyBac dox-inductible et le système de sélection EOS (Tsukiyama et al., 2011).

Enfin, ces 5 dernières années de nouvelles techniques sont apparues n'utilisant pas de vecteurs viraux ou plasmidiques. Des lignées iPSCs murines ont été établies en apportant les facteurs de reprogrammation par le biais de protéines recombinantes (Zhou et al., 2009) ou d'ARNs messagers (Warren et al., 2010). La transfection de microARNs spécifiquement surexprimés dans les lignées iPS ou ES a également permis de reprogrammer des cellules murines et humaines (Miyoshi et al., 2011). Certains composés chimiques ayant un effet très positif sur la reprogrammation cellulaire, il est aussi possible de dériver des lignées iPS uniquement par l'action de ces composés (Li and Ding, 2010).

b. Le mécanisme de la reprogrammation cellulaire

i. Un processus stochastique ou déterministe ?

La faible efficacité de la reprogrammation cellulaire, associée au développement de différentes techniques, ont amené plusieurs équipes à se pencher sur les mécanismes moléculaires sous-tendant ce phénomène complexe. Deux modèles sont apparus pour expliquer le faible taux de reprogrammation (Yamanaka, 2009). Le modèle élite fait l'hypothèse que seul un faible pourcentage des cellules a la capacité de se reprogrammer. D'une part, ces cellules pourraient être une sous population de cellules moins différenciées présentes dans la population initiale, avec un taux de présence correspondant au taux de reprogrammation observé. Cette idée est mise en défaut par la reprogrammation réussie de cellules fortement différenciées de type lymphocytes B et T (Hanna et al., 2008; Loh et al., 2010) ainsi que par l'obtention d'efficacités de reprogrammation bien supérieures à celles initialement décrites au fur et à mesure de l'évolution des techniques. D'autre part, le modèle élite suggère que certaines cellules pourraient acquérir la compétence à se reprogrammer au cours de l'intégration virale. Ici encore, l'apparition de nouvelles techniques telles que les systèmes non-intégratifs, ainsi que l'absence de sites d'intégration communs entre différentes lignées reprogrammées par le même protocole, viennent contredire ce modèle. Néanmoins, les cellules les moins différenciées sont reprogrammées avec une efficacité supérieure.

Le second modèle, le modèle stochastique, est capable d'expliquer ce phénomène. Il fait l'hypothèse que toutes les cellules ont la capacité de se reprogrammer mais que seule une faible partie d'entre elles réussit à franchir toutes les étapes conduisant à un état pluripotent stable. Ce modèle permet d'expliquer pourquoi les cellules les moins différenciées sont reprogrammées plus facilement : ayant déjà certaines caractéristiques des cellules souches, elles auraient moins d'étapes à franchir pour atteindre l'état pluripotent. Pour appuyer cette hypothèse, une étude a montré que les cellules

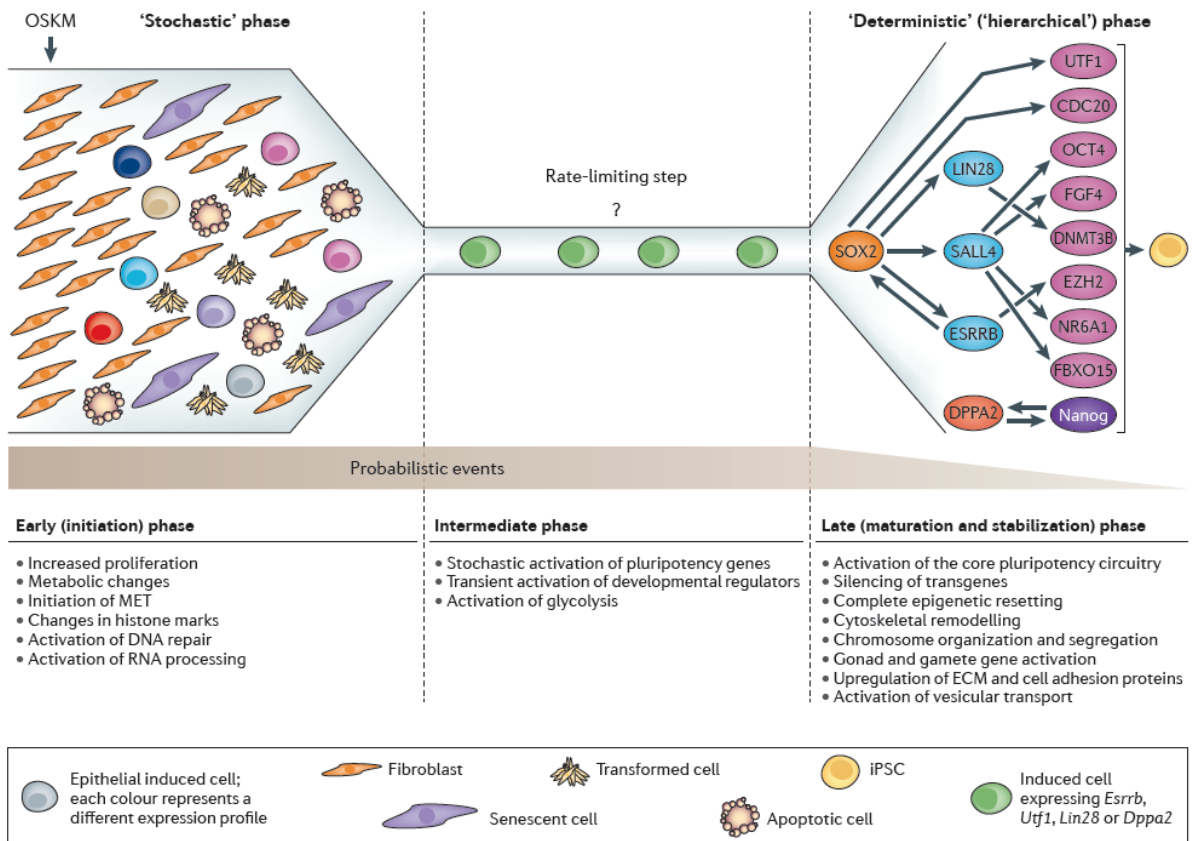


Figure 21 : Le modèle stochastique et les différentes étapes de l'établissement de la pluripotence stable lors de la reprogrammation cellulaire (Buganim et al., 2013).

exprimant déjà certains des quatre facteurs comme *SOX2* peuvent être reprogrammées sans leur surexpression (Kim et al., 2009). De plus, ce modèle explique aussi pourquoi l'ajout de différentes molécules permet d'augmenter l'efficacité de la reprogrammation, car ces molécules aideraient les cellules à franchir les différentes étapes nécessaires. Mais quelles sont ces différentes étapes ?

ii. *Les étapes fondamentales de la reprogrammation cellulaire*

La reprogrammation cellulaire est un phénomène complexe qui peut être découpé en trois stades : initiation, maturation et stabilisation (Buganim et al., 2013) (Figure 21). L'étape d'initiation est caractérisée par l'augmentation de la prolifération cellulaire et l'extinction de gènes caractéristiques de la lignée somatique (Mikkelsen et al., 2008). La sur-régulation de gènes tels que *c-Myc* permet de diminuer la longueur de cette étape en augmentant la prolifération cellulaire (Mikkelsen et al., 2008; Smith et al., 2010), corroborant le fait que *c-Myc* est dispensable au processus de reprogrammation mais en accroît l'efficacité. Un fort taux de prolifération cellulaire, caractérisé par une phase G1 du cycle cellulaire courte, est en effet requis pour l'établissement de la pluripotence en favorisant notamment la déméthylation passive (Ruiz et al., 2011). De plus, l'inactivation de la voie p53 qui régule la mort cellulaire fait partie des premières étapes du processus de reprogrammation (Stadtfield and Hochedlinger, 2010). La première phase est également caractérisée par un changement de morphologie rapide des cellules. Dès les premières divisions cellulaires, leur taille diminue (Smith et al., 2010). Elles subissent une transition mésenchymo-épithéliale (MET), à l'inverse de ce qui se produit au cours du développement, au cours de laquelle l'expression de la cadherine E, un marqueur épithélial, est augmentée via l'action de KLF4, tandis que des marqueurs mésenchymateux sont réprimés (Li et al., 2010). Les facteurs de reprogrammation OCT4, SOX2, KLF4 et c-MYC bloquent l'activation de la voie TGFβ qui s'oppose à cette transition. Cela appuie le fait que la voie de signalisation TGFβ/Activin soit spécifique des cellules amorcées et non de la pluripotence naïve (Hanna, 2010). Enfin, les quatre facteurs sur-régulent des miARNs favorisant la MET et sous-régulent des acteurs favorisant transition opposée (Li et al., 2010). Ces premiers phénomènes sont observables avant l'apparition de l'expression de marqueurs de pluripotence précoces tels que *SSEA1* (Stadtfield et al., 2008a; Li et al., 2010).

Les cellules en cours de reprogrammation subissent un remodelage épigénétique profond de manière à établir un épigénome similaire à celui observé dans les ESCs (Maherali et al., 2007). L'étude de lignées partiellement reprogrammées a permis d'identifier ce remodelage comme une barrière importante à franchir pour atteindre l'état de pluripotence stable (Mikkelsen et al., 2008), ce qui est confirmé par l'augmentation de l'efficacité de la reprogrammation en présence de 5-azacytidine (Huangfu et al., 2008a; Mikkelsen et al., 2008). La répression de gènes associés à la différenciation est possible via la liaison des marqueurs de reprogrammation sur les promoteurs de ces gènes et le recrutement consécutif de complexes modelant la chromatine pour apposer la marque H3K27me3,

signe de répression, sur les promoteurs (Mikkelsen et al., 2008; Stadtfeld and Hochedlinger, 2010). Ailleurs, l'apparition de méthylation sur l'histone H3 en K4 indique un état de chromatine ouverte (Mikkelsen et al., 2008). Ce phénomène s'accompagne de l'expression stochastique de nombreux gènes de pluripotence. Parmi ces gènes précoces, certains ne sont pas prédictifs du succès de la reprogrammation tels que *Sall4*, *Fbxo15*, *Oct4* et *Fgf4* (Buganim et al., 2012), c'est pourquoi les lignées sélectionnées sur l'expression de *Fbxo15* ne sont pas totalement pluripotentes (Maherali et al., 2007; Stadtfeld et al., 2008a).

L'analyse sur cellule unique menée par Buganim et collaborateurs (Buganim et al., 2012) a permis d'identifier les gènes dont l'expression chez la souris est prédictive d'une reprogrammation stable : *Esrrb*, *Utf1*, *Lin28* et *Dppa2*. L'activation du circuit central de la pluripotence se fait tardivement, au cours de la phase de maturation, c'est pourquoi l'expression des transgènes est nécessaire jusqu'à 10 jours après leur surexpression (Stadtfeld et al., 2008a). Les gènes endogènes sont alors activés, comme l'indique la déméthylation du promoteur du gène *Oct4*, ainsi que celle de promoteur du gène *Nanog* dont l'activation est une étape-clé (Maherali et al., 2007; Buganim et al., 2012). Mais d'après Buganim et collaborateurs, c'est surtout l'activation du *Sox2* endogène qui active le circuit central des gènes nécessaires au maintien de la pluripotence, et ce de façon déterministe (Buganim et al., 2012). Une fois ce circuit activé, dans les dernières étapes de la reprogrammation cellulaire, les transgènes sont éteints et le remodelage épigénétique s'achève avec l'inactivation d'un chromosome X pour les lignées femelles (Mikkelsen et al., 2008; Stadtfeld et al., 2008a; Buganim et al., 2013).

Cette activation stochastique de gènes de pluripotence, suivie d'un processus déterministe entraînant la stabilisation des lignées, explique la possibilité de reprogrammer des cellules en utilisant des cocktails de reprogrammation différents. Ces différents facteurs fonctionnent en réseau et permettent d'activer le même programme. Ils contrôlent aussi des microARNs, comme la famille let-7 dont l'expression bloque la reprogrammation en interférant avec l'activation de *Lin28* et de *Sall4* ou *miR-294* dont l'effet est inverse (Viswanathan et al., 2008; Stadtfeld and Hochedlinger, 2010; Thornton and Gregory, 2012). Enfin, les lincARN (« large intergenic non coding RNA ») jouent également un rôle en réprimant l'activation des voies de stress et de mort cellulaire telles que la voie p53, dont l'inactivation représente une autre étape importante au cours de la reprogrammation (Loewer et al., 2010).

Récemment, l'équipe de Hanna (Rais et al., 2013) a obtenu une efficacité de reprogrammation proche de 100% en levant la barrière épigénétique via la déplétion de *Mbd3*. Le gène *Mbd3* code pour une sous-unité du complexe NurRD (« Nucleosome remodeling and histone deacetylase »). Son rôle dans la pluripotence a été décrit par l'étude de cellules mES *Mbd3*^{-/-} dont les capacités de différenciation sont réduites (Kaji 2006). Ce remodeleur de la chromatine est recruté par les facteurs de reprogrammation et bloque alors la conversion de cellules amorcées en cellules naïves en réprimant la réactivation de gènes de la pluripotence. Ce résultat met en exergue l'importance du franchissement de

la barrière épigénétique et identifie le complexe répresseur Mbd3/NuRD comme une cible majeure pour l'amélioration de la reprogrammation cellulaire.

c. Un outil à fort potentiel

i. *De nombreuses applications thérapeutiques envisageables*

Les propriétés d'auto-renouvellement et de différenciation des cellules souches pluripotentes en font un outil d'intérêt majeur dans le développement de nouvelles applications thérapeutiques. Cependant, l'utilisation de cellules ES humaines est limitée par le besoin de recourir à des embryons pour isoler les lignées cellulaires. Les cellules souches pluripotentes induites possèdent les mêmes capacités de différenciation que les cellules ES et sont capables de se différencier en de nombreux types cellulaires tels que la lignée hématopoïétique, cardiovasculaire, neurale ou adipocytaire (Boué et al., 2010). Elles offrent donc une alternative aux cellules ES particulièrement intéressante car elles permettent d'utiliser du matériel biologique du patient concerné et pourraient permettre le développement de greffes autosomales (Saha and Jaenisch, 2009). Plusieurs lignées issues de patients porteurs de maladies ont déjà été établies (Park et al., 2008a) parmi lesquelles on peut citer les hiPS dérivées de patients souffrant de la maladie de Parkinson (Soldner et al., 2009), des syndromes d'Angelman et de Prader-Willi (Chamberlain et al., 2010), de désordres métaboliques du foie (Rashid et al., 2010), etc... Ces cellules représentent un atout majeur pour modéliser les maladies génétiques, tester des molécules thérapeutiques, mais aussi pour corriger *in vitro* une mutation causale, différencier les cellules et les réinjecter dans le tissu malade pour régénération (Yusa et al., 2011).

ii. *Un exemple d'application : la différenciation germinale mâle*

Parmi les diverses voies de différenciation possibles des cellules pluripotentes, plusieurs équipes se sont concentrées sur la différenciation en cellules germinales. Un premier protocole établi à partir de cellules mES utilise la formation de corps embryoïdes suivi d'un traitement à l'acide rétinoïque, qui stimule la prolifération des PGCs et initie la méiose (Bowles and Koopman, 2007), pour induire la formation de cellules primordiales germinales mâles (Geijsen et al., 2004). L'utilisation de gènes rapporteurs spécifiques de cellules germinales ou de la méiose tels que *Prmd1*, *Ddx4*, ou *Stra8* (Toyooka et al., 2003; Nayernia et al., 2006; Tilgner et al., 2010) a ensuite permis d'isoler ces cellules et, grâce à l'induction continue par l'acide rétinoïque, d'obtenir des cellules souches spermatogoniales qui, transplantées dans un testicule, ont poursuivi leur différenciation jusqu'au stade du spermatozoïde (Nayernia et al., 2006). L'injection intracytoplasmique de cellules haploïdes dans des ovocytes a permis le développement d'embryons et la naissance de souriceaux viables (Nayernia et al., 2006). La différenciation en cellules germinales peut être accrue par l'ajout de BMP (« Bone Morphogenetic Protein ») (Kee et al., 2006) et la surexpression de *Dazl* (Kee et al., 2009; Yu et al., 2009b). Des expériences similaires utilisant les miPSCs ont démontré leur capacité à se différencier en cellules

germinales primordiales ayant une spermatogénèse complète une fois transplantées dans les tubes séminifères d'une souris sans spermatogonies. L'utilisation de ces spermatozoïdes lors d'une ICSI (« Intracytoplasmic Sperm Injection ») suivi du transfert des embryons obtenus dans une mère porteuse a donné naissance à une progéniture normale et fertile (Hayashi et al., 2011).

Chez l'homme, la capacité des hESCs à différencier en cellules exprimant des marqueurs méiotiques a été prouvée par Clark et al. en 2004 (Clark et al., 2004). En 2011, la différenciation germinale des cellules hiPS a également été décrite (Panula et al., 2011) en supplémentant le milieu de différenciation par BMP-4,-7 et -8b, en utilisant *DDX4* comme gène rapporteur pour isoler les cellules d'intérêt et en sur-exprimant des gènes de la famille DAZ (« Deleted in AZoospermia ») pour déclencher l'entrée en méiose de ces cellules germinales (Panula et al., 2011). La même année, la méiose complète et l'obtention de cellules haploïdes à partir d'hiPSCs par traitement à l'acide rétinoïque et culture des cellules sélectionnées dans un milieu supplémenté en LIF, bFGF, Forskolin et un inhibiteur de CYP26 a été décrite (Eguizabal et al., 2011).

d. Un outil à améliorer

i. Instabilité génomique des lignées pluripotentes

Avant leur utilisation en thérapie cellulaire, il est important de vérifier que les cellules iPS ne soient pas une source potentielle de danger pour les patients, notamment par un risque cancéreux accru. En effet, en tant que lignées se répliquant rapidement et sans point de contrôles des dommages de l'ADN, les cellules souches pluripotentes sont plus à même d'accumuler des aberrations chromosomiques (Peterson and Loring, 2013). En 2008, Lefort et collaborateurs ont identifié une amplification dans la région q11.21 du chromosome 20 humain concernant 4 lignées hES sur 5 testées (Lefort et al., 2008). Bien que sans association avec une transformation oncogénique, cette amplification est constatée dans plusieurs lignées cancéreuses. Cette région contient des gènes contrôlant la survie cellulaire, la croissance et le cycle cellulaire ; elle conférerait ainsi un avantage prolifératif aux cellules. D'autres régions ont ensuite été identifiées comme couramment amplifiées dans les lignées hESCs ou hiPSCs, en particulier une trisomie du chromosome 12 contenant les gènes de pluripotence *NANOG* et *GDF3*, et des duplications sur le chromosome 17 (Mayshar et al., 2010; Taapken et al., 2011; Martins-Taylor and Xu, 2012). Pour les iPS, bien qu'une étude montre des aberrations chromosomiques près du site d'insertion des transgènes (Kadaja 2009), ces aberrations ont également été identifiées dans des lignées issues de reprogrammation non intégrative (Mayshar et al., 2010). Elles ne dépendraient pas non plus du milieu de culture employé ni de l'utilisation de cellules nourricières (Taapken et al., 2011). Ces anomalies, qui touchent entre 10 et 20% des lignées étudiées, sont récurrentes et rappellent l'instabilité chromosomique de l'embryon (Peterson et al., 2011). Cependant, dans l'embryon, cette instabilité ne présente pas d'anomalies récurrentes comme constaté dans les cellules pluripotentes ; les deux phénomènes ne seraient donc pas liés (Taapken et al., 2011).

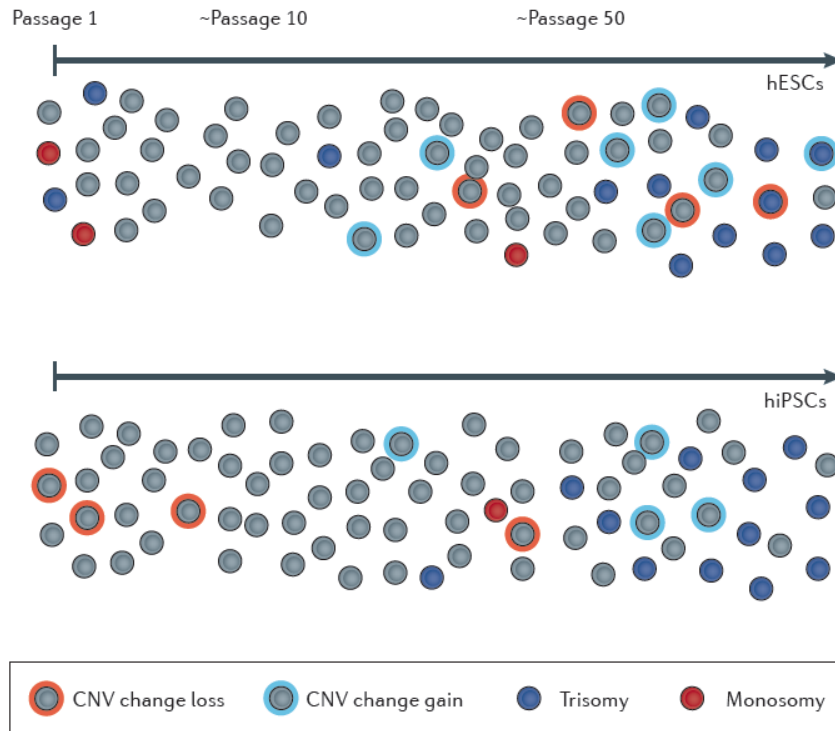


Figure 22 : Incidence des modifications génomiques dans les cellules souches pluripotentes humaines au cours des passages (Lund et al., 2012). Dans les deux types de lignées, les aneuploidies s'accumulent avec le nombre de passages. Les CNVs sont eux déjà présents dès les premiers passages dans les hiPSCs à cause de la reprogrammation cellulaire et s'accumulent avec le temps dans les hESCs.

Les cellules pluripotentes exhibent également une augmentation de la fréquence des variations du nombre de copies (CNV) de gènes ou régions subchromosomiques (Hussein et al., 2011; Laurent et al., 2011). Ici encore, la distribution de ces CNVs est spécifique de certaines régions chromosomiques, et des différences sont observées entre cellules ES et cellules iPS, avec un nombre de CNV deux fois plus important dans les iPSCs (Hussein et al., 2011) ce qui indique un effet propre à la reprogrammation cellulaire. Ces CNVs ciblent des régions spécifiques telles que celles contenant les gènes *OCT4*, *NANOG*, *SOX2*, *c-MYC* et *KLF4* (Laurent et al., 2011; Martins-Taylor et al., 2011). Ils sont acquis soit au cours de la reprogrammation cellulaire, avec notamment la délétion de gènes suppresseurs de tumeur, soit au fil des passages avec par exemple la duplication d'oncogènes (Laurent et al., 2011) (Figure 22). Il est toutefois important de noter que certaines de ces anomalies apparaissant au cours de la reprogrammation peuvent dériver directement du mosaïcisme des cellules présentes dans la population somatique de départ, la reprogrammation étant un procédé clonal qui ne sélectionne que quelques cellules (Ronen and Benvenisty, 2012). Cependant, bien que l'intégration virale ne semble pas être liée à l'apparition des CNVs (Martins-Taylor and Xu, 2012), la dose de facteurs de reprogrammation utilisée pourrait jouer un rôle important (Chen et al., 2014). Les hiPSCs présentent aussi un fort taux de variation de nucléotide (SNV) mais sans récurrence entre les différents clones (Ji et al., 2012; Young et al., 2012). Ces mutations apparaissent au cours de la reprogrammation, bien que certaines soient présentes dans la population de départ et capturées clonalement (Ji et al., 2012; Ronen and Benvenisty, 2012; Young et al., 2012). Cette instabilité chromosomique et génomique pourrait augmenter la tumorigénécité des cellules mais peu d'études à ce jour ont relié ces aberrations génomiques avec la formation de tumeur. L'instabilité chromosomique pourrait même être un avantage, permettant de corriger naturellement dans les lignées iPSCs des anomalies présentes dans les cellules somatiques comme les chromosomes en anneaux (Bershteyn et al., 2014).

ii. *Reprogrammation partielle des iPSCs*

Avant de considérer toute application thérapeutique, il est également important de s'assurer que les cellules iPS possèdent bien des caractéristiques similaires aux cellules ES. Sur le plan fonctionnel, les lignées iPSCs murines sont capables tout comme les mESCs de participer au développement embryonnaire, en permettant la formation de chimères à transmission germinale (Okita et al., 2007) et en récapitulant le développement embryonnaire complet par complémentarité tétraploïde (Zhao et al., 2009). Chez l'homme, des tests fonctionnels aussi approfondis n'existent pas, il est donc difficile de garantir la pluripotence totale des cellules. Bien que les lignées iPS et ES partagent des profils d'expression très proches notamment au niveau du réseau de gènes de pluripotence, des voies métaboliques, de la réparation de l'ADN et du cycle cellulaire, il existe des disparités entre le transcriptome de ces deux populations cellulaires (Boué et al., 2010). De la même manière, bien que le profil de méthylation de l'ADN, l'état du chromosome X et les modifications des queues d'histones soient similaires entre iPS et ES (Mikkelsen et al., 2008; Hanna et al., 2010), il existe des différences

sur le plan épigénétique entre les deux types de cellules souches pluripotentes. La perte de l’empreinte parentale au niveau du locus *Dlk1-Dio3* (Stadtfeld et al., 2010) ou l’hypométhylation du gène soumis à empreinte *Zrsr1* (Chang et al., 2013) sont deux exemples illustrant les variations épigénétiques retrouvées dans les iPS. Les lignées iPS sont suspectées de conserver une mémoire de leur tissu d’origine qui favoriserait leur différenciation vers ce tissu plutôt que vers un autre (Kim et al., 2010; Polo et al., 2010; Zwaka, 2010). Chaque lignée iPS étant un peu différente d’une autre (Boué et al., 2010; Bock et al., 2011), l’hypothèse d’une reprogrammation incomplète des iPS est la plus probable pour expliquer les variations observées. En effet, la reprogrammation épigénétique est un élément clé du processus de reprogrammation cellulaire : la récupération de l’empreinte parentale sur le locus *Dlk1-Dio3* en présence de vitamine C, connue pour augmenter l’efficacité de reprogrammation cellulaire, appuie cette hypothèse (Stadtfeld et al., 2012).

L’étude du profil d’expression et de méthylation de l’ADN est donc envisageable pour prédire leur potentiel de pluripotence et les sélectionner pour l’application en thérapie (Bock et al., 2011). De plus, le remplacement des cellules nourricières murines par du Matrigel ou par des lignées humaines fibroblastiques produisant du bFGF permet l’obtention de lignées libres de protéines animales (Liu et al., 2014). Grâce au développement des conditions de culture et malgré les doutes subsistant encore sur la sécurité sanitaire de l’utilisation des iPS en thérapie, les premiers essais cliniques utilisant des hiPS ont été autorisés en 2013, particulièrement pour corriger la dégénérescence maculaire liée à l’âge (DMLA) (Melville et al., 2013).

3. Le challenge des lignées pluripotentes porcines

a. La pluripotence chez le porc

i. Faible succès de l’établissement de lignées ES porcines

Le développement des cellules souches embryonnaires murines et humaines laissait supposer un développement rapide de lignées de même type chez les espèces d’élevage, et notamment chez le porc. Dans les deux décennies suivant cette découverte, un effort important a été fourni, notamment afin d’utiliser les cellules ES pour le développement d’animaux transgéniques ou en tant que donneurs de noyaux. Le porc est un modèle animal largement utilisé pour l’étude de maladies humaines, et très reconnu dans le domaine des transplantations (Lunney, 2007; Litten-Brown et al., 2010). Sa morphologie, sa physiologie, la taille de ses organes et sa durée de vie en font un modèle plus proche de l’homme que ne l’est la souris. Le développement de cellules ES porcines est donc d’un grand intérêt pour modéliser les maladies génétiques humaines ou générer des organes pour la transplantation, mais également pour approfondir notre connaissance du développement de cet animal et ainsi améliorer les techniques d’élevage.

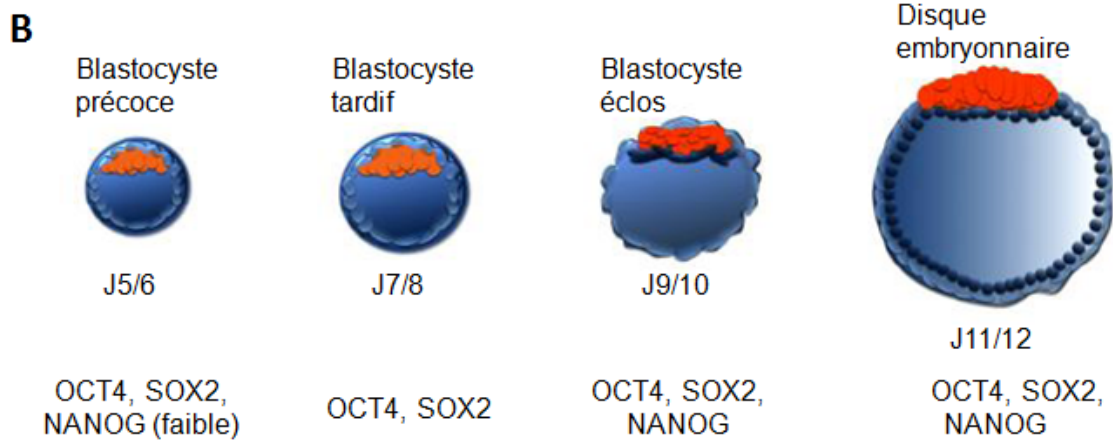
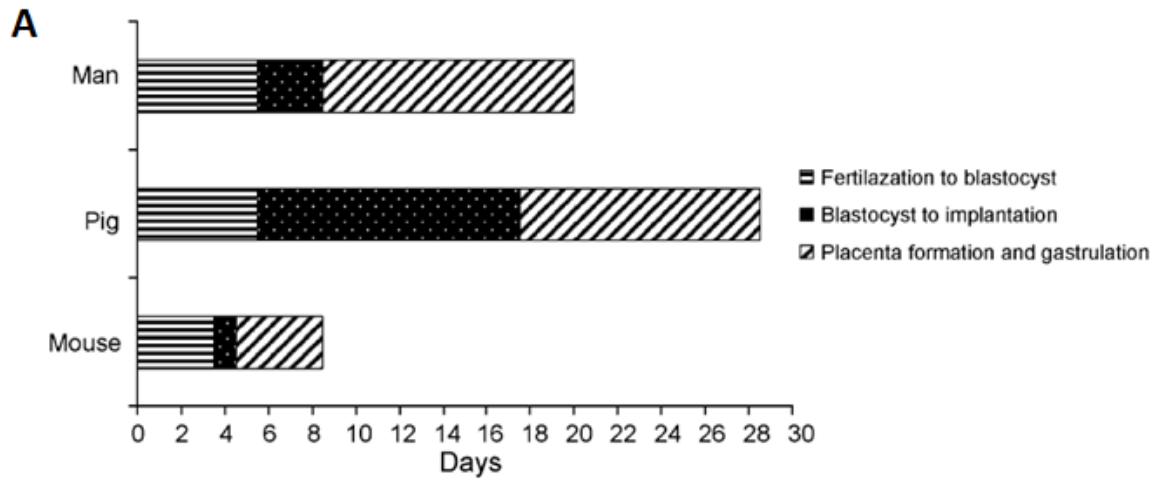


Figure 23 : Développement embryonnaire précoce chez le porc **A.** La phase pré-implantatoire est plus longue chez le porc que chez l'homme et la souris (Brevini et al., 2007a). **B.** Le blastocyste porcine se forme à J5. Il est composé de l'ICM (orange) et du trophectoderme. Il s'étend jusqu'à J8 puis éclos de la zone pellucide. L'ICM se sépare alors en l'épiblaste (rouge) et l'endoderme primitif (noir). Après J10, le trophectoderme polaire se désintègre pour former le stade du disque embryonnaire où les cellules de l'épiblaste sont en contact avec le milieu utérin. L'expression des gènes OCT4 et SOX2 est détectable dans les quatre stades tandis que celle de NANOG évolue au cours de développement.

Les premières lignées porcines isolées à partir de blastocystes entiers prélevées entre J7 et J10 sont apparues en 1990 (Notarianni et al., 1990). A l'exception de l'expérience de Chen et collaborateurs (Chen et al., 1999), aucune des lignées dérivées jusqu'à aujourd'hui n'a été capable de former des chimères, avec ou sans transmission germinale (Brevini et al., 2007a; Brevini et al., 2008). De plus, ces lignées ne sont pas stables dans le temps en raison d'une forte différenciation spontanée. Elles ont été caractérisées principalement sur leur morphologie et sur l'expression de quelques marqueurs de pluripotence tels que *OCT4* et *NANOG* (Brevini et al., 2007b; Blomberg and Telugu, 2012) et ont été développées en utilisant les conditions de cultures des ESCs murines. Ces différents essais ont amené de nombreuses questions : les conditions de cultures des mESCs sont-elles adaptées pour la dérivation de pESCs ? à quel stade est-il plus efficace d'isoler les pESCs ? les marqueurs de pluripotence sont-ils conservés entre les espèces ? la pluripotence porcine dépend-elle des mêmes voies de signalisation ?

ii. *Un développement embryonnaire précoce particulier*

Pour répondre à ces questions, il est important de se pencher sur le développement embryonnaire porcine, qui diffère de celui de l'homme et de la souris à la fois sur le plan temporel et expressionnel (Figure 23). Chez le porc et l'homme, le blastocyste se forme à J5. L'éclosion du blastocyste humain a lieu au cinquième jour, tandis que le blastocyste porcine s'étend et éclos de la zone pellucide à J7-8. Lors de cette éclosion, la masse cellulaire interne (ICM) est séparée en deux populations cellulaires : l'hypoblaste, ou endoderme primitif, et l'épiblaste (Ostrup et al., 2009). La formation de cet épiblaste s'achève à J12, contre J6 chez l'homme. A J9-10, le trophoctoderme polaire du blastocyste porcine se désintègre et l'épiblaste est directement exposé à l'environnement utérin, formant le disque embryonnaire (ED) (Brevini et al., 2007a; Ostrup et al., 2009) (Figure 23). Ce disque s'étend, devient ovale, montre les premiers signes d'une polarité antéro-postérieure et initie la gastrulation avant l'implantation à J14, contre J8 pour l'homme (Hall et al., 2009). La phase pré-implantatoire porcine est donc plus longue que celle de l'homme ou de la souris et présente différentes étapes auxquelles il est possible de prélever de potentielles cellules souches embryonnaires : dans l'ICM avant l'éclosion à J5-7, dans l'épiblaste après l'éclosion entre J8 et J10 ou dans le disque embryonnaire à partir de J10.

Afin de caractériser ces différents stades, plusieurs études se sont intéressées à l'expression de gènes de pluripotence connus chez l'homme et la souris tels que *OCT4*, *NANOG* et *SOX2* (Hall et al., 2009; Ostrup et al., 2009; Cao et al., 2014) (duPuy et al., 2011; Wolf et al., 2011; Rodríguez et al., 2012; Hall and Hyttel, 2014). Dans l'ICM du blastocyste précoce sont exprimés *OCT4*, *SOX2*, *KLF4*, *c-MYC*, et *SSEA1*. *NANOG* est quant à lui exprimé très faiblement, ainsi que d'autres gènes de la pluripotence naïve tels que *REX1* et *DAX1* qui sont très faiblement observés dans l'ICM et le trophoctoderme (TE) (Hall et al., 2009; Rodríguez et al., 2012; Cao et al., 2014; Hall and Hyttel, 2014) Chez la souris et l'homme, la spécification du trophoctoderme est contrôlée par la répression spécifique d'*OCT4* dont l'expression se restreint alors à l'ICM, tandis que *CDX2* permet

spécifiquement le maintien du trophectoderme. Chez le porc, à partir de J7/8, l'expression d'*OCT4* est plus faible dans le TE comparé à l'ICM mais n'est pas totalement réprimée (du Puy; duPuy et al., 2011; Cao et al., 2014). A l'exception d'*OCT4*, peu d'autres gènes sont exprimés dans ce blastocyste tardif, indiquant un état réticent à la pluripotence (Hall and Hyttel, 2014). L'expression de *NANOG* et de *SOX2* augmente à partir de J9 (Hall et al., 2009). Le stade épiblaste, entre J9 et J10, présente de nombreux gènes caractéristiques de la pluripotence naïve - *OCT4*, *NANOG*, *SSEA1*, *SOX2*, *REX1* et *DAX1*. Néanmoins l'expression de certains gènes reste faible, non-spécifique de l'épiblaste ou bien mosaïque comme pour *SSEA1* (Hall et al., 2009; Hall et al., 2010). Dans le disque embryonnaire, le profil d'expression ressemble à celui de cellules de pluripotence amorcée avec l'expression marquée d'*OCT4*, *NANOG*, *SOX2*, *KLF4* et c-MYC et l'extinction de *SSEA1* et *DAX1*. De plus, *REX1* et *KLF2* sont également observables dans une sous-population cellulaire.

L'ajout de bFGF ou de LIF ne permet pas de diminuer la différenciation spontanée des pES, indiquant que les voies du FGF et JAK/STAT pourraient être moins primordiales que chez la souris ou l'homme (Brevini et al., 2007b). Malgré ce, l'ajout d'inhibiteurs de ces deux voies a un effet important sur la prolifération cellulaire. L'absence du récepteur LIFr indique que la voie JAK/STAT est bien présente dans le blastocyste mais serait dépendante d'une autre cytokine que le LIF et son récepteur LIFr (Hall et al., 2009). La voie FGF est bien active au stade du disque embryonnaire, rappelant les cellules ES humaines ou les cellules EpiSCs murines et leur état dit amorcé (Hall et al., 2009). Des acteurs de la voie Activin/Nodal sont également exprimés au cours du développement pré-implantatoire (Hall and Hyttel, 2014). Ces phénomènes corroborent l'hypothèse d'une dépendance des cellules dérivées de l'épiblaste porcin à J10.5 à ces deux voies de signalisation (Alberio et al., 2010; Rodríguez et al., 2012). Bien que les principales voies de régulation de la pluripotence murine se retrouvent dans le développement embryonnaire précoce porcin (bFGF, JAK/STAT, WNT, TGF β /Activin/Nodal), elles ne semblent pas être activées par les mêmes marqueurs moléculaires (Cao et al., 2014; Hall and Hyttel, 2014). D'une manière plus générale, les ICMs murins et porcins partagent l'expression de gènes du cycle cellulaire, du développement, de la régulation de la transcription, mais l'ICM du porc active spécifiquement la voie des acides gras, suggérant un rôle particulier de cette voie métabolique dans le développement embryonnaire porcin (Cao et al., 2014).

Malgré toutes ces avancées permettant de mieux caractériser les acteurs de la pluripotence porcine, les seules lignées isolées à partir de l'ICM montrant des caractères de la pluripotence naïve et pouvant être maintenues en culture dans du milieu contenant du LIF ont été obtenues en sur-exprimant *KLF4* et *OCT4* par infection lentivirale afin de reprogrammer en partie les cellules vers un état de pluripotence stable (Telugu et al., 2011).

Tableau 4 : Caractéristiques des différentes expérimentations de reprogrammation cellulaire chez le porc

Reference	Système de reprogrammation	Cytokines	Extinction des transgènes	Tératomes	Chimère	Etat
Kues 2013	Sleeping Beauty transposon OSKM	bFGF	non	oui	NR	bFGF-dépendant
Kwon 2013	lentivirus OSKMLN	bFGF LIF	non	oui	NR	LIF-dépendant
Fujishiro 2012	rétrovirus OSKM	LIF forskolin	non	immature	fœtus chimériques	LIF-dépendant
De Cheng 2012	rétrovirus OSKM	LIF bFGF VPA	non	oui	blastocystes chimériques	LIF-dépendant
Montserrat 2012	rétrovirus SKM	bFGF LIF	non	oui	NR	NR
Ruan 2011	rétrovirus OSKM	bFGF	non	oui	NR	bFGF-dépendant
Montserrat 2011	plasmide OSKM	LIF bFGF	non	oui	NR	NR
Telugu 2010	plasmide épisomal OSKMNL	LIF+2i VPA 5aza	non	oui	NR	LIF-dépendant
West 2010	lentivirus OSKMNL	bFGF	non	NR	porcelets chimériques stériles	bFGF-dépendant
Ezashi 2009	lentivirus OSKM	bFGF	non	oui	NR	bFGF-dépendant
Wu 2009	lentivirus dox-inductible OSKM	aucun	non (DOX-dépendant)	oui	NR	NR
Esteban 2009	retrovirus OSKM	bFGF	non	oui	NR	bFGF-dépendant

O : *OCT4* ; S : *SOX2* ; K : *KLF4* ; M: *c-MYC* ; N: *NANOG* ; L : *LIN28* ; 2i : inhibiteur des voies ERK et GSK3 β ; VPA : acide valproïque ; 5aza : 5-azacytidine ; NR : Non Renseigné

b. Dérivation de lignées iPSCs porcines

L'intérêt du porc en tant qu'animal agronomique et en tant que modèle pour des maladies humaines, la difficulté à obtenir des pESCs stables et l'apparition de la technique de reprogrammation cellulaire par Takahashi et Yamanaka en 2006 (Takahashi and Yamanaka, 2006) ont conduit plusieurs équipes à établir des lignées de cellules iPSC porcines à partir de 2009 (Tableau 4).

Dans un premier temps, plusieurs groupes ont dérivé des lignées cellulaires présentant les caractéristiques de pluripotence (expression de plusieurs marqueurs, capacité de se différencier en 3 feuilletts embryonnaires) en utilisant différentes techniques : infection rétrovirale (Esteban et al., 2009), infection lentivirale (Ezashi et al., 2009) et infection lentivirale dox-inductible (Wu et al., 2009). Dans ces trois premières études, les lignées obtenues sont cultivées en présence de bFGF et ressemblent plus aux lignées iPS humaines que murines, indiquant un état amorcé de pluripotence. De plus, malgré une bonne différenciation en corps embryoïdes et la formation de tératomes, ces lignées n'ont pas permis d'obtenir de chimères probablement à cause de l'expression continue des transgènes (Esteban et al., 2010a). En 2010 cependant, des lignées obtenues par infection lentivirale d'un cocktail de 6 facteurs (*OCT4*, *SOX2*, *KLF4*, *c-MYC*, *NANOG*, et *LIN28*) et cultivées avec du bFGF se sont révélées capables de former des chimères avec la naissance de porcelets chimériques, qui se sont avérés stériles (West et al., 2010), et ce malgré l'absence d'extinction des transgènes.

Tout comme chez l'homme et la souris, des méthodes non intégratives sont alors apparues pour limiter l'intégration des transgènes, telles que l'utilisation de plasmides (Montserrat et al., 2011), d'épisomes (Telugu et al., 2010) ou de transposons (Kues et al., 2013). Néanmoins, la plupart des expérimentations sont encore réalisées avec des systèmes intégratifs (Tableau 4). Ces nouvelles techniques n'ont pas permis l'extinction des transgènes.

Les premières lignées exhibant des caractéristiques de pluripotence naïve comme l'expression de *SSEA1* et la dépendance au LIF ont été établies en 2010 (Telugu et al., 2010) par l'utilisation du milieu LIF double inhibiteur ayant déjà montré sa capacité à convertir les lignées mEpiSCS en lignées naïves (Silva et al., 2008). Les auteurs ont également ajouté de l'acide valproïque et de la 5-azacytidine connus pour leur effet positif sur le remodelage épigénétique en cours de reprogrammation (Huangfu et al., 2008a). L'absence de certitudes quand aux voies de signalisation régulant la pluripotence chez le porc a amené plusieurs équipes à travailler avec des milieux de culture supplémentés à la fois en LIF et en bFGF, rendant difficile la classification des lignées obtenues (Montserrat et al., 2011; Cheng et al., 2012; Montserrat et al., 2012; Kwon et al., 2013). Néanmoins, deux de ces équipes ont confirmé que leurs lignées étaient plutôt LIF-dépendantes (Cheng et al., 2012; Kwon et al., 2013) ce qui permettrait de qualifier les cellules de naïves. De plus, les lignées cultivées en LIF obtenues par Fujishiro et al. (Fujishiro et al., 2012) exhibent l'expression de gènes de la pluripotence naïve. Il est toutefois important de noter que les marqueurs de cet état tels que *REX1*, *DAX1* ou *KLF2* ne sont peut-être pas de bons marqueurs pour le porc, puisque leur expression n'est pas clairement restreinte à

l'ICM dans le développement pré-implantatoire porcin (Hall and Hyttel, 2014). Afin de caractériser au mieux les voies actives dans les piPSCs établies à ce jour, une étude a comparé le transcriptome de plusieurs de ces lignées. Cette équipe a ainsi mis en évidence la sur-régulation de molécules d'adhésion cellulaire épithéliale, et a confirmé la plus grande ressemblance des lignées porcines avec les lignées humaines au niveau des voies de signalisations JAK/STAT, TGFB, WNT et VEGF ainsi qu'au niveau de l'expression de gènes de l'état amorcé tels que *OTX2* et *FABP7* et de l'absence de gènes de l'état naïf comme la famille KLF et *TBX3* (Liu et al., 2013). De plus, la surexpression des quatre facteurs *OCT4*, *SOX2*, *KLF4* et *c-MYC* ainsi que la maintenance de voie de signalisation spécifiques pourrait favoriser chez le porc la reprogrammation vers d'autres types cellulaires, comme observé lors de l'apparition d'une sous-population cellulaire ayant la signature expressionnelle du trophoblaste (Ezashi et al., 2011). Enfin, les difficultés à établir de véritables cellules pluripotentes naïves avec formation de chimères à transmission germinale se retrouvent chez d'autres espèces d'élevage. La reprogrammation cellulaire chez le bovin nécessite l'ajout de *NANOG* et de *LIN28* dans le cocktail de reprogrammation ainsi que l'utilisation de transgènes bovins (Ezashi et al., 2012). Bien que des lignées ayant la capacité de se différencier dans les trois feuilletts embryonnaires aient été produites chez l'ovine, le caprin et le bovin, seule une lignée ovine est capable de contribuer au développement de chimères.

MATERIELS et METHODES

I. Composition des solutions utilisées

Les produits et matériels utilisés au cours de ce projet sont regroupés dans l'Annexe 2.

1. Solutions

Tampon TE : 10mM Tris-HCl pH 8.0, 1mM EDTA

Tampon IP : 10 mM Na-Phosphate pH 7.0, 140 mM NaCl, 0.05% Triton-X100

NTMT : 100mM Tris-HCl pH 9.5, 100mM NaCl, 50mM MgCl₂, 0.1% Triton X-100

KTBT : 50mM Tris-HCl pH 7.5, 150mM NaCl, 10mM KCl, 0.1% Triton-X100

PFA 4% : 4g de Paraformaldéhyde dans 100mL de PBS chauffé à 60°C.

Gélatine 0.1% : 0.5g de gélatine dans 500mL d'eau

PS : 5g de Streptomycine et 3g de Pénicilline dans 500mL d'eau. Filtrer.

Fixateur : 2 volumes d'acide acétique pour 3 volumes d'éthanol absolu.

dNTP 10X : 1mM dATP, 1mM dGTP, 1mM dCTP, 0.3mM dTTP dans 10mM Tris-HCl pH 7.5, 1mM EDTA

Solution de phosphate disodique (Na₂HPO₄) 0.1M : 17.9g de phosphate disodique dans 250mL d'eau distillée.

Solution tampon d'acide citrique (C₃H₈O₇) 0.2M : 5.25g d'acide citrique dans 250mL d'eau distillée.

Solution de coloration Giemsa : H₂O, 3% Giemsa microscopy, 3% méthanol, 3% solution tampon d'acide citrique. Ajuster à pH=6.8 avec du phosphate disodique.

2. Milieux de cultures

Milieu F : DMEM, 10% Sérum de veau fœtal, 1% Sodium Pyruvate, 1% Pénicilline-Streptomycine, 5µg/mL Plasmocin Treatment

Milieu piPS : DMEM/F12, 10% Sérum de veau fœtal ES, 10% KnockOut Serum Replacement, 1% Sodium Pyruvate, 1% Acides aminés non essentiels, 1% Pénicilline-Streptomycine, 1% L-Glutamine, 100µM 2-mercaptoéthanol, 6ng/mL bFGF, 5µg/mL Plasmocin Treatment

Milieu LIF+2i : DMEM/F12, 20% KnockOut Serum Replacement, 1% Sodium Pyruvate, 1% Acides aminés non essentiels, 1% Pénicilline-Streptomycine, 1% L-Glutamine, 100µM 2-mercaptoéthanol, 5µg/mL Plasmocin Treatment, 1000U/mL ESGRO Mouse LIF Medium Supplement, 1% N-2 Supplement 100X, 1% B-27 Supplement 50X, 1µM PD0325901, 3µM 1-Azakenpaullone

Milieu DIFF : DMEM, 10% Sérum de veau fœtal décomplémenté, 1% Sodium Pyruvate, 1% Pénicilline-Streptomycine, 1% Acides aminés non essentiels, 5µg/mL Plasmocin Treatment

Tableau 5 : Caractéristiques des animaux étudiés normaux et à mauvaise qualité de semence.

A : asthénospermie ; T : tératospermie ; O : oligospermie.

	Nom	Caryotype	Qualité de semence	Genotype	Volume (mL)	Nombre total (sptz/éjaculat)	Concentration (10 ⁶ /mL)	Mobilité (%)	Motilité (1 à 5)	Vivants normaux (%)
Valeur moyenne					262	95	402	88	3.3	
Valeur limite						>15		>70	> 2.5	> 60
groupe contrôle	N1	38, XY	Normal	Large White	75	22	303	90	4	90
	N2	38, XY	Normal	Large White	226	144	660	90	4	85
	N3	38, XY	Normal	Piétrain	160	91	585	75	2.5	80
	N4	38, XY	Normal	Piétrain	165	76	475	85	3.5	90
	N5	38, XY t(3;4)	Normal	Large White	269	50	200	90	3.7	
Groupe expérimental	A1	38, XY	A	Large White			490	30	1	
	T1	38, XY	T	Piétrain	100	46	342	50	3	40
	O1	38, XY	O	Piétrain	93	7	81	85	4	85
	O2	38, XY	O	Piétrain	179	11	59	91	3.4	91
	AT1	38, XY	A T	P76*	290	99	340	5	0.5	43
	AT2	38, XY	A T	Piétrain	95	62	677	20	2	0
	AT3	38, XY t(13;17)	A T	GP1062*	115	31	282	5	0.5	0
	OAT1	38, XY t(1;14)	O A T	Piétrain	196	4	19	38	1.5	

* Lignées synthétiques commerciales

Milieu mES : DMEM/F12, 15% Sérum de veau fœtal ES, 1% Sodium Pyruvate, 1% Acides aminés non essentiels, 1% Pénicilline-Streptomycine, 1% L-Glutamine, 100 μ M 2-mercaptoéthanol, 500U/mL ESGRO Mouse LIF Medium Supplement, 5 μ g/mL Plasmocin Treatment

II. Etude de la méthylation de l'ADN spermatique porcin

1. Choix des animaux

Les animaux employés dans l'étude proviennent de l'unité expérimentale INRA UEICP GenESI de Rouillé. La qualité de la semence est évaluée en observant la couleur de l'éjaculat, particulièrement la présence de sang, en calculant le volume de l'éjaculat (moyenne sur 12000 animaux : 262mL) puis la concentration en spermatozoïdes par photométrie. Cinq animaux sont classés dans le groupe contrôle, à qualité de semence normale, et 8 dans le groupe expérimental qui rassemble les animaux dont la semence est de mauvaise qualité (Tableau 5). Trois animaux sont porteurs de remaniements chromosomiques, l'un dans le groupe contrôle et les deux autres dans le groupe expérimental.

2. Extraction d'ADN à partir de semence

Les cellules spermatiques congelées issues d'éjaculat dilué ainsi que de paillettes sont au préalable lavées au PBS. Elles sont ensuite lysées dans une solution de lyse (10mM Tris-HCl pH 8.0, 10mM EDTA, 100mM NaCl, 2% SDS) en présence de 10mM DTT pendant 1h30 à 65°C puis incubées sur la nuit à 37°C en présence de protéinase K 10mg/mL. L'ADN est ensuite extrait par une solution de phénol/chloroforme/alcool isoamylique (25:24:1) puis précipité et lavé à l'éthanol 70%. Le culot d'ADN est alors repris et dilué dans du tampon TE. La quantité d'ADN est mesurée par spectrométrie au NanoDrop 1000.

3. Mesure du niveau global de méthylation de l'ADN : LUMA

La technique LUMA a été décrite en 2006 par Karimi et collaborateurs (Karimi et al., 2006a; Karimi et al., 2006b). Elle utilise la différence de sensibilité à la méthylation des enzymes de restriction HpaII et MspI qui ciblent la séquence CCGG. HpaII clive les sites non méthylés alors que MspI clive les sites qu'ils soient méthylés ou non. Pour chaque échantillon, 500ng d'ADN génomique est digéré par les enzymes EcoRI+HpaII ou EcoRI+MspI puis dilué dans du « PyroMark Annealing Buffer ». Les échantillons sont ensuite analysés directement par pyroséquençage grâce au PyroMark Q24 (Qiagen) avec les réactifs associés. Chaque incorporation de nucléotide renvoie un pic dont la hauteur est représentative de la quantité de nucléotide incorporé. Au cours du pyroséquençage, les nucléotides C et G sont incorporés simultanément. Les résultats sont normalisés en divisant la valeur obtenue pour

l'incorporation C+G, qui reflète le nombre de site CCGG digérés, par celle obtenue pour l'incorporation d'un A qui reflète la digestion au site EcoRI, commune aux deux échantillons. En calculant ensuite le ratio des hauteurs de pic pour la double incorporation C+G entre l'échantillon digéré par HpaII et l'échantillon digéré par MspI, on obtient une valeur reflétant le pourcentage de méthylation global du génome. La différence du niveau de méthylation entre groupe contrôle et groupe expérimental est évaluée par un test de Student (*t*-test).

4. Immunoprécipitation d'ADN méthylé (MeDIP) et hydroxyméthylé (hMeDIP)

L'ADN génomique issu du sperme est fragmenté en morceaux de 300 à 500 pb par sonication (Bioruptor, Diagenode). Puis, 3µg (MeDIP-Seq) ou 1µg (MeDIP-qPCR) de cet ADN fragmenté est dilué dans du tampon TE, dénaturé dans l'eau bouillante pendant 10 minutes et immédiatement refroidi dans la glace 10 minutes. Ensuite 50µL de tampon IP concentré 10X sont ajoutés au mélange ainsi que 1µg d'anticorps monoclonal anti-5-methylcytosine (clone 33D3) ou 0.5µg d'anti-5-hydroxymethylcytosine (MAB-31HMC, Diagenode). Après incubation à 4°C sur la nuit avec agitation, les complexes ADN-anticorps sont purifiés en utilisant 40µL de « Dynabeads-ProteinG » préalablement lavées dans du PBS. Les complexes ADN-anticorps sont lavés deux fois dans du tampon IP puis l'ADN est purifié à l'aide du kit iPure (Diagenode). Des échantillons de 200ng d'ADN témoin, ou input, ne sont pas couplés aux anticorps et ne sont donc pas enrichis en marques épigénétiques. Le MeDIP et le hMeDIP sont réalisés en 2 réplicats techniques pour l'ADN spermatique des animaux normaux N1 et N2 (réplicats biologiques), qui seront ensuite séquencés. Le MeDIP est également réalisé sur l'ensemble des animaux étudiés (cf Tableau 5) afin de réaliser une étude locale de la méthylation par MeDIP-qPCR.

5. Séquençage haut-débit de l'ADN méthylé et hydroxyméthylé d'animaux contrôles

a. Préparation des bibliothèques et séquençage haut débit

Les bibliothèques ont été préparées avec le kit NEXTflex MethylSequencing Kit à partir d'ADN double brin, préalablement soniqué (Bioruptor Diagenode) pour obtenir des fragments d'environ 300 pb. Une fois les adaptateurs fixés, les fragments sont soumis au MeDIP comme précédemment. Les fractions enrichies en ADN méthylé ou hydroxyméthylé ainsi que les fractions témoins (input) sont ensuite amplifiées par PCR puis séquencées en « paired-end » par le séquenceur haut débit Illumina Hi-Seq 2000 de la plateforme GeT-PlaGe à hauteur de 30 millions de lecture par bibliothèques.

Tableau 6 : Amorces utilisées pour la PCR quantitative suivant le MeDIP (F : Forward, R : Reverse)

Région		Séquence 5'-3'	Région		Séquence 5'-3'
ACTB	F	TCCATGGGAACAGCTTCTCT	MEST A	F	ACTCTAGTCCGCACCCCATT
	R	TTGCTAAGGGGGCACATTAC		R	ATGGGATAAGACGCCACCAT
RTL1	F	CCTCCTGGCTCATGGAGTAG	MEST B	F	TGCGTGATTCAACTTCGTTT
	R	CCAGAGCGTGATTCACTTCA		R	CTCCCAGAACCTCGCTGTAG
DLK1 / MEG3	F	CTGCATCTTCGTGCACTTGT	POU5F1	F	GCCCTAGGAAGTGGAAGGAC
	R	AACGTTTAAATGCGCAGGAG		R	CTAGGCGCCGAATTTAGAGA
DLK1 A	F	TGTCCCCGATGACAAAGAAT	RASGRF1	F	TCCACACAGAACATGGCTTC
	R	TGGAGATGCTTCTGATTGA	DMR	R	ACGGTTTGCGCATTTTTAAG
DLK1 B	F	GCGTGACGTGTGAAAATCAA	RASGRF1	F	ACTCCACGCGCTTACATCTT
	R	CCGTTGACCTGAGAAGGAAA		R	CCCGGAGTGTAGAGATTTGC
MEG3 DMR	F	GTTAGTGCAAAGCCGACTCC	DDX4	F	CTGAAGCCTCAAAGCCAAC
	R	CCTAGCAGCCCTTCTCACAC		R	TCAACAACAAATGGCACCAG
NESP55 A	F	ACTCCTTCCCTTGAGCTTCC	UBE3A	F	GTCTGCAACCTCCACCACA
	R	GCTGCGGTAAGAGCAAGTCT		R	TCCGACTAGGAACCATGAGG
NESP55 B	F	TCTTACCTCGCAAGGAGCAC	IMPACT A	F	GTCAGGGATCCTTCATCCAA
	R	GGGCACAGATCGCTGTAAGT		R	AGCTGCCATTGCTTCAATTT
NESP55 C	F	CACAAAACCTCACCCGACTT	IMPACT B	F	AATGATGAGTCTCGGGCTGT
	R	TTTACCCCGCAAACATTCTC		R	CCGCTTGTCTTGTCTCTTC
GNASXL ICR	F	CCCTGTGTTAGGTGGGTTTG	DAZL	F	GAACACTGGCTGTGCGAGTA
	R	AACGCCCTAAGAGGTTGTGA		R	CACTCCCAGCCTCTCGAC
GRB10	F	GTGGCAGAAGGTGAGGAAGA	OSBPL1A	F	GCTCCTAGCCAGCATCTC
	R	GGATCAAATTTTCGGATCCTGT		R	TTCCCCTCTTGCTTTGTAG
H19 ICR A	F	GTGCAGGCTTACACATCTCG	ACAD8	F	CGGGACTACAACCTCCAGAC
	R	CTAAGGTGGCTGGAGAGAGC		R	AGGCTCTAGCCCTCACACCT
H19 ICR B	F	GCGGCTTTAGGGGATAC	QPCT	F	AAAGCTGGTCTCCATTTACGC
	R	CCGCTTCTACCTCCACACC		R	GCTGAAGTTTCCAGGCAGTC
IGF2 DMR1 A	F	TGACAACCCTGCCCTGTTAG	WT1	F	CCTTACCCTGGTTGCGAATA
	R	GGACCTTGGAGGTAGGAAGG		R	CGCCTTCACTGTCCATTTCT
IGF2 DMR1 B	F	GGGAGACAGGCCATATGTTT	HM13	F	AGCAACACAGGCAGGAAAAG
	R	GGTCCCTTAAAGCCTGACT		R	AAGCGAATCGTTCAAAGGAA
IGF2 DMR2 A	F	CAGCGGGATACCGTAATGAT	CNTN3	F	GTTGAATCGGAGCTGTAGCC
	R	TTCTCCTCTCCCCACTTCT		R	TCCGACTAGGAACCATGAGG
IGF2 DMR2 B	F	GGAAGAGTGCTTCCGTA	C10ORF67	F	TCATTACGTACAGGGCATCG
	R	AGAGAGGGGCTGCCTTAGG		R	CGTGTGGATTCTGGACTTT
IGF2R DMR	F	TTAAGATGCCAACGGAGAGG	XIST	F	ATTCCTGAGGTTTGGGTA
	R	CGCCTATTTGGTCGATTTTC		R	AGTGCAGTTTGCCAAATTAT
IGF2R A	F	GACCAGCAACTGCAATCAGA	NNAT	F	CGACAATACCAGATTCCTTC
	R	AAAGCTATGCGTCCGCTAAA		R	CTTGGTCCAGATCAGAATGT
IGF2R B	F	AGGTCTCACCTCTTCAGGTT	PEG10	F	GTTGTTAATGGCTGGAAGAG
	R	CTGTGCAAATTAAGGCTTCT		R	AGTCACTTCCCCTTCCTAAG
LINE1	F	CACTGTTAGCCCCATCAACA	PEG3	F	AACTGCCCAAGCAGTACCAC
	R	GCAGTGTTCCTTGTCACT		R	TGGAGATGGAGCCTTACCAC

b. Analyse des résultats

i. Alignement des lectures sur le génome porcin

Les lectures sont alignées sur l'assemblage du génome porcin Sscrofa10.2 par l'outil bwa (Burrows-Wheeler Alignment Tool). Les lectures sont ensuite triées pour ne conserver que celles qui forment des paires orientées de façon convergente sur le même chromosome, et les duplicats sont supprimés (Picard MarkDuplicates). La couverture génomique est calculée via l'outil bedtools et les résultats sont visualisés par les logiciels IGV (Robinson et al., 2011) et Seqmonk (Babraham Bioinformatics).

ii. Détection des pics de méthylation

Après fusion des réplicats techniques pour chaque animal, l'utilisation de l'algorithme MACS v1.4 permet l'identification de pics de méthylation et d'hydroxyméthylation dans le génome de l'ADN spermatique porcin (Zhang et al., 2008; Feng et al., 2011). L'outil DIFFBIND, du package Bioconductor, permet de visualiser la corrélation entre les listes de pics obtenues, la marque épigénétique considérée et l'animal étudié (Stark and Brown, 2011). Les annotations génomiques de la dernière version du génome porcin sont obtenues via la base de données de l'UCSC Genome Browser (www.genome.ucsc.edu) : gènes, introns, exons, 5'UTR, 3'UTR, cds, îlots CpGs et éléments répétés. Les promoteurs sont définis comme la région de 1000pb en amont du gène. Plusieurs listes de gènes spécifiques de fonctions biologiques sont créées en se basant notamment sur les termes définis par Gene Ontology : biologie des cellules germinales (n=160), spermatogénèse (n=356), fertilisation (n=83), développement du blastocyste (n=59), implantation de l'embryon (n=24) et gastrulation (n=100). L'utilisation de bedtools permet de quantifier et d'identifier les annotations génomiques qui se superposent avec une zone enrichie en méthylation ou hydroxyméthylation. Un test de χ^2 est utilisé pour évaluer les différences entre les pourcentages d'annotations méthylées des différentes listes.

6. Analyse locale de la méthylation des animaux à qualité de semence dégradée

a. PCR quantitative sur les fractions d'ADN méthylé et témoin

i. Choix des gènes et PCR quantitative

Les gènes étudiés sont des gènes décrits comme soumis à empreinte chez le porc, l'homme ou la souris, auxquels nous avons ajouté trois gènes de développement précoce, le gène de ménage Actine- β (*ACTB*) et la séquence répétée *LINE1* utilisée comme référence de la méthylation globale du génome. Les régions d'intérêt ont été définies soit par homologie de séquence avec les DMRs gamétiques connus chez la souris, soit en recherchant les îlots CpG présents dans la région du gène d'intérêt à

**Tableau 7 : Amorces utilisées pour le pyroséquençage (R : Forward, R : Reverse, S : amorce de séquençage).
Les couples G4-G14 sont localisés dans le locus complexe GNAS (voir Figure 32).**

Région		Séquence 5' - 3'	Région		Séquence 5' - 3'
IGF2 DMR1 A	F	TTGGAAGTTGGGGAGGTT	G4	F	AGTTTTTTAGTTAAATGGGAAAGTTTAT
	R	[Btm]CACCAACCCTAAAAAACTTCTACTA ATAC		R	[Btm]AACCCATCCCTAAAAAAACCAATAAAA T
	S	GTAGTGATTTGGATT		S	AGTAGAGATTTTAGGAAGAAATAG
MEST B	F	GAGTTTTTGGAGGGGTGGTTAA	G5	F	AAGGGTTTGAGTTGAAAGGA
	R	[Btm]ACCAACCCCAAAACCAAACTT		R	[Btm]AACCTCCAAATTATTTTTCTCTACTC
	S	AGTTAATTGAGTGATTTAAT		S	AGTTGAAAGGATGGT
H19 ICR	F	TTTTTGGTAGTGGGTTGATTTTTATAGG	G6	F	[Btm]AGGTGGGGTTGGGATTTT
	R	[Btm]AACACCCCATCCCCTAAATA		R	ACCCACCCTCTAACTCTACAAA
	S	GAGGAGGTAGTGTAGG		S	CTAACTCTACAAAAACAACCT
IGF2 DMR2 B	F	[Btm]GTTGAGATTTGGTTTTGTTGGAGATT TAT	G7	F	GTAATGAGAGAAAAGTAATTAGATTGGTAT
	R	AAACCCCTCTAACAACACTTAA		R	[Btm]CCCAAATTAACAACCACTAAACCTA
	S	CTACCCAAAAAAAACCTACCTT		S	AAAAGTAATTAGATTGGTATT
IGF2R DMR	F	TGGGGTGAAGGGGTGAAA	G8	F	TGTGTTTTTTTAGTTAATTTGGGGGAGAA
	R	[Btm]CCCCCTTCTAACCTCAAC		R	[Btm]AAACCCTTAACCTATATTCCTTCTA
	S	GTGAAGGGGTGAAA		S	GGGAGAAGATAGATTAGTGAAAT
IMPACT A	F	TGAGGAAATTGAAGTAATGGTAGT	G9	F	[Btm]GTTATTAGAGAGAGTAAGGGTTGGTT
	R	[Btm]ATATACCTACAAACAAAATATCCACT TA		R	CCCTAAAAAATTATAAACTAAAATAAATT
	S	GAAATTGAAGTAATGGTAGTT		S	ATAACCACCAAACCTAAAATAC
RASGRF1 DMR	F	GGGAAAATGGATATAGGGATGTAT	G10	F	[Btm]GTTGGGGAGGGGAGGGTTTTAGT
	R	[Btm]CAACCCTAAATCAAAAATCAAAAAA ACCTT		R	CTCCTTCCAATCCTCTCCCTAACTAATTT
	S	TGTTGAGTTATTGGAGG		S	TCTCCCTAACCAAA
MEG3 DMR	F	[Btm]GGTTAGAAATTAGAGGGAAAAT	G12	F	AGAAAGAGAAGATGGGTTATATGTG
	R	CACTAAAATCCAAACCCCAAATT		R	[Btm]AAACCCCTTCACTAAAACCTTCTAAC
	S	CCAAACCCCAAATTTT		S	GATTTTGTAGTGGAGAGTA
DLK1 B	F	TTTGGGGAGAGAAGAGGAAGTT	G13	F	TTAATAGGGAAGTTGGAGGTAAGG
	R	[Btm]ATCCTAAAATCACACCTCCTAATATT C		R	[Btm]CCAAACCTACCTCCCAATTTTC
	S	AAGAGGAAGTTAGTGAT		S	AGTTGGAGGTAAGGA
RTL1	F	[Btm]GGTTTTTGTTTAGGGAGTAGTAG	G14	F	GGTATAGTGGGTATGAGTAGGATT
	R	AACTACCAACCCTTTCTAAT		R	[Btm]AACCCCAACCAATACCCCTAAA
	S	AAACCCCATCATCCC		S	ATTTGGGTAGAGTGGGAA
DAZL	F	AGGAGAGGGTGAGAGTTG	LINE1	F	GGGAGAAAAGAGGAGAGTGAGAGA
	R	[Btm]CTCTCTACCCCCCTCAAACCCCTATA		R	[Btm]ACTCCCCTTACTACTACTCTTCTCT AC
	S	TTTTTGTTTTGTTTTTTTGGGA		S	AGAGTGAGAGAAGGT

l'aide de CpGplot (Rice et al., 2000). Les amorces (Tableau 6) sont dessinées dans les régions d'intérêt grâce au logiciel Primer3. Elles sont tout d'abord testées par PCR conventionnelle sur un ADN témoin puis leur spécificité est testée et leur efficacité est calculée en réalisant une PCR quantitative sur une série de dilution d'un pool d'ADN (1/4 ; 1/8 ; 1/16 ; 1/32) grâce à l'appareil de PCR quantitative LightCycler 480 de Roche suivie d'une cinétique de melting. Une fois les amorces validées, chaque échantillon d'ADN témoin ou enrichi pour la méthylation est amplifié sur le même appareil en triplicat en présence des amorces ciblant la région d'intérêt et de FastStart SYBR Green Master mix (Roche).

ii. *Exploitation des résultats*

En se basant sur les valeurs de cycle seuil obtenues par la PCR quantitative, le pourcentage d'enrichissement en méthylation est calculé en utilisant la formule ci-dessous, dans laquelle le facteur de compensation 2.322 est utilisé pour prendre en compte la dilution 1/5 de l'input.

$$R = 2^{((Ct_{input\ DNA} - 2.322) - Ct_{met\ hylDNA})} \times 100$$

Ce résultat obtenu pour chaque région d'intérêt est ensuite normalisé sur la valeur obtenue pour la séquence LINE1. La différence de niveau de méthylation entre le groupe contrôle et le groupe expérimental pour chaque région du génome est évaluée par un test non paramétrique de Wilcoxon, l'hypothèse d'homogénéité des variances ne pouvant pas être vérifiée dans toutes les régions.

b. Conversion bisulfite et pyroséquençage

i. *Dessin des amorces*

Le pyroséquençage nécessite trois amorces : deux sont des amorces d'amplification et la troisième est une amorce de séquençage. L'une des deux amorces d'amplification doit être biotinylée en 5' afin de garantir la fixation de l'amplicon sur des billes de sépharose, étape nécessaire à la purification du produit PCR avant pyroséquençage. Les amorces sont dessinées à l'aide du logiciel Pyromark Assay Design 2.0 (Tableau 7)

ii. *Conversion bisulfite de l'ADN et amplification par PCR*

500ng d'ADN génomique issus de cellules spermatiques ou sanguines sont convertis en duplicat au bisulfite selon les instructions du kit MethylEdge Bisulfite Conversion kit (Promega). L'ADN est amplifié dans les régions d'intérêt grâce aux amorces d'amplification décrites précédemment, en utilisant le kit Pyromark PCR selon le programme PCR suivant : 95°C 15s (activation de la polymérase) ; 45 cycles [94°C 30s ; 56°C 30s ; 72°C 30s] ; 72°C 10 min. L'efficacité de la PCR est validée par électrophorèse sur gel d'agarose.

Tableau 8 : Amorces d'amplification (F,R) et de séquençage (S) utilisées pour la détection de l'empreinte parentale.

Gènes		Sequence 5'-3'
<i>CNTN3</i> exon 31	F	[B _{tn}]CAATGGTCATATGAAATCTACCTG
	R	AACTTTCTCTGCAAGGTCTGTAAA
	S	CTCTGCAAGGTCTGTAAA
<i>WT1</i> exon 9	F	AACGACTTTGCAAGCTGACTT
	R	[B _{tn}]CAAACCTGGTAGGATCCAGTG
	S	CTTCATCCAGCGTCC
<i>RASGRF1</i> exon 15	F	[B _{tn}]TTGCTGAAAGGCGACATG
	R	AATATCCCCATCATCACAGGC
	S	TCAGCTGCAACTCCAA
<i>IMPACT</i> exon 11	F	GCACAACCTGTAGATGAAAGAAGA
	R	[B _{tn}]GTTATGTAATTGTGGTGCCTGTTT
	S	AGGAAGATTTGATAAGAATG
<i>QPCT</i> exon 8	F	[B _{tn}]GCCATACTTCAGGGAAAGGAGAT
	R	ATTTGGTTTGCTCAAGGATCACTC
	S	TGTGATTCAGGATGACC
<i>GRB10</i> exon 7	F	GTGGGATTTTTGAAAACTCATAT
	R	[B _{tn}]GCGAGCGAAAGCAAATTTTAT
	S	ATTTTTGAAAACTCATATT

iii. Pyroséquençage

Après fixation sur les billes de sépharose en présence de Binding Buffer, le produit PCR est dénaturé et lavé grâce à la station de travail du PyroMark Q24 en présence des réactifs associés, puis couplé à l'amorce de séquençage en présence d'« Annealing Buffer ». Le produit final est pyroséquencé en présence des réactifs du kit « Pyromark Gold Reagents » sur le Pyromark Q24 de Qiagen. Les résultats sont analysés via le logiciel de l'appareil Pyromark Q24 qui permet de juger de la qualité du séquençage. Le pyrogramme obtenu donne le pourcentage de méthylation de chaque CpG présent dans la région séquencée. L'incorporation nulle d'un nucléotide T avant ou après un C dans un contexte non CpG sert de contrôle de la bonne conversion de l'échantillon. Les variations entre les deux groupes d'animaux ou entre le groupe contrôle et chaque animal du groupe expérimental sont évaluées par un test de Student.

c. Recherche d'empreinte parentale chez le porc

Pour certains gènes pour lesquels le statut d'empreinte n'était pas encore décrit chez le porc, l'utilisation de croisements réciproques entre animaux de race Large White et Meishan développés dans le cadre du projet ANR Porcinet porté par une autre équipe de notre unité nous a permis d'étudier l'empreinte parentale. Tout d'abord, l'ADN issu de quatre parents est analysé par séquençage Sanger pour détecter des SNPs entre les deux races. Pour cela, l'ADN est amplifié par PCR dans les régions géniques sélectionnées, puis le produit PCR est purifié par un traitement à l'exonucléase I (10U) et la phosphatase alcaline (2U) pendant 45 minutes à 37°C. Après inactivation des enzymes pendant 30 minutes à 80°C, 0.5µM d'amorce et les réactifs du « BigDye Terminator v3.1 Cycle sequencing kit » (Applied Biosystem) sont ajoutés au produit PCR purifié. Le séquençage est alors réalisé par amplification PCR et analysé sur le séquenceur ABI3730. La lecture des séquences par le logiciel ChromasPro permet de définir le génotype des parents et de détecter les SNPs dans les exons des gènes. Une fois les SNPs validés, l'ARN est extrait à partir de broyats (TissueLyser (Retsch) en présence d'azote liquide) de quatre tissus (foie, muscle, poumon et cerveau) issus de fœtus croisés réciproques LWxMS et MSxLW. L'extraction d'ARN est réalisée en présence de TriZol puis les ARNs sont traités par la rDNase et purifiés en utilisant les colonnes et les réactifs du kit Nucleospin RNA (Macherey-Nagel). La qualité et la quantité des ARNs sont validées par électrophorèse sur gel d'agarose et dosage au NanoDrop 1000. Ces ARNs sont ensuite rétrotranscrits en ADNc en présence de Superscript Transcriptase II. Le pyroséquençage de ces ADNc par des amorces dessinées autour des SNPs identifiés (Tableau 8) permet de déterminer l'origine parentale ou non de l'expression des gènes. Le pyroséquençage est réalisé de la même manière que précédemment, avec une température d'hybridation de 59°C au cours de la PCR.

Tableau 9 : Couples d'amorces utilisés pour l'analyse de l'expression des transcrits GNAS dans le testicule porcin.

Gènes		Séquence 5'-3'
NESP55	F	TCTTACCTCGCAAGGAGCAC
	R	GGGCACAGATCGCTGTAAGT
GNAS	F	TGCTCGCTACACTACTCCTG
	R	CAGGTGAAGTGAGGGTAGCA
GNASXL	F	CAACTTCCTCGTCCAGGCTT
	R	CAGGATCCTCATCTGCTTCACA
IA	F	GGACACTCAGTCGCGTCG
	R	CAGGATCCTCATCTGCTTCACA
GSA	F	CAGTAAGACCGAGGACCAGC
	R	CAGGATCCTCATCTGCTTCACA
POU5F1	F	AGTGCCCAAAGCCCACTCT
	R	TTCTGGCGACGGTTGCA
TBP	F	AACAGTTCAGTAGTTATGAGCCAGA
	R	AGATGTTCTCAAACGCTTCG

Tableau 10 : Caractéristiques des animaux choisis pour la reprogrammation cellulaire. *Lignée synthétique commerciale

Nom	Caryotype	Qualité de semence	Race
N1	38, XY	Normale	Large White
N2	38, XY	Normale	Large White
PEF	38, XY	Normale	Meishan
Y14	38, XY, t(Y;14)	Azoospermie	Duroc
AT3	38, XY, t(13;17)	Asthenospermie, Tératospermie	GP1062*
OAT1	38, XY, t(1;14)	Oligospermie, Asthenospermie, Tératospermie	Piétrain
XXY	38, XY / 39, XXY	Oligospermie, Tératospermie	Landrace Français

d. Expression des transcrits du locus GNAS dans le testicule

L'ARN est extrait de tissu testiculaire selon les instructions du kit Nucleospin RNA (Macherey-Nagel). Après validation de la qualité et de la quantité d'ARN, la conversion en ADNc et la PCR quantitative sur le LC480 de Roche sont réalisées de la même manière que précédemment. Les amorces sont dessinées pour amplifier spécifiquement les quatre transcrits majoritaires du locus GNAS qui ne se différencient que par leurs premiers exons : NESP55, GNASXL, 1A et GSA. Un couple d'amorces commun aux quatre transcrits est également ajouté (Tableau 9).

III. Dérivation et caractérisation de lignées pluripotentes induites porcines

1. Choix des animaux

Grâce à l'activité de la plateforme de contrôle chromosomique qui fait partie intégrante de l'équipe nous avons identifié plusieurs animaux porteurs de remaniements chromosomiques et présentant des défauts de qualité de semence. Les animaux sélectionnés pour la dérivation de lignée iPS sont décrits dans le Tableau 10. Nous avons également utilisé comme lignée contrôle de la reprogrammation des fibroblastes adultes ou embryonnaires (PEF) d'animaux fertiles à caryotype normal.

2. Reprogrammation cellulaire

a. Culture de fibroblastes

Après dilacération au scalpel, le tissu subit une digestion enzymatique en présence de trypsine 2X pendant 1h à 37°C. Les fibroblastes ainsi obtenus sont cultivés dans le milieu F à 37°C et 5% CO₂, avec changements de milieu réguliers, et sont repiqués en utilisant la trypsine 0.25X quelques minutes à 37°C pour décoller le tapis cellulaire. Les fibroblastes embryonnaires proviennent d'un embryon à J52 de race Meishan.

b. Systèmes de reprogrammation cellulaire

i. Infection par le lentivirus PL-SIN-EOS-C(3+)-EiP

Le lentivirus PL-SIN-EOS-C(3+)-EiP nous a été fourni par Marielle Afanassieff de la plateforme PrimaStem du laboratoire SBRI (Stem-cell and Brain Research Institute). Entre 50000 et 100000 cellules sont infectées avec le lentivirus avec une concentration finale de 200000pV/mL dans 1mL de milieu OptiMEM (Multiplicity of Infection (MOI) de 4). Une MOI de 4 représente théoriquement une

infection de 98% des cellules avec au moins une particule virale. Après 5h d'incubation à 37°C 5% CO₂, 2mL de milieu F sont ajoutés. 24h après l'infection, les cellules sont lavées 3 fois au PBS et remises en culture dans du milieu F à 37°C 5% CO₂.

ii. Infection par les virus pMXs

Les virus pMXs-hOCT4, pMXs-hSOX2, pMXs-hKLF4 et pMXs-hc-MYC nous ont été fournis par Marielle Afanassieff de la plateforme PrimaStem du laboratoire SBRI. A J0, entre 80000 et 100000 fibroblastes sont infectés simultanément par les quatre virus à une concentration finale de 200000pV/mL/virus dans 1mL de milieu OptiMEM. Après 5h d'incubation à 37°C 5% CO₂, 2mL de milieu F sont ajoutés. A J2, les cellules sont lavées 3 fois au PBS et remises en culture dans du milieu F. A J3, les cellules infectées sont repiquées ($5 \cdot 10^5$, 10^5 et $5 \cdot 10^4$ cellules par boîte) vers des boîtes de culture 55cm² prétraitées par de la gélatine 0.1% et contenant des cellules nourricières STO irradiées (13000 cellules/cm²). Le milieu de culture est alors remplacé par du milieu piPS et est changé quotidiennement. Entre J4 et J30, l'apparition de clones est surveillée quotidiennement et les clones sont sélectionnés sur leur morphologie et le cas échéant (cellules EOS+) sur l'expression de la protéine GFP (microscope Leica DM IRBE statif inversé). Dans le cas de cellules EOS+, l'ajout de puromycine prévient contre la colonisation trop rapide par les fibroblastes non reprogrammés. Les clones sont repiqués de manière individuelle sous microscope et transférés dans un puits de culture de 0.95cm² pour amplification progressive. Des lignées de fibroblastes embryonnaires murins (MEFs) ont été reprogrammés de la même manière en utilisant le milieu spécifique aux cellules souches pluripotentes murines mES.

iii. Système de reprogrammation non intégratif

Les fibroblastes sont reprogrammés en utilisant le kit CytoTuneTM-iPS Reprogramming Kit. Rapidement, à J0 100000 fibroblastes sont transduits par $3 \cdot 10^6$ CIU de chaque vecteur Sendai, chacun contenant l'un des quatre même facteurs de reprogrammation que précédemment, dans 2mL de milieu F. Le milieu est changé à J2 puis les cellules sont diluées à J7 de la même manière que précédemment.

3. Culture des lignées de cellules pluripotentes

Les lignées piPS-like sont cultivées dans des boîtes de culture prétraitées à la gélatine 0,1 %, en présence d'une couche de cellules STO irradiées (13000/cm²), dans du milieu piPS. Le milieu est changé quotidiennement. Les passages se font par traitement au « StemPro Accutase Cell Dissociation Reagent ». L'effet du changement de milieu sur le profil d'expression des lignées piPS-like est étudié de MEFs et les lignées de mESCs ROSA fournies par le laboratoire SBRI, qui ont servi de témoin pour la reprogrammation cellulaire et la morphologie des lignées obtenues, ont été cultivées à l'identique en présence de milieu mES.

4. Validation des caractères de pluripotence

a. Expression de la phosphatase alcaline

Après lavage au PBS et fixation à la PFA 4%, les cellules sont lavées et incubées dans la solution KTBT (3 minutes à température ambiante) puis dans la solution NTMT (5 minutes à température ambiante). On incube ensuite au minimum 20 minutes dans une solution de NTB/BCIP (12 μ L/mL NTMT). Après apparition de la coloration bleue caractéristique de l'expression de la phosphatase alcaline, les cellules sont lavées et incubées dans la solution de KTBT, fixées à nouveau dans une solution de PFA 2% et conservées dans du PBS à 4°C.

b. Immunocytochimie

Les cellules sont mises en culture sur des lamelles traitées à la gélatine 0.1%. Elles sont ensuite fixées à la PFA 4%. Après plusieurs lavages au PBS puis au PBT (PBS, 0.1% Triton X-100), les cellules sont incubées 30 minutes dans du PBT 10% SVF. Les anticorps primaires sont alors ajoutés : anti-Nanog (0.5 μ g/mL, Peprotech 500-P236), anti-Lin28 (2 μ g/mL, ab63740), anti-Oct4 (2 μ g/mL, ab19587), anti-Sox2 (2.5 μ g/mL, R&D Systems MAB2018), anti-SSEA4 (4 μ g/mL, ab16287), anti-E-cadherin (1.25 μ g/mL, BD 610181) et anti-SSEA1 (1/100, Hybridoma Bank MC-480). Après une nuit d'incubation à 4°C et un second cycle de lavage et incubation dans du PBT 10%SVF, les anticorps secondaires sont ajoutés (Anti-Rabbit Alexa 568, Anti-Mouse Alexa 568, Anti-Mouse IgM) pour 1h à température ambiante à l'abri de la lumière. Le montage des lamelles sur lame de microscope s'effectue dans le milieu de montage VECTASHIELD contenant du DAPI (Eurobio). Les lames sont conservées à 4°C jusqu'à l'analyse et la capture d'image (microscope Zeiss, logiciel Volocity).

La détection de SSEA4 par immunomarquage en suspension a été réalisée comme suit : les cellules sont suspendues à une densité de 4.10⁶ cellules/mL dans du milieu F. Ensuite, 5.10⁵ cellules sont prélevées et incubées 15 minutes sur roue avant l'ajout de 0.5 μ g d'anticorps primaire anti-SSEA4 (SSEA4 (813-70), SC21704). Après incubation 30 minutes sur roue à température ambiante et à l'abri de la lumière, les cellules sont lavées 3 fois dans du PBS. L'anticorps secondaire est alors ajouté (Anti-Mouse IgG F(ab')₂-APC, SC3818) avant une seconde étape d'incubation sur roue 30 minutes à température ambiante. Les cellules sont lavées 3 fois dans une solution de PBS - BSA 0.5% puis resuspendues dans du PBS. L'immunomarquage de SSEA1 a été réalisé de la même manière en utilisant un anticorps primaire déjà couplé à un fluorochrome (SSEA1 (480), SC21702 AF647) et ne nécessitant donc pas l'ajout d'un anticorps secondaire. L'analyse du nombre de cellules fluorescentes est réalisée au FACS (FAC Scanto II, SBRI).

Tableau 11 : Amorces et paramètres de PCR pour la détection des marqueurs de pluripotence

Gène		Séquence 5'-3'	Taille de l'amplicon (pb)
NANOG	F	CCCCGAAGCATCCATTTCC	214
	R	CGAGGGTCTCAGCAGATGACAT	
LIN28	F	CGGAACTGGCCACTGTAAAT	390
	R	TCTGACAGCAATGGCACTTC	
OCT4	F	CTTTCCTCGGTGTCTGTCA	125
	R	GTACAAACCCCAGGCTTCTCTC	
SOX2	F	GGTACACTGCCTCTCTCTCACAT	109
	R	ACTGTTTCTTACTCTCCTCCCATTT	
KLF4	F	CCGTCGGCCGGCCCTTAGAG	426
	R	TGTGGGTTTCGCAGGTGTGCC	
c-MYC	F	CAACGACAGCAGCTCGCCCA	758
	R	GGGCCATTCTAGTTCCTCCCTCCA	
ACTB	F	CACGCCATCCTGCGTCTGGA	380
	R	AGCACCGTGTGGCGTAGAG	

Paramètres PCR : 35 cycles, Tm=59°C

Tableau 12 : Amorces et paramètres de PCR pour la détection des transgènes

Nom		Séquence 5'-3'	Taille de l'amplicon (pb)
pMXs-AS3200	R	TTATCGTCGACCACTGTGCTGGCG	
hOct4-S947	R	CAGGGCCCCATTTTGGTACC	283
hSox2-S807	R	CCTCCGGGACATGATCAGC	247
hKlf4-S1282	R	TTCGCCCGCTCAGATGAACTG	231
hCmyc-S1199	R	CAGCATACATCCTGTCCGTCC	221
SeV	F	GGATCACTAGGTGATATCGAGC	181
	R	ACCAGACAAGAGTTTAAGAGATATGTATC	
Oct4-SeV	F	CCCGAAAGAGAAAGCGAACCAG	483
	R	AATGTATCGAAGGTGCTCAA	
Sox2-SeV	F	ATGCACCGCTACGACGTGAGCGC	451
	R	AATGTATCGAAGGTGCTCAA	
Klf4-SeV	F	TTCCTGCATGCCAGAGGAGCCC	410
	R	AATGTATCGAAGGTGCTCAA	
cMyc-SeV	F	TAACTGACTAGCAGGCTTGTCG	532
	R	TCCACATACAGTCCTGGATGATGATG	

Paramètres PCR pour les transgènes pMXs : 30 cycles, Tm=60°C
 Paramètres PCR pour les transgènes SeV : 35 cycles, Tm=55°C

Tableau 13 : Lignées cellulaires analysées par PCR quantitative

Lignée cellulaire somatique	Système de reprogrammation	Milieu de culture	Etat	Nom de la lignée
t(Y;14)	Intégratif	piPS	Passage précoce	I3
t(Y;14)	Intégratif	piPS	Passage tardif	I3
t(Y;14)	Intégratif	LIF+2i	LIF+2i	I3
t(Y;14)	Intégratif	piPS	Passage précoce	I4
t(Y;14)	Intégratif	piPS	Passage tardif	I4
t(Y;14)	Intégratif	LIF+2i	LIF+2i	I4
(38,XX)	Intégratif	piPS	Passage précoce	PB20
t(Y;14)	Non intégratif	piPS	Passage précoce	NI12
t(Y;14)	Non intégratif	piPS	Passage précoce	NI13
t(Y;14)	Non intégratif	Diff	Corps embryoïde	NI13
t(Y;14)	Non intégratif	piPS	Passage précoce	NI20
t(Y;14)	Non intégratif	Diff	Corps embryoïde	NI20
PEF (38,XY)	Non reprogrammé	F		PEF
t(Y;14)	Non reprogrammé	F		Y14

c. Analyse du cycle cellulaire

Les cellules sont resuspendues dans du PBS puis fixées par l'ajout goutte à goutte d'éthanol 70 % froid et conservées au minimum 2h à -20°C. Après un lavage au PBS, elles sont traitées à la RNase A 0,1mg/mL pendant 1h à 37°C. L'ADN est marqué par l'ajout de 25 µg/mL d'Iodure de Propidium.

Pour étudier la présence de points de contrôle dans le cycle cellulaire des lignées pluripotentes, les cellules sont traitées au préalable 24h avec 66 ng/µL de doxorubicine hydrochloride (DH) qui induit des cassures à l'ADN.

Le ratio de cellules dans chaque phase du cycle est mesuré par cytométrie de flux sur l'appareil BD FACScalibur de la plateforme d'imagerie Genotoul TRI du plateau de l'IPBS (Institut de Pharmacologie et Biologie Structurale, Toulouse).

d. Expression des gènes de pluripotence

i. Extraction d'ARN et transcription inverse

L'ARN total est extrait à partir de culots cellulaires en suivant les instructions du kit Nucleospin RNA (Macherey-Nagel). La quantité et la qualité des ARNs sont évaluées par un dosage (NanoDrop 1000) et une électrophorèse sur gel d'agarose. 1µg d'ARN est converti en ADNc d'après les instructions du kit Superscript II (Invitrogen). La concentration des échantillons d'ADNc est mesurée au NanoDrop 1000.

ii. PCR conventionnelle

L'expression des facteurs de pluripotence *NANOG*, *LIN28*, *OCT4*, *SOX2*, *KLF4* et *c-MYC* est vérifiée par PCR conventionnelle et électrophorèse sur gel d'agarose 1%. L'ADN copie est amplifié sur 40 cycles en présence de GoTaq DNA Polymérase (Promega) et d'amorces décrites dans le Tableau 11. La température d'hybridation est de 59°C.

L'expression des facteurs humains de reprogrammation est également étudiée par PCR conventionnelle. Les amorces, décrites dans le Tableau 12, sont spécifiques du système de reprogrammation utilisé, ainsi que le nombre de cycle d'amplification et la température d'hybridation.

iii. PCR quantitative

Afin de caractériser au mieux l'état pluripotent des différentes lignées obtenues dans les différentes conditions de culture (milieu piPS ou milieu LIF+2i) et par différentes méthodes (intégrative ou non-intégrative), et à différents stades (passages précoce, tardif, différenciation), une PCR quantitative a été réalisée sur 41 gènes recensés comme marqueurs de la pluripotence naïve ou amorcée (Tableau 13 et 14), 3 gènes de référence et la *GFP*. Les amorces sont tout d'abord validées par PCR

Tableau 14 : Amorces utilisées pour la PCR quantitative

Gène		Séquence 5'-3'	Gène		Séquence 5'-3'
<i>ACTA2</i>	F R	TGTGCTGGACTCTGGAGATG GTGGTCACGAAGGAGTAGCC	<i>LEFTY2</i>	F R	CAGGGACTATGGAGCTCAGG ACACTCATAGGCCAGGAAGC
<i>ACVR2B</i>	F R	GCTGCTGGCTAGACGACTTC TGGCTCGTACGTGACTTCTG	<i>LIF</i>	F R	TCAATCCTGGTGTCTGTGAAC CAGCCCAGCTTCTTCTCTG
<i>CDC20</i>	F R	TTCCTCTGCAGACATTACAG ACCAGAGAATGGAGCACACC	<i>LIN28B</i>	F R	CGGAACTGGCCACTGTAAAT TCTGACAGCAATGGCACTTC
<i>cMYC</i>	F R	GCTGGATTTCTTCGGATAG TTGGTGAAGCTGACGTTGAG	<i>MHC1 / SLA1</i>	F R	CCTCTTCTGCTGCTGTCTG AGCGTGTCTTCCCCATCT
<i>DAX1 / NROB1</i>	F R	CTTCCTTGCCAAGTGCTGGA CTGCTGAGTTCCCCACTGAA	<i>NANOG</i>	F R	AAGAAGCAGAAGATCAGAAGTGTGT TGTTTGTAGCTAAGGTTTCAGGATGT
<i>DAZL</i>	F R	GCATTTTCACGACATCATGG TAGCCACCTATTGCCTGTCC	<i>NODAL</i>	F R	CGTCTCCAGATGGACCTGTT CTGCTCTGGAGAGAGGTTGG
<i>DKK1</i>	F R	GAGGGGAAATTGAGGAAACC GTGGCACAGTCTGATGATCG	<i>NR6A1</i>	F R	TGTCCTCCCTCACCGTTTAC TGCAAGCATACTCCTCGTTG
<i>DNTM3B</i>	F R	ATGAAGAAGAGGGTGCCAG GCCTCTGATATCTGGGGTGC	<i>OCT4 / POU5F1</i>	F R	AGTGCCAAAGCCCCTCT TTCTGGCGACGGTTGCA
<i>DPPA2</i>	F R	CTTCAAGAGCCGTTACCCCT GGCGAACCAACCTTCTGTA	<i>OTX2</i>	F R	TTGTCCACCTCCTTCTCCTG ACTGAGTGTGGTCCCTGGTC
<i>DPPA3 / STELLA</i>	F R	TGAACAGAGACCCGCCTTC TGACATCTTCCCCAAGTCGC	<i>PECAM1</i>	F R	TTCACAACGTCTCCTCCACG CACCCCTGGGATTGGACTC
<i>DPPA5</i>	F R	GATGCTCCAGTCTATGGCAGAGT GTGAATTCATGGCTTCTCAAGTC	<i>PIWIL2</i>	F R	AGGTCGTGGTATTGGCAGAG TGATCCAGAGAGTGCAGTGG
<i>CDH1</i>	F R	CTTCATAACCCGGAGATTAGTTTTTCT TCTCTTTTCTGTCTTCTGAGACCAT	<i>REX1 / ZFP42</i>	F R	CTTCAAGGAGAGCGCAAAAC GACAGCCTTCAAAGGGACAC
<i>ERAS</i>	F R	ATGTTTGATCTGGGCCTGGG ACCACCGCTTGTACTCAGG	<i>SALL4</i>	F R	CAGGAGTACCAGAGCCGAAG GCTTCGACCTTCCATCTCAG
<i>ESRRB</i>	F R	AGGCTCTGACCACTCTCTGT ACACCAGCTTGTCTGCTGATG	<i>SOX2</i>	F R	GGTACACTGCCTCTCTCTCACAT ACTGTTTCTTACTCTCCTCCCATT
<i>EZH2</i>	F R	AATGGCGATCACAGGATAGG GTTCCAGAGAGGAGGGGAAG	<i>TBX3</i>	F R	GACCATGGAGCCTGAAGAAG TTTATCCAGCCCAGAACC
<i>FBXO15</i>	F R	AACAGACCTCGATGGCATT ATGTGCTCCTCTCCACAAGG	<i>TERT</i>	F R	GAGGAACCCCTCATTTTTTCC GAACCAGCTTGAGCAGGAAG
<i>FGF4 3UTR</i>	F R	GCGATGAGTGCAAGTTCAAAA GTCCGAAGAAAGTGCACCAC	<i>UTF1</i>	F R	CTAAGTTTGGGGTGGGGAAG GTGACATCGTGTCCATCCTG
<i>GATA6</i>	F R	CAGGAAACGAAAACCTAAGAGCAT TTCTCGGGATTAGCGCTCTC	<i>ACTB</i>	F R	CTCCTCCCTGGAGAAGAGCTA CTTCATGATGGAGTTGAAGGTAGTT
<i>GBX2</i>	F R	AATGCCAACTCCAAGACAGG GGTGGCTGTGAGTTTTGCTT	<i>GAPDH</i>	F R	TCGGAGTGAACGGATTG CCTGGAAGATGGTGATGG
<i>GDF3</i>	F R	TAAGAACCATCCCCAAGCTG CACACATGTGGCTCCTGAAC	<i>TBP</i>	F R	AACAGTTCAGTAGTTATGAGC AGATGTTCTCAAACGCTTCG
<i>KLF2</i>	F R	GCCTTCGGTATCTTTGACGA GAGGTGCGAGCTCTTTGTGT	<i>GFP</i>	F R	AGATCCGCCACAACATCGAG TCTCGTTGGGGTCTTTGCTC
<i>KLF4</i>	F R	CCGTCGGCCGGCCCTTAGAG TGTGGGTTTCGAGGTGTGCC	<i>NESTIN</i>	F R	TTGAGGACCTAGGGACGGAG TCCTCATCTGCAAACCCGTC
<i>KLF5</i>	F R	AAGTCCAGATAGACAAGCAGAGATG CTATTGTATCTGACAGGCTGGATGT	<i>PRDM1</i>	F R	CGCCAAGTTCACCCAGTTT GAGGCAACTTCAGGAGGGAC

conventionnelle puis la PCR quantitative est réalisée en triplicat en mélangeant 2 μ L d'ADNc, les amorces et du SYBR Green sur l'appareil BioMark DH System de Fluidigm. L'analyse de l'efficacité des amorces est faite simultanément sur une série de dilution d'un pool d'ADNc (1/4 ; 1/8 ; 1/16 ; 1/32). Les résultats sont visualisés grâce au logiciel « Real-Time PCR Analysis » de Fluidigm. Les gènes de référence choisis sont *ACTB*, *GAPDH* et *TBP*. Les différentiels d'expression sont calculés par la méthode du $2\Delta\Delta C_t$ (Livak and Schmittgen, 2001) en utilisant la moyenne des Cts des trois gènes de ménage comme référence pour chaque gène et la lignée I3 en milieu LIF+2i comme calibrateur : c'est en effet pour cette lignée que le plus de gènes sont détectés (tous sauf *MHCI*). L'expression des gènes *KLF2* et *FOXA2* n'a été détectée dans aucune lignée cellulaire, et les résultats pour *DNMT3B* et *NESTIN* ont été supprimés en raison d'un trop grand nombre de pics parasites. Les comparaisons entre passages précoce et tardif et entre cultures en milieu piPS et en milieu LIF+2i pour les lignées I3 et I4 sont réalisées par un test de Student, tout comme la comparaison entre les lignées NI13 et NI20 et leurs corps embryoides respectifs. Les différences entre systèmes de reprogrammation intégratif et non intégratif sont évaluées en comparant les lignée I3 et I4 avec les 3 lignées non intégratives par une ANOVA à deux facteurs prenant en compte l'effet lignée (les lignées étant cultivées en bFGF et prises à passage précoce).

5. Différenciation des piPSCs

a. Différenciation en corps embryoides

La formation de corps embryoides est observée après le passage de 2 millions de cellules en suspension et sous agitation en milieu DIFF dans des boîtes de Pétri non traitées. Après 6 jours, la moitié des corps embryoides obtenus est transférée en culture adhérente dans des boîtes de culture cellulaire, tandis que l'autre moitié reste en suspension pendant 10 jours environ avant d'être récupérée sous forme de culot par centrifugation et conservée à -20°C.

b. Formation de tératomes

La formation des tératomes a été réalisée par injection intramusculaire de 2 à 3 x 10⁶ cellules au niveau de la patte arrière de souris immunodéficientes (NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ). Les tumeurs ainsi obtenues ont été réséquées 6 à 12 semaines après injection puis fixées dans du PBS/PFA 4%. Les coupes de tissus réalisées après inclusion en paraffine ont été colorées à l'hématoxyline/éosine avant leur analyse histologique. Ces expérimentations ont été réalisées par les membres de l'unité U935 de la plateforme de cellules souches embryonnaires humaines (Villejuif).

Tableau 15 : Lignées cellulaires dans différents états de culture utilisées pour l'analyse transcriptomique

Condition	Système de reprogrammation	Lignée reprogrammée	Lignée cellulaire initiale
A	Intégratif	I3 passage précoce	Fibroblastes t(Y;14)
B	Intégratif	I3 passage tardif	Fibroblastes t(Y;14)
C	Intégratif	I4 passage précoce	Fibroblastes t(Y;14)
D	Intégratif	I4 passage tardif	Fibroblastes t(Y;14)
E	Intégratif	PB20	Amniocytes primaires
F	Non reprogrammé	PEF	PEF
G	Non intégratif	NI12 passage précoce	Fibroblastes t(Y;14)
H	Non intégratif	NI12 corps embryoides	Fibroblastes t(Y;14)

6. Analyse transcriptomique

a. Hybridation sur la puce porcine Agilent 60K customisée

Les extractions d'ARNs sont menées en triplicat pour les différentes lignées cellulaires étudiées (Tableau 15) à l'aide du kit Nucleospin RNA (Macherey-Nagel). Les échantillons d'ARNs sont transmis à la plateforme BioPuces de la Génopole Toulouse Midi-Pyrénées. La validation quantitative et qualitative des échantillons d'ARNs est réalisée au NanoDrop et au Bioanalyser (Agilent Technologies) et permet de s'assurer que les échantillons ont une concentration supérieure à 50ng/μL, des rapports 260/280 et 260/230 supérieurs à 1.8, un RIN (RNA Integrity Number) supérieur ou égal à 9 et un rapport ARN28S/ARN18S supérieur à 1.7. Les échantillons sont ensuite marqués par le fluorochrome Cyanine et convertis en ARNc puis hybridés sur la puce porcine 60K d'Agilent customisée. En effet, la puce porcine 60K produite par Agilent Technologies (Pig Gene Expression Microarray, 020109) a été modifiée afin d'y ajouter des gènes spécifiques du développement des cellules germinales, de la spermatogénèse, de la fertilisation et du développement embryonnaire précoce. Le rendement minimum accepté est de 6pmol Cy/μg ARNc. Pour hybrider les 24 échantillons, 3 lames ont été nécessaires.

b. Exploitation des résultats

i. Analyse statistique

Dans un premier temps, la qualité de chaque hybridation est vérifiée afin de ne conserver que les échantillons bien hybridés. Ainsi, l'un des triplicats de l'échantillon B est supprimé. Dans un second temps, les données sont filtrées afin de ne conserver que celles pour lesquelles la qualité et l'intensité du signal est suffisante dans au moins 2 des 3 réplicats. Sur les 62209 sondes présentes sur la puce, 38114 sont conservées. Les valeurs d'intensité sont ensuite transformées en log₂ puis normalisées par la méthode de normalisation par les quantiles de manière à rendre les distributions d'intensité similaires pour toutes les conditions. L'analyse différentielle est réalisée à l'aide de l'outil Limma du package Bioconductor (Smyth, 2004) pour 19 comparaisons. On considère les oligonucléotides différentiellement exprimés entre deux conditions lorsque la p-valeur ajustée (Benjamini & Hochberg) est inférieure à 0.01 et le logarithme du « fold change » (lfc) supérieur à 2.

ii. Analyse fonctionnelle

Uniquement les sondes annotées par un nom de gène officiel ont été conservées pour l'analyse fonctionnelle. Nous nous sommes concentrés sur 5 listes de gènes différentiellement exprimés - les gènes communément différentiellement exprimés dans les lignées I3, I4, PB20 et NI12 par rapport aux fibroblastes : comparaison lignée reprogrammée / lignée non reprogrammée

Tableau 16 : Mélange pour la réaction de marquage des ADN

	Quantité	Concentration finale
ADN	500ng	10 ng/ μ L
2.5X Random Primers	20 μ L	1X
Solution de dNTPs 10X	5 μ L	1X
Solution de dUTPs marqués 1mM	3.5 μ L	5,7 μ M
Klenow 40U/ μ L	1 μ L	0,8U/ μ L
H2O	qsp 50 μ L	

Tableau 17 : Marquage des ADN

Marquage rouge (Alexa 568)	Marquage vert (Alexa 488)
ADN génomique porc	ADN fibroblaste t(Y ;14)
ADN fibroblaste t(Y;14)	ADN I3 précoce
ADN fibroblaste t(Y;14)	ADN I3 tardif
ADN fibroblaste t(Y;14)	ADN I4 précoce
ADN fibroblaste t(Y;14)	ADN I4 tardif

- les gènes communément différentiellement exprimés entre la lignée NI12 et les lignées I3 et I4 : comparaison des systèmes de reprogrammation
- les gènes différentiellement exprimés entre les corps embryoides et la lignée NI12 dont ils sont issus
- les gènes différentiellement exprimés entre passages tardif et précoce pour les lignées I3 et I4, séparément.

L'analyse fonctionnelle préliminaire a été menée en utilisant les programmes GeneCodis (Carmona-Saez et al., 2007; Nogales-Cadenas et al., 2009; Tabas-Madrid et al., 2012), GOrilla (Eden et al., 2009) et Ingenuity Pathway Analysis (IPA, www.qiagen.com/ingenuity) pour mettre en évidence l'enrichissement statistiquement significatif (FDR>0.05) d'annotations fonctionnelles (notamment celles de Gene ontology : « Biological process », « Molecular function » et « Cellular compartment ») et de voies de signalisation (notamment « KEGG Pathways ») dans ces listes. Pour cela, ces programmes comparent la proportion de gènes associés à une annotation spécifique dans la liste soumise à celle de la même annotation dans un plus grand ensemble de gènes issus du génome humain, chaque site ayant sa base de donnée : Ensembl pour GeneCodis et Ingenuity Knowledge Database pour IPA.

7. Etude de l'instabilité génomique

a. Caryotypes en bandes G

Les cellules sont repiquées à une densité de 50000 cellules/cm² dans des flasques 75cm² traitées à la gélatine 0,1% et contenant des cellules nourricières STO irradiées. A 80% de confluence environ, les cellules sont bloquées en métaphase par 0.4µg/mL de Colcémide pendant 3h puis centrifugées 10 minutes à 1500 rpm. Elles subissent ensuite un choc hypotonique de 2 minutes dans du KCl 0.075M à 37°C et sont pré-fixées grâce à l'ajout de quelques gouttes de fixateur frais (Ethanol:Acide acétique 3:2) avant une nouvelle centrifugation de 10 minutes à 1500 rpm. Le culot est resuspendu dans du fixateur et conservé à 4°C jusqu'à l'étalement sur lame de microscope froide. Pour obtenir des bandes G, les lames sont traitées 1 minute et 15 secondes dans un bain de trypsine 0.025% à température ambiante avant de passer rapidement dans un bain de PBS et enfin d'être colorées grâce à un bain de 8 à 9 minutes dans la solution de coloration contenant le Giemsa. La capture d'image et l'analyse sont réalisées à l'aide du système CYTOVISION et de son logiciel d'analyse d'image Genus.

b. Hybridation génomique comparative (CGH)

L'ADN est extrait de culots cellulaires par une incubation dans un tampon de lyse (10mM Tris-HCl

Tableau 18 : Marquage par peintures chromosomiques (SSC : Sus Scrofa)

Peintures chromosomiques	Marquage
SSC3	direct dUTP - Alexa 568
SSC4	direct dUTP- Alexa 568
SSC5	indirect dUTP-Biotine + Streptavidine-Alexa 488
SSC8	indirect dUTP-Biotine + Streptavidine-Alexa 488
SSC10	direct dUTP - Alexa 568
SSC12	indirect dUTP-Biotine + Streptavidine-Alexa 488

Tableau 19 : Marquage par BAC

Clone de BAC	Cible	Localisation	Marquage
PigI-858F5	CD1	4q15-4q16	direct dUTP- Alexa 568
PigI-736D9	télomère	9pter	indirect dUTP-Biotine + Streptavidine-Alexa 488

pH 8.0, 1mM EDTA, 150mM NaCl, 1% SDS) en présence de protéinase K (0.01 μ M) puis une extraction sur colonne en présence d'une solution de phénol/chloroforme/alcool isoamilique (25 :24 :1). Après l'ajout d'acétate de sodium (0.3M final) et d'éthanol en excès, l'échantillon est centrifugé, lavé dans de l'éthanol 70% et resuspendu dans du tampon TE.

L'ADN est ensuite marqué par la méthode d'amorçage aléatoire, ou « random priming », en incorporant des dUTPs fluorescents Alexa 488 ou 568 (Tableau 16 et 17). Pour cela, 500ng d'ADN additionnés d'amorces aléatoires (1mM final) sont dénaturés 5 minutes dans l'eau bouillante puis refroidis rapidement dans la glace. Un mélange de dNTPs est ensuite ajouté ainsi que la solution de dUTPs marqués et les Fragments de Klenow. Après au minimum 5h à 37°C, on ajoute 5 μ L de Stop Buffer. Les amorces, l'enzyme (Klenow), et le Buffer Stop proviennent du kit Bioprime DNA Labeling SYStem (Invitrogen).

L'ADN marqué est précipité en présence de 35 μ g d'ADN de sperme de saumon (10 μ g/ μ L), de 20 μ g d'ADN compétiteur (1 μ g/ μ L, Cot 1 Porcine Hyblock DNA), de NaCl (0.1M final) et d'éthanol en excès à -20°C. Les sondes ainsi obtenues sont centrifugées et lavées à l'éthanol 70% froid, séchées sous cloche à vide puis resuspendues dans du milieu d'hybridation (Hybridization solution, Sigma) et incubées au minimum 4h à 37°C.

Les lames de métaphases contrôles sont dénaturées entre 2 et 5 minutes à 73°C dans une solution de 2X SSC 70% formamide. Les sondes sont dénaturées 7 minutes dans l'eau bouillante puis pré-hybridées pendant 45 minutes à 37°C. Chaque couple de sonde est déposé sur une lame de métaphase contrôle et l'hybridation se déroule pendant 3 jours à 37°C. A l'issue de l'incubation, les lames hybridées sont lavées plusieurs fois à 45°C dans une solution de 2X SSC 50% formamide. Enfin, une goutte de milieu de montage Vectashield contenant du DAPI est déposée entre lame et lamelle et les métaphases sont observées au microscope. Les métaphases sont capturées puis caryotypées à l'aide du logiciel Genus de Cytovision. Le calcul du rapport de fluorescence entre ADN contrôle en rouge et ADN cas en vert le long des chromosomes permet d'identifier des zones de duplication ou de délétion de l'ADN dans les lignées cellulaires analysées.

c. Hybridation in situ en fluorescence (FISH)

2 μ g de chaque sonde de peinture chromosomique (Yerle et al., 1993), ou 1 μ g d'ADN de BAC, sont marqués en suivant le même protocole que pour la CGH, soit par marquage direct soit par marquage indirect en utilisant des dUTPs couplées à de la biotine (Tableau 18 et 19).

Les sondes de peintures chromosomiques ou BACs sont ensuite purifiées au travers d'une colonne G50 Illustra et les couples de sondes sont précipités ensemble. Après centrifugation et reprise dans du milieu d'hybridation, les sondes sont incubées 4h au bain-marie à 37°C. Les lames sont traitées de la façon suivante : lavage de 30 minutes dans du 2X SSC puis incubation 1h à 37°C dans une solution de RNase diluée 20 fois dans du 2X SSC puis 9 minutes dans une solution de protéinase K diluée dans du

2X SSC (0.1µg/mL), deux rinçages dans du 2X SSC, et déshydratation dans des bains successifs d'éthanol (70%, 80%, 90%) de 3 minutes chacun et séchage à froid. La sonde est alors déposée sur la lame, recouverte d'une lamelle que l'on scelle au Rubber Cement. Les lames sont placées dans un hybridizer (Dako) programmé comme suit : 5 minutes de dénaturation à 75°C et 24h à 48h d'hybridation à 37°C. Après l'hybridation et retrait des lamelles, les lames sont lavées 2 fois 15 minutes dans du SSC 2X à température ambiante sous légère agitation, 3 fois 3 minutes dans une solution de 2X SSC 50% formamide pH=7 à 45°C et finalement 3 fois 3 minutes dans du 2X SSC à 45°C. Le marquage ADN dUTP-Biotine est révélé en déposant 5µg/mL de Streptavidine couplée à un fluorochrome, ici l'Alexa 488, pendant 2h à 37°C. Enfin, une goutte de milieu de montage contenant du DAPI est déposée entre lame et lamelle et les métaphases sont observées au microscope à fluorescence.

**ETUDE DE LA METHYLATION
DE L'ADN SPERMATIQUE
PORCIN**

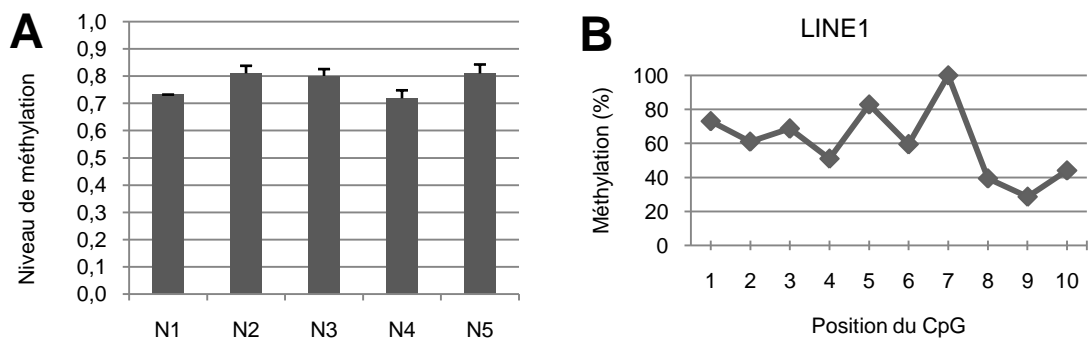


Figure 24 : A. Niveau de méthylation global de l'ADN spermatique dans 5 animaux à qualité de semence normale. B. Profil de méthylation de la séquence répétée LINE1 dans les verrats contrôles.

Tableau 20 : Filtre des lectures alignées sur le génome après le séquençage haut-débit.

Echantillon	Nombre total de lectures	Lectures alignées	Lectures alignées (%)	Lectures appariées	Lectures appariées (%)	Lectures appariées uniques	Lectures appariées unique (%)
N1 5hmC1	74 079 452	59171381	79.88	55 783 138	75.30	48 345 701	65.26
N1 5hmC2	50 569 156	41123712	81.32	38 973 540	77.07	35 977 710	71.15
N1 5mC1	130 762 846	111054092	84.93	106 879 970	81.74	100 595 962	76.93
N1 5mC2	87 139 854	74095867	85.03	71 148 798	81.65	67 474 474	77.43
N1 input	84 128 318	71851395	85.41	69 269 834	82.34	68 124 786	80.98
N2 5hmC1	58 205 894	44 027 134	75.64	39 168 092	67.29	17 944 425	30.83
N2 5hmC2	76 448 658	58 585 282	76.63	52 204 056	68.29	10 553 085	13.80
N2 5mC1	77 171 612	66 263 069	85.86	62 045 322	80.40	55 932 294	72.48
N2 5mC2	81 843 080	70 269 494	85.86	65 693 526	80.27	58 637 741	71.65
N2 input	91 357 860	77 982 786	85.36	72 171 728	79.00	70 117 843	76.75

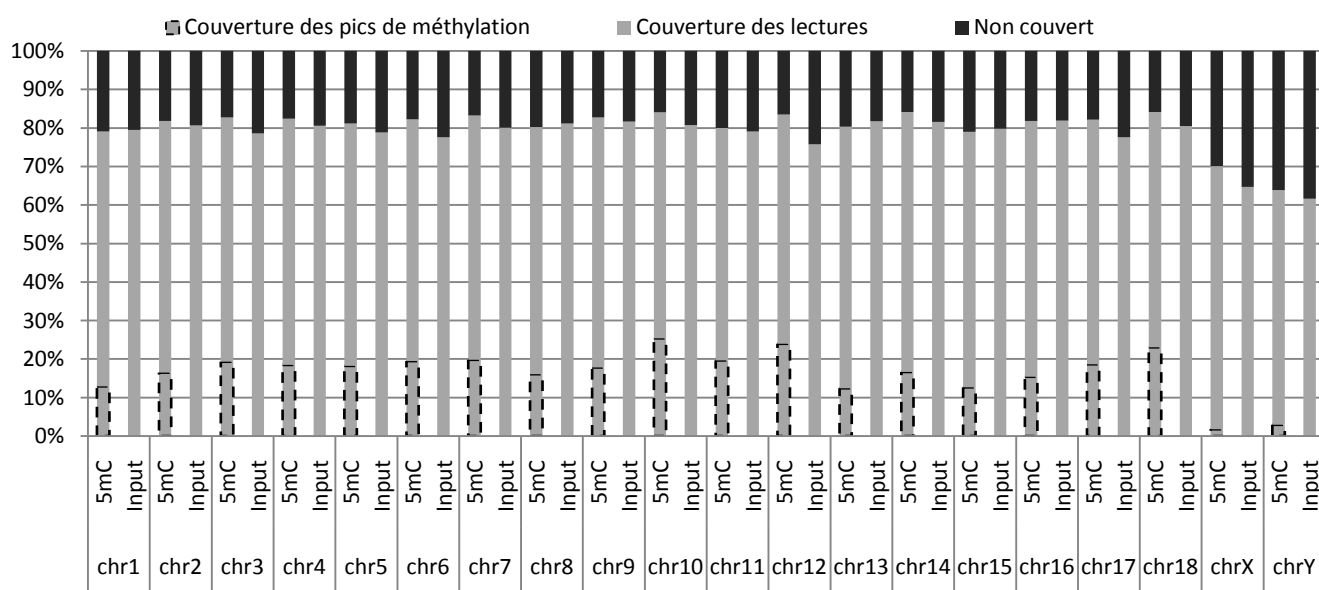


Figure 25 : Pourcentage de couverture des lectures et des pics de méthylation sur chaque chromosome.

Nous décrivons ici dans une première partie les résultats obtenus lors de l'étude de la méthylation de l'ADN spermatique porcin chez des verrats fertiles et infertiles, puis ces résultats sont discutés dans les paragraphes II (comparaison inter-espèce) et III (comparaison entre animaux fertiles et infertiles). Les résultats de l'analyse comparative de la méthylation de l'ADN entre animaux fertiles et infertiles font l'objet d'un article en premier auteur paru en ligne dans le journal « Biology of Reproduction » le 15 octobre 2014 (Annexe 1), tandis que les résultats de l'analyse tout-génome par MedIP-Seq feront l'objet d'un second article en co-auteur.

I. Présentation des résultats

1. Un ADN spermatique fortement méthylé et hydroxyméthylé

Dans un premier temps, l'utilisation de la technique LUMA (Luminometric-based Methylation Assay) nous a permis de quantifier le niveau global de méthylation de l'ADN spermatique porcin. En effet, cette technique qui associe digestion enzymatique et pyroséquençage de l'ADN génomique digéré permet de quantifier la méthylation de toutes les séquences CCGG du génome, cibles des enzymes HpaII et MspI. Sur un groupe d'animaux contrôles composé de 5 individus, le niveau de méthylation atteint $77.83\% \pm 4.7\%$ (Figure 24A).

Ensuite, le pyroséquençage d'une région de la séquence répétée LINE1 comportant 10 CpGs a permis d'évaluer le niveau de méthylation de cette région à $61.5\% \pm 3.2\%$, bien que le niveau soit variable le long de la séquence analysée, allant de 31% pour un CpG à 100% pour un autre (Figure 24B). Même si cette région n'est pas méthylée de manière uniforme sur toute sa longueur, elle est souvent utilisée comme référence pour les analyses de méthylation car la séquence LINE1 est fortement présente dans le génome (Yang et al., 2004). Ce résultat nous permet de confirmer que le niveau de méthylation des CpGs est élevé sur l'ensemble du génome.

Enfin, le MeDIP-seq réalisé sur deux animaux contrôle révèle également la forte présence de marques de méthylation sur l'ADN spermatique. Pour chaque animal, les deux réplicats techniques immunoprécipités pour la méthylation ou l'hydroxyméthylation et l'échantillon contrôle (ou input) ont été séquencés à une profondeur d'environ 30 millions de lectures par librairie. Sur le nombre de lectures obtenues, entre 75% et 86% selon les échantillons ont été alignées sur la séquence du génome porcin (version 10.2). Parmi ces lectures, nous n'avons conservé que celles étant alignées en paires convergentes sur le même chromosome, et les duplicats ont été supprimés. Ainsi, entre 65% et 85% des lectures ont été conservées, à l'exception des échantillons N2 enrichis en hydroxyméthylation qui présentent un taux de lectures conservées beaucoup plus bas (Tableau 20).

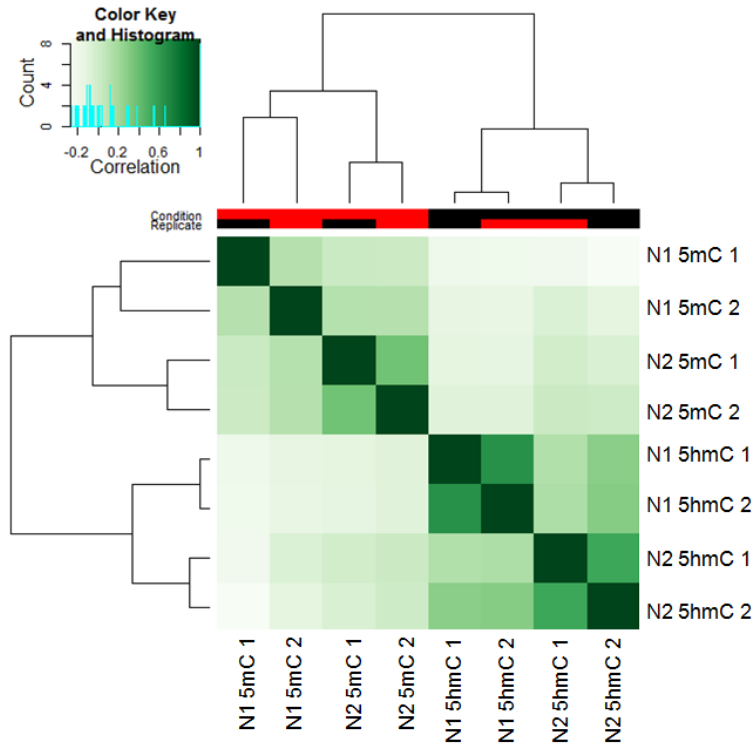


Figure 26 : Corrélation entre les réplicats techniques (répétitions du MeDIP et hMeDIP) et biologiques (animaux N1 et N2) établie par DiffBind. Le vert foncé indique une corrélation de 1.

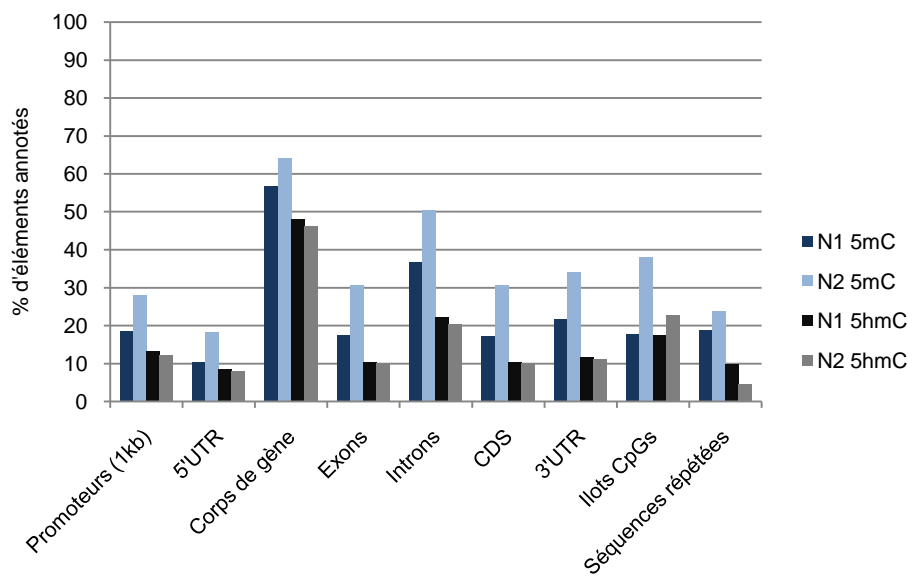


Figure 27 : Pourcentage d'éléments annotés sur le génome porcin ayant une intersection avec les pics de méthylation (Corps de gène : ensemble du gène du 5'UTR au 3'UTR ; CDS : Coding DNA Sequence).

Le MeDIP-seq est une approche basée sur l'affinité d'un anticorps pour les cytosines méthylées de l'ADN. En comparant les résultats du séquençage pour l'échantillon ayant subi l'immunoprécipitation et l'échantillon contrôle, il apparaît que le taux de couverture moyen du génome est similaire (Figure 25). On peut en conclure que le génome porcin est en grande partie méthylé dans le sperme, puisque la fraction enrichie en méthylation recouvre une portion du génome aussi grande que la fraction totale (non enrichie). Il est intéressant de noter que le taux de couverture des chromosomes X et Y est diminué à la fois dans les échantillons immunoprécipités et dans l'échantillon contrôle.

2. Profil de méthylation dans le génome des cellules spermatiques

L'utilisation d'un outil de recherche de pics, macs1.4, a ensuite permis de détecter les zones du génome les plus enrichies en méthylation ou en hydroxyméthylation, en utilisant les données de l'échantillon contrôle comme référence. Tout d'abord, des analyses menées sur chaque réplicat technique ont permis de s'assurer de leur similitude. En effet, lorsqu'on effectue la recherche de pics dans l'un des réplicats en utilisant le second réplicat comme référence au lieu de l'échantillon contrôle, le nombre de zones détectées diminue drastiquement. De plus, l'utilisation de Diffbind, qui permet de classer les échantillons par leur niveau de corrélation, indique que les réplicats techniques sont plus proches entre eux que les réplicats biologiques (Figure 26).

Les réplicats techniques ont donc été fusionnés pour la recherche de pics. Chez l'animal N1, 356794 pics de méthylation et 133577 pics d'hydroxyméthylation ont été identifiés. Chez l'animal N2, ce nombre est un peu plus élevé en méthylation, avec 383872 pics, et un peu plus faible en hydroxyméthylation, avec 107048 pics. En moyenne, les pics de méthylation recouvrent 15% du génome (Figure 25) et les pics d'hydroxyméthylation en recouvrent 7%. Les différences entre animaux peuvent s'expliquer par la différence du nombre de lectures conservées après séquençage, et notamment par le faible nombre de lectures conservées pour l'analyse de l'hydroxyméthylation de l'animal N2. Néanmoins, plus de 65% des pics recensés sont communs aux deux animaux.

En croisant les coordonnées chromosomiques des pics avec les coordonnées de certaines classes d'annotation du génome, on peut établir un profil de méthylation du génome (Figure 27). Il apparaît que plus de 50% des gènes comportent un pic de méthylation, majoritairement dans les introns, tandis que les régions promotrices et les 5'UTR ne sont que faiblement méthylées. On constate également une méthylation moindre dans les îlots CpGs, souvent associés aux promoteurs. L'étude locale par MeDIP-qPCR de la méthylation des promoteurs des gènes *OCT4* (ou *POU5F1*), *DDX4*, *DAZL*, *ACAB8* et *C10ORF67* vient confirmer ce résultat, ces cinq promoteurs présentant un niveau de méthylation très bas en comparaison de la séquence *LINE1* (voir Figure 29). Ces résultats sont comparables à ceux observés chez l'homme et la souris, à l'exception du promoteur de *POU5F1* qui

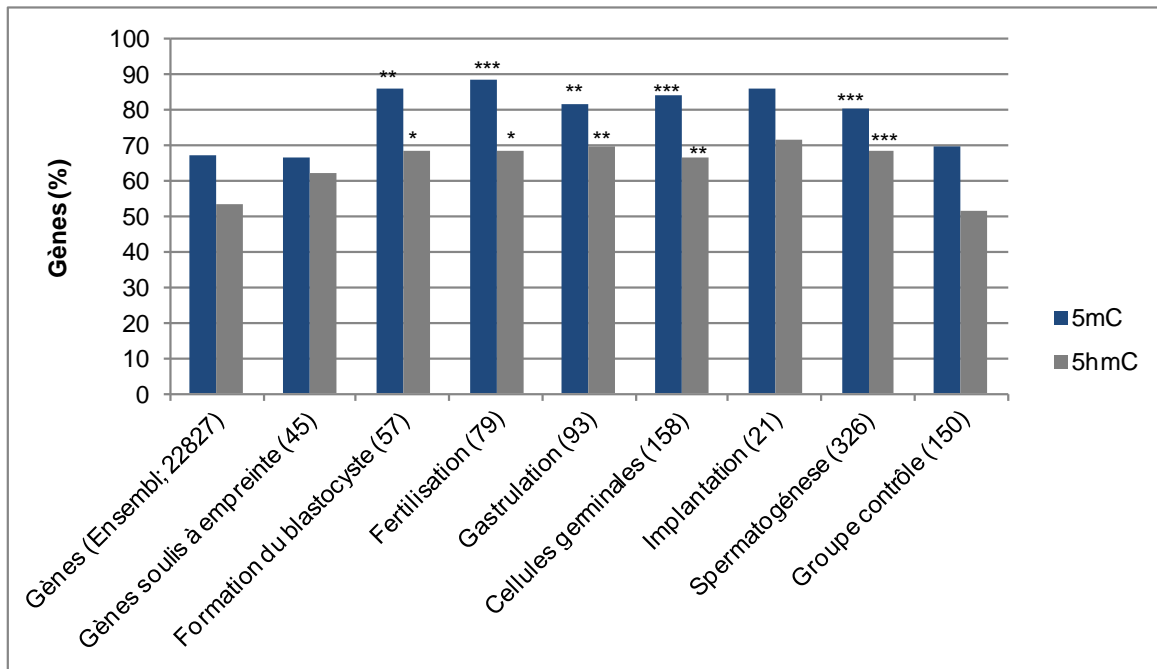


Figure 28 : Pourcentage des gènes impliqués dans la biologie des cellules germinales et du développement embryonnaire précoce ayant une intersection avec les pics de méthylation. Test de χ^2 : * $p < 0.05$; ** $p < 0.01$; *** $p < 0.01$.

Tableau 21 : Détection de l'expression biallélique ou monoallélique dans des tissus fœtaux.

	Position du SNP (Assemblage Sscrofa10.2)	Référence du SNP (Ensembl)	Génotype des parents				Profil d'expression		Expression
			♂LW	♀MS	♂MS	♀LW	LWxMS	MSxLW	
<i>WT1</i> exon 9	chr2:30945103	rs55619464	G/A	A/A	A/A	G/G	G+A	G+A	Biallélique
<i>CNTN3</i> exon 31	chr13:61558099	rs339003919	G/G	A/A	A/A	A/A	G+A	A	Biallélique
<i>IMPACT</i> exon 11	chr6:102037866	rs329889650	T/T	G/G	G/G	T/T	T+G	T+G	Biallélique
<i>RASGRF1</i> exon 15	chr7:53697957	rs337793313	G/A	G/G	G/G	A/A	A ou G	G ou G+A	Paternelle
<i>GRB10</i> exon 7	chr9 :150298033	rs330552874	C/C	C/T	T/T	C/C	C ou C+T	C+T	Biallélique
<i>QPCT</i> exon 8	chr3 :109334945	rs81214310	A/A	G/G	G/G	A/A	A+G	A+G	Biallélique

apparaît être spécifiquement méthylé chez l'homme, indiquant une possible différence pour la réactivation de ce gène au cours de l'embryogénèse précoce entre les espèces.

Nous avons ensuite analysé la méthylation des gènes spécifiques de la gamétogénèse et du développement embryonnaire précoce. Les listes de gènes par fonction biologique ont été constituées en utilisant Gene Ontology : 326 gènes associés à la spermatogénèse, 158 gènes spécifiques de la lignée germinale, 79 gènes impliqués dans la fertilisation, 57 gènes dans la formation du blastocyste, 21 gènes dans l'implantation de l'embryon et 93 gènes dans la gastrulation. Un groupe de 150 gènes pris au hasard a été utilisé comme contrôle, et un groupe de 45 gènes soumis à empreinte parentale a été ajouté. On constate que ces groupes de gènes spécifiques, à l'exception des gènes soumis à empreinte et du groupe contrôle, sont plus méthylés que l'ensemble des gènes (Figure 28, test du χ^2 avec correction de Yates).

3. Focus sur les gènes soumis à empreinte parentale

a. Choix des gènes et recherche d'empreinte parentale chez le porc

Les gènes soumis à empreinte étant généralement différenciellement méthylés, soit sur leur promoteur soit sur une zone spécifique (DMR), entre l'allèle paternel et l'allèle maternel, nous nous sommes penchés sur le profil de méthylation de ces gènes dans les gamètes mâles. Chez le porc, de nombreux gènes ont été décrits comme soumis à empreinte, mais leur nombre est encore réduit comparé aux études menées chez l'homme et la souris. Les gènes *IGF2*, *DLK1*, *MEST* (ou *PEG1*), *PEG3*, *PEG10*, *NNAT*, *GNASXL*, *RTL1*, *RASGRF1*, *XIST* et *IMPACT* sont connus pour leur expression paternelle chez l'homme et/ou la souris, tandis que les gènes *H19*, *IGF2R*, *NESP55*, *GRB10*, *OSBPL1A*, *MEG3*, *HM13*, *CNTN3*, *WT1*, *UBE3A*, et *QPCT* sont décrits comme exprimés sur l'allèle maternel. Sur ces 22 gènes, seuls 11 ont été confirmés comme soumis à empreinte chez le porc : *HM13* et *UBE3A* ont été décrits comme bialléliquement exprimés tandis que les autres n'ont pas encore été analysés dans cette espèce. Les régions étudiées pour chacun de ces gènes (voir Article en Annexe 1) ont été sélectionnées soit par homologie de séquence avec les DMRs définis chez la souris, soit par recherche d'îlots CpGs.

Pour certains gènes dont l'empreinte n'a pas encore été étudiée chez le porc, l'utilisation de croisements réciproques entre animaux de race Large White (LW) et Meishan (MS) nous a permis de rechercher l'empreinte parentale dans plusieurs tissus fœtaux grâce au polymorphisme génétique existant entre ces deux races (Tableau 21). Tout d'abord, le séquençage de l'ADN des parents a permis l'identification de SNPs dans les exons des gènes étudiés. En séquençant l'ADNc chez des descendants LWxMS et MSxLW de ces quatre animaux, il est donc possible de déterminer si l'expression est biallélique ou s'il n'y a qu'un seul allèle exprimé, et dans ce cas de connaître son

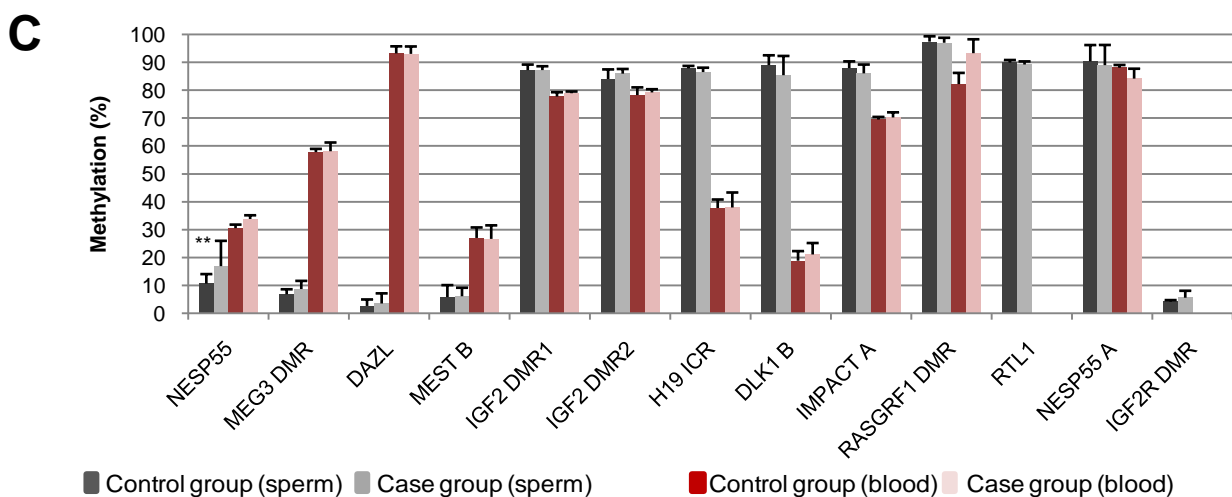
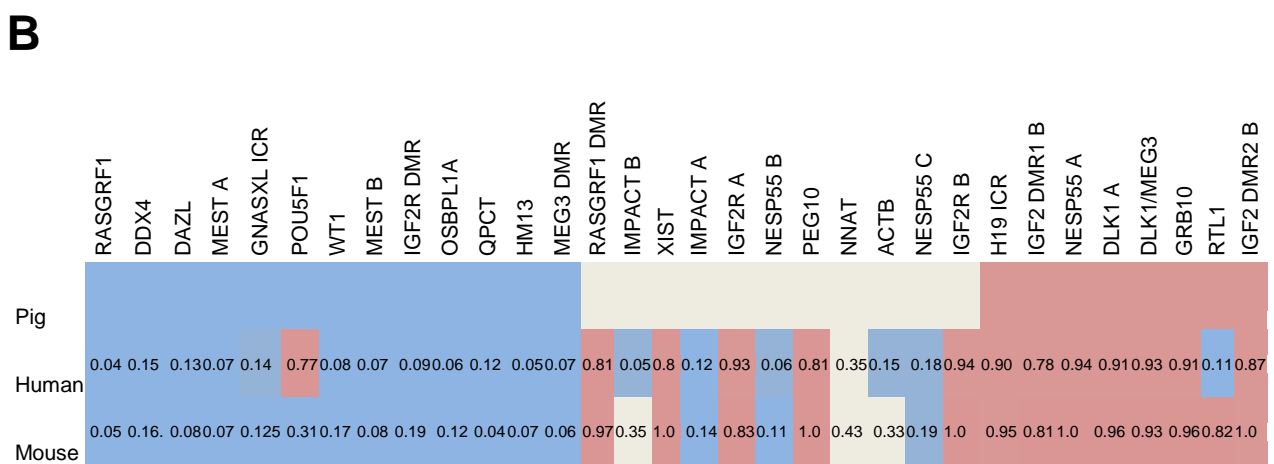
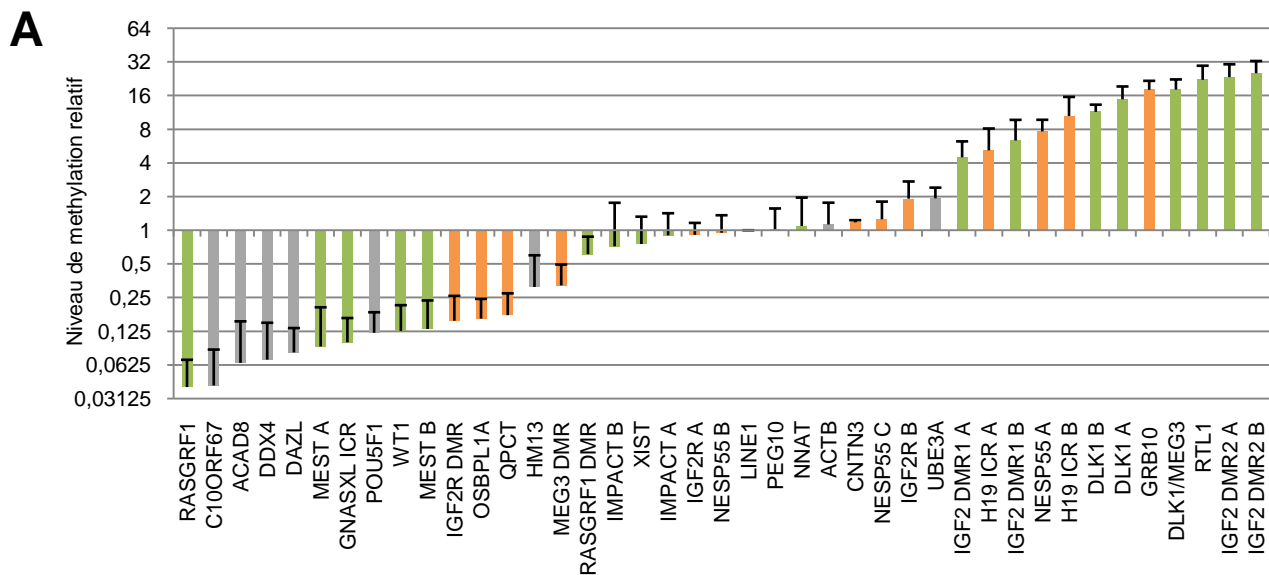


Figure 29 : La méthylation locale de l'ADN spermatique est majoritairement conservée entre espèces et varie peu en fonction du statut de fertilité des animaux. **A.** Niveau de méthylation relatif par rapport à la référence LINE1 de gènes paternellement exprimés (vert), bialléliques (gris) ou maternellement exprimés (orange) dans 5 animaux à qualité de semence normale par MeDIP-qPCR. Barre d'erreur : écart-types. **B.** Comparaison des données de méthylation de l'ADN spermatique porcin obtenu par MeDIP-qPCR avec les données établies chez l'homme et la souris par séquençage tout-génome d'ADN converti au bisulfite. Bleu : faiblement méthylé (<33% pour l'homme et la souris ou fold change inférieur à 0,5 pour le porc) ; gris : modérément méthylé (33-66% ou fold change entre 0,5 et 2) ; rouge : fortement méthylé (>66% ou fold change supérieur à 2). **C.** Niveau de méthylation absolu calculé par pyroséquençage dans le sperme (gris) et le sang (rouge) d'animaux contrôles (couleur foncée) ou à mauvaise qualité de semence (couleur claire). Welsh test ** $p < 0,01$

origine parentale. Il est toutefois important de noter que la mutation est parfois hétérozygote chez l'un des deux parents (*WT1*, *RASGRF1*, *GRB10*), ou présente dans un seul sens de croisement (*CNTN3*), rendant certains résultats non informatifs quand à l'origine parentale de l'allèle exprimé. Les gènes *IMPACT* et *QPCT*, dont l'expression a été détectée dans les quatre tissus fœtaux testés, se sont révélés exprimés de manière biallélique dans ces quatre tissus. Pour les gènes restants, l'expression n'a pas été détectée dans l'ensemble des tissus fœtaux étudiés. Les gènes *CNTN3* et *GRB10*, exprimés maternellement chez la souris, se sont également révélés exprimés de façon biallélique dans le muscle, les poumons et le cerveau. De la même manière, l'expression paternelle de *WT1* chez la souris ne semble pas être conservée chez le porc, l'expression simultanée de séquences contenant un G et un A indiquant une expression biallélique du gène dans le muscle, seul tissu dans lequel l'expression de *WT1* a été détectée. A l'inverse, nos résultats corroborent l'expression paternelle du gène *RASGRF1* chez la souris dans le poumon et le cerveau porcine puisque la moitié des animaux LWxMS expriment l'allèle A présent uniquement chez le père, l'autre moitié exprimant l'allèle G non informatif.

b. Profil de méthylation des gènes soumis à empreinte conservé chez le porc

Le niveau de méthylation relatif des gènes par rapport à celui de la séquence répétée LINE1 a été évaluée par immunoprécipitation de l'ADN méthylé et PCR quantitative pour 5 animaux contrôles (Figure 29A). Les valeurs ont ensuite été comparées aux données tout-génome disponibles chez l'homme et la souris (Figure 29B). Certains des résultats obtenus ont ensuite fait l'objet d'une validation par conversion bisulfite et pyroséquencage, ce qui permet d'obtenir une quantification absolue sur un nombre précis de CpGs (Figure 29C). Outre les promoteurs de gènes non soumis à empreinte, on retrouve parmi les régions faiblement méthylées *RASGRF1*, *MEST*, *WT1*, *OSBPLIA*, *QPCT*, *HM13* ainsi que les DMRs de *GNASXL*, *IGF2R*, et *MEG3*, conformément à ce qui est observé chez l'homme et la souris. Si le 5'UTR du gène *RASGRF1* apparaît très faiblement méthylé et son DMR assez modérément, l'analyse par pyroséquencage a révélé que le DMR est en réalité très fortement méthylé, en accord avec les données établies chez la souris (Figure 29C).

Du côté des régions fortement méthylées, on retrouve en accord avec la littérature les DMRs du locus soumis à empreinte *IGF2/H19*. En effet, l'expression réciproque de ces deux gènes suivant l'origine parentale est dépendante de la méthylation sur l'allèle paternel de tous les DMRs. On retrouve également *DLK1* dont la méthylation de l'allèle paternel est connue chez le porc, l'homme et la souris. *RTL1*, un gène proche du locus *DLK1/MEG3*, est également reporté comme exprimé et méthylé sur l'allèle paternel chez la souris. Nos données indiquent que ce profil est conservé dans le sperme porcine où *RTL1* est fortement méthylé. A l'inverse, la région homologue humaine est déméthylée. Tout comme chez l'homme et la souris, la région *NESP55* A située très en amont du gène est fortement

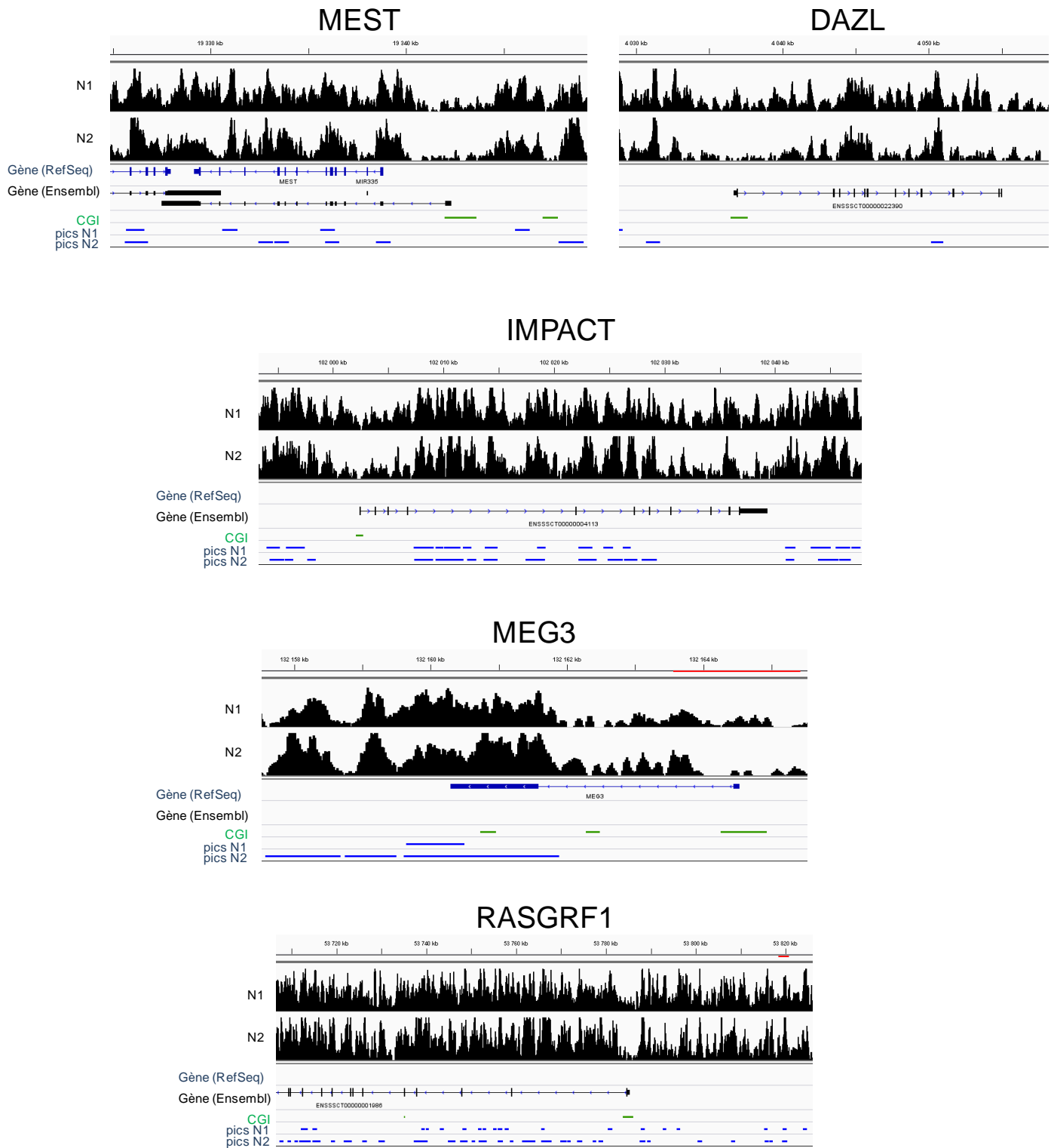


Figure 30 : Profil de méthylation de l'ADN spermatique porcin autour des gènes *MEST*, *DAZL*, *IMPACT*, *RASGRF1*, *MEG3* et *GNAS* obtenu par MeDIP-Seq révélant la faible méthylation spécifique des TSS et des îlots CpGs (barres vertes) correspondants aux régions promotrices ainsi que la méthylation sexe-spécifique des DMRs (trait rouge) des gènes soumis à empreinte. Barres bleues: pics de méthylation.

méthylée chez le porc, reflétant le fort niveau de méthylation global de l'ADN spermatique. Le niveau élevé de méthylation de toutes ces régions a été confirmé par l'analyse en pyroséquençage. Enfin, de la même manière que chez l'homme et la souris, le corps du gène de *GRB10* s'est avéré fortement méthylé.

Le MeDIP-qPCR ne permettant qu'une quantification relative, il est difficile de tirer des conclusions sur les régions dont le niveau de méthylation ne varie que peu (« fold change » compris entre 0.5 et 2) par rapport à la séquence de référence. Parmi ces régions se trouvent le promoteur et le corps du gène *IMPACT*, dont l'analyse par pyroséquençage indique une très forte méthylation dans le corps du gène. Chez l'homme et la souris, le corps du gène est faiblement voire modérément méthylé tandis que le promoteur est très faiblement méthylé. Les régions sélectionnées pour les gènes *NNAT*, *PEG10* et *IGF2R* correspondent au corps du gène ou au 3'UTR et non au DMR : leur niveau de méthylation n'est donc pas informatif sur le statut d'empreinte parentale du gène chez le porc. Toutefois, le niveau de méthylation des régions choisies est proche de celui indiqué dans les régions correspondantes chez l'homme et la souris, suggérant encore une forte conservation du profil de méthylation de l'ADN spermatique entre espèces. Enfin, la faible méthylation de la région promotrice du gène *NESP55* décrite chez l'homme et la souris dans le sperme, qui n'apparaît pas lors de l'analyse par MeDIP-qPCR, est bien conservée chez le porc d'après les données de pyroséquençage.

c. Etude locale ou globale : des résultats comparables

La combinaison de deux techniques d'analyses des résultats du MeDIP, l'une se focalisant sur quelques régions uniquement (MeDIP-qPCR) et la seconde se généralisant à l'ensemble du génome (MeDIP-Seq), nous permet de contrôler nos résultats (Figure 30). Par exemple, la faible méthylation dans les régions promotrices, recouvertes par un îlot CpG, est retrouvée à la fois en PCR quantitative et en séquençage pour les gènes *MEST*, *DAZL*, *MEG3*, *IMPACT*, *RASGRF1* et *GNAS*. A l'inverse, le corps des gènes *IMPACT*, *RASGRF1*, *MEST* et *GNAS* apparaît fortement méthylé. Le DMR de *RASGRF1* coïncide avec un pic de méthylation détecté dans les deux animaux, tandis que le DMR de *MEG3* apparaît faiblement méthylé.

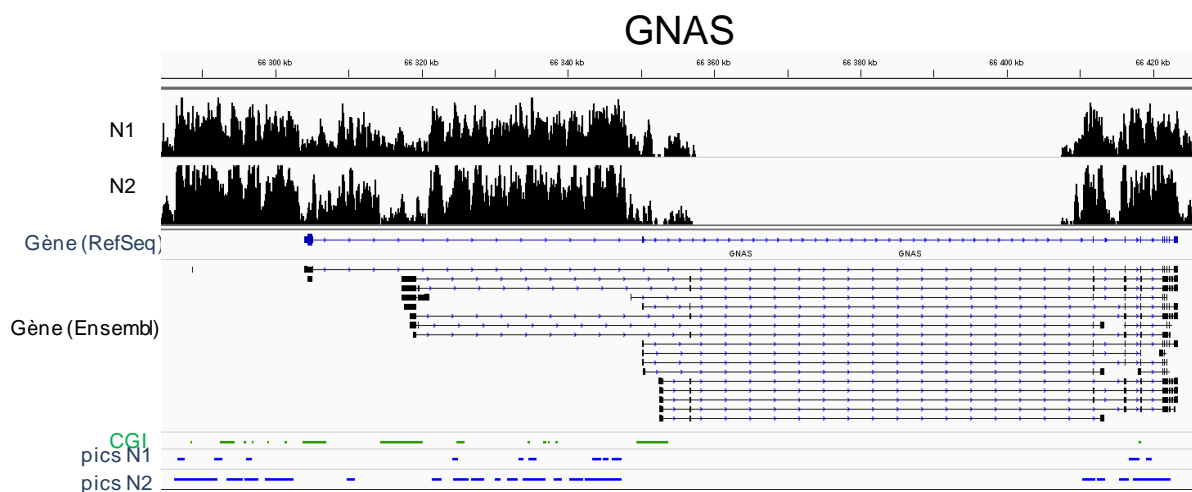


Figure 30 - suite.

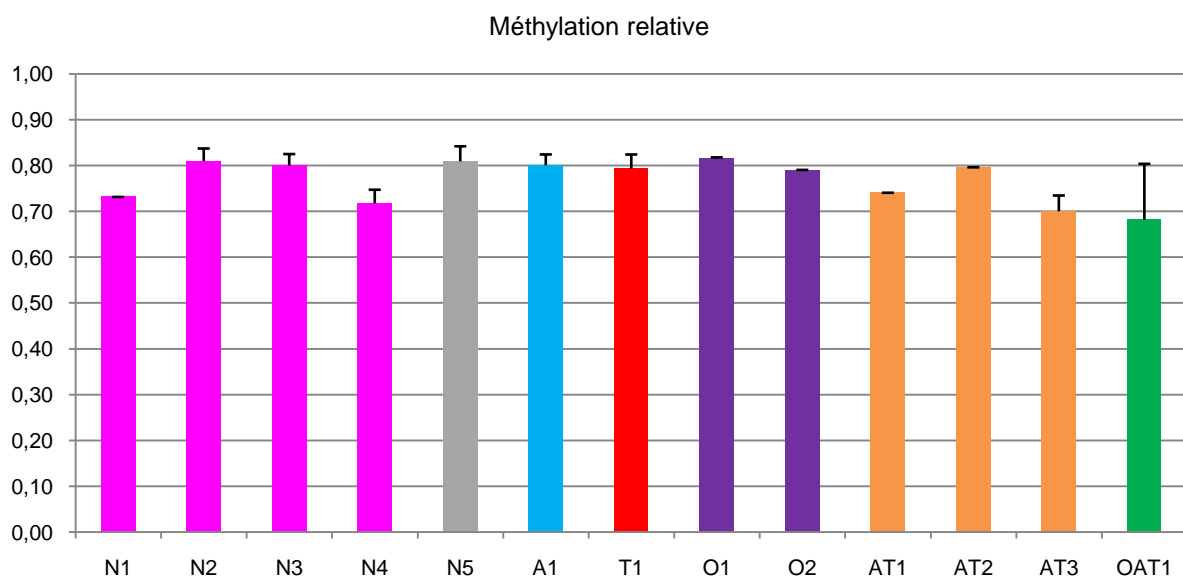


Figure 31 : Le niveau de méthylation global de l'ADN spermatique ne dépend pas de la qualité de la semence chez le porc. N1-N4 : animaux fertiles normaux (rose) ; N5 : fertile transloqué $t(3;4)$ (gris) ; A1 : asthénospermique (bleu) ; T1 : tératospermique (rouge) ; O1-2 : oligospermiques (violet) ; AT1-AT3 : asthénotératospermiques (orange) dont un transloqué $t(13;17)$ (AT3); OAT1 : oligo-asthéo-tératospermique transloqué $t(1;14)$.

4. Etude de la méthylation de l'ADN spermatique chez des verrats à qualité de semence dégradée

a. Un niveau de méthylation global conservé chez les verrats infertiles

La comparaison du niveau de méthylation global de l'ADN spermatique entre 5 animaux contrôles et 8 animaux présentant des paramètres de semence dégradés (Cf Tableau 5) indique que les deux groupes arborent un niveau de méthylation similaire dans le sperme, le groupe expérimental se révélant méthylé à $76.6\% \pm 6.1\%$, valeur très proche de celle observée dans le groupe contrôle (77.83% , t -test p -valeur = 0.611) (Figure 31).

Au niveau local, l'analyse sur 38 régions n'a révélé que peu de différences entre les deux groupes dans l'ADN spermatique. En effet, bien que plusieurs loci présentent une différence significative par l'analyse de MeDIP-qPCR (voir article en Annexe 1) la grande variabilité des résultats due à la faible reproductibilité de la méthode et les faibles valeurs de « fold change » observées n'en font pas un outil suffisamment adapté à l'analyse comparative de nos deux conditions. Ainsi, en utilisant le pyroséquençage, la grande majorité de ces régions se sont avérées présenter un niveau constant de méthylation entre animaux fertiles et animaux à qualité de semence dégradée (Figure 29C). Cela concerne entre autres des régions telles que l'ICR de *H19* ou le promoteur de *MEST* qui sont reportés dans la littérature comme hypo ou hyperméthylées chez des patients à spermogramme dégradé.

b. Variations spécifiques dans le locus GNAS

La région du gène *GNAS* est la seule présentant une augmentation significative de la méthylation dans plusieurs CpG la composant, et ce quelle que soit la méthode employée. Le niveau de méthylation augmente en effet de 4 à 8 fois dans les îlots CpG couvrant les promoteurs et les premiers exons de *NESP55* et de *GNASXL* dans une sous-population du groupe à qualité de semence dégradée. Les animaux concernés sont un tératospermique, un asthéo-tératospermique porteur du remaniement $t(13;17)$ et l'animal oligo-asthéo-tératospermique porteur de la translocation $t(1;14)$. Le dessin de plusieurs amorces supplémentaires a permis d'évaluer le niveau de méthylation dans les principaux îlots CpG recouvrant le locus *GNAS* (Figure 32A). Très en amont de *NESP55*, le niveau de méthylation est très élevé et ne varie pas entre les deux groupes, conformément aux résultats obtenus sur l'ensemble du génome spermatique porcin (Figure 32B). Dans l'îlot CpG recouvrant le promoteur du transcrit *NESP55*, ainsi que dans celui recouvrant le promoteur de *GNASXL*, le niveau de méthylation basal est très bas chez les animaux contrôles et quelques animaux à qualité de semence dégradée, mais il augmente significativement dans le sperme de 3 animaux (Figure 32B). Il est

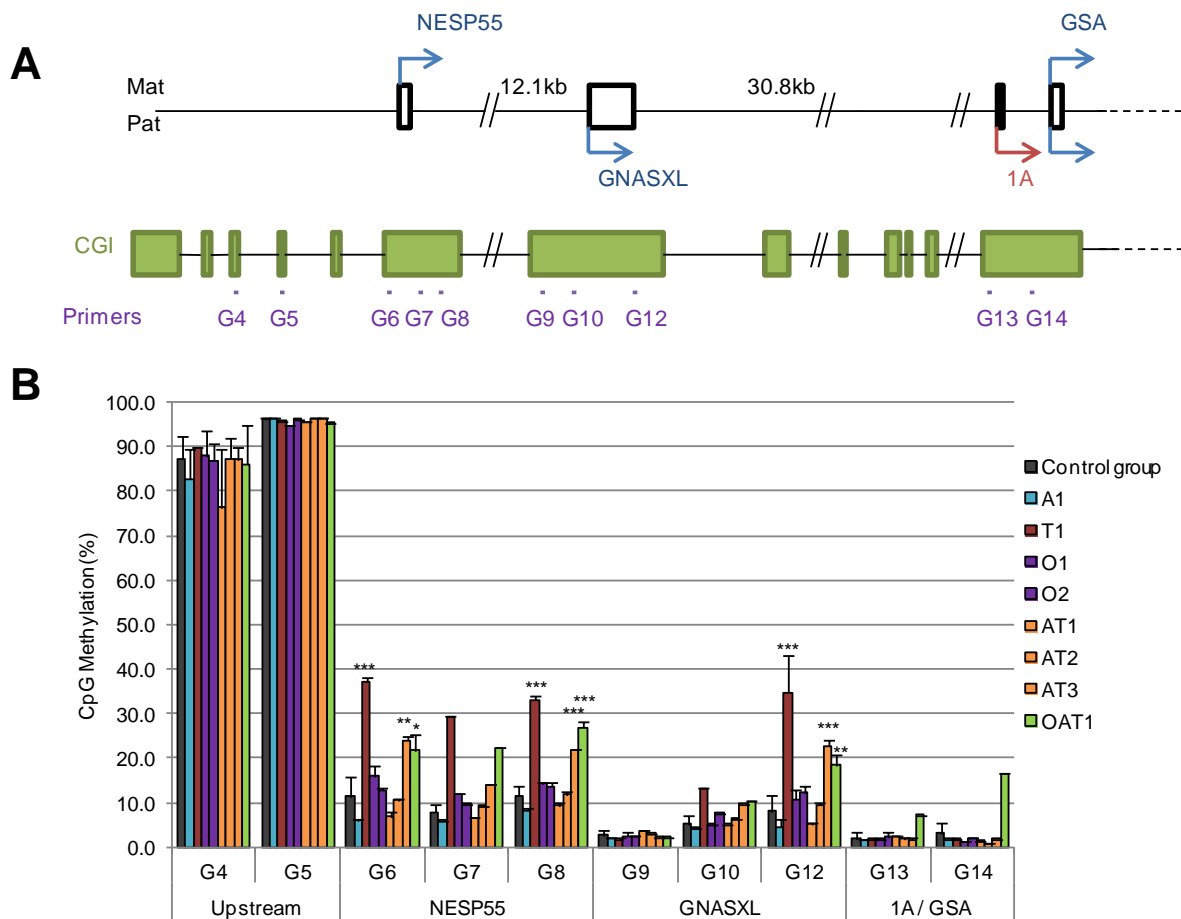


Figure 32 : Profil de méthylation du locus *GNAS* par conversion bisulfite et pyroséquence. **A.** Schéma du locus *GNAS* soumis à empreinte composé de transcrits codants (flèches bleues) ou non codants (flèches rouges) dans lequel des amorces de pyroséquence ont été dessinées (points violets) au niveau des îlots CpGs (rectangles verts). **B.** Le niveau de méthylation augmente pour 3 animaux dans les îlots CpGs couvrant le promoteur et le premier exon de *NESP55* et de *GNASXL*. *t*-test * $p < 0.05$; ** $p < 0.011$; *** $p < 0,001$.

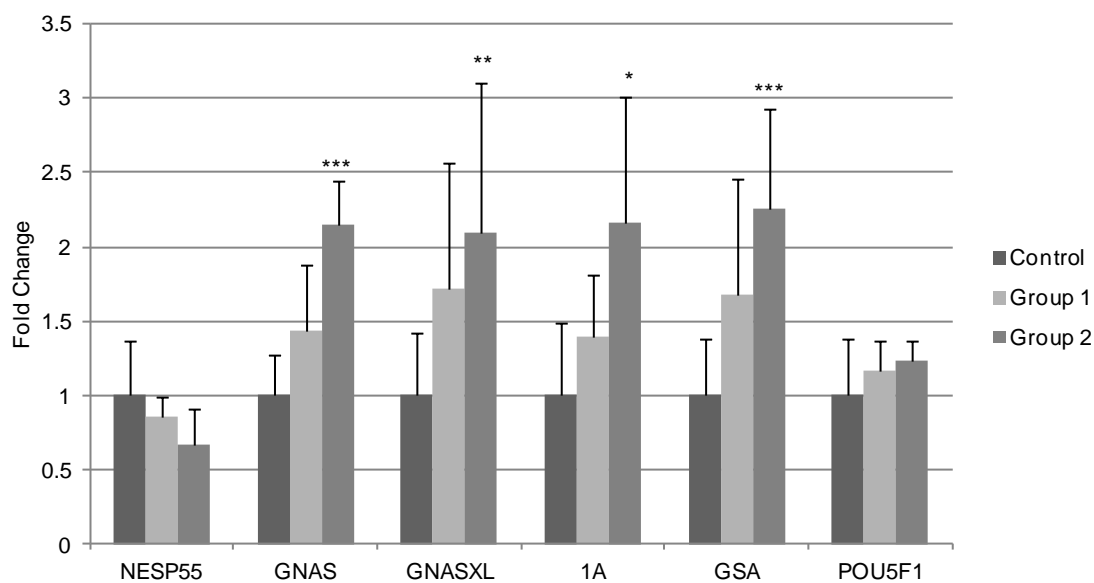


Figure 33 : Niveau d'expression des transcrits du locus *GNAS* dans les testicules d'animaux à bonne ou mauvaise qualité de semence. Groupe contrôle : animaux N1 et N5 fertiles; Groupe 1 : animaux O1 et AT2 à mauvaise qualité de semence; Groupe 2 : animaux AT3 et OAT1 à mauvaise qualité de semence et variation du niveau de méthylation dans le locus *GNAS*. *t*-test * $p < 0.05$; ** $p < 0.01$ *** $p < 0.001$.

intéressant de constater que l'augmentation de méthylation dans l'îlot CpG de *GNASXL* est spécifiquement retrouvée dans la partie recouvrant le premier exon, tandis que le début de l'îlot n'est pas affecté. Enfin, les promoteurs des transcrits *IA* et *GSA* ne présentent aucune variation du niveau de méthylation dans le sperme (Figure 32B). L'analyse par MeDIP-Seq nous permet de valider le profil de méthylation de ce locus en mettant en évidence l'absence de méthylation spécifique des îlots CpG correspondants aux promoteurs des différents transcrits tandis que ceux situés en amont du promoteur apparaissent fortement méthylés (Figure 30).

Cette variation de la méthylation dans le locus *GNAS* s'associe à une variation de l'expression de certains transcrits du locus dans le testicule (Figure 33). En effet, l'analyse par PCR quantitative des transcrits *NESP55*, *GNASXL*, *IA* et *GSA* par des couples d'amorces dessinés spécifiquement dans le premier exon alternatif entre ces transcrits, auxquels a été ajouté un couple d'amorces dessinées dans les exons communs, indique que l'expression du transcrit *IA*, de *GNASXL* et de *GSA* augmente significativement dans le groupe composé des animaux AT3 et OAT1. A l'inverse, l'expression de *NESP55* montre une tendance à la diminution dans ce groupe d'animaux, bien que cette tendance ne soit pas significative. L'utilisation d'un couple d'amorces communs aux quatre transcrits du locus révèle également une augmentation de l'expression, reflétant la variation observée dans le transcrit majoritairement exprimé *GSA*.

II. Conservation du profil de méthylation spermatique entre mammifères

1. Un ADN spermatique fortement méthylé à l'exception des régions promotrices

Notre étude de la méthylation de l'ADN spermatique porcin a révélé dans cette espèce le fort niveau de méthylation des gamètes mâles. Ces résultats sont comparables à ceux obtenus chez l'homme et les primates, ainsi que chez la souris. En effet, le sperme de ces espèces est généralement décrit comme fortement méthylé, avec un niveau moyen de méthylation des CpGs de 70% chez l'homme (Molaro et al., 2011) et de 89.4% chez la souris (Kobayashi et al., 2012). De plus, la faible méthylation des îlots CpGs et des promoteurs (par exemple ceux des gènes *DAZL*, *DDX4* et *MEST*) décrite à la fois par le MeDIP-Seq et par l'analyse locale de régions en MeDIP-qPCR est également conservée entre ces espèces (Molaro et al., 2011; Kobayashi et al., 2012), à l'exception du promoteur de *POU5F1* qui apparaît fortement méthylé dans le sperme humain. Dans les cellules germinales souches adultes, l'hypométhylation spécifique des promoteurs concerne en particulier les gènes de la lignée germinale, actifs dans ces cellules, ainsi que ceux régulant le développement embryonnaire précoce, en particulier des facteurs de transcription. Sur ces gènes, l'absence de méthylation est associée à des marques bivalentes sur la chromatine (H3K4me3 et H3K27me3) dont le rôle serait de les préparer à leur

réactivation rapide après la fécondation (Hammoud et al., 2014). Le niveau de transcription étant très faible dans les spermatozoïdes, suite au remplacement des histones par des protamines leur conférant un état chromatinien très condensé et répressif (Grootegoed et al., 2000), l'absence de méthylation dans ces promoteurs dans le sperme est en accord avec l'hypothèse de la préparation à l'activation rapide. La méthylation spécifique du promoteur de *POU5F1* dans le sperme humain pourrait alors être un indicateur d'un rôle différent dans les premières étapes suivant la fécondation, bien que son expression dans le blastocyste puis les cellules germinales soit conservé (Hansis et al., 2000). Nos résultats indiquent que les gènes du développement embryonnaire précoce sont particulièrement fortement méthylés dans le sperme, correspondant à un profil promoteur déméthylé / corps de gène méthylé souvent caractéristique de l'activation transcriptionnelle.

Enfin, les analyses préliminaires sur l'hydroxyméthylation de l'ADN spermatique porcin indiquent que cette marque épigénétique est plus faiblement présente dans le sperme que la méthylation. Les pics d'hydroxyméthylation sont localisés en majorité dans les mêmes régions que les pics de méthylation, les îlots CpGs et les 5'UTR étant les régions contenant le moins ces marques épigénétiques. Ces résultats corroborent ceux de Wang et collègues (Wang et al., 2014). En raison de l'utilisation d'un pool de cellules, ainsi d'une technique d'enrichissement n'étant pas résolutive à la base près, nous n'avons pas pu déterminer si ces deux marques coexistent sur l'ADN spermatique porcin ou s'il s'agit d'un mosaïcisme au sein de la population cellulaire analysée, avec une fraction des cellules portant des marques de méthylation et une autre fraction portant des marques d'hydroxyméthylation. Il est toutefois intéressant de noter que la distribution des 5hmCs est asymétrique entre les deux brins d'ADN, se situant préférentiellement sur le brin le plus riche en guanine, comme cela a déjà été observé dans les cellules ES murines (Ficz et al., 2011; Yu et al., 2012) ou sur le génome paternel (Wang et al., 2014).

2. Profils de méthylation spécifiques des gènes soumis à empreinte

Nous nous sommes ensuite intéressés plus particulièrement au profil de méthylation des gènes soumis à empreinte parentale dans le sperme. En effet, l'étude de l'ADN spermatique permet l'identification des régions différenciellement exprimées entre allèles maternel et paternel dans le génome des cellules somatiques. C'est en particulier le cas pour les régions méthylées sur l'allèle maternel, qui apparaissent comme des îlots d'hypométhylation dans l'ADN spermatique (Tomizawa et al., 2011). A l'inverse, les régions méthylées sur l'allèle paternel le sont également dans le sperme et leur étude nécessite alors une localisation précise du DMR. Il serait donc intéressant de comparer les résultats obtenus dans le sperme avec une étude de la méthylation de l'ovocyte, afin de confirmer la présence des DMRs.

a. Evaluation du statut d’empreinte de 5 gènes chez le porc

En se basant sur les DMRs gamétiques décrits chez l’homme et la souris et en y ajoutant des îlots CpGs, souvent caractéristiques des DMRs, nous avons évalué le niveau de méthylation de 42 régions génomiques correspondant à 29 gènes. Parmi les gènes décrits comme soumis à empreinte chez l’homme et/ou la souris, le statut d’empreinte chez le porc n’était pas toujours connu. Nous avons donc utilisé des croisements réciproques d’animaux Large White et Meishan pour étudier leur statut d’empreinte dans des tissus fœtaux. L’existence de SNPs entre les deux races sur les gènes *CNTN3*, *QPCT*, *WT1*, *IMPACT*, *GRB10* et *RASGRF1* nous a permis de caractériser leur statut d’empreinte.

Les cinq premiers se sont avérés exprimés de façon biallélique dans les tissus testés. Le statut d’empreinte de *QPCT*, *CNTN3* et *WT1* reste controversé, les données portant uniquement sur le placenta murin, et une contamination maternelle ne pouvant être exclue (Proudhon and Bourc’his, 2010). L’expression paternelle d’*IMPACT* chez la souris et le rat n’est pas conservée chez l’homme (Okamura et al., 2000), et des données sur le cerveau porcin suggèrent également une expression biallélique dans cette espèce (Okamura et al., 2005), confirmant nos résultats dans le cerveau, le foie, le poumon et le muscle fœtal. Le statut d’empreinte de *GRB10* est par contre bien décrit chez l’homme, le mouton et la souris, où il est exprimé par l’allèle maternel dans la majorité des tissus, en particulier le muscle, et par l’allèle paternel dans le cerveau (Miyoshi et al., 1998; Arnaud et al., 2003). Il semble donc que ce statut ne soit pas conservé chez le porc, l’expression étant biallélique dans le muscle, les poumons et le cerveau. Néanmoins, la possible présence de plusieurs isoformes ayant des statuts d’empreinte différents n’a pas été prise en compte dans notre analyse (Arnaud et al., 2003). Enfin, l’analyse de la séquence du transcrit du gène *RASGRF1* exprimé dans le poumon et le cerveau indique une expression paternelle de ce gène dans ces deux tissus chez le porc, corroborant les données murines (Pearsall et al., 1999) et les premières données chez le porc (Ding et al., 2014). Chez l’homme, l’empreinte de ce locus est encore incertaine et les données de méthylation dans le sang et le sperme suggèrent qu’il n’y a pas de DMR lié à l’origine parentale dans cette région (Pitamber et al., 2012).

b. Profil de méthylation des gènes soumis à empreinte conservé entre les espèces

La comparaison du niveau de méthylation de l’ensemble des régions par rapport à la séquence répétée LINE1, puis la quantification absolue de la méthylation des CpGs, nous ont permis de classer ces régions en trois catégories : peu méthylé, modérément méthylé et fortement méthylé. Cette classification nous a permis de comparer nos données avec celles obtenues chez l’homme et la souris. Dans la majorité des régions étudiées, le niveau de méthylation est comparable entre les trois espèces.

i. Régions faiblement méthylées

Les régions RASGRF1, MEST, WT1, OSBPL1A, QPCT et HM13 apparaissent comme faiblement méthylées dans les trois espèces, ainsi que les DMRs de *GNASXL*, *IGF2R* et *MEG3*. *MEST* étant exprimé par l'allèle paternel, il est conforme de retrouver sa zone promotrice déméthylée dans le sperme. Les résultats corroborent également la déméthylation paternelle des DMRs de *GNASXL* et *IGF2R*. De la même façon, le DMR de *MEG3* apparaît faiblement méthylé dans les trois espèces. Pourtant, l'étude du cluster soumis à empreinte *DLK1/MEG3* indique que ce promoteur est méthylé sur l'allèle paternel pour une expression sur l'allèle maternel du gène. De plus, le niveau de méthylation proche de 50% dans le sang au niveau de ce DMR reflète bien son statut de gène soumis à empreinte (cf Figure 29C). Il est donc envisageable que l'empreinte ne soit pas conservée dans le sperme et s'établisse tardivement au cours du développement embryonnaire. C'est également le cas pour le promoteur d'*OSBPL1A*, décrit comme maternellement exprimé chez le porc et l'homme, mais faiblement méthylé dans les gamètes mâles des trois espèces. Cependant, le système de régulation de ce gène n'a pas encore été décrit, il est possible que le promoteur ne soit pas la région la plus informative pour le caractériser. La très faible méthylation du promoteur de *WT1* corrobore les données murines et humaines. Néanmoins, il est probable que ce résultat soit plus à associer à la fonction de promoteur de cette séquence plutôt qu'à la présence d'un DMR, les analyses d'empreinte confirmant l'expression biallélique du gène chez le porc et l'homme.

ii. Régions fortement méthylées

Plusieurs régions apparaissent comme fortement méthylé dans les trois espèces. C'est le cas entre autres du DMR de *RASGRF1*, dont l'expression paternelle est associée chez la souris à une méthylation du DMR sur l'allèle paternel (Pearsall et al., 1999; Yoon et al., 2002). Comme attendu, nos données ont confirmé la méthylation de l'ICR de *H19* sur l'allèle maternel et la méthylation sur l'allèle paternel des deux DMRs décrits pour *IGF2*, ce cluster étant largement décrit dans la littérature comme conservé entre les trois espèces. Ici encore, le niveau de méthylation d'*H19* dans le sang reflète son statut de gène soumis à empreinte. Au niveau du cluster *DLK1/MEG3*, il est intéressant de noter que l'îlot CpG nommé *DLK1/MEG3* dans notre étude pourrait correspondre chez le porc à l'IG DMR de la souris de part sa localisation génomique, bien que l'homologie de séquence n'ait pas permis de localiser ce DMR sur le génome porcin. Cet îlot apparaît comme fortement méthylé chez le porc, ainsi que d'autres îlots CpGs de la région, en conformité avec les données murines et humaines. En revanche, la forte méthylation au niveau du gène *RTL1*, issu de l'insertion d'un rétrotransposon près de ce locus, est observée chez le porc et la souris, mais la région homologue apparaît comme hypométhylée chez l'homme, suggérant une fonction spécifique chez l'homme de ce gène connu pour son rôle dans la croissance fœtale chez les eutheriens (Edwards et al., 2008). Une autre différence concerne le gène *IMPACT*, dont le niveau de méthylation est élevé chez le porc tandis qu'il l'est peu chez l'homme et la souris. Cependant, la faible densité en CpG rend cette région difficile à évaluer par

les méthodes d'analyses locales. L'analyse par MeDIP-Seq nous a permis de confirmer que chez le porc, le promoteur du gène est peu méthylé alors que le corps du gène l'est fortement (cf Figure 30).

3. Différentes approches d'analyse de la méthylation de l'ADN

Parmi les techniques existantes aujourd'hui pour étudier la méthylation de l'ADN, les principales sont basées soit sur l'enrichissement de la fraction d'ADN méthylé, soit sur la conversion de l'ADN par le bisulfite. La seconde méthode est plus résolutive que la première et ne souffre pas de biais dû à l'affinité de l'anticorps pour les régions riches en GC, c'est pourquoi nous l'avons employé, couplée au pyroséquençage, pour quantifier de manière absolue des régions spécifiques du génome, quand le MeDIP-qPCR ne nous permettait qu'une classification relative. Néanmoins, la technique du MeDIP, basée sur l'immunoprécipitation de l'ADN méthylé, couplée au séquençage haut débit permet d'établir un profil sur l'ensemble du génome dont la qualité a pu être vérifiée en plusieurs points par l'approche locale.

La recherche de pics de méthylation est un bon outil permettant de localiser rapidement les régions les plus enrichies en méthylation, mais elle ne permet pas d'établir un profil de méthylation par CpG. Pourtant, en comparant cette méthode avec l'utilisation du package MEDIPS (Chavez et al., 2013; Lienhard et al., 2014) qui est plus résolutif, nous avons constaté que les résultats obtenus par les deux méthodes se recourent fortement : la grande majorité des pics détectés par MACS1.4 se superposent avec des zones dont le score de méthylation est supérieur à 66% par l'analyse MEDIPS. De plus, nous avons également comparé nos résultats aux données de « Reduced Representation Bisulfite Sequencing » (RRBS), obtenues chez le porc par l'équipe de Michael Weber (IREBS, Strasbourg), qui permet de quantifier de manière absolue la méthylation de régions riches en CpGs sur l'ensemble du génome, et avons retrouvé les mêmes régions enrichies en méthylation.

III. Identification d'une région différentiellement méthylée dans le sperme de qualité dégradée

1. Augmentation du niveau de méthylation dans le locus GNAS

a. Description du locus complexe GNAS

La quantification du niveau de méthylation par conversion bisulfite et pyroséquençage dans le locus soumis à empreinte GNAS nous a permis de mettre en évidence une augmentation de la méthylation de l'ADN pour 3 animaux à qualité de semence dégradée sur les 8 analysés. Le locus GNAS est une région complexe situé chez l'homme sur le chromosome 20, chez la souris sur le chromosome 2, et dont l'homologue porcin se situe sur le chromosome 17. Il est composé de plusieurs variants de

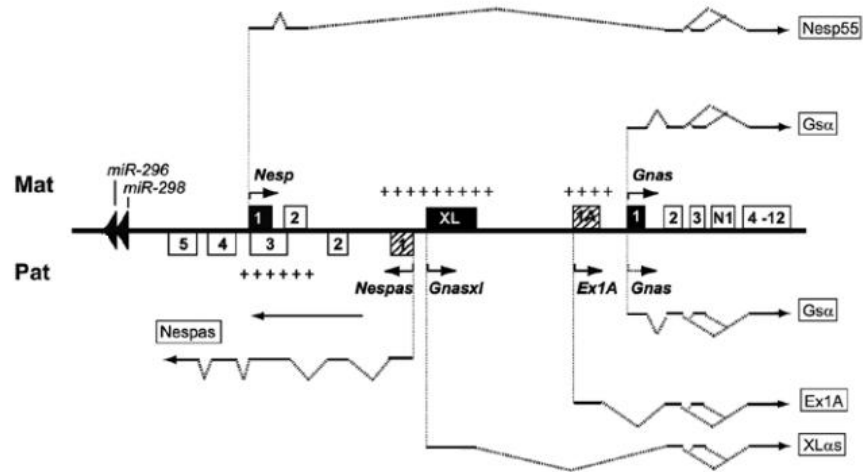
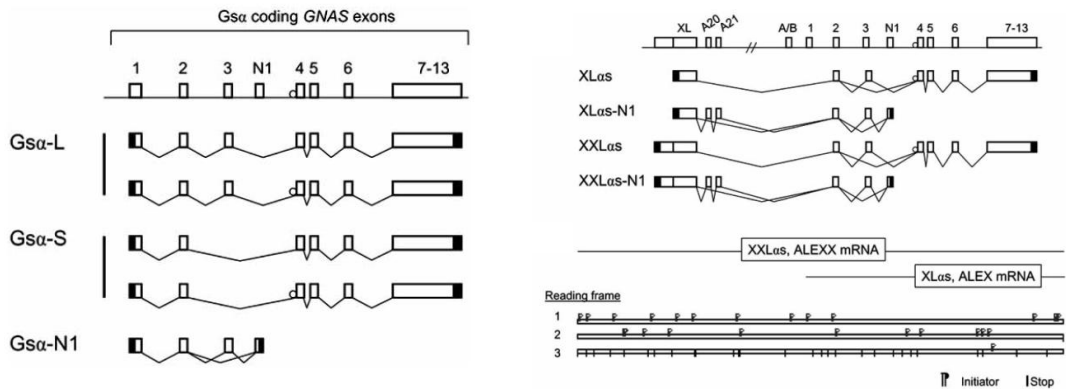
A**B**

Figure 34 : **A.** Représentation schématique du locus *GNAS* chez la souris. Les carrés noirs représentent les premiers exons des transcrits codants et les carrés hachurés les premiers exons des transcrits non codants. La méthylation différentielle entre allèle paternel et maternel est indiquée par le signe +. Adapté de (Peters and Williamson, 2007) **B.** Représentation des différents isoformes de GSA et GNASXL (Bastepe, 2007) : Parmi eux, les isoformes longs (L) et courts (S) dépendent de l'épissage alternatif de l'exon 3, et possèdent ou non l'inclusion d'une sérine supplémentaire au début de l'exon 4, l'isoforme N1 est tronqué dans sa partie C-terminale, l'isoforme XXLas est étendu du côté 5' et les isoformes ALEX et ALEXX utilisent un « Open Reading Frame » (ORF) alternatif.

transcrits qui se différencient en priorité par l'utilisation d'un promoteur et d'un premier exon alternatif tandis que les exons 2 à 13 sont communs à tous les isoformes (Figure 34A).

Le premier de ces exons alternatifs permet la production de la protéine NESP55, ou « Neuroendocrine Secretary Protein », dont l'expression maternelle est contrôlée par la méthylation sur l'allèle paternel de son DMR, situé au niveau de sa région promotrice (Kelsey et al., 1999; Peters et al., 1999). Ce DMR n'est pas gamétique, et la méthylation paternelle ne s'effectue qu'après l'implantation, expliquant la faible méthylation de l'îlot CpG recouvrant ce promoteur dans le sperme (Liu et al., 2000b).

L'utilisation du second promoteur, et donc du premier exon alternatif GNASXL, est caractéristique du gène codant *GNASXL* (ou *XLas*), pour lequel plusieurs isoformes ont été décrits chez l'homme (Figure 34B). Ces isoformes sont exprimés par l'allèle paternel grâce à la méthylation sur l'allèle maternel de la région promotrice (Bastepe, 2007).

L'utilisation d'un troisième premier exon alternatif, l'exon 1A, permet la transcription, par l'allèle paternel, d'un transcrit non codant. Le DMR gamétique de l'exon 1A a une fonction d'ICR, car il permet de réguler l'empreinte du transcrit *GSA* dans plusieurs tissus (Williamson et al., 2004; Liu et al., 2005).

En effet, le dernier des promoteurs permet la transcription d'un gène codant pour la protéine « Guanine Nucleotide-binding Alpha-Subunit » et intitulé *GSA* dans notre étude. Ce transcrit existe également en plusieurs isoformes, similaires à ceux décrits pour *GNASXL* (Figure 34B). Il est exprimé de manière biallélique dans la majorité des tissus, et par l'allèle maternel dans quelques tissus tels que le cortex rénal, la thyroïde et les ovaires (Yu et al., 1998; Bastepe, 2007) : ce procédé est régulé par la répression de l'expression sur l'allèle paternel par le transcrit non codant 1A (Williamson et al., 2004). Enfin, un transcrit antisens, *NESPAS*, qui n'a pas encore été décrit chez le porc, est connu chez l'homme et la souris pour être exprimé par l'allèle paternel. Son expression monoallélique est liée à la méthylation de son promoteur, qui est un DMR gamétique situé juste en amont du promoteur de *GNASXL* (Figure 34A). Le rôle principal de cet antisens non codant est la répression de l'expression de *NESP55* sur l'allèle paternel. Il exerce aussi un rôle de régulation de l'expression de *GNASXL* dont l'expression diminue drastiquement lorsque le DMR de *NESPAS* est supprimé, ainsi que sur les transcrits *1A* et *GSA* (Williamson et al., 2006).

Chez le porc, la grande quantité d'isoformes présents dans ce locus complexe a rendu l'étude de l'empreinte parentale par l'utilisation d'un modèle parthénogénétique très difficile (Bischoff et al., 2009). De plus, nous n'avons pas trouvé de SNPs dans les premiers exons *GNASXL* et *1A* qui auraient permis de déterminer l'origine parentale de l'allèle exprimé dans des fœtus issus de croisements LWxMS. Néanmoins, l'empreinte maternelle de *NESP55* dans le cerveau, le muscle, le foie et le rein chez le porc a déjà été confirmée par d'autres auteurs (Oczkowicz et al., 2012). Cette conservation de l'empreinte de *NESP55* ainsi que la structure très proche du locus *GNAS* entre porc, souris et homme nous suggère que son mode de régulation pourrait être conservé dans l'espèce porcine.

b. Une région liée à des maladies du développement et à la croissance

Sur le plan fonctionnel, le locus GNAS encode donc la sous-unité alpha des protéines G, un élément clé des voies de transduction du signal et notamment de la voie des AMP cycliques dont le rôle est primordial dans l'action des hormones et des neurotransmetteurs : chez la souris, un « knock-out » du gène est léthal (Yu et al., 1998). Chez l'homme, les mutations hétérozygotes sont liées à de nombreuses maladies. Des mutations de ce gène sont ainsi retrouvées dans les adénomes endocriniens (Lyons et al., 1990). Les mêmes mutations ont été retrouvées chez des patients souffrant du syndrome de McCune-Albright (MAS), dont les symptômes sont une sexualité précoce, une dysplasie fibreuse des os et une hyperpigmentation de la peau, ainsi que chez les patients atteints de pseudohypoparathyroïdisme de type Ia (PHPT-Ia), ou de pseudopseudohypoparathyroïdisme (PPHPT). Ces deux maladies sont caractérisées par la résistance aux hormones y compris l'hormone parathyroïde (PTH) dont la voie de signalisation dépend du bon fonctionnement de GNAS. Enfin, l'hétéroplasie osseuse progressive (POH) associée à une inactivation de *GSA* est transmise de façon héréditaire par la voie paternelle, suggérant un rôle de l'empreinte paternelle du locus GNAS dans cette maladie, en particulier des transcrits paternellement exprimés tels que *GNASXL*. En effet, *GNASXL* partageant une grande identité avec *GSA*, la protéine résultante est capable *in vitro* d'agir sur les mêmes fonctions biologiques et sa dérégulation pourrait donc être à l'origine de dysfonctionnement des voies de signalisation (Bastepe, 2007). La régulation de l'empreinte du locus GNAS est également associée à la croissance et au développement (Plagge et al., 2004; Ball et al., 2013; Eaton et al., 2013; Richard et al., 2013). Ainsi, la transmission paternelle d'une mutation de l'exon 1A provoque un retard de croissance après la naissance via la perte de l'empreinte de *GSA* tandis que la transmission maternelle de cette même mutation entraîne une augmentation de la masse corporelle via la diminution de l'expression de *NESP55* (Eaton et al., 2013). Les mutations transmises via l'allèle paternel sont également liées à des retards de croissance intrautérins et indiquent un rôle important de *GNASXL* dans le développement fœtal (Richard et al., 2013). Chez le porc, le locus GNAS est situé dans un locus à caractères quantitatifs (QTLs) avec un effet de l'origine parentale influençant la masse corporelle et à la croissance des porcelets (Thomsen et al., 2004).

c. Un profil de méthylation modifié chez certains animaux à spermogramme anormal

Notre étude a révélé une augmentation de la méthylation dans l'îlot CpG couvrant le promoteur et le premier exon de *NESP55*, ainsi que dans l'îlot CpG couvrant le promoteur et le premier exon de *GNASXL*, ici plus spécifiquement dans la partie exonique. Cette variation est spécifique du sperme, le niveau de méthylation dans le sang étant constant. Les animaux concernés présentent des défauts sévères de qualité de sperme : tératospermie (animal T1), asthénotératospermie (AT3) et

oligotératospermie (OAT1). Deux d'entre eux sont porteurs de remaniements chromosomiques, dont l'un sur le chromosome 17 portant le locus *GNAS* chez le porc. Néanmoins, les remaniements chromosomiques ne semblent pas être liés à la variation du niveau de méthylation, l'animal contrôle porteur d'une translocation n'étant pas affecté et l'animal T1 n'étant pas porteur de remaniement chromosomique. Toutefois, la présence d'un remaniement chromosomique pouvant être l'indicateur d'une qualité de semence particulièrement dégradée, il serait donc intéressant d'évaluer sur un plus grand nombre d'animaux si ces variations de méthylation sont présentes dans les cas les plus sévères, comme cela a déjà été montré au niveau du gène soumis à empreinte *MEST* chez l'homme (Marques et al., 2008). De plus, le locus *GNAS* n'ayant pas été reporté comme variant dans les cas d'infertilité mâle humaine, il serait intéressant de comparer nos données avec l'analyse de la méthylation de ce locus dans le sperme anormal humain. En effet, une étude préliminaire chez le bovin indique une augmentation significative de la méthylation de l'ADN dans le locus *GNAS* d'animaux hypofertiles (Eli Sellem, communication personnelle), espèce dans laquelle des polymorphismes de ce locus sont associés à des performances de croissance et de développement, ainsi que de fertilité (Sikora et al., 2011). L'analyse de l'expression des 4 transcrits (*NESP55*, *GNASXL*, *IA* et *GSA*) du locus *GNAS* chez le porc dans le testicule d'animaux contrôles ou à qualité de semence dégradée révèle que les variations de méthylation sont associées à une augmentation significative de l'expression de *GNASXL*, *IA* et *GSA* tandis que l'expression de *NESP55* montre une tendance à la diminution. Chez l'homme, la perte de l'expression de *NESP55* résulte de l'augmentation de la méthylation de son promoteur, mais ce défaut n'apparaît lié à aucun phénotype particulier (Liu et al., 2000a). Chez la souris, la délétion du DMR de *NESP55* entraîne une augmentation de l'expression de *GNASXL* et *IA* (Fröhlich et al., 2010). Ainsi, nos résultats semblent indiquer un nouveau rôle pour le locus *GNAS* dans l'établissement de gamètes fonctionnels chez le porc, bien que les mécanismes soient encore à déterminer. Il serait en particulier intéressant d'étudier le profil d'expression de ces transcrits directement dans les spermatozoïdes puis de mener la même étude (méthylation puis expression) sur des méiocytes afin de déterminer la dynamique d'établissement du profil de méthylation anormal dans les gamètes porcins.

2. Stabilité du niveau global de méthylation

Outre la variation dans le locus complexe *GNAS*, nous n'avons retrouvé aucune variation significative du niveau de méthylation chez les animaux à qualité de semence dégradée en comparaison avec le groupe d'animaux fertiles. C'est le cas à la fois au niveau global, où le niveau de méthylation du génome reste stable, ainsi qu'au niveau local, où le niveau de méthylation de plusieurs gènes soumis à empreinte ou de gènes de la lignée germinale et du développement embryonnaire est constant. L'absence de variation globale est en accord avec des résultats obtenus chez l'homme. En effet, le niveau de méthylation de la séquence répétée *LINE1* ne varie pas (Marques et al., 2008), et bien que

de nombreux défauts aient été observés à l'échelle du génome, ils concernent principalement des régions soumises à empreinte (Houshdaran et al., 2007) ou un sous-ensemble des CpGs présents dans le génome (environ 30%) (Pacheco et al., 2011). Il est cependant surprenant de ne retrouver aucune variation dans les loci soumis à empreinte *IGF2/H19* ou *MEST* qui ont été abondamment décrits dans la littérature comme corrélés à la qualité de la semence (Marques et al., 2004; Houshdaran et al., 2007; Kobayashi et al., 2007; Marques et al., 2008; Boissonnas et al., 2010; Hammoud et al., 2010; Poplinski et al., 2010; Sato et al., 2011; Li et al., 2013). Ces résultats pourraient être enrichis par l'étude d'un panel plus large d'animaux présentant des spermogrammes plus variés et de gravité plus sévère. Enfin, la comparaison du méthylome des animaux fertiles et infertiles par MeDIP-Seq ou RRBS permettrait d'identifier des DMRs entre les deux phénotypes sur l'ensemble du génome, sans se restreindre à quelques gènes seulement.

**DERIVATION DE LIGNEES iPS
PORCINES A PARTIR
D'ANIMAUX INFERTILES
PORTEURS DE
REMANIEMENTS
CHROMOSOMIQUES**

Tableau 22 : Bilan des expériences de reprogrammation cellulaire par système intégratif (bleu) ou non intégratif (orange).

Expérience	Méthode de reprogrammation	Stoechiométrie des facteurs	Sérum	Cellules nourricières	Cytokine	Animal	Caryotype	Tissu d'origine	EOS	Nombre de clones analysés
V1	Rétrovirus pMX OSKM	1 : 1 : 1 : 1	SVF	iSTO	bFGF+LIF	N1	2n=38, XY	Peau	oui	32
					bFGF+LIF	N1	2n=38, XY	Peau	oui	26
					bFGF+LIF	N1	2n=38, XY	Peau	non	5
V2	Rétrovirus pMX OSKM	1 : 1 : 1 : 1	SVF	iSTO	LIF	MEF	2n=40	Lignée murine	non	5
					bFGF	N1	2n=38, XY	Testicule	non	0
					bFGF	AT3	2n=28, XY, t(13;17)	Peau	oui	7
					LIF	AT3	2n=28, XY, t(13;17)	Peau	oui	4
V3	Rétrovirus pMX OSKM	1 : 1 : 1 : 1	SVF+KOSR	Gélatine 0,1% + iSTO ou Matrigel + Gélatine 0,1%	LIF	MEF	2n=40	Lignée murine	non	2
					bFGF	AT3	2n=28, XY, t(13;17)	Peau	oui	0
					bFGF	Y14	2n=38, XY, t(Y;14)	Testicule	oui	19
					bFGF	XXY	38 XY ; 39 XXY	Testicule	oui	24
V4	Rétrovirus pMX OSKM	1 : 1 : 3:0	SVF+KOSR	Matrigel + Gélatine 0,1%	bFGF ou LIF	N1	2n=38, XY	Testicule	oui	0
					bFGF ou LIF	AT3	2n=28, XY, t(13;17)	Peau	oui	0
					bFGF ou LIF	Y14	2n=38, XY, t(Y;14)	Testicule	oui	0
					bFGF ou LIF	XXY	38 XY ; 39 XXY	Testicule	oui	0
V5	Rétrovirus pMX OSKM	1 : 1 : 1 : 1	KOSR	Gélatine 0,1% + iSTO	LIF	MEF	2n=40	Lignée murine	oui	0
					bFGF + VPA	N2	2n=38, XY	Testicule	oui	8
V6	Rétrovirus pMX OSKM	1 : 1 : 1 : 1	SVF+KOSR	Gélatine 0,1% + iSTO	bFGF + VPA	OAT1	2n=38, XY, t(1;14)	Testicule	oui	6
					bFGF	N2	2n=38, XY	Testicule	oui	5
V7	SeV OSKM	1 : 1 : 1 : 1	SVF+KOSR	Gélatine 0,1% + iSTO	bFGF	N1	2n=38, XY	Peau	oui	8
					bFGF	PEF	2n=38, XY	Lignée porcine	oui	24
					bFGF	AT3	2n=28, XY, t(13;17)	Peau	non	5
					bFGF	Y14	2n=38, XY, t(Y;14)	Testicule	non	14
					bFGF	PEF	2n=38, XY	Lignée porcine	non	0

Ce chapitre est découpé en plusieurs parties. La première expose les différentes expérimentations de reprogrammation cellulaire réalisées au cours de ce projet de thèse. Les paragraphes II à V présentent les résultats de l'analyse de clones iPS-like obtenus par deux systèmes de reprogrammation différents. Enfin, ces résultats sont discutés et les perspectives envisagées sont énoncées dans la sixième partie.

I. Bilan des différentes expériences de reprogrammation cellulaire

Afin de dériver des lignées piPS chez des animaux contrôles ou à qualité de semence dégradée et porteurs de remaniements chromosomiques, nous avons sélectionné différentes populations cellulaires initiales. Pour les contrôles, des fibroblastes issus de testicule ou de peau de deux animaux contrôles adulte (N1 et N2) ou des fibroblastes embryonnaires (PEF) d'un animal Meishan à J52 de développement ont été utilisés. Pour les animaux à qualité de semence dégradée, des fibroblastes de tissu testiculaire adulte ont été dérivés à partir d'animaux identifiés par l'activité de contrôle chromosomique de l'équipe Cytogene : un verrat azoospermique transloqué t(Y;14), un verrat AT3 transloqué t(13;17) , un verrat OAT1 transloqué (1;14) et enfin le porc porteur d'une trisomie partielle des chromosomes sexuels (38 XY ; 39 XXY) correspondant au syndrome de Klinefelter humain, oligo-tératospermique. En parallèle de plusieurs expériences de reprogrammation, nous avons dérivé à partir de fibroblastes embryonnaires murins (MEF) des lignées iPS murines comme contrôle.

Afin d'améliorer le protocole de production de lignée iPS porcines, nous avons fait varier de nombreux paramètres à la fois au niveau de la technique de reprogrammation employée (rétrovirus, système non intégratif) mais aussi au niveau des conditions de culture (utilisation de cellules nourricières, milieu de culture). L'ensemble des expériences menées est regroupé dans le Tableau 22. En se basant sur les premières lignées de cellules piPS décrites dans la littérature, notre milieu de culture initial était composé à la fois de bFGF et de LIF, et le sérum employé était le sérum de veau fœtal (SVF). En observant les caractéristiques des lignées produites en combinant bFGF, LIF et VPA, nous avons finalement sélectionné, pour la culture des lignées piPS-like obtenues, un milieu enrichi en bFGF uniquement (6 ng/ μ L) en présence d'un mélange de sérum SVF et KOSR (Knock Out Serum Replacement) (milieu piPS, voir Matériel et Méthodes). Pour la majorité des clones, nous avons également observé une faible adhérence des cellules au support de culture, que ce soit en présence de cellules nourricières (constituées de cellules murines STO irradiées (iSTO)) ou du Matrigel uniquement. Nous avons finalement décidé de conserver le système avec les cellules nourricières en pré-traitant les boîtes de cultures par de la gélatine 0,1%. Avec ce traitement, certaines lignées peuvent pousser indifféremment sur cellules nourricières ou en présence de Matrigel. Enfin, une variation de la stœchiométrie des facteurs de reprogrammation (expérience V4) a été expérimentée, en triplant la dose de *KLF4* humain et en supprimant totalement *c-MYC*, dont l'action n'est pas strictement nécessaire à la reprogrammation cellulaire. Dans ces conditions, nous n'avons pas pu obtenir de clones présentant

une morphologie similaire à celle des cellules ES après l'injection du cocktail de reprogrammation, que ce soit dans le cas de cellules porcines ou murines.

Au cours de la reprogrammation cellulaire, les clones apparaissant au milieu des fibroblastes ont été sélectionnés sur la base de leur ressemblance avec les colonies de cellules ES murines et humaines : de petites cellules poussant en colonies serrées, en dôme ou à plat. Dans le cas de fibroblastes préalablement infectés par le vecteur EOS (voir Matériel et Méthodes), l'expression de la GFP visualisée au microscope à fluorescence ainsi que l'ajout de puromycine dans le milieu de culture ont permis d'améliorer la sélection des clones en reprogrammation. Une fois ces clones repiqués, des analyses préliminaires ont permis de ne conserver que ceux présentant des caractéristiques de pluripotence. La première étape-clé est l'évaluation visuelle des clones : morphologie, homogénéité de la population cellulaire, et surtout conservation de la morphologie au cours des passages, un changement dans la morphologie étant souvent un bon indicateur de l'entrée en différenciation des cellules. Ensuite, pour chaque clone, l'expression de la phosphatase alcaline a été évaluée ainsi que l'expression de quatre marqueurs de pluripotence porcins, *NANOG*, *LIN28B*, *OCT4* et *SOX2* (par PCR conventionnelle), tout comme celle des quatre facteurs de reprogrammation humains. Ces derniers sont en effet censés s'éteindre une fois les cellules totalement reprogrammées. Pourtant, aucune des lignées dérivées par le système de rétrovirus intégratif pMX n'a éteint ces exogènes et ce malgré un grand nombre de passages. Le même problème ayant été observé par d'autres équipes chez le porc (Esteban et al., 2010a; West et al., 2010; Hall et al., 2012), ce critère a été mis de côté pour l'évaluation des lignées porcines. Sur le plan expressionnel, l'immunocytochimie a été utilisée pour évaluer l'expression de *NANOG*, *LIN28B*, *SOX2*, *OCT4*, *SSEA4* et *CDH1*. Parmi les lignées exprimant les gènes de pluripotence, certaines se sont avérées incapables de former des corps embryoides lorsqu'elles ont été placées en suspension dans un milieu ne contenant pas de bFGF (milieu DIFF). La grande majorité des lignées placées en milieu LIF+2i n'ont pas supporté ce changement, entamant soit une phase de différenciation cellulaire soit une phase de mort cellulaire. Ce résultat indique que les lignées ne seraient pas suffisamment reprogrammées pour pouvoir évoluer vers un état naïf de pluripotence tel que décrit chez la souris. Enfin, l'analyse caryotypique a révélé de nombreuses anomalies dans des lignées à passage précoce (10-20), suggérant une forte instabilité chromosomique liée à la reprogrammation cellulaire. Les principaux résultats pour chaque clone analysé sont regroupés dans l'Annexe 3.

En se basant sur tous ces critères, seules quelques lignées ont été conservées pour une analyse plus poussée de la pluripotence porcine. En effet, une très grande majorité des clones obtenus s'est révélée très instable dans le temps, différenciant au bout de quelques passages, ou ne possédant pas suffisamment de caractères de pluripotence. C'est par exemple le cas pour l'ensemble des lignées analysées issues d'animaux normaux pour lesquelles, malgré de nombreuses tentatives et l'expression

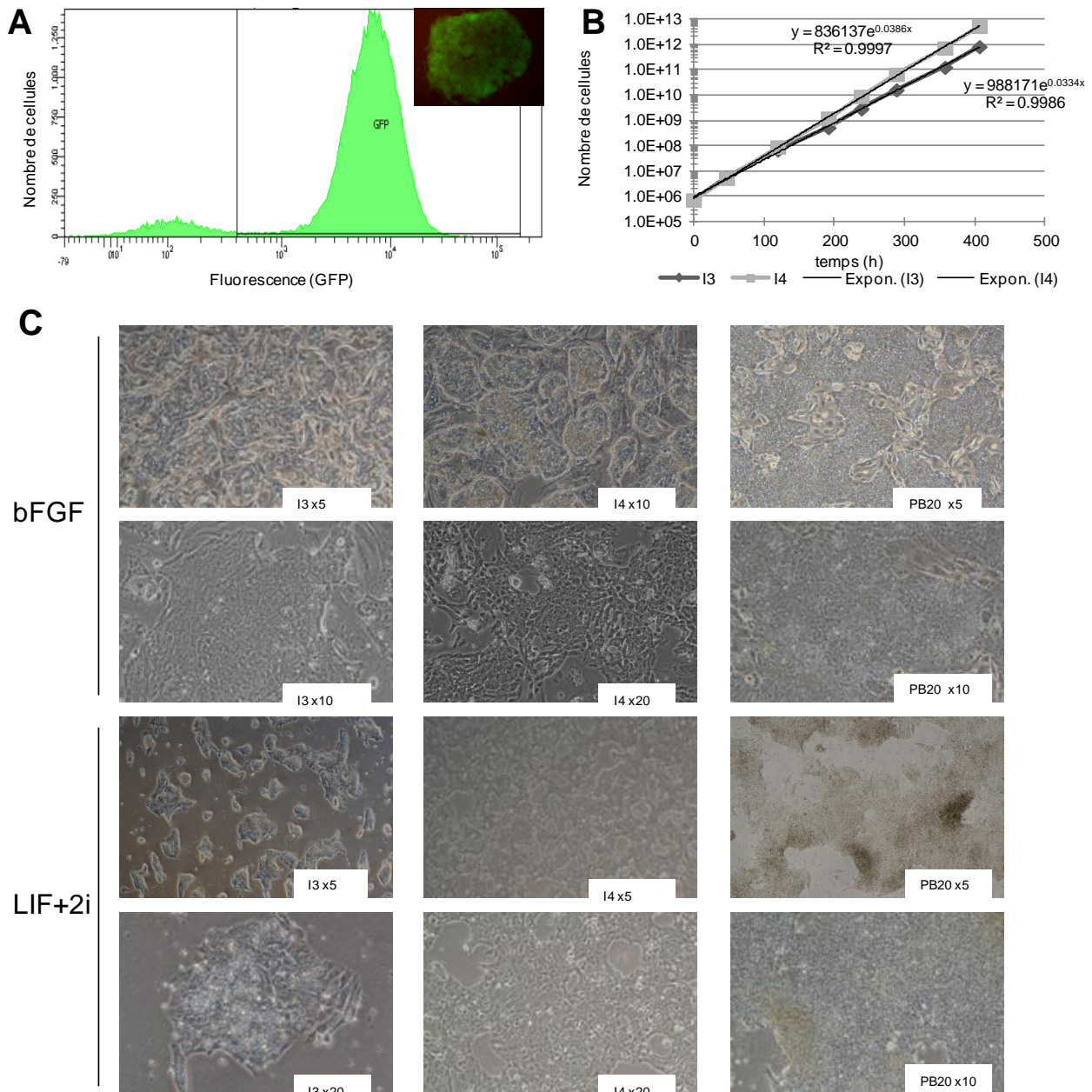


Figure 35 : *A. Le profil obtenu en cytométrie de flux (FACS) et l'observation au microscope à fluorescence confirment l'expression de la GFP dans les lignées I3 et I4. B. Les courbes de croissance exponentielles des lignées I3 et I4 reflètent leur forte prolifération cellulaire. C. Morphologie iPS-like des lignées I3, I4 et PB20 en milieu piPS et en milieu LIF+2i.*

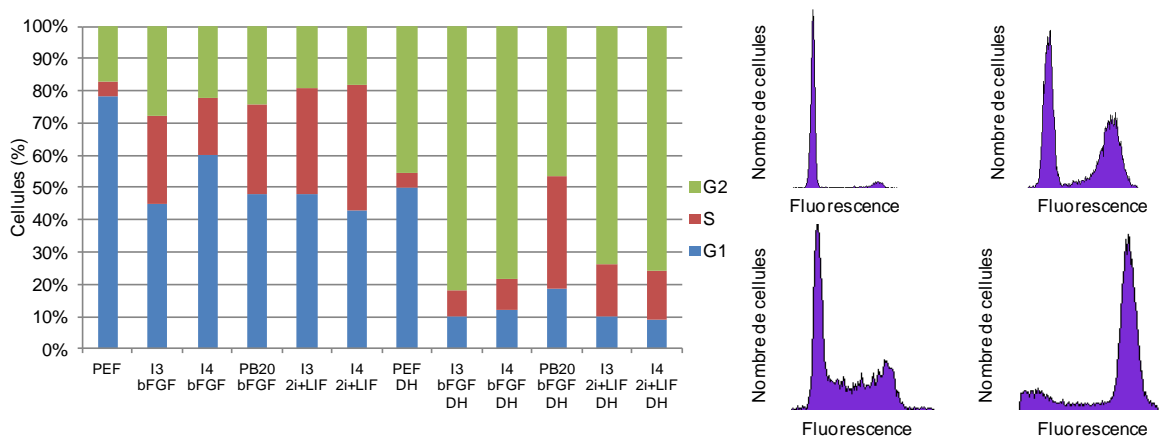


Figure 36 : *Analyse du cycle cellulaire par FACS. : faible durée de la phase G1 et absence de point de contrôle G1/S dans les lignées reprogrammées par rapport aux PEFs quelque soit le milieu de culture. Profils de FACS : 1^{ère} ligne lignée PEF et PEF traitée, 2^{ème} ligne lignée I3 et I3 traitée.*

de quelques marqueurs de pluripotence, nous n'avons pas obtenu de population cellulaire répondant à suffisamment de critères. Pour remédier à ce problème, une collaboration avec l'équipe du Pr Ali Thuran (U935, Villejuif) nous a permis d'accéder à une lignée cellulaire porcine contrôle, nommée PB20, que nous avons utilisé pour la suite des expérimentations. Cette lignée a été dérivée à partir d'amniocytes primaires par un système de lentivirus intégratifs avec 6 facteurs (NLOS KM) en présence de bFGF et se maintient en culture dans notre milieu piPS. Chez les animaux infertiles porteurs de remaniements chromosomiques, nous avons décidé de nous concentrer sur les lignées dérivées de fibroblastes porteurs de la translocation réciproque t(Y;14) qui présentent un plus grand nombre de critères de pluripotence, et qui ont fait l'objet d'une reprogrammation à la fois par le système rétroviral intégratif mais aussi par un système non intégratif basé sur l'utilisation de virus Sendai, nous permettant ainsi de comparer les résultats obtenus avec les deux techniques. Dans la suite de ce manuscrit, nous nous concentrerons donc sur les résultats obtenus pour la lignée contrôle PB20, les lignées I3 et I4 dérivées de l'animal transloqué réciproque t(Y;14) par rétrovirus dans le milieu piPS, et les lignées NI12, NI13 et NI20 dérivées des mêmes fibroblastes par le système non intégratif.

II. Caractéristiques de pluripotence des clones transloqués I3 et I4 et de la lignée contrôle PB20

1. Morphologie et cycle cellulaire ES-like

Les lignées I3 et I4 sont caractérisées tout d'abord par l'expression de la protéine fluorescente GFP, indiquant que le processus de reprogrammation s'est bien déroulé (Figure 35A). Ces deux lignées présentent une morphologie ressemblant à celles des cellules ES, c'est-à-dire que les cellules sont de petite taille et poussent en colonies très denses où les frontières entre cellules sont peu visibles (Figure 35C). Cette morphologie et leur propension à pousser en dôme en font des cellules assez peu adhérentes sur la surface de la boîte de culture. Néanmoins, en utilisant un pré-traitement à la gélatine 0.1%, ces cellules se maintiennent en culture à la fois sur une couche de cellules nourricières ou sur du Matrigel. Le temps de doublement, calculé sur 7 passages, est de 20h45mn pour la lignée I3 et de 18h pour la lignée I4 (Figure 35B). Plus généralement, ces lignées nécessitent une dilution au huitième tous les trois jours avant d'atteindre la confluence. Les cellules PB20 quand à elles présentent un cycle cellulaire encore plus rapide, une dilution au dixième tous les trois jours étant nécessaire, et possèdent une morphologie caractéristique des lignées pluripotentes (Figure 35C).

L'étude du cycle cellulaire par marquage de l'ADN à l'iodure de propidium montre que ces trois lignées possèdent un cycle cellulaire similaire à celui des mESCs avec une phase G1 beaucoup plus courte que celle observée dans les fibroblastes embryonnaires : à un instant donné, près de 80% des PEFs sont dans la phase G1 du cycle tandis que 45% seulement des cellules I3 sont en G1, 60% pour

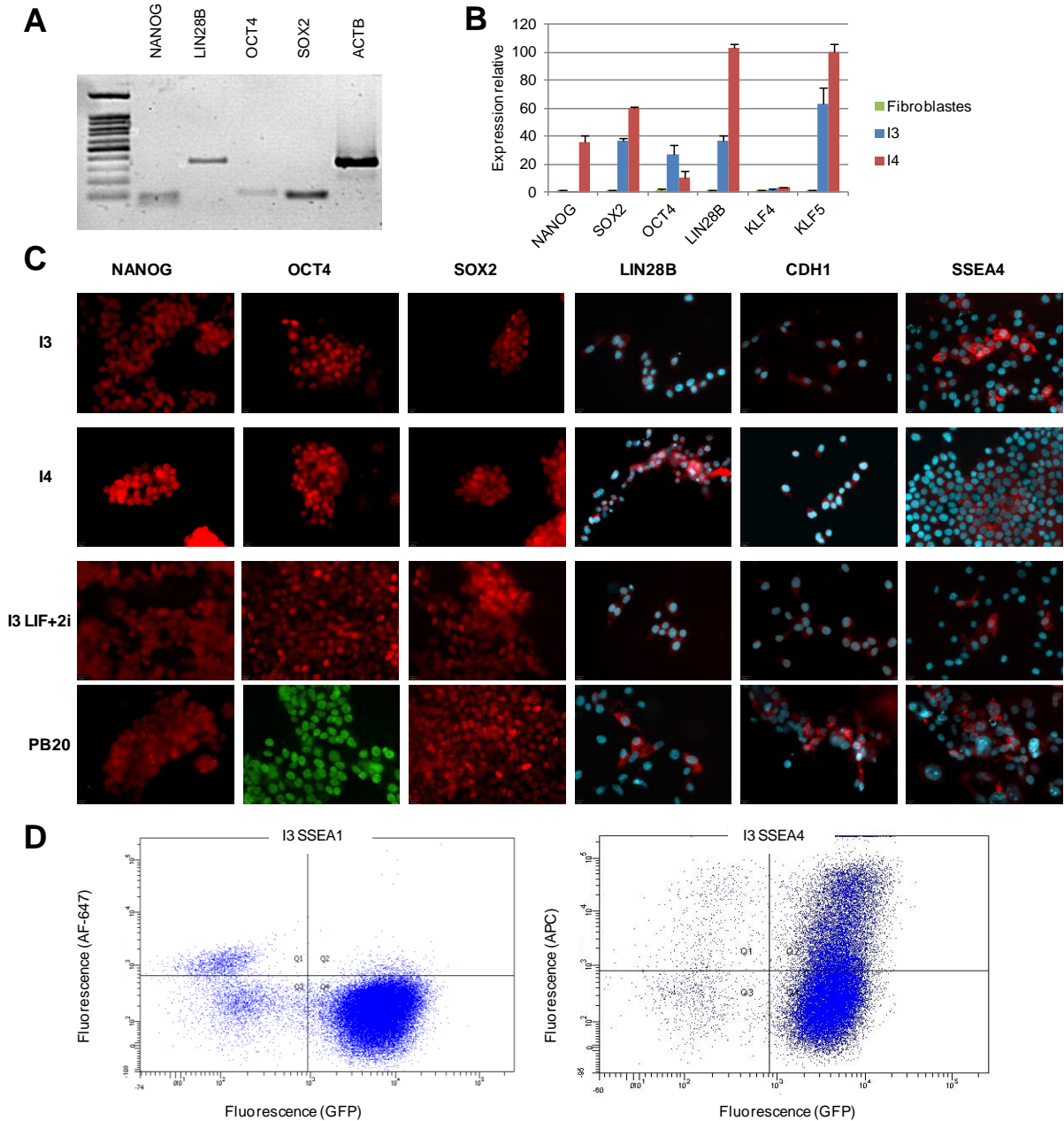


Figure 37 : Expression des facteurs-clés de pluripotence dans les lignées I3 et I4. **A.** PCR conventionnelle et gel d'agarose indiquant l'expression des endogènes de NANOG, LIN28B, OCT4 et SOX2 ainsi que du contrôle ACTB dans la lignée I3. **B.** Niveau d'expression relatif des endogènes NANOG, SOX2, OCT4, LIN28B, KLF4 et KLF5 dans les lignées I3 et I4 comparé à la lignée de fibroblastes avant reprogrammation cellulaire. **C.** Immunocytochimie des marqueurs nucléaires NANOG, OCT4 et SOX2 (Alexa anti-rabbit 568 ou Alexa anti-rabbit 488) et des marqueurs cytoplasmique et membranaires LIN28B, CDH1 et SSEA4 (Alexa anti-rabbit 568) autour des noyaux marqués au DAPI. **D.** Marquage en suspension des cellules par les anticorps de SSEA1 (gauche) et SSEA4 (droite) couplés à un fluorochrome rouge et analyse au FACS : les cellules sont positives pour la GFP (Q4), négatives pour SSEA1 (Q2), et une sous-population cellulaire seulement est positive pour SSEA4 (Q2).

les I4 et 48% pour les PB20 (Figure 36). De plus, le traitement par de la doxorubicine hydrochloride (DH) induisant des cassures doubles brins de l'ADN nous a permis d'évaluer la présence de points de contrôle du cycle cellulaire dans les cellules. Comme leurs équivalentes humaines et murines, les lignées porcines étudiées ici ne possèdent pas de point de contrôle entre les phases G1 et S. En effet, en présence de DH on observe chez les PEFs un double blocage des cellules en G1 et en G2 (50,2% et 45,5% des cellules respectivement) tandis que les lignées I3 et I4 sont majoritairement bloquées en G2 (82% et 78,5%) suggérant l'absence de point de contrôle stoppant les cellules entre les phases G1 et S (Figure 36). Les résultats pour la lignée PB20, plus mitigés, indiquent un plus fort blocage en phase G2 (46,6%) qu'en phase G1 (18,6%) et d'une manière surprenante un assez fort pourcentage de cellules en phase S (34,8%) suggérant qu'une partie des cellules étaient encore en train de cycler au moment de l'étude. Toutefois, on peut considérer que ces cellules ont franchi sans obstacle le point de contrôle G1/S malgré des cassures doubles brins de l'ADN, suggérant ici encore l'absence de ce point de contrôle.

Ces trois lignées supportent bien le remplacement du milieu piPS au profit du milieu LIF+2i avec lequel les cellules sont cultivées exclusivement sur du Matrigel. Bien qu'on observe des changements morphologiques (Figure 35C), elles peuvent en effet être maintenues plus de 30 jours dans ce milieu sans apparition visible de signes de différenciation. De plus, la vitesse de croissance montre une tendance à l'accélération, avec une dilution 1/10 nécessaire tous les 2 à 3 jours. Les caractéristiques du cycle cellulaire sont conservées pour les lignées I3 et I4, avec une phase G1 contenant 48% et 43% des cellules des deux lignées respectivement, et un blocage majoritaire des cellules en phase G2 après induction de cassures double brins à l'ADN caractérisant l'absence de point de contrôle G1/S au cours du cycle cellulaire (74% et 76% des cellules respectivement)(Figure 36).

2. Expression des gènes de la pluripotence

a. Activation des gènes du réseau central de pluripotence

Les lignées I3 et I4 sont toutes deux positives pour l'expression de la phosphatase alcaline, ainsi que pour celle de marqueurs endogènes faisant partie du circuit principal de la pluripotence : *NANOG*, *LIN28B*, *OCT4* et *SOX2* (Figure 37A et 37B). Il est toutefois intéressant de noter que le gène *KLF4*, faisant partie du cocktail de reprogrammation, est bien moins surexprimé dans les cellules reprogrammées comparées aux fibroblastes que les autres facteurs de reprogrammation *OCT4* et *SOX2* (Figure 37B). De plus, bien que l'analyse par PCR quantitative n'ait pas permis de détecter une augmentation de *NANOG* dans la lignée I3, contrairement à celle observé dans la lignée I4, les études suivantes ont confirmé l'augmentation de l'expression de ce gène dans les deux lignées. L'analyse par immunocytochimie a aussi confirmé l'expression nucléaire de *NANOG*, *OCT4* et *SOX2*, et

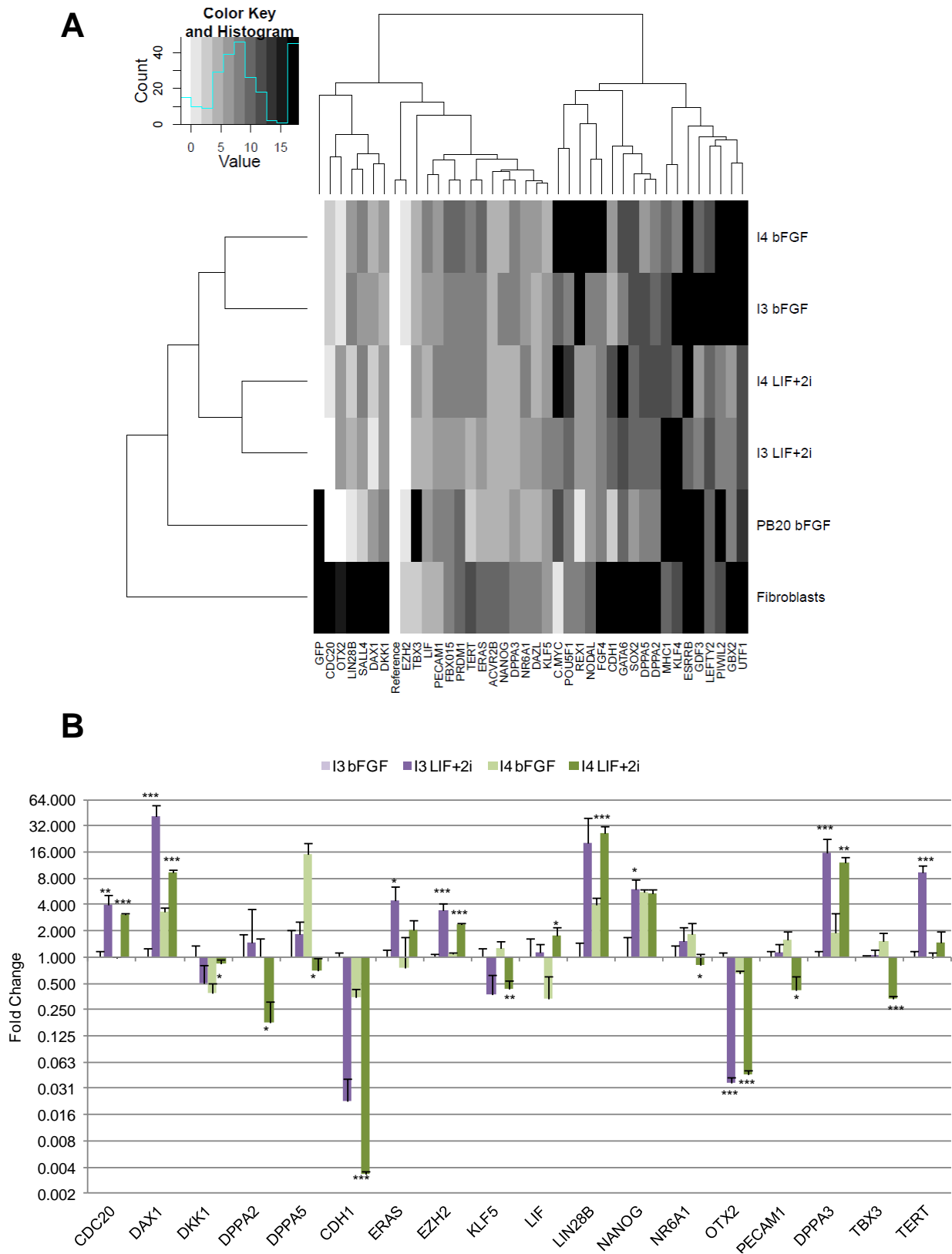


Figure 38 : Expression de nombreux gènes de pluripotence dans les lignées I3, I4 et PB20 **A.** Heatmap comparant les lignées I3 et I4 cultivées en présence de bFGF (milieu piPS) ou de LIF et d'inhibiteurs des voies GSK3 β et MEK (milieu LIF+2i) avec la lignée PB20 et une lignée de fibroblastes embryonnaires. L'échelle de gris représente les ΔC_t entre chaque gène cible et la référence, du blanc, gène exprimé autant ou plus que le gène de référence, au noir, expression non détectée. **B.** Niveau d'expression différentiel pour un sous-ensemble de gènes de la pluripotence naïve ou amorcée en présence de LIF+2i par rapport à la culture en bFGF dans les lignées I3 et I4. Les valeurs sont normalisées par rapport à la lignée I3 en bFGF. t-test * $p < 0.05$; ** $p < 0.01$ *** $p < 0.001$.

l'expression cytoplasmique et membranaire de CDH1 et LIN28B (Figure 37C). Seule une faible partie de la population cellulaire exprime le marqueur membranaire SSEA4 et l'expression de SSEA1 n'a pas été détectée, ni par marquage sur cellules fixées ni par marquage sur cellules en suspension suivi d'une analyse par cytométrie de flux FACS (Figure 37C et D). L'expression des quatre marqueurs sélectionnés pour valider les lignées cellulaires (NANOG, LIN28B, OCT4 et SOX2) est conservée lorsque les lignées sont transférées dans le milieu LIF+2i (Figure 37C). La lignée contrôle exprime elle aussi ces quatre marqueurs ainsi que CDH1 et SSEA4 (Figure 37C).

b. Etude d'un ensemble de gènes de la pluripotence naïve et amorcée

Afin de mieux caractériser le profil d'expression des gènes de pluripotence dans les lignées I3, I4 et PB20, et ce en culture dans le milieu piPS contenant du bFGF et dans le milieu LIF+2i, nous avons élargi les analyses à un plus grand nombre de gènes décrits dans la littérature comme caractéristiques des lignées pluripotentes. Parmi eux, certains gènes sont caractéristiques de la pluripotence dite amorcée et sont exprimés dans les lignées pluripotentes humaines et dans les lignées EpiSCs murines (*OTX2*, *NODAL*, *DKK1*, *NESTIN*, *DNMT3B*, *GATA6*, *LEFTY2*, *ACVR2B* et *MHC1*), tandis que d'autres sont plus représentatifs de la pluripotence naïve décrite dans les lignées mES (*DAX1*, *DPPA2*, *DPPA3*, *DPPA5*, *DAZL*, *REX1*, *KLF2*, *KLF4*, *KLF5*, *ERAS*, *LIN28B*, *EZH2*, *PECAM1*, *GBX2*, *FBX015*, *FGF4*, *PIWIL2* et *ESRRB*).

Dans un premier temps, on constate que de nombreux gènes sont éteints dans les lignées de fibroblastes (*LIN28B*, *DKK1*, *FGF4*, *SOX2*, *ESRRB*, *DPPA5*, *DPPA2*, *CDH1*, *GFP*, *CDC20*, *DAX1*, *SALL4*, *GDF3*, *UTF1*, *GBX2* et *GATA6*) ou que leur expression est plus faible que dans la majorité des lignées reprogrammées (*ACVR2B*, *DAZL*, *EZH2*, *KLF5*, *NANOG*, *NODAL*, *DPPA3*, *TERT*, *OTX2*) (Figure 38A). Néanmoins, certains des gènes étudiés s'expriment à un niveau équivalent voir supérieur dans les fibroblastes, comme par exemple le gène utilisé comme facteur de reprogrammation *c-MYC* dont l'expression est très forte dans ces cellules, ou encore *TBX3* (Figure 38A). Ces résultats suggèrent que ces gènes ne sont peut-être pas de bons indicateurs de la pluripotence porcine.

En comparant les résultats obtenus pour les lignées I3 et I4 cultivées en présence de bFGF et cultivées en présence de LIF et des inhibiteurs de GSK3 β et de MEK, on constate que de nombreux gènes montrent une variation statistiquement significative du niveau d'expression entre les deux conditions (Figure 38B). Tout d'abord, les gènes *UTF1*, *PIWIL2*, *REX1* et *GBX2* sont spécifiquement activés dans les deux lignées lors du passage en milieu LIF+2i (Figure 38A). De plus, les gènes *NODAL*, *LEFTY2* et *GDF3* se réactivent dans l'une des deux lignées (I3 pour *GDF3* et *LEFTY2*, I4 pour *NODAL*) et leur expression augmente dans la seconde, de façon significative pour *NODAL* et *GDF3*. Le gène *ESRRB* quant à lui est uniquement réactivé dans la lignée I3. L'utilisation du milieu LIF+2i

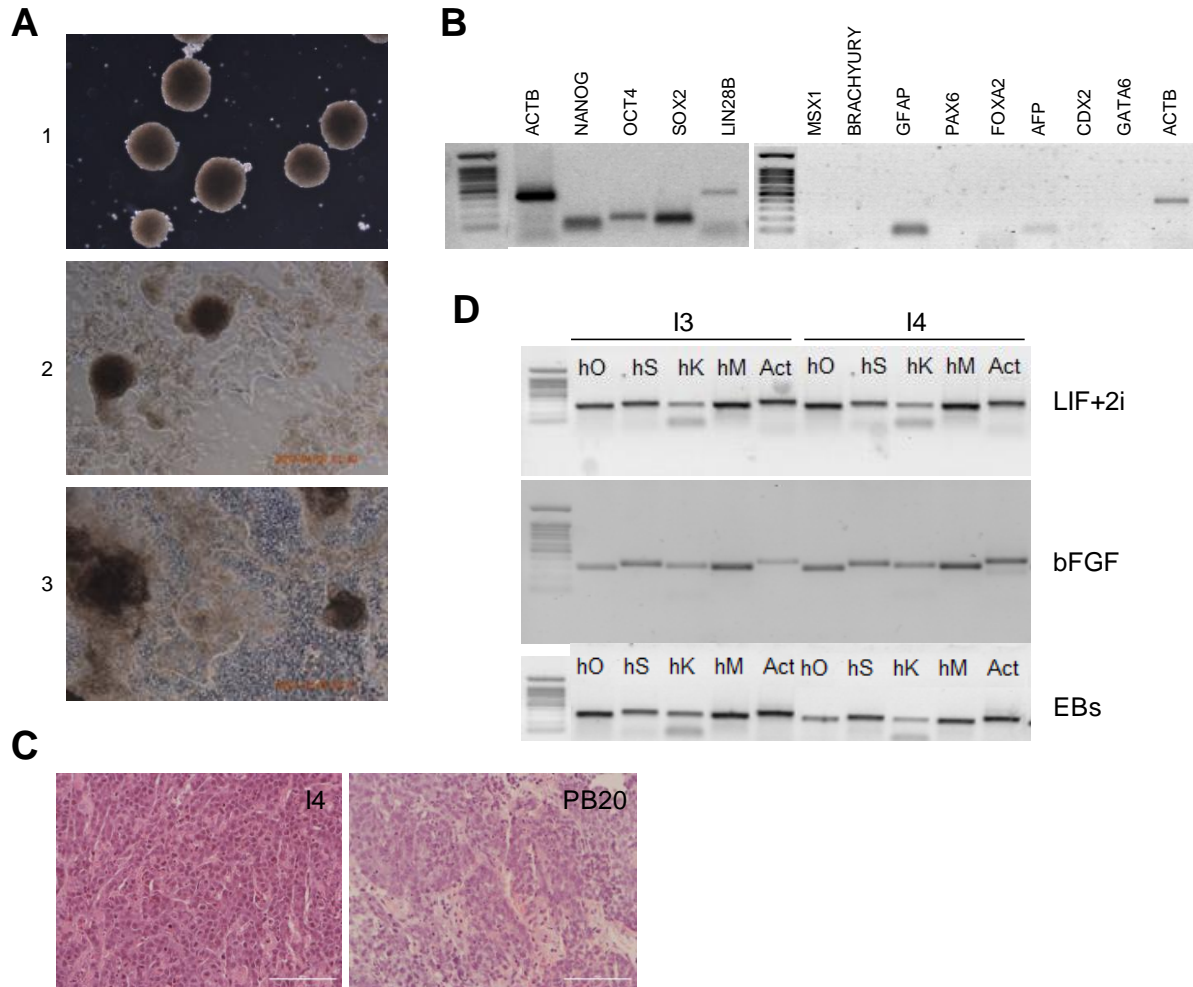


Figure 39 : Faible capacité de différenciation des lignées I3 et I4. **A.** Formation de corps embryoides (EBs) en culture en suspension dans le milieu DIFF (1) proliférant mais ne différenciant pas après 10 jours de culture adhérente (2) contrairement aux EBs de lignées ES murines (3). **B.** Expression continue des marqueurs endogènes de pluripotence NANOG, OCT4, SOX2 et LIN28B et répression des marqueurs de différenciation précoce dans les EBs (lignée I4). **C.** Coloration hématoxyline/éosine de coupe histologique des tumeurs formées par les lignées I4 et PB20 après injection intramusculaire dans des souris immunodéficiente. Echelle 100µm. **D.** Expression continue des 4 facteurs de reprogrammation humains à passage élevé dans les lignées I3 et I4 quelque soit le milieu de culture, ainsi que dans les EBs.

entraîne également une augmentation significative de l'expression de *CDC20*, *DAX1*, *EZH2* et *DPPA3* dans les deux lignées (Figure 38B). L'expression des gènes *LIN28B*, *TERT* et *ERAS* augmente elle aussi dans les deux lignées bien que le résultat ne soit statistiquement significatif que pour l'une des deux lignées. De la même manière, le changement de milieu déclenche également une diminution significative de l'expression d'*OTX2* dans les lignées I3 et I4 et une diminution de *CDH1* et *KLF5* significative dans la lignée I4 seulement accompagnée d'une tendance forte à la diminution dans la lignée I3. Enfin, certains gènes varient spécifiquement dans une lignée : par exemple, l'expression de *NANOG* augmente dans la lignée I3 pour rejoindre le niveau d'expression détecté dans la lignée I4 quelque soit le milieu de culture. Il est intéressant de souligner ici que parmi les gènes surexprimés en milieu LIF+2i, 9 sont caractéristiques de la pluripotence naïve (*ERAS*, *LIN28B*, *PIWIL2*, *REX1*, *GBX2*, *DAX1*, *EZH2*, *DPPA3*, *ESRRB*), indiquant une possible évolution des cellules vers l'état naïf de pluripotence au cours de la culture en présence de LIF et d'inhibiteurs.

Enfin, bien que plusieurs gènes de pluripotence tels que *GDF3*, *TBX3* ou *PIWIL2* soient éteints dans la lignée PB20, le niveau d'expression de plusieurs autres gènes est plus élevé dans cette lignée que dans les lignées I3 et I4, et ce, quelque soit le milieu de culture. C'est en particulier le cas pour *CDC20*, *CDH1*, *DPPA5*, *ERAS*, *LIN28B*, *OTX2*, *NR6A1*, *REX1*, *SALL4*, *TERT* et *SOX2* (Figure 38A).

3. Des capacités de différenciation limitées

La capacité de différenciation des lignées I3, I4 et PB20 a été évaluée par la formation de corps embryoides. Placées en culture en suspension et sous légère agitation dans le milieu DIFF, les cellules forment rapidement des structures en 3 dimensions (Figure 39A). Toutefois, certaines de ces structures ne sont pas stables et se délitent rapidement. De plus, bien que les cellules se rassemblent, les structures obtenues n'arborent pas toutes l'aspect homogène et arrondi des corps embryoides. Ces structures se fixent bien une fois replacées en culture adhérente, mais la croissance cellulaire ne diminue pas, comme c'est le cas pour les cellules miPS. De plus, la population reste très homogène et on ne remarque que peu de sous-populations présentant une morphologie différente caractéristique de la différenciation cellulaire, à l'inverse des résultats obtenus avec des lignées ES murines pour lesquelles on observe au bout de 15 jours de culture l'apparition de cellules contractiles (Figure 39A). Enfin, ces structures expriment de manière continue *NANOG*, *OCT4*, *SOX2* et *LIN28B* mais sont négatives pour des marqueurs de la différenciation précoce tels que *BRACHYURY* et *MSX1*, spécifiques du mésoderme, *GFAP* et *PAX6*, spécifiques de l'ectoderme, ou encore *FOXA2* et *AFP*, spécifiques de l'endoderme (Figure 39B).

Après injection intramusculaire dans des souris immunodéficientes, ni les lignées I3 et I4 ni la lignée PB20 n'ont permis le développement de tératomes matures. Les cellules continuent de proliférer mais forment des carcinomes et ne différencient pas dans les trois feuilletts embryonnaires (Figure 39C).

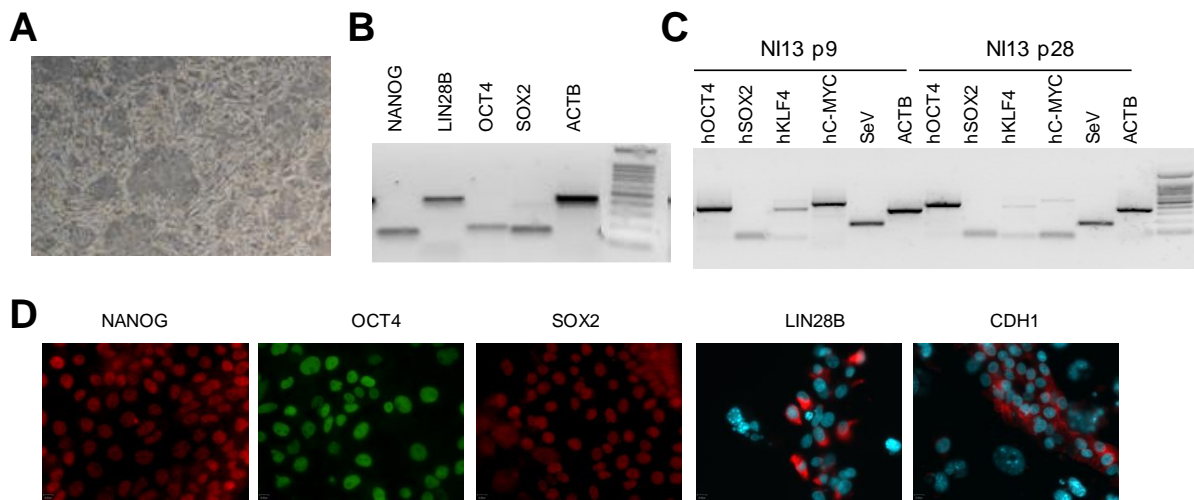


Figure 40 : Caractères de pluripotence des lignées NI. **A.** Morphologie iPS-like des lignées NI en culture. **B.** Expression des marqueurs de pluripotence NANOG, LIN28, OCT4 et SOX2 endogènes. **C.** Diminution de l'expression des facteurs de reprogrammation exogènes au cours des passages en culture dans la lignée NI13. **D.** Immunocytochimie des marqueurs nucléaires NANOG, OCT4, SOX2 (anticorps secondaire Alexa 568 ou 488) et des marqueurs cytoplasmiques et membranaires LIN28B et CDH1 (anticorps secondaire Alexa 568 et DAPI pour contre-colorer les noyaux).

Tableau 22: Comparaison de l'expression de plusieurs gènes dans les lignées obtenues par système non intégratif par rapport aux lignées I (ANOVA : * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$)

Gene	Sens de variation	p-valeur
POU5F1	+	2.09E-09
CDC20	+	4.07E-09
c-MYC	+	3.41E-03
DAX1	+	5.84E-05
DAZL	+	6.89E-03
DPPA2	+	7.07E-10
DPPA5	+	2.67E-05
CDH1	+	2.31E-05
ERAS	+	4.82E-06
EZH2	+	8.88E-10
FBX015	+	4.82E-02
GATA6	+	9.10E-05
KLF5	+	3.65E-02
LEFTY2	+	2.34E-02
LIN28B	+	7.81E-08
NANOG	+	1.54E-05
NODAL	+	5.62E-06
NR6A1	+	6.23E-05
PRDM1	+	2.41E-02
SALL4	+	3.67E-09
SOX2	+	3.52E-06
DPPA3	+	1.70E-10
TERT	+	3.77E-06
GFP	-	1.44E-15
OTX2	-	6.18E-06
TBX3	-	3.19E-02

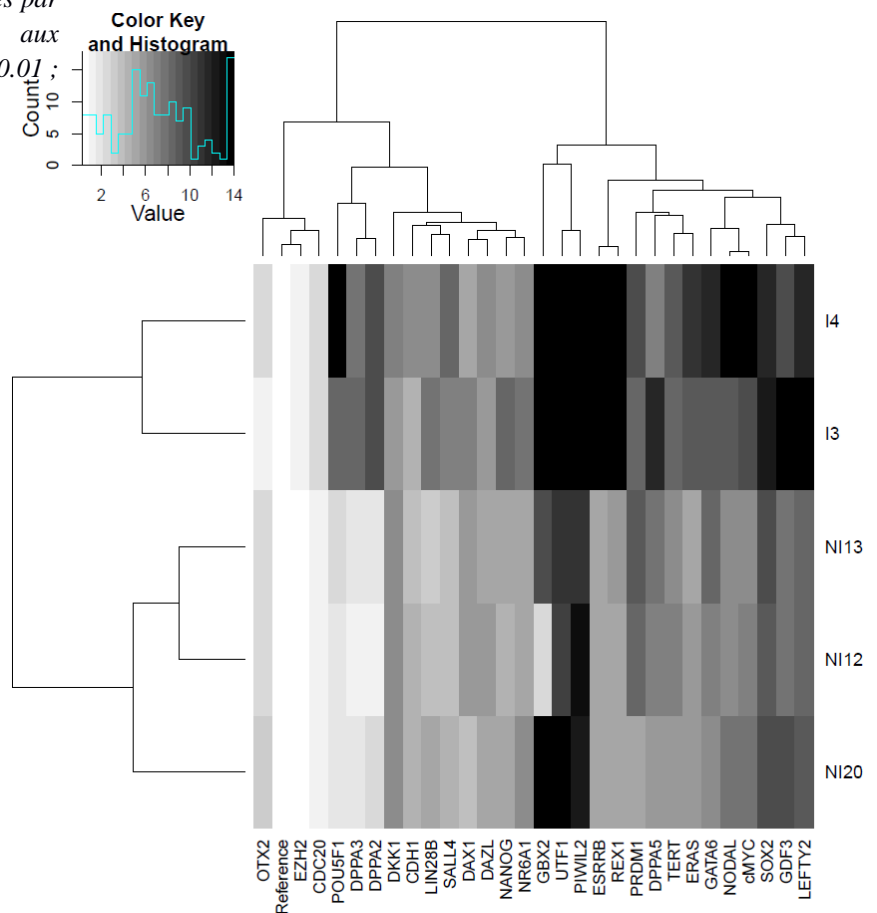


Figure 41 : Heatmap représentant les ΔC_t entre gènes cibles et gènes de références, de blanc, expression équivalente au gène de référence, à noir, gène non détecté : de nombreux gènes de pluripotence sont plus fortement exprimés dans les lignées issues du système non intégratif.

L'expression des quatre facteurs humains utilisés comme cocktail de reprogrammation des cellules porcines a été évaluée par PCR conventionnelle. Après plus de 70 passages en culture, on constate que ces quatre facteurs sont toujours actifs dans les lignées I3 et I4 (Figure 39D). Le remplacement du milieu de culture par le milieu LIF+2i ne semble pas avoir d'incidence sur l'expression de ces marqueurs qui est conservée après plus de 30 jours de culture dans ce nouveau milieu, soit plus de 10 passages (Figure 39D). Cette expression continue des marqueurs humains pourrait constituer un obstacle à la reprogrammation totale des cellules mais également bloquer leur capacité de différenciation. Il est d'ailleurs intéressant de noter que ces marqueurs sont toujours exprimés après 15 jours de culture dans du milieu de différenciation DIFF, en suspension puis en adhérence (Figure 39D). La lignée contrôle PB20 dérivée par infection lentivirale est elle aussi dépendante de l'expression de ces exogènes.

III. Reprogrammation cellulaire par un système non intégratif

1. Expression de nombreux gènes de pluripotence

Afin d'éviter l'expression continue des transgènes, des fibroblastes de l'animal transloqué t(Y;14) ont été reprogrammés par le biais d'un système non intégratif (NI), basé sur l'utilisation d'un virus Sendai (voir Tableau 22). Dès le cinquième jour suivant la reprogrammation, des foyers cellulaires ressemblant à des clones iPS sont apparus. Ils ont été repiqués entre le 15^{ème} et le 22^{ème} jour suivant l'infection. Ces clones à morphologie iPS-like (Figure 40A) sont positifs pour l'expression de la phosphatase alcaline, ainsi que pour celle des marqueurs NANOG, LIN28B, OCT4, SOX2 et CDH1 (Figure 40B et D). De plus, de nombreux marqueurs de pluripotence sont exprimés plus fortement dans les lignées non intégratives que dans les lignées intégratives et ce dans les mêmes conditions de culture (Figure 41). Ainsi, le niveau d'expression des gènes recensés dans le Tableau 23 est statistiquement différent entre les deux types de systèmes de reprogrammation, avec des valeurs de fold change pouvant aller jusqu'à 80 pour *DPPA3*. Enfin, les lignées issues du système non intégratif expriment plusieurs gènes non détectés dans les lignées I3 et I4 dans les mêmes conditions de culture (Figure 41).

Malgré l'utilisation d'un système non intégratif, l'expression des facteurs exogènes de reprogrammation est détectée de manière continue jusqu'au passage 10, à l'exception de *SOX2* qui est indétectable (Figure 40C). Néanmoins, cette expression semble plus limitée que dans les lignées I3 et I4, et s'atténue au fil des passages, en particulier pour *KLF4* et *c-MYC* (Figure 40C). De plus, l'expression des marqueurs porcins *OCT4*, *SOX2* et *c-MYC* est plus forte dans les lignées NI, suggérant une meilleure réactivation du circuit endogène (Figure 41 et Tableau 23).

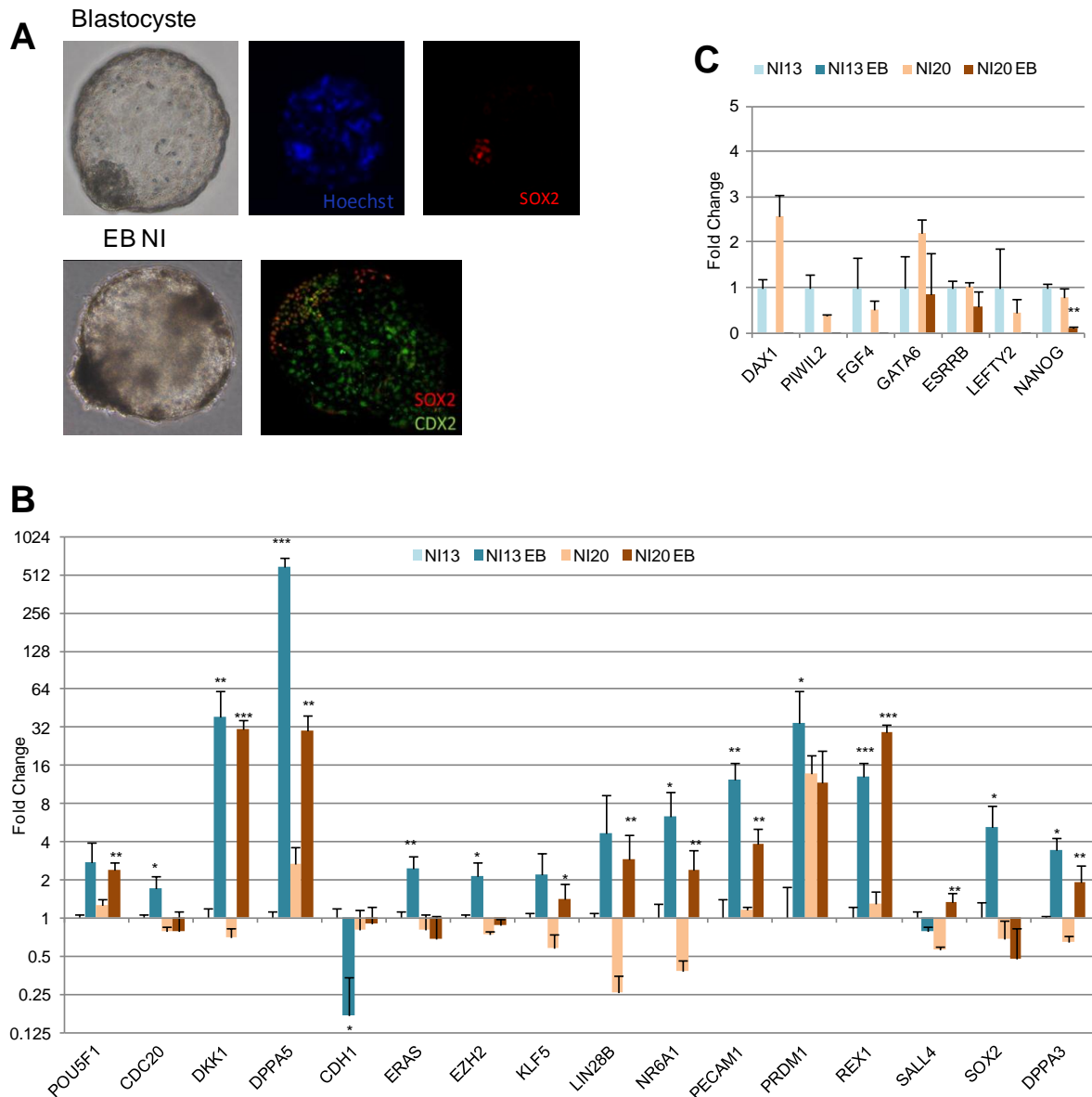


Figure 42 : A. La morphologie des EBs issus des lignées NI est similaire à celle des blastocystes porcins et ces structures expriment des marqueurs des territoires embryonnaires (SOX2, rouge) et extraembryonnaires (CDX2, vert) B et C. Evolution de l'expression d'un sous-ensemble de gènes lors de la formation de corps embryoïdes à partir des lignées NI13 et NI20 (normalisation : NI13). t-test : * $p < 0.05$; ** $p < 0.01$ *** $p < 0.001$.

2. Formation de structures « blastocyst-like »

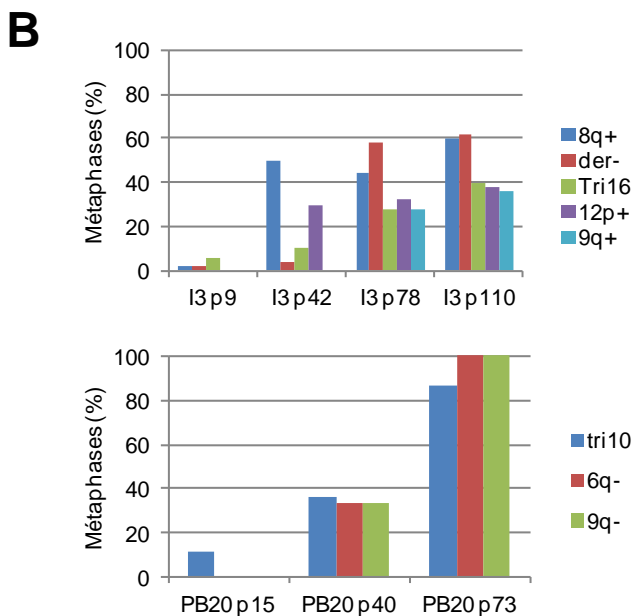
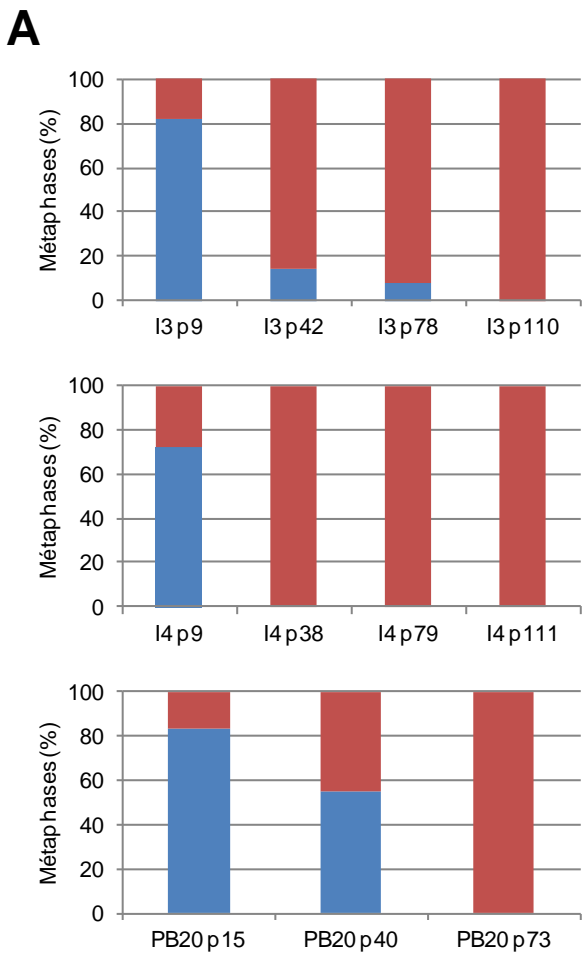
Les lignées cellulaires issues du système de reprogrammation non intégratif forment en culture en suspension dans le milieu DIFF des structures ressemblant fortement à des blastocystes porcins, constituées d'une membrane cellulaire sphérique et d'une masse cellulaire polarisée (Figure 42A). De plus, ces corps polarisés possèdent le même profil d'expression de SOX2 que les blastocystes, c'est-à-dire une expression restreinte à l'ICM (Figure 42A). Sur le plan expressionnel, ces corps expriment de façon continue, voire même intensifiée, plusieurs gènes de pluripotence. En effet, les gènes *DKK1*, *DPPA5*, *NR6A1*, *PECAM1*, *REX1* et *DPPA3* sont significativement surexprimés dans les deux lignées de corps embryoides avec un fold change pouvant aller jusqu'à plus de 500 pour *DPPA5* (Figure 42B). Les gènes *POU5F1*, *KLF5*, *LIN28B* montrent également une forte tendance à la surexpression, de même que *GBX2* qui est réactivé dans les corps embryoides de la lignée NI20. Enfin, les gènes *DAX1*, *PIWIL2*, *FGF4*, *LEFTY2*, *GATA6*, *ESRRB* et *NANOG* sont spécifiquement éteints ou sous-régulés dans les corps embryoides (Figure 42C).

IV. Instabilité génomique des piPSCs

1. Accumulation d'anomalies chromosomiques au fil des passages

L'étude du caryotype des lignées I3 et I4 indique une forte instabilité génomique dans les lignées iPSC-like porcines. Ainsi, bien que ces lignées soient stables en culture jusqu'à des passages élevés (>100), de nombreuses anomalies chromosomiques apparaissent au cours des passages. L'analyse en bandes G sur 50 métaphases à quatre passages différents (environ p10, p40, p80 et p110) (Figure 43A) a révélé que le nombre de cellules normales dans la lignée I3 chute brusquement entre les passages 9 et 42, et qu'au delà du centième passage toutes les cellules présentent au moins une anomalie. Pour la lignée I4, les anomalies sont généralisées dès le quarantième passage.

Parmi les anomalies identifiées, de nombreuses sont récurrentes dans la majorité des métaphases (Figure 43B). C'est le cas par exemple d'une addition d'un fragment de chromosome sur le bras long du chromosome 8, notée 8q+, qui est retrouvée dans la moitié des métaphases analysées à partir du passage 42 pour la lignée I3. On retrouve également en grande fréquence dès ce passage la perte du chromosome der14, c'est-à-dire du chromosome dérivé de la translocation réciproque t(Y;14). Au fil des passages suivants, les fréquences d'une trisomie du chromosome 16, d'une addition chromosomique sur le bras long du chromosome 9, notée 9q+, et d'une addition sur le bras court du chromosome 12, notée 12p+, augmentent considérablement pour atteindre respectivement 40%, 36% et 38% des métaphases à passage tardif (p110). Dans la lignée I4, on retrouve l'anomalie 8q+, la trisomie 16 et la perte du chromosome dérivé der14, auxquelles s'ajoute un fragment supplémentaire



C

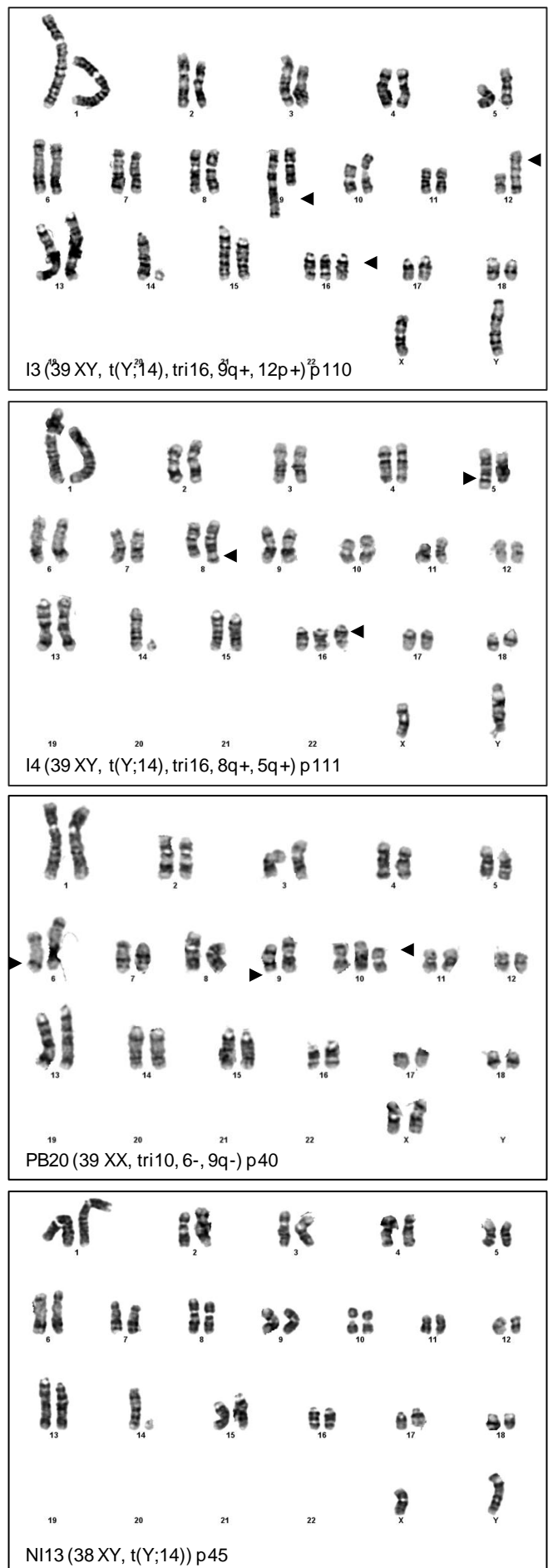


Figure 43 : Evolution du caryotype des lignées reprogrammées au cours du temps. **A.** Diminution drastique du nombre de métaphases à caryotype normal au fil des passages, soit ($2n=38XY, t(Y;14)$) pour les lignées I3 et I4 et ($2n=38XX$) pour la lignée PB20 (bleu : normal, rouge : anormal). **B.** Remaniements majoritaires observés dans la lignée I3 et dans la lignée PB20. **C.** Exemples de caryotype en bande G obtenus pour les lignées I3, I4 et PB20 avec anomalies et la lignée NI13 sans anomalie. Les flèches indiquent les anomalies (hormis la translocation $t(Y;14)$ attendue).

sur le chromosome 5 noté 5q+. Ces anomalies coexistent et forment des profils caryotypiques. Ainsi, la lignée I3 abrite deux populations majoritaires, la première de caryotype (37XY, t(Y;14), der14-, 8q+) concernant 38% des cellules et la seconde de caryotype (39XY, t(Y;14), tri16, 9q+, 12p+) (Figure 43C) concernant 28% des cellules au passage le plus tardif. Les 34% de métaphases restantes présentent ces mêmes anomalies combinées différemment ainsi que d'autres anomalies de fréquence moindre. De la même façon dans la lignée I4, les deux populations majoritaires ont pour caryotype (39XY, t(Y;14), tri16, 8q+, 5q+) (Figure 43C) observé dans 42% des métaphases et (39XY, t(Y;14), tri16, 8q+) observé dans 32% des métaphases. L'absence de ces anomalies dans les cellules somatiques avant la reprogrammation, à l'exception de la trisomie 16 détectée 1 fois sur 17 métaphases analysées, suggère que ces réarrangements apparaissent dans les lignées iPS où elles deviennent rapidement majoritaires, probablement en conférant aux cellules remaniées un avantage prolifératif.

La lignée contrôle PB20 montre quant à elle plusieurs anomalies apparaissant au cours des passages, bien que de manière moins rapide. Néanmoins, après plus de 70 passages en culture, les 15 métaphases analysées se sont toutes révélées porteuses de remaniements chromosomiques (Figure 43A). Les principaux remaniements observés sont une trisomie du chromosome 10 et des délétions sur les chromosomes 6 et 9 (Figure 43B et C).

2. Cas du système non intégratif

La lignée NI13 a été cultivée jusqu'au passage 45 afin de déterminer l'évolution de son caryotype au fil des passages. Au passage 22, les 27 métaphases analysées possèdent un caryotype « normal », c'est-à-dire porteur de la translocation réciproque t(Y;14) présente dans les lignées somatiques. Au passage 45, les 16 métaphases analysées présentent elles aussi ce caryotype (Figure 43C). Ces résultats suggèrent que le génome de la lignée NI13 est stable dans le temps, contrairement à celui des lignées I3, I4 et PB20. Il est donc possible que la technique de reprogrammation ait un effet direct sur la stabilité du génome, ou un effet indirect en permettant une reprogrammation cellulaire plus aboutie et donc des lignées cellulaires plus stables.

3. Caractérisation des remaniements des lignées I3 et I4

a. Apparitions de duplications dans le génome

L'étude par CGH et caryotypes en reverse DAPI des lignées de fibroblastes t(Y;14), et des lignées I3 et I4 a mis en évidence l'apparition de zones dupliquées dans les lignées I3 et I4 (Figure 44A). Ces duplications apparaissent dès les premiers passages et semblent s'intensifier dans les lignées aux

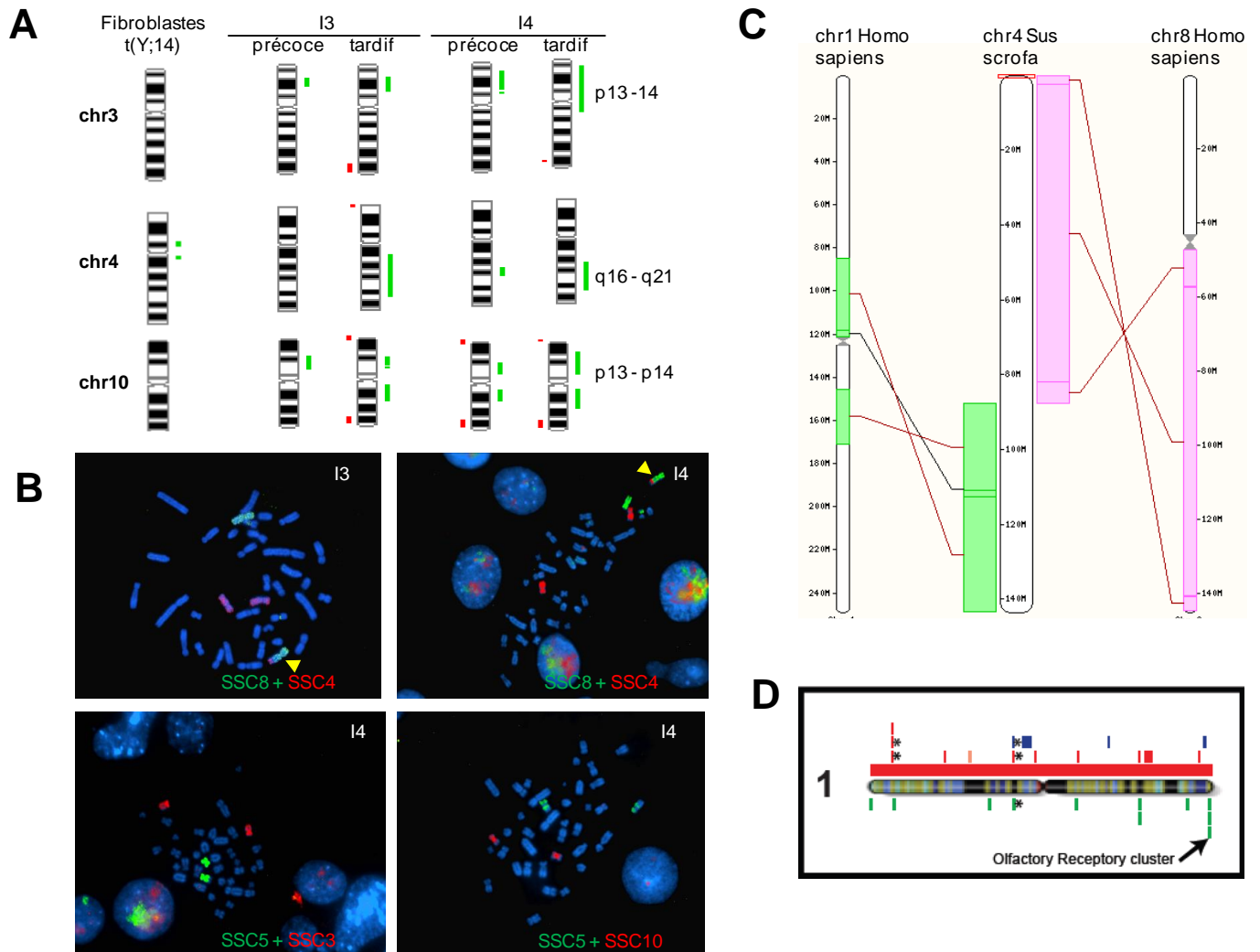


Figure 44 : Caractérisation des remaniements majoritaires dans les lignées I3 et I4. **A.** Analyse CGH des lignées I3, I4 et de la lignée de fibroblastes t(Y;14) révélant des duplications (vert) et des délétions (rouge) apparaissant dans les lignées I3 et I4 sur les chromosomes 3, 4 et 10 **B.** La combinaison des sondes de peinture (SSC8+SSC4) (SSC5+SSC3) (SSC5+SSC10) en hybridation in situ en fluorescence (FISH) sur les métaphases des lignées I3 et I4 permet de déterminer (flèches jaunes) que le fragment transloqué sur le bras long du chromosome 8 (vert) provient de la duplication du chromosome 4 (rouge) et que l'addition d'un fragment sur le chromosome 5 (vert) est sans association avec les duplications observées en CGH des chromosomes 3 et 10 (rouge). **C.** Régions de synténie entre le chromosome 4 porcine et les chromosomes 8 et 1 humains (extrait de Ensembl). **D.** Zones dupliquées sur le chromosome 1 humain dans les lignées hESCs (rouge) et hiPSCs (bleu) (Laurent et al., 2011).

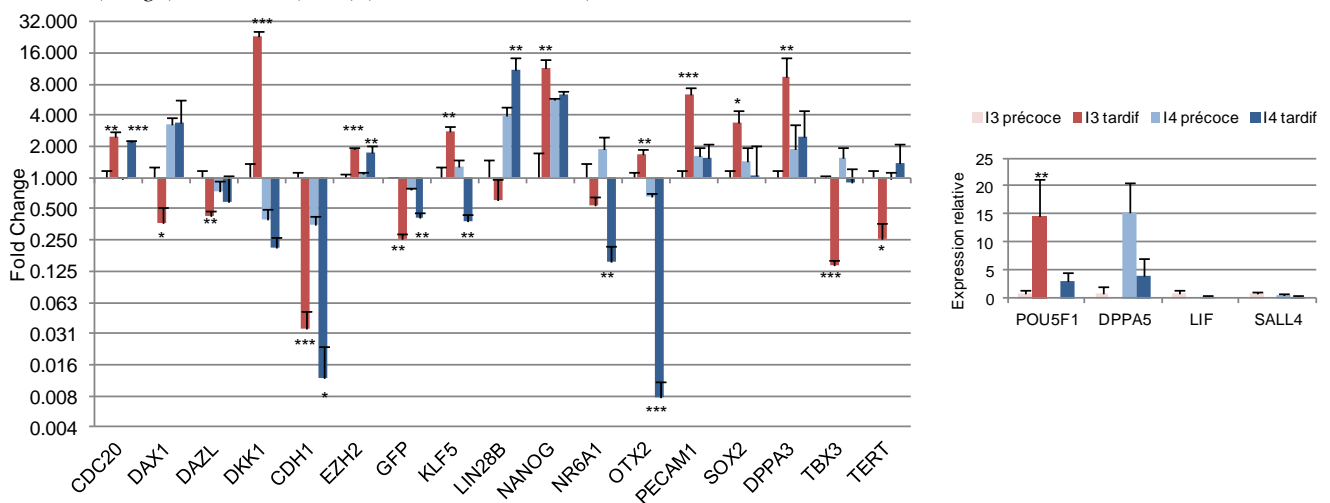


Figure 45 : Variation du niveau d'expression de gènes de pluripotence entre passages précoce et tardif (normalisation : I3 précoce) t-test: * $p < 0.05$; ** $p < 0.01$ *** $p < 0.001$.

passages tardifs, notamment dans la lignée I4. Elles concernent le bras court du chromosome 3 (p13-p14), le bras long du chromosome 4 (q16-q21) et le chromosome 10. La difficulté d'orienter ce chromosome lors de l'analyse en reverse DAPI ne permet pas de déterminer si la duplication autour du centromère concerne le bras p ou le bras q.

Afin de relier ces duplications aux anomalies chromosomiques majoritaires recensées dans les lignées I3 et I4, nous avons réalisé une étude par FISH en hybridant des sondes de peinture chromosomique et des sondes de BACs des chromosomes remaniés ou dupliqués sur les métaphases des lignées I3 et I4. Cette étude a révélé un marquage spécifique du chromosome 4 sur l'extrémité du bras long du chromosome 8, correspondant à l'anomalie 8q+, dans les lignées I3 et I4 (Figure 44B). L'utilisation comme sonde d'un BAC localisé en 4q15-16 (clone PigI-858F5) indique que cette région n'est pas transloquée sur le chromosome 8, bien qu'elle soit frontalière de la zone apparaissant dupliquée en CGH. Ce résultat suggère que la région dupliquée et transloquée se situe entre les bandes G q21 et q25. L'utilisation d'une sonde de peinture du chromosome 5 a confirmé la présence d'une anomalie en 5q+ sur l'un des deux chromosomes dans plusieurs métaphases de la lignée I4 (Figure 44B). En effet, l'un des deux chromosomes apparaît plus long que l'autre dans la moitié des métaphases capturées, mais cette extrémité n'est marquée ni par la peinture du chromosome 5, ni par celle des chromosomes 3 et 10, ne nous permettant pas ici de relier duplication génomique détectée par CGH et remaniement chromosomique (Figure 44B).

b. Caractérisation de la zone dupliquée du chromosome 4

La duplication et translocation d'une partie du chromosome 4 sur le chromosome 8 étant très conservée dans nos deux lignées iPS porcines putatives, nous nous sommes intéressés aux gènes qui la composent. L'analyse de la synténie entre le chromosome 4 porcine et le génome humain révèle que ce chromosome correspond aux chromosomes humains 8 (pour la région p15-q15) et 1 (pour la région q16-q25) (Figure 44C). Or, le chromosome 1 a été décrit comme souvent dupliqué dans les lignées de cellules souches pluripotentes humaines, y compris dans les régions chr1 :103908591-104012569 et chr1 :111296543-116660879 (Figure 44D). Chez le porc, les régions homologues se trouvent sur le bras q du chromosome 4 (chr4 : 110922275-143465848) et pourraient faire partie de la zone dupliquée dans les lignées I3 et I4. Les gènes présents dans la région de 5.36MB du chromosome 1 humain sont listés dans le Tableau 24 avec leurs homologues porcins. Parmi eux, on retrouve deux gènes impliqués dans la voie de signalisation WNT importante pour la maintenance de la pluripotence (*WNT2B* et *VANGLI*), deux gènes impliqués dans le déroulement du cycle cellulaire et de la mitose (*MOV10* et *SYCP1*) et plusieurs oncogènes (*NRAS*, *RAP1A*, *RHOC*). Enfin, en effectuant une analyse fonctionnelle sur 533 gènes identifiés sur le bras long du chromosome 4 porcine (sur un total de 1216 transcrits Ensembl), plusieurs fonctions biologiques sont enrichies avec un nombre de gènes supérieur

Tableau 23 : Gènes contenus dans la région dupliquée 111296543-116660879 du chr1 humain (Ensembl, Assemblage GRCh38).

Homo sapiens genes	Location	Sus scrofa gènes homologues	Information
<i>CD53</i>	1:111415775-111442550	4:119674096-119703427	Leukocyte surface antigen.
<i>LRIF1</i>	1:111489807-111506701	4:119652638-119659834	Ligand-dep nuclear-receptor interacting factor.
<i>DRAM2</i>	1:111659955-111682838	4:119550712-119574480	DNA damage regulated autophagy.
<i>CEPT1</i>	1:111682249-111727724	4:119320217-119544247	Choline/ethanolaminephosphotransferase 1
<i>DENND2D</i>	1:111729796-111747157	4:119302443-119319206 4:119501409-119515710	DENN domain-containing protein 2D.
<i>CHI3L2</i>	1:111743393-111786062	4:119457085-119474711	Chitinase-3-like protein 2
<i>CHIA</i>	1:111833484-111863188	4:119410879-119422056	Acidic mammalian chitinase..
<i>OVGP1</i>	1:111956936-111970399	4:119112229-119124274	Oviduct glycoprotein.
<i>WDR77</i>	1:111982512-111991998	4:119089787-119098186	Methylosome protein50.
<i>ATP5F1</i>	1:111991486-112005395	4:119078679-119090067	ATP synthase, mitochondrial.
<i>C1orf162</i>	1:112016414-112021134	4:119057620-119064131	Transmembrane protein C1orf162
<i>ADORA3</i>	1:112025970-112106584	4:118996434-119053012	Adenosine receptor A3
<i>RAP1A</i>	1:112084840-112259313	4:118826791-118897331	RAS-related protein.
<i>FAM212B</i>	1:112223252-112298446	4:118808703-118810452	FAM212B
<i>DDX20</i>	1:112297867-112310638	4:118772038-118783544	DEAD box polypeptide 20.
<i>KCND3</i>	1:112313284-112531777	4:118751271-118763120 4:118549776-118550885	Potassium voltage-gated channel subfamily D
<i>CTTNBP2NL</i>	1:112938803-113006078	4:118026418-118081234	CTTNBP2 N-terminal-like protein
<i>WNT2B</i>	1:113009163-113072787	4:117972643-117990368	wingless-type MMTV integration site family, member 2B
<i>ST7L</i>	1:113066140-113163447	4:117891640-117957340	Suppressor of tumorigenecity 7 like
<i>CAPZA1</i>	1:113161795-113214241	4:117802092-117818131	F-actin-capping protein subunit alpha-1
<i>MOV10</i>	1:113215763-113243368	4:117822828-117848382	Moloney leukemia virus 10 homolog. Helicase
<i>RHOC</i>	1:113243728-113250056	4:117848597-117854870	Ras homolog gene family, member C
<i>PPM1J</i>	1:113252616-113258099	4:117792236-117797886	protein phosphatase 1J.
<i>FAM19A3</i>	1:113263041-113269857	4:117783000-117785015	FAM19A3. Chemokine-like
<i>SLC16A1</i>	1:113454469-113499635	4:117603634-117614944	Monocarboxylate transporter 1
<i>LRIG2</i>	1:113615831-113674882	4:117506415-117511417	Leucine-rich repeats and immunoglobulin-like domains protein 2
<i>MAGI3</i>	1:113933371-114228545	4:117024784-117266451	Membrane-associated guanylate kinase
<i>PHTF1</i>	1:114239453-114302111	4:116955747-117021813	Putative homeodomain transcription factor 1
<i>RSBN1</i>	1:114304454-114355098	4:116938181-116953629	Round spermatid basic protein1
<i>PTPN22</i>	1:114356433-114414381	4:116918294-116927387	Tyrosine-protein phosphatase non-receptor type 22 (lymphoid)
<i>BCL2L15</i>	1:114420790-114430169	4:116879922-116887481	Bcl-2-like protein 15
<i>AP4B1</i>	1:114437370-114447823	4:116862014-116874408	Adaptator-related protein complex 4 subunit beta-1
<i>DCLRE1B</i>	1:114447763-114456708	4:116854619-116862006	5' exonuclease
<i>HIPK1</i>	1:114471814-114520426	4:116708892-116821627	Homeodomain-interacting protein kinase1.
<i>OLFML3</i>	1:114522063-114524876	4:116704502-116707235	Olfactomedin-like 3
<i>SYT6</i>	1:114631913-114696541	4:116574903-116619861	synaptotagmin 6.
<i>TRIM33</i>	1:114935399-115053781	4:116244045-116351102	Tripartite motif-containing 33. E3 ubiquitin protein ligase
<i>BCAS2</i>	1:115110178-115124260	4:116180381-116190995	Pre-mRNA splicing factor
<i>AMPD1</i>	1:115215719-115238239	4:116057655-116080925	AMP deaminase.
<i>NRAS</i>	1:115247090-115259515	4:116034666-116042522	Neuroblastoma RAS viral oncogene homolog. GTPase
<i>CSDE1</i>	1:115259534-115301297	4:115990249-116034509	Cold-shock protein domain-containing protein E1.
<i>SIKE1</i>	1:115312100-115323306	4:115968345-115974919	Suppressor of IKK epsilon
<i>SYCP1</i>	1:115397424-115537991	4:115731360-115868963	Synaptonemal complex protein 1
<i>TSHB</i>	1:115572415-115576941	4:115691689-115692634	Thyrotropin subunit beta.
<i>TSPAN2</i>	1:115590632-115632121	4:115641713-115680720	Tetraspanin 2
<i>NGF</i>	1:115828539-115880857	4:115445499-115446230 4:115357141-115357866	Nerve growth factor
<i>VANGL1</i>	1:116184574-116240845	4:115085548-115134327	VANG-like protein1 : tetraspanin family
<i>CASQ2</i>	1:116242628-116311402	4:115007632-115070760	Calsequestrin 2.
<i>NHLH2</i>	1:116378998-116386538	4:114947020-114947427	Nescient helix loop helix 2
<i>SLC22A15</i>	1:116519119-116612675	4:114752826-114845145	Solute carrier family 22 member 15
<i>MAB21L3</i>	1:116654376-116677861	4:114691058-114716240	MAB21 like 3

à 10 et un FDR inférieur à 5%. Ces gènes interviennent dans la régulation de la transcription (51 gènes, FDR = 0.0009), la transduction du signal (39 gènes, FDR = 0.0040) et la division cellulaire (13 gènes, FDR = 0.0393). La voie de signalisation la plus enrichie est celle de la transduction du signal olfactif, en accord avec la présence sur le chromosome 4 porcine d'un cluster de récepteurs du système olfactif, que l'on retrouve également sur le chromosome 1 humain (Figure 44D).

c. Expression des gènes de pluripotence entre passages précoces et tardifs

En comparant les résultats d'expression dans les lignées I3 et I4 entre les passages précoces (~p10) et les passages plus tardifs (~p110), on remarque que plusieurs gènes du réseau de pluripotence sont modifiés. Nous avons ainsi identifié plusieurs gènes dont le niveau d'expression varie à la fois dans la lignée I3 et dans la lignée I4, tandis que certaines variations sont spécifiques de chaque lignée cellulaire (Figure 45). Ainsi, les gènes *CDC20* et *EZH2* sont surexprimés au fil des passages. Le rôle bien connu de *CDC20* dans la promotion du cycle cellulaire est en accord avec l'augmentation de la fréquence des repiquages. On remarque que l'expression d'*OCT4* (*POU5F1*) est activée dans la lignée I4 et augmentée dans la lignée I3, indiquant une possible sélection des cellules ayant réactivé le circuit endogène de pluripotence. Il est également intéressant de noter que l'expression de la GFP, caractéristique de l'expression de la cassette EOS qui permet de suivre l'apparition des clones reprogrammés, diminue avec le temps. C'est également le cas pour *CDH1* ainsi que pour *DPPA5*, *LIF* et *SALL4* dont l'expression est réprimée. Enfin, les passages tardifs sont caractérisés par l'absence d'expression de *c-MYC*, *FGF4*, *NODAL* et *KLF4* qui étaient pourtant activés dans au moins une des deux lignées à passage précoce.

V. Etude transcriptomique

Après l'approche gène candidat se concentrant sur les gènes connus de la pluripotence murine et humaine, une étude transcriptomique a été réalisée en triplicat sur plusieurs lignées cellulaires dans différentes configurations, soit au total 8 conditions différentes (cf Matériel et Méthodes, Tableau 15), pour mettre en avant les variations globales du profil d'expression entre différentes lignées. Après filtration des données, 38114 sondes de la puce Agilent ont été conservées. Les données ont ensuite été normalisées puis la corrélation linéaire entre les répliquats techniques a été validée. Une analyse en composante principale (ACP) a révélé que plus de 50% de la variabilité du jeu de données final est expliquée par les 8 conditions analysées et qu'il n'y a pas d'effet visible de la lame d'hybridation (Figure 46A). Il est intéressant de noter ici que la condition la plus éloignée des autres est celle des fibroblastes embryonnaires (F) et que les lignées I3 et I4 sont proches (A,B,C,D) (Figure 46A). Le nombre de sondes différentiellement exprimées entre chaque comparaison a été déterminé en utilisant le package Bioconductor, en fixant un logarithme du « fold change » (lfc) minimum de 2 et une

Tableau 24 : Nombre de sondes sur et sous-régulées pour chaque comparaison entre deux conditions. A : I3 précoce ; B : I3 tardif ; C : I4 précoce ; D : I4 tardif ; E: PB20 ; F: PEF ; G: NI12 ; H: NI12 EBs

	A-F	B-F	C-F	D-F	E-F	G-F	H-F	B-A	D-C	C-A	D-B
Nombre de sondes sur-régulées	3044	3435	3189	3524	4067	3559	3820	1129	642	541	927
Nombre de sondes sous-régulées	2598	3102	2639	2889	3726	3393	3697	854	365	508	701
	A-E	B-E	C-E	D-E	G-E	H-G	G-A	G-C			
Nombre de sondes sur-régulées	1999	1594	2067	1845	740	1754	2018	1922			
Nombre de sondes sous-régulées	2625	1956	2442	2059	1409	2030	1930	2099			

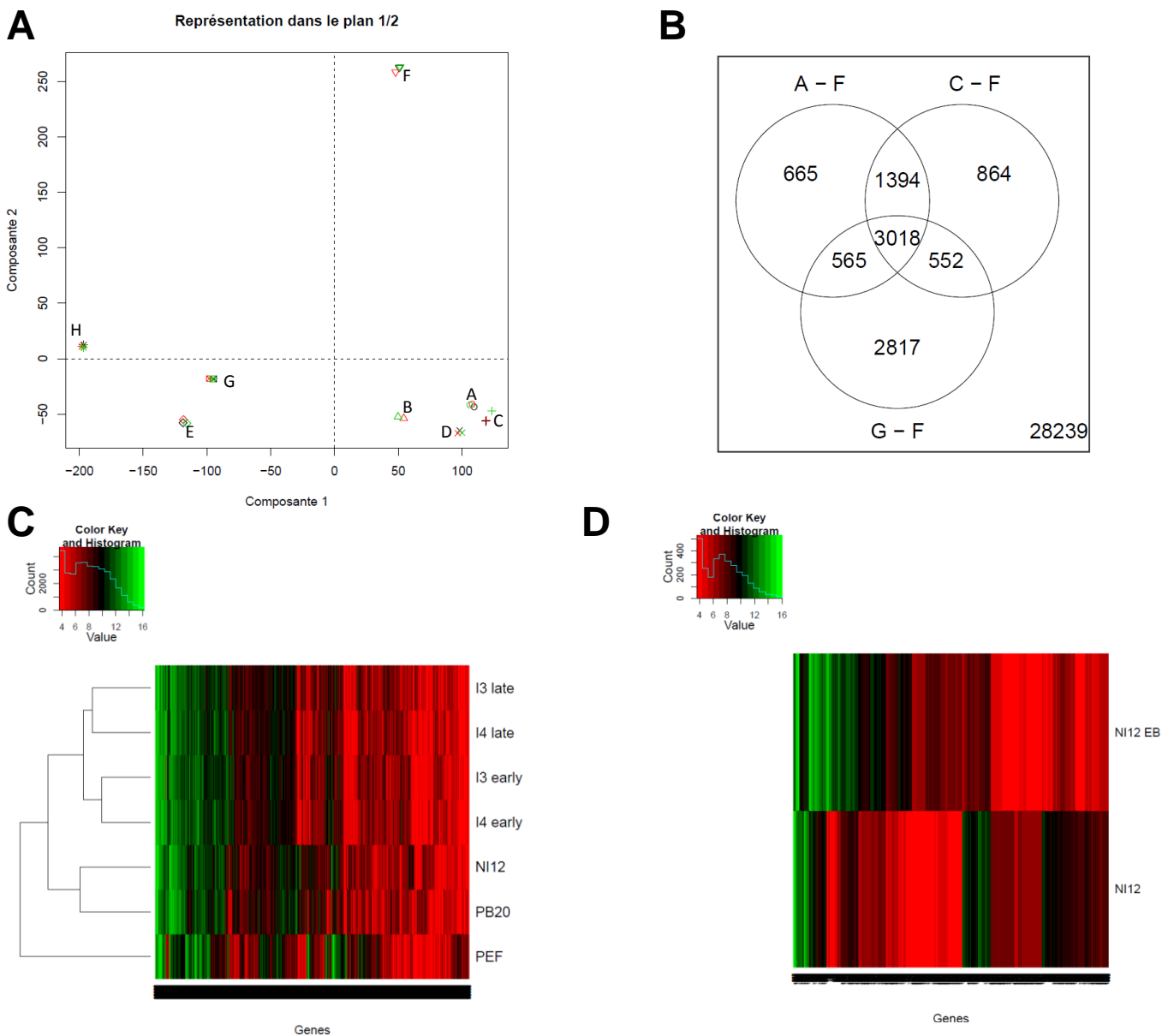


Figure 46 : **A.** Analyse en composante principale montrant le fort regroupement des données par répliquats techniques (marqueurs) et l'absence d'effet de la lame d'hybridation (couleurs) ainsi que la grande variabilité de la condition F (triangle inversé) représentant la lignée fibroblastique contre les autres conditions. **B.** Diagramme de Venn des comparaisons des lignées I3 (A), I4 (C) et NI12 (G) par rapport à la lignée de fibroblastes (F) **C.** Heatmap de l'expression (log) de 5136 gènes présentant au moins un différentiel d'expression ($p < 0.01$, $lfc > 2$) entre les 7 conditions. **D.** Heatmap des gènes différentiels ($p < 0.01$, $lfc > 2$) entre la lignée NI12 et les EBs.

p-valeur ajustée de 0.01 (Tableau 25). Toutefois, de nombreuses sondes différentiellement exprimées sont conservées entre plusieurs comparaisons : c'est par exemple le cas lorsque l'on compare les lignées I3, I4 et NI12 à la lignée de fibroblastes (Figure 46B). Ainsi, le nombre de sondes présentant un différentiel d'expression dans au moins une des comparaisons s'élève à 15817. Pour la suite de l'analyse, nous ne conserverons que les sondes correspondant à un gène annoté chez le porc. Entre les lignées reprogrammées (I3 précoce, I3 tardif, I4 précoce, I4 tardif, PB20 et NI12) et la lignée de fibroblastes, le nombre de gènes annotés différentiellement exprimés dans au moins une des 7 conditions s'élève à 5136 (Figure 46C).

1. Comparaison des lignées reprogrammées

Dans un premier temps, nous avons comparé les différentes lignées pluripotentes entre elles ainsi qu'avec la lignée de fibroblastes. En sélectionnant les gènes différentiellement exprimés dans au moins une de ces comparaisons, il s'avère que le profil d'expression de ces gènes est proche dans les lignées I3 et I4 qui se regroupent en fonction de leur passage (Figure 46C). La lignée non intégrative est plus proche de la lignée contrôle PB20, et la lignée de fibroblastes, comme précédemment indiqué, possède le profil d'expression le plus éloigné (Figure 46C). Ces résultats sont en accord avec les données de PCR quantitative qui indiquaient déjà une plus grande similarité entre lignées issues du système non intégratif et la lignée contrôle.

Il est également intéressant de constater que parmi les 425 gènes communément surexprimés dans les lignées reprogrammées I3, I4, PB20 et NI12 par rapport aux fibroblastes, on retrouve les deux marqueurs *SOX2* et *OCT4*, ainsi que *CDH1*, *SALL4*, *LIN28A* et *OTX2*. Au moins une des deux lignées I ainsi que les lignées PB20 et NI12 surexpriment également les gènes *KLF5*, *DAX1* et *DNMT3B*. Les lignées PB20 et NI12 surexpriment un plus grand nombre de marqueurs de pluripotence, avec *NR6A1*, *LIN28B* et *ERAS*, ainsi que *REX1*, *TERT* et *EZH2* spécifiquement pour la lignée PB20 et *DPPA2*, *DPPA3*, *GDF3*, *NANOG* et *DKK1* spécifiquement pour la lignée NI12. Le niveau de ces gènes surexprimés dans plusieurs lignées reprogrammées en comparaison aux fibroblastes varie également d'une lignée reprogrammée à l'autre. Ainsi, en accord avec les données de PCR quantitative, *SOX2* est plus fortement exprimé dans la lignée NI12 que dans les lignées I3 et I4, tout comme *OCT4* et *CDH1*. Il est également intéressant de noter l'augmentation de l'expression de *MYCN*, *ZFP64*, *CENPA* et *CDT1*, gènes cibles des facteurs de reprogrammation, dans les lignées reprogrammées. Enfin, les gènes *TBX3*, *KFL2*, *PRDM1* et *KLF4* apparaissent sous-régulés dans les lignées reprogrammées par rapport à la lignée fibroblastique, indiquant ici que ces gènes ne sont pas caractéristiques de la pluripotence porcine, ou dans le cas de *KLF4* que l'activation de l'endogène est faible.

Tableau 25 : Annotations fonctionnelles les plus significativement enrichies dans les gènes sur-exprimés par les lignées reprogrammées (GeneCodis).

Nombre de genes	FDR	Annotations
25	5.86E-13	GO:0051301: cell division (BP)
25	1.22E-12	GO:0000278: mitotic cell cycle (BP)
14	5.17E-10	GO:0000087: M phase of mitotic cell cycle (BP)
25	1.92E-09	GO:0007049: cell cycle (BP)
17	3.50E-09	GO:0007067: mitosis (BP)
12	1.84E-08	GO:0000236: mitotic prometaphase (BP)
13	4.42E-08	GO:0000086: G2/M transition of mitotic cell cycle (BP)
22	7.90E-07	GO:0045893: positive regulation of transcription, DNA-dependent (BP)
17	2.95E-06	GO:0042493: response to drug (BP)
21	0.00014835	GO:0006915: apoptotic process (BP)

En utilisant les programmes d'analyse fonctionnelle GeneCodis et GOrilla, nous avons pu identifier les annotations significativement enrichies dans les gènes communément sur-exprimés ou sous-exprimés dans les lignées reprogrammées par rapport aux fibroblastes. Ainsi, les annotations les plus enrichies par les 425 gènes surexprimés dans les lignées reprogrammées concernent la régulation du cycle cellulaire (Tableau 26 et Figure 47). Parmi les annotations suivantes, moins enrichies et moins significatives, se trouvent l'organisation chromosomique et chromatinienne, et plusieurs fonctions du développement embryonnaire : développement des systèmes respiratoires, neuronaux et reproductifs. Ce dernier point est particulièrement intéressant car il est associé à des fonctions de régulation du développement des organes et des cellules germinales dans lesquels de nombreux acteurs se retrouvent dans les cellules pluripotentes. Du côté des 609 gènes dont l'expression décroît avec la reprogrammation cellulaire, on trouve plusieurs gènes spécifiques de l'organisation de la matrice extracellulaire et de l'adhésion cellulaire, de la régulation du développement embryonnaire, de l'organogénèse, de la régulation de la prolifération cellulaire, et de voies plus spécifiques comme la biosynthèse et l'organisation du collagène fibrillaire, la régulation de la réponse aux stimuli externes (réponse inflammatoire, réparation des lésions) ou encore la contraction musculaire (Figure 48). D'une manière globale, les gènes différentiellement exprimés dans toutes les lignées reprogrammées par rapport aux fibroblastes sont caractéristiques des procédés biologiques de développement des organismes et sont impliqués dans les voies du cancer (Figure 49)

Les lignées dérivées de l'animal transloqué t(Y:14) par système de reprogrammation intégratif et non intégratif se différencient par l'expression de 1126 gènes annotés. Ces gènes appartiennent en grande partie aux voies du développement (différenciation, organogénèse), de l'organisation cellulaire (prolifération, migration, adhésion) et de la régulation de l'expression génique. Parmi les principales voies contenant ces gènes, on retrouve les voies de régulation du cancer, de l'adhésion focale, et de la diapédèse (Figure 50). Les voies de signalisations JAK/STAT, MAPK, WNT et TGF β caractéristiques des cellules souches sont également modulées entre la lignée NI12 et les lignées I3 et I4. Ainsi, l'expression du récepteur TGF β de type II est diminuée dans la lignée NI12 par rapport aux lignées I3 et I4 (et celle du récepteur de type I augmentée) tandis que celle des effecteurs de la voie WNT augmente, de même que celle du récepteur FGFR (Figure 51).

2. Comparaison des passages précoces et tardifs

Les profils d'expression des lignées I3 et I4 dépendent de leur nombre de passages en culture. En observant l'expression moyenne dans les 5136 gènes sélectionnés, on constate que le profil d'expression est plus proche d'abord en fonction du nombre de passage, puis en fonction de la lignée cellulaire (Figure 46C). Néanmoins, les lignées I3 et I4 n'évoluent pas tout à fait de la même manière. On retrouve pour la lignée 4 la diminution de *CDH1* et d'*OTX2* révélée par l'analyse en PCR

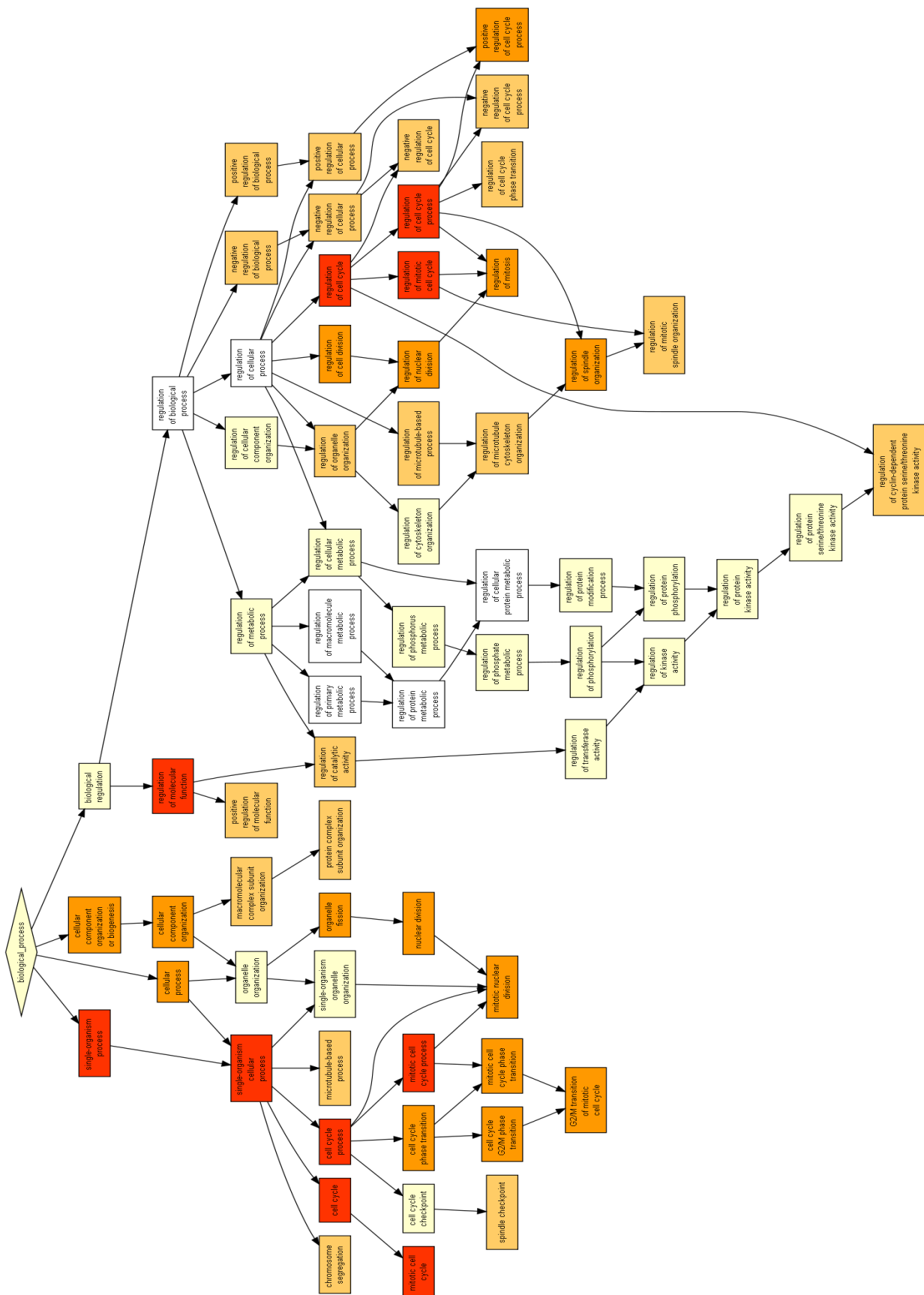


Figure 47 : Les quatre lignées reprogrammées I3, I4, PB20 et NI12 sur-expriment par rapport aux fibroblastes un ensemble de gènes impliqués dans les fonctions régulant le cycle cellulaire. Les couleurs représentent la p-valeur pour chaque procédé biologique (analyse GOrilla) : rouge $p < 10^{-9}$; orange foncé $p < 10^{-7}$; orange clair $p < 10^{-5}$; jaune $p < 10^{-3}$ et blanc $p > 10^{-3}$.

quantitative, tandis que la lignée I3 sur-régule *PECAMI*, et, de manière contradictoire aux données de PCR quantitative, *DAX1* et *LIN28B*.

De nombreuses sondes sont différenciellement exprimés entre les passages précoces et tardifs : 1983 pour la lignée I3 (soit 828 gènes annotés) et 1007 pour la lignée I4 (soit 458 gènes annotés) (Tableau 25). Parmi elles, 402 sont communes aux deux lignées. D'une manière générale, les fonctions biologiques impactées par ces gènes sont celles impliquées dans le cancer et le développement des organismes.

Sur les 388 gènes spécifiquement sur-exprimés dans la lignée I3 à passage tardif en comparaison avec la lignée I3 à passage précoce, 27 sont localisés sur le bras long du chromosome 4 parmi lesquels *WNT2B*. Ce chiffre élevé tend à confirmer la duplication de la région. Néanmoins, pour la lignée I4, seulement 5 gènes sur 180 sont retrouvés dans cette région génomique, dont 2 communs avec ceux de la lignée I3 : *FCGR1A* et *Clorf114*.

3. Profil d'expression des corps embryoides

Les données de transcriptome issues de la lignée NI12 et des corps embryoides formés par cette lignée révèlent dans un premier temps qu'un grand nombre de gènes, près de 1500, sont différenciellement exprimés entre ces deux conditions (Figure 46D). Elles nous permettent également de confirmer que l'expression des gènes *OCT4*, *DKK1*, *DPPA5*, *PECAMI*, *PRDM1*, *REX1* et *DPPA3* augmente lors de la formation des corps embryoides tandis que celle de *DAX1* et *NANOG* diminue.

Dans un second temps, l'analyse fonctionnelle préliminaire indique que les gènes différenciellement exprimés dans les corps embryoides sont impliqués dans l'adhésion, la communication et la prolifération cellulaire, la différenciation et l'organogénèse, la régulation de la transcription mais aussi dans la réponse aux stimuli et dans le métabolisme et le transport des lipides (Figure 52). Les voies de signalisation MAPK, JAK/STAT, TGF β , p53 et WNT sont elles aussi impactées. La voie JAK/STAT en particulier est fortement dérégulée, affectant l'expression de *NANOG* (Figure 53).

VI. Discussion et perspectives

1. Bilan moléculaire et fonctionnel

La dérivation de lignées de cellules iPS à partir de fibroblastes porcins s'est avérée être un défi plus complexe que ne le laissait présager les données bibliographiques. L'inefficacité du système de reprogrammation intégratif utilisé dans diverses conditions sur plusieurs lignées de cellules

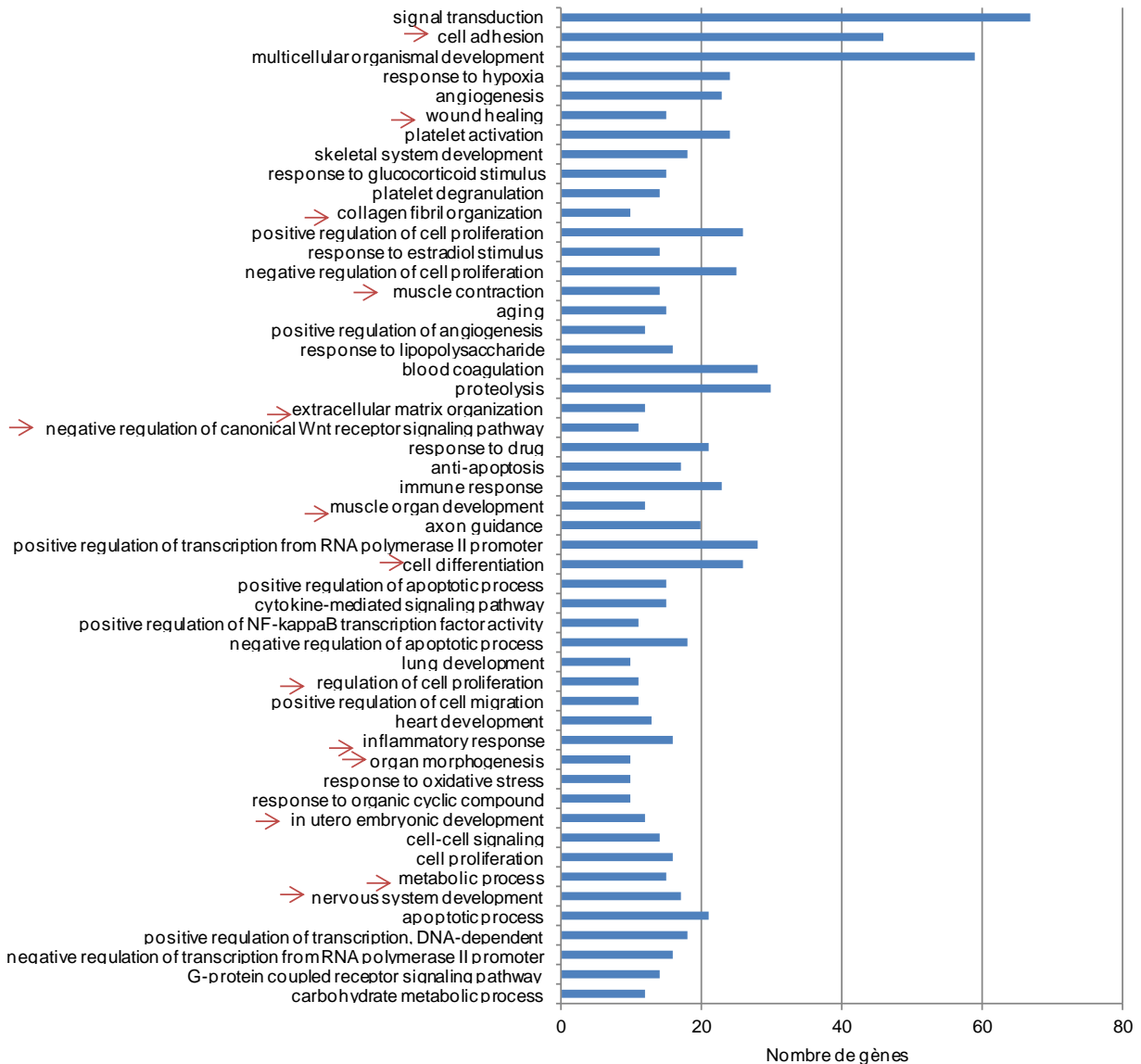


Figure 48 : Procédés biologiques (BP) les plus significativement enrichis ($FDR < 0.01$ et nombre de gènes >10) dans la liste de gènes communément sous-régulés dans les quatre lignées reprogrammées par rapport aux fibroblastes.

somatiques issues de tissus, races, phénotypes et caryotypes différents indique que la reprogrammation cellulaire porcine est un procédé encore mal maîtrisé. Néanmoins, la reprogrammation de lignées à caryotype anormal est à notre connaissance la première dans cette espèce, et nous suggère que la présence du remaniement t(Y;14) n'impacte pas le mécanisme de reprogrammation. Il est au contraire possible que cette translocation ait un effet positif, au vu du nombre de clones obtenus en comparaison avec d'autres lignées, bien que le remaniement soit équilibré. Une définition précise du point de cassure et l'analyse des gènes impliqués, ou l'analyse fine de l'organisation nucléaire en présence de ce remaniement, seraient à envisager pour dégager des pistes expliquant ce phénomène.

Les lignées obtenues possèdent plusieurs caractéristiques de la pluripotence : elles arborent une morphologie ressemblant à celles des lignées de cellules souches humaines, elles sont clonales, et possèdent un cycle cellulaire proche de celui de mES avec une phase G1 raccourcie et l'absence du point de contrôle G1/S. L'expression de nombreux gènes de pluripotence confirme leur potentiel sur le plan moléculaire. De plus, l'expression de la GFP dans les lignées I3 et I4 provenant de cellules somatiques porteuses de la cassette EOS contenant l'enhancer distal d'*OCT4* suggère l'utilisation de cet enhancer, qui est spécifiquement utilisé dans les lignées mESCs naïves tandis que l'expression d'*OCT4* dans les mEpiSCs est sous le contrôle de l'enhancer proximal (Tesar et al., 2007). Malgré cela, ces lignées rappellent les lignées ES humaines plutôt que les lignées ES murines par leur maintien en présence de bFGF, de sérum et de cellules nourricières. Elles possèdent cependant une grande plasticité leur permettant de s'adapter à une culture sans sérum ni cellules nourricières, en présence de LIF et d'inhibiteurs de *MEK* et *GSK3β*. L'expression de plusieurs gènes de pluripotence est alors modulée vers l'activation de marqueurs spécifiques de la pluripotence naïve, à l'exception de *CDH1* qui, bien qu'activé dans les deux états, est décrit comme plus fortement exprimé dans les lignées naïves (Bao et al., 2009).

Malgré ces caractéristiques moléculaires prometteuses, les lignées I3 et I4 possèdent un potentiel de différenciation limité. La lignée contrôle PB20, pour laquelle l'expression des gènes de pluripotence est plus marquée, est également réticente à la formation de corps embryoides en culture en suspension. De plus, aucune de ces trois lignées n'a conduit à la formation de tératomes matures après injection intramusculaire dans des souris immunodéficientes. Ces trois lignées expriment de manière continue les transgènes utilisés pour la reprogrammation, comme l'ensemble des lignées iPSCs établies à ce jour chez le porc. Bien que ces dernières semblent posséder une capacité de différenciation plus grande, avec la formation de corps embryoides et de tératomes, elles ne contribuent que faiblement au développement du blastocyste, et à ce jour seule une lignée a permis la formation de chimères avec transmission germinale (West et al., 2010).

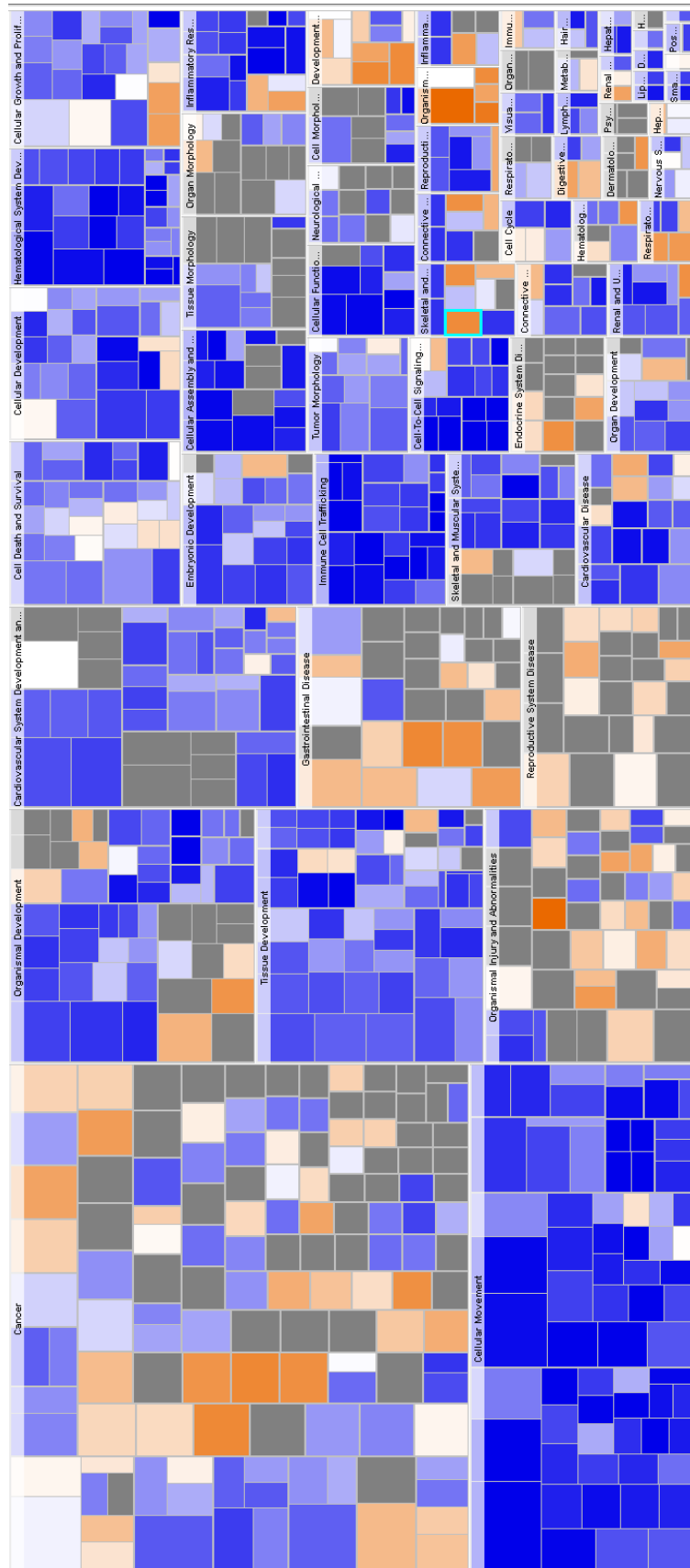


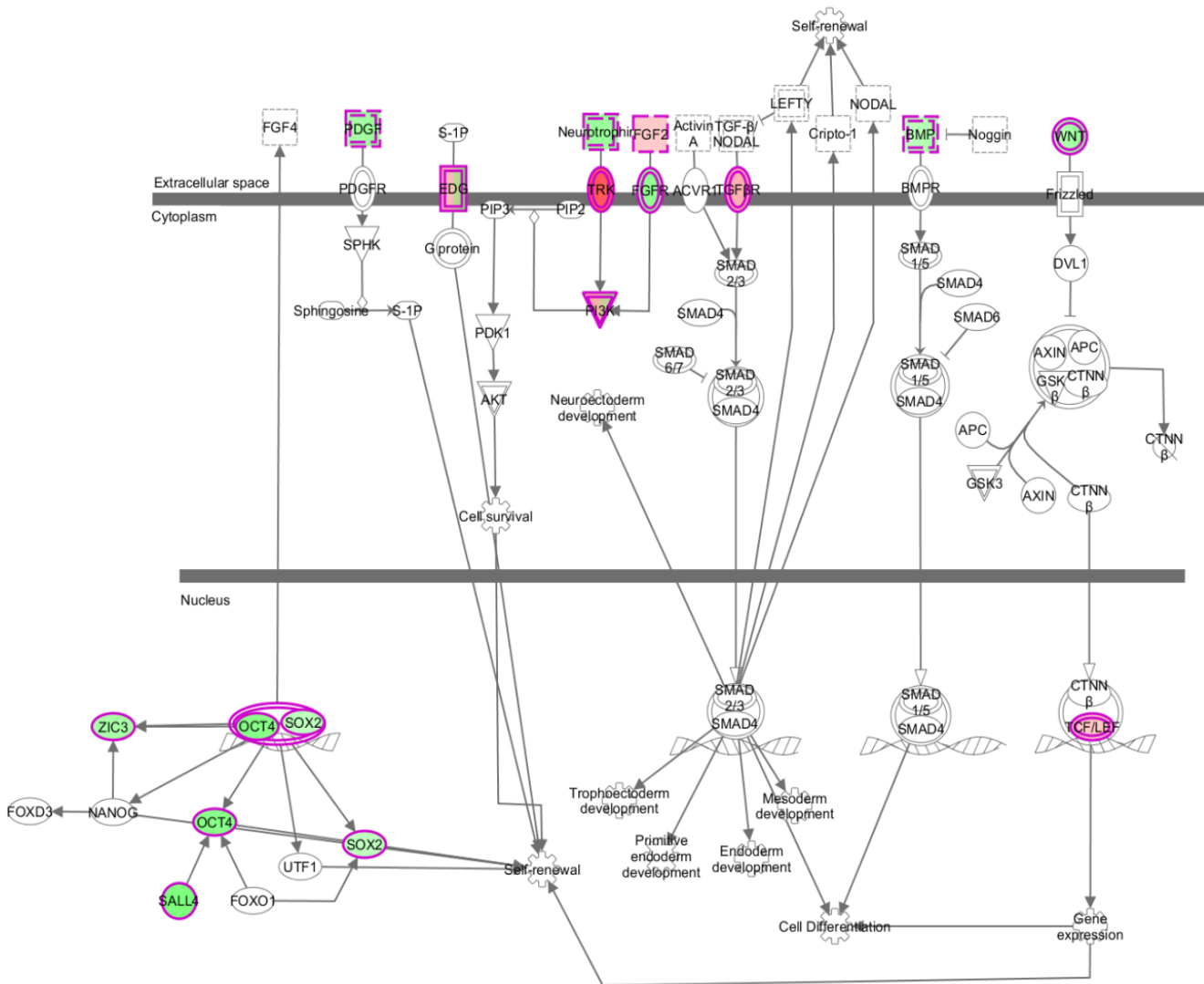
Figure 49 : Principales fonctions biologiques (Ingenuity Knowledge Database) représentées dans les gènes communément sur- ou sous-régulés dans les quatre lignées reprogrammées par rapport aux fibroblastes. La taille des carrés est fonction de la p-valeur pour chaque fonction et la couleur est fonction du z-score indiquant une tendance à l'activation (orange) ou à l'inactivation (bleu)

L'utilisation d'un système de reprogrammation non intégratif nous a permis d'obtenir des lignées possédant une morphologie ES-like et dont le profil d'expression indique une plus forte activation de nombreux marqueurs de la pluripotence et une diminution de l'expression des facteurs exogènes. Le transfert de ces lignées en milieu LIF+2i, l'étude de leur cycle cellulaire, ainsi que la formation de tératomes sont envisagés afin de les caractériser de manière plus approfondie. Cependant, la formation de corps embryoïdes ressemblant à des blastocystes porcins nous permet d'envisager une plus grande potentialité pour ces nouvelles lignées. En effet, outre leur morphologie, l'analyse transcriptomique révèle que ces corps embryoïdes réactivent les voies du métabolisme des lipides. Or, ces voies ont été décrites comme spécifiquement impliquées chez l'espèce porcine dans le développement du blastocyste, en plus des voies de signalisation classiques (Cao et al., 2014). Ces dernières, qui sont elles aussi modulées lors de la formation des corps embryoïdes, sont impliquées dans le développement embryonnaire précoce porcin mais les effecteurs et les récepteurs par lesquels ces voies sont régulées semblent être différents d'une espèce à l'autre (Cao et al., 2014; Hall and Hyttel, 2014). C'est également pour cette raison qu'il est difficile de définir pour les cellules pluripotentes porcines le statut de pluripotence naïf ou amorcé tel que décrit chez l'homme et la souris. Une étude transcriptomique réalisée sur plusieurs lignées piPSCs dérivées par plusieurs équipes de recherche a récemment révélé que les piPSCs sont plus proches de l'état amorcé décrit pour les hiPSCs et les mEpiSCs plutôt que de l'état naïf (Liu et al., 2013). Bien que de nombreux marqueurs de pluripotence décrits dans ces espèces soient spécifiquement surexprimés dans les lignées porcines, certains ne semblent pas entrer en jeu, laissant présager des différences dans le réseau de régulation de la pluripotence. C'est par exemple le cas des gènes de la famille KLF, notamment *KLF2*, et de *TBX3* qui ne sont pas exprimés dans les piPSCs (Liu et al., 2013) alors que l'expression de ce dernier gène augmente la capacité des miPSCs à former des chimères (Han et al., 2010). L'étude transcriptomique réalisée sur plusieurs lignées piPSCs dérivées à ce jour (Liu et al., 2013) indique que les lignées possédant le plus fort potentiel de développement *in vivo* sont celles exprimant *EPCAM*, qui pourrait donc constituer un nouveau marqueur permettant d'évaluer les piPSCs. Dans notre étude, ce gène apparaît surexprimé dans les lignées PB20 et NI12 par rapport aux PEFs, et son expression augmente au cours du temps dans la lignée I3.

En dehors d'une définition encore incomplète des gènes maintenant la pluripotence porcine, une autre hypothèse est que les lignées piPSCs n'aient pas achevé le processus de reprogrammation cellulaire et se trouvent dans un état déséquilibré et dépendant des transgènes. Le franchissement de la barrière épigénétique est une étape clé de la reprogrammation cellulaire, ainsi que l'a démontré l'expérience de Rais et collaborateurs (Rais et al., 2013) avec la déplétion de *MBD3* augmentant de façon spectaculaire l'efficacité de la reprogrammation cellulaire dans les cellules humaines. En effet, le complexe MBD3/NuRD est recruté par les facteurs de reprogrammation pour réprimer l'expression des gènes cibles de ces facteurs. Les cellules ES déficientes pour *Mbd3* chez la souris sont capables

Jak-STAT signaling pathway Natural killer cell mediated cytotoxicity Amoebiasis Protein digestion and absorption Toll-like receptor signaling pathway Chemokine signaling pathway MAPK signaling pathway Calcium signaling pathway Measles Axon guidance Regulation of actin cytoskeleton Focal adhesion Osteoclast differentiation Cytokine-cytokine receptor interaction Tuberculosis Arrhythmogenic right ventricular cardiomyopathy (ARVC) Dilated cardiomyopathy Pathways in cancer Endocytosis Leukocyte transendothelial migration Melanogenesis Neuroactive ligand-receptor interaction Hepatitis C ECM-receptor interaction Tight junction Hypertrophic cardiomyopathy (HCM) Neurotrophin signaling pathway TGF-beta signaling pathway Wnt signaling pathway Toxoplasmosis Cell adhesion molecules (CAMs)

Figure 50 : Représentation en nuage des 30 voies de signalisation (KEGG pathways) les plus significativement enrichies dans les gènes différentiellement exprimés entre la lignée NI12 et les lignées I3 et I4. La taille représente le nombre de gènes impliqués dans la voie.



© 2000-2014 QIAGEN. All rights reserved.

Figure 51 : Activation (vert) et inactivation (rouge) d'acteurs des voies de signalisation régulant la pluripotence des hESCs (Ingenuity Pathways Analysis) dans la lignée NI12 par rapport aux lignées I3 et I4.

d'auto-renouvellement mais pas de différenciation et le gène doit être réinséré pour récupérer la capacité de différenciation (Kaji et al., 2006; Rais et al., 2013). *MBD3*, tout comme *MBD2*, un autre membre du complexe NuRD dont l'expression limite l'efficacité de la reprogrammation, n'est pas spécifiquement sous-régulé dans les lignées iPS-like porcines que nous avons dérivées, ni dans celles d'autres équipes (Liu et al., 2013). La méthylation de l'ADN réprime l'expression de gènes de pluripotence dans les cellules différenciées (Borgel et al., 2010; Smallwood et al., 2011). Ces marques épigénétiques sont donc supprimées dans les cellules reprogrammées, y compris au niveau de *NANOG* et *OCT4*. Ces gènes étant réactivés dans nos lignées, il serait intéressant d'étudier leur profil de méthylation. Cependant, il serait également intéressant d'étudier la méthylation de l'ADN d'une manière plus globale, soit au niveau de l'ensemble du génome par RRBS ou MeDIP-Seq, soit au niveau de 36 gènes décrits comme hypométhylés à la fois dans les piPSCs et les mESCs, certains étant la cible des facteurs de reprogrammation (Arai et al., 2013). Une approche par immunoprécipitation de la chromatine (ChIP-Seq) pourrait également permettre d'identifier ou non l'apparition de marques bivalentes sur la chromatine, spécifiques des cellules ESCs et des cellules germinales, ou encore de déterminer le profil de déposition du variant d'histone H2A.X qui a récemment été décrit comme un marqueur clé de la véritable pluripotence, son absence entraînant la différenciation vers les tissus extraembryonnaires (Wu et al., 2014).

L'étude du locus soumis à empreinte parentale *DLK1/MEG3* est communément employé pour valider le profil épigénétique des lignées pluripotentes. En effet, ce locus possède dans de nombreuses lignées miPSCs une méthylation aberrante qui réprime l'expression des gènes (Stadtfeld et al., 2010), et qui est corrélée à la faible capacité de ces lignées à former des chimères avec transmission germinale. Chez le porc, l'expression des gènes de ce locus s'est également révélée réprimée (Liu et al., 2013), confirmant l'hypothèse d'altérations du profil de méthylation dans ces lignées en comparaison d'authentiques lignées pluripotentes, et expliquant la faible capacité des lignées porcines à former des chimères.

Enfin, l'obtention de structures ressemblant à des blastocystes mais ne différenciant que peu *in vitro* nous laisse présager que d'autres facteurs seraient nécessaires à l'activation des voies de différenciation dans les différents feuilletts embryonnaires. La faible capacité des lignées iPS porcines à contribuer au développement embryonnaire pourrait provenir des spécificités de développement embryonnaire précoce rencontrées dans cette espèce. Le blastocyste porcin entre plus précocement dans l'utérus, et se développe plus longuement au travers d'une phase d'élongation avant son implantation. De plus, la formation d'un disque embryonnaire avant l'implantation permet la mise en contact des cellules de l'épiblaste avec l'environnement intra-utérin. Les facteurs extrinsèques présents dans cet environnement ainsi que l'implantation du blastocyste pourraient constituer des éléments clés permettant l'engagement des cellules vers différentes lignées cellulaires. Il est intéressant de noter que la phase d'élongation est aussi présente au cours du développement embryonnaire bovin. Or, les

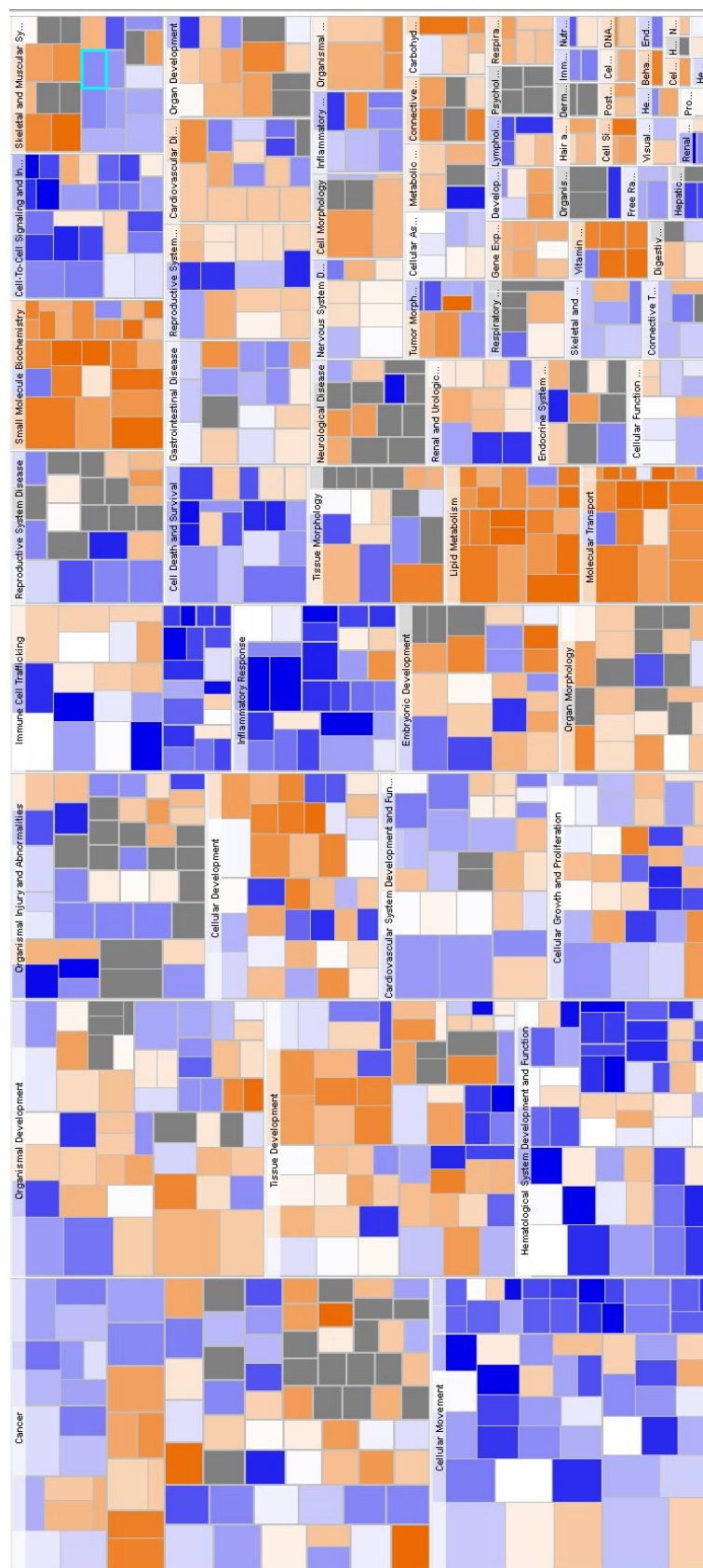


Figure 52 : Principales fonctions biologiques (Ingenuity Knowledge Database) représentées dans les gènes communément sur- ou sous-régulés dans les corps embryoides par rapport aux cellules NI12. La taille des carrés est fonction de la p-valeur pour chaque fonction et la couleur est fonction du z-score indiquant une tendance à l'activation (rouge) ou à l'inactivation (bleu)

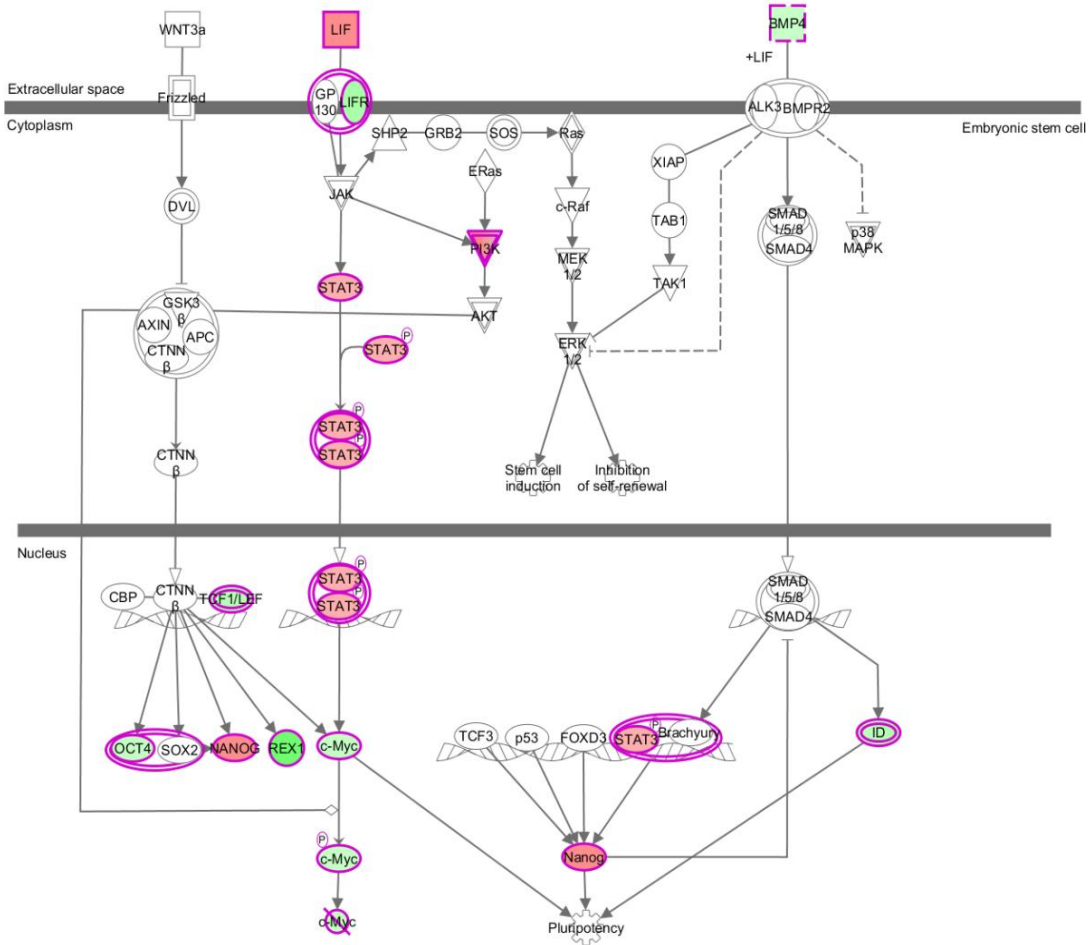
lignées ES et iPSCs capables de produire des chimères restent à ce jour difficiles à obtenir pour ces deux espèces. A l'inverse, le développement préimplantatoire est similaire chez la souris, l'homme, les primates et le lapin, espèces dans lesquelles les lignées ESCs sont plus facilement établies.

2. Bilan caryotypique

L'analyse cytogénétique détaillée des lignées I3 et I4 a révélé la forte instabilité du génome de ces lignées, caractérisée par une forte propension à accumuler des réarrangements chromosomiques et la présence de zones dupliquées sur le génome. Cette instabilité s'accroît au fil des passages, suggérant la sélection de populations cellulaires ayant un avantage prolifératif dans les conditions de cultures établies. Si les deux lignées présentent des caryotypes différents, il est toutefois intéressant de noter que la duplication identifiée sur le chromosome 4 porcin dans les deux lignées semble liée à la présence d'une anomalie commune, un ajout de fragment chromosomique sur la partie terminale du bras q du chromosome 8.

La comparaison avec les données humaines indique que cette région dupliquée a déjà été observée dans des lignées hiPS et pourrait donc présenter un avantage sélectif (Laurent et al., 2011). Une analyse plus fine de la région en utilisant plusieurs clones de BACs situés sur le bras long du chromosome 4 est envisagée pour définir au mieux les bordures de la zone dupliquée se transloquant sur le chromosome 8. Ces aberrations ne sont que faiblement présentes dans les passages précoces et n'ont pas été identifiées dans les lignées de fibroblastes, réduisant l'hypothèse d'une amplification clonale d'anomalies présentes de façon mosaïque dans les lignées de départ (Ronen and Benvenisty, 2012). La présence d'un remaniement t(Y;14) dans les fibroblastes ne semble pas non plus être à l'origine de cette accumulation d'anomalies par effet inter-chromosomique comme il a déjà été observé au cours du développement précoce (Alfarawati et al., 2012), puisque l'utilisation du système non intégratif sur les mêmes lignées porteuses de la translocation t(Y;14) n'induit pas d'instabilité chromosomique. De plus, la lignée contrôle PB20 révèle elle aussi une instabilité chromosomique s'installant au cours du temps, avec des remaniements différents.

Chez l'homme et la souris, l'instabilité génomique des lignées ES et iPS a également été constatée. Elle se manifeste par la présence d'anomalies chromosomiques souvent récurrentes qui peuvent concerner des gènes jouant un rôle dans le maintien de la pluripotente, ainsi que par une augmentation de la fréquence des variations du nombre de copies de gènes (Peterson and Loring, 2013). Cette instabilité est plus fortement observée dans les lignées iPS, indiquant un rôle de la reprogrammation cellulaire. Elle rappelle les phénomènes observés au cours des premières divisions de l'embryon, avant l'éclosion du blastocyste. Les cellules de l'embryon arborent alors de façon mosaïque un fort taux



© 2000-2014 QIAGEN. All rights reserved.

Figure 53 : Dérégulation de la voie LIF/JAK/STAT3 dans les corps embryoides de la lignée NI2 et activation de la voie BMP menant à la répression de NANOG (Ingenuity Pathway Analysis). vert : activation ; rouge : inactivation.

d'anomalies (Delhanty et al., 1993; Ambartsumyan and Clark, 2008; Mantzouratou and Delhanty, 2011). Ce taux diminue ensuite, la présence continue d'anomalies pouvant entraîner la perte spontanée de l'embryon ou, lorsque ces anomalies ne sont pas létales, des désordres génétiques. En dehors des aberrations provenant d'erreurs méiotiques au cours de la gamétogénèse, la forte fréquence des anomalies chromosomiques post-zygotiques peut s'expliquer par la permissivité du cycle cellulaire à ces stades, phénomène qui se retrouve dans les lignées cancéreuses connues pour leur capacité à accumuler des anomalies, ainsi que dans les cellules ES et iPS (Ambartsumyan and Clark, 2008). En effet, l'absence de réparation des cassures doubles-brins de l'ADN au cours du cycle cellulaire de ces cellules n'induit pas l'apoptose comme dans les cellules somatiques (Mantel et al., 2007). De plus, une étude sur les lignées ES et iPS murines a révélé leur grande plasticité, caractérisée par leur capacité à évoluer ponctuellement vers le stade de développement 2 cellules : au fil des passages, la majorité des cellules passe par cette étape transitoire s'accompagnant de la réactivation de rétrovirus endogènes spécifiques du stade zygote/2C connu pour sa haute propension à l'instabilité chromosomique (Macfarlan et al., 2012).

Enfin, la découverte récente du phénomène de chromothripsis ouvre de nouvelles perspectives sur l'origine des réarrangements de structures complexes survenant au cours du développement des lignées cancéreuses ou au cours de l'embryogénèse précoce (Stephens et al., 2011). Ce phénomène est caractérisé par un évènement génomique catastrophique unique menant à la génération de dizaines de réarrangements chromosomiques, anomalies structurales ou variations du nombre de copies (CNV), colocalisés sur un même chromosome ou sur un faible nombre de chromosomes. Le profil des CNVs est spécifique, avec une alternance entre conservation et perte d'hétérozygotie (Zhang et al., 2013a) (Figure 54). Une des origines les plus probables du mécanisme de chromothripsis est la présence d'un grand nombre de cassures de l'ADN, que l'on retrouve dans l'embryon pré-implantatoire suite à la décondensation du pronucléus paternel (Pellestor, 2014). Il serait donc intéressant de d'étudier les mécanismes de réparation d'ADN et d'apoptose dans les lignées piPS putatives, notamment le taux de recrutement de γ H2AX et des protéines de réparation comme RAD51 et la fréquence de l'apoptose.

Bien que l'hypothèse d'un rôle de la technique de reprogrammation, intégrative ou non, sur la stabilité des lignées iPS humaines ait été écartée par Mayshar et collaborateurs (Mayshar et al., 2010; Martins-Taylor and Xu, 2012), notre étude révèle qu'une lignée dérivée par système non intégratif possède un caryotype stable jusqu'au passage 45. La caractérisation d'autres lignées non intégratives et l'évaluation du caryotype à des passages plus élevés sont nécessaires pour confirmer ce résultat. Néanmoins, il est possible que si l'intégration virale en elle-même ne soit pas liée à l'apparition de CNVs, la dose de facteurs de reprogrammation employée soit déterminante (Chen et al., 2014). Or, les lignées iPS porcines décrites à ce jour ont la spécificité de dépendre des marqueurs exogènes de reprogrammation. Cette expression continue peut être à l'origine d'une reprogrammation cellulaire incomplète, menant à l'établissement de lignée iPS-like n'ayant pas acquis la pluripotence de manière

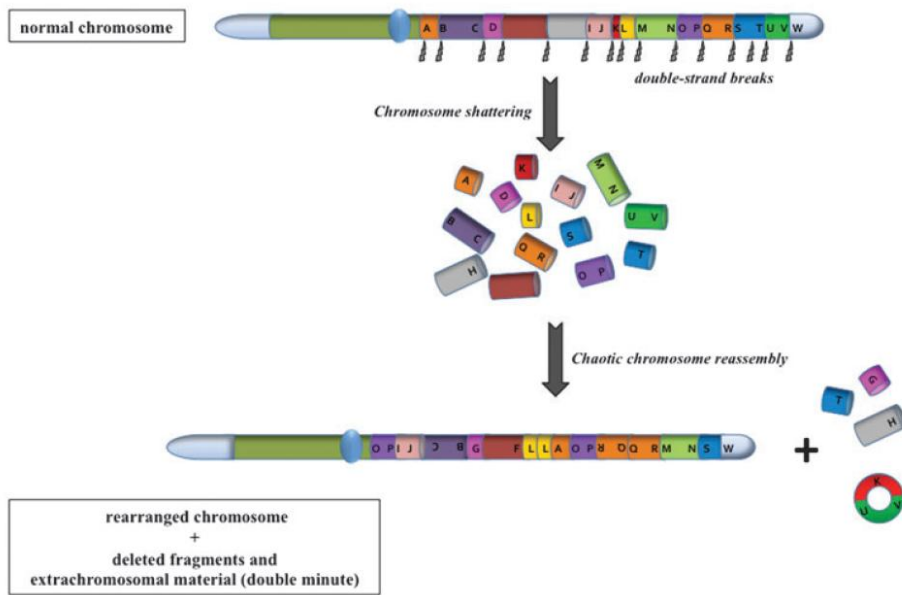


Figure 54 : Phénomène de chromothripsis : le chromosome est fragmenté et reconstruit au cours d'un même évènement catastrophique, entraînant la perte de matériel génétique et de nombreux réarrangements dans le chromosome reconstitué (Pellestor, 2014).

stable. L'utilisation d'un système non intégratif chez cette espèce pourrait représenter un atout majeur dans l'établissement de lignées plus stables, en garantissant une expression des facteurs de reprogrammation plus limitée au cours des passages et une meilleure réactivation du circuit endogène. De plus, l'expression de *TERT* est plus forte dans les lignées non intégratives ainsi que dans les lignées lorsqu'elles sont cultivées en milieu LIF+2i, suggérant une meilleure protection des télomères et donc une meilleure maintenance de l'intégrité du génome au cours de la réplication (King et al., 2006) dans le cadre d'une reprogrammation cellulaire plus aboutie. Il serait intéressant d'étudier le niveau d'expression de *ZSCAN4* dans les deux types de lignées, son expression ayant été reliée à la stabilité génomique des cellules ES (Zalzman et al., 2010). Enfin, plusieurs groupes ont décrit l'effet de la technique de dissociation enzymatique sur la fréquence des altérations caryotypiques dans les lignées hESCs quelque soit l'agent utilisé (trypsine, collagénase ou accutase) (Brimble et al., 2004; Mitalipova et al., 2005; Hasegawa et al., 2006). Néanmoins, l'accutase permet d'augmenter le taux de réadhérence lors du repiquage des cellules hESCs dont la faible clonogénicité limite la possibilité d'individualiser les cellules (Bajpai et al., 2008). De plus, les lignées passées mécaniquement exhibent aussi un nombre croissant de remaniements au fil des passages (Lefort et al., 2008; Hovatta et al., 2010). Dans notre étude, le même protocole de dissociation à l'accutase a été employé pour les lignées dérivées des systèmes intégratifs et non intégratifs, sans apparition d'anomalies dans la lignée NI13, suggérant un effet de la technique de reprogrammation ou du niveau de reprogrammation atteint plutôt que de la technique de dissociation.

3. Perspectives

Les lignées cellulaires obtenues ne répondent donc pas à tous les critères des lignées iPS murines. Néanmoins, l'utilisation d'un système de reprogrammation non intégratif laisse présager de l'obtention de lignées possédant une plus grande stabilité génomique et dont les corps embryoïdes possèdent une structure ressemblant à celle de blastocystes. L'analyse du méthylome de ces lignées permettra de déterminer si la barrière épigénétique de la reprogrammation cellulaire a bien été franchie, notamment au niveau des gènes régulateurs de la pluripotence. Associée à l'analyse détaillée du transcriptome, le profil de méthylation pourrait également nous permettre d'identifier des gènes candidats à l'établissement et au maintien de la pluripotence porcine. Le développement récent de la technologie CRISPR (Koike-Yusa et al., 2014; Shalem et al., 2014) qui permet des analyses fonctionnelles sans génération de mutants est un atout majeur pour les espèces d'élevage. Cette nouvelle technique d'édition du génome pourrait être mise à profit pour développer une stratégie de gain/perte de fonction permettant de valider le rôle des gènes candidats dans le maintien ou la différenciation des lignées NI. De plus, ce système pourrait être utilisé pour faciliter le suivi et la sélection des clones reprogrammés en permettant d'intégrer un gène rapporteur comme la GFP en aval du promoteur de *NANOG* par exemple. Avec la formation *in vitro* de corps embryoïdes ressemblant à des blastocystes mais sans

différenciation apparente, les capacités de différenciation des lignes NI dans les trois feuillets embryonnaires restent à valider. Outre la formation de tératomes, l'hypothèse d'un besoin de facteurs extrinsèques provenant de l'environnement intra-utérin pour la différenciation de ces structures pourrait être étudiée en cultivant ces corps dans un milieu contenant des protéines sécrétées dans l'utérus au cours de l'implantation ou en les implantant directement dans des truies pseudogestantes.

Dans le futur, des lignées dérivées de fibroblastes t(Y;14) ou de fibroblastes normaux par le système non intégratif pourront être cultivées dans un milieu stimulant la différenciation germinale. Pour cela, nous avons construit un vecteur rapporteur en couplant le promoteur du gène *DDX4* au gène exprimant la protéine fluorescente DsRed dans un vecteur pLKO.1 résistant à la puromycine. Une fois testé, par transfection dans des méiocytes ou des PGCs, ce rapporteur pourra être utilisé pour suivre, puis purifier, l'apparition de cellules de la lignée germinales au cours d'une culture en présence de BMP4 et d'acide rétinoïque qui orientent la différenciation des cellules souches vers la voie germinale.

CONCLUSION GENERALE

Les modèles de dérivation *in vitro* de cellules germinales à partir de cellules souches embryonnaires ou induites récemment développés chez la souris et l'homme représentent un outil majeur pour l'analyse de la biologie des cellules germinales. L'utilisation d'iPS autosomales de patients infertiles en particulier pourrait permettre une meilleure définition des mécanismes génétiques et épigénétiques à l'origine de cas d'infertilité encore inexpliqués. Chez le porc, l'infertilité est aussi une problématique récurrente pour les centres d'insémination artificielle. Au cours de ce projet de thèse, je me suis donc penchée sur l'étude de verrats infertiles, porteurs ou non de remaniements chromosomiques connus pour avoir un effet sur le déroulement de la méiose, en utilisant deux approches différentes. La première est l'étude de la méthylation de l'ADN spermatique issu de semence de mauvaise qualité, et la seconde concerne la dérivation de lignées pluripotentes induites porcines à partir de fibroblastes d'animaux infertiles et porteurs de remaniements chromosomiques, afin de pouvoir par la suite reconstituer le procédé de développement des cellules germinales chez ces animaux.

L'étude de l'ADN spermatique de verrats à qualité de semence dégradée m'a permis de mettre en évidence la forte conservation du profil de méthylation de l'ADN spermatique entre cette espèce et deux autres mammifères, l'homme et la souris. Ainsi, l'ADN spermatique porcine est fortement méthylé à l'exception de promoteurs et des îlots CpGs, et le niveau de global de méthylation ne varie pas dans le cas d'animaux infertiles. Une étude locale sur un sous-ensemble de gènes a révélé quelques différences entre les trois espèces (*POU5F1*, *RTL1*) chez les animaux normaux, ainsi qu'une augmentation de la méthylation dans le locus complexe *GNAS* pour quelques animaux infertiles, suggérant pour la première fois un rôle de ce locus dans l'établissement de gamètes fonctionnels. Il serait intéressant de mener une étude spécifique du locus *GNAS* chez des hommes à spermogramme anormal afin de valider ou d'infirmer ce nouveau marqueur de l'infertilité chez l'homme. Une analyse à l'échelle du génome nous permettrait d'identifier d'autres zones différentiellement méthylées entre espèces ou entre animaux fertiles et infertiles.

A notre connaissance, cette étude est la première à décrire la reprogrammation cellulaire d'individus porteurs de remaniements chromosomiques et infertiles chez le porc. Les lignées iPS-like produites m'ont permis de mettre en évidence les verrous de la reprogrammation cellulaire chez le porc. Ainsi, l'utilisation d'un système de reprogrammation intégratif a mené à la production de lignées avec un profil d'expression se rapprochant de celui des cellules pluripotentes, mais dont les capacités de différenciation sont limitées et qui sont caryotypiquement instables. L'expression continue des transgènes dans ces lignées est un élément majeur pouvant expliquer le fait que ces lignées ne soient pas complètement reprogrammées. L'utilisation d'un système non intégratif a quant à lui mené à la dérivation de lignées présentant elles aussi un profil d'expression proche de celui des cellules

pluripotentes, avec en particulier l'expression de plus de gènes de la pluripotence naïve. Ces lignées sont caryotypiquement stables, et sont capables de former des corps embryoïdes ressemblant à des blastocystes porcins. Des analyses supplémentaires sont nécessaires pour confirmer leur pluripotence, néanmoins ces lignées pourraient constituer un outil plus abouti pour l'étude des mécanismes régulant la pluripotence porcine au cours de l'embryogénèse précoce, qui restent à ce jour mal définis.

Afin d'étudier de manière plus approfondie l'évolution du profil de méthylation et d'expression au cours de la spermatogénèse, la purification de cellules germinales à différents stades (spermatogonies, spermatocytes et spermatides) par tri cellulaire est envisagée pour les animaux dont la spermatogénèse n'est pas arrêtée trop précocement. En effet, la culture de spermatogonies ou la dérivation cellules germinales embryonnaires (EG) nécessite l'accès à du matériel souvent indisponible ou de mauvaise qualité chez les animaux à qualité de semence dégradée. De plus, les spermatogonies sont peu stables en culture et ne nous permettraient pas d'étudier le développement précoce des PGCs. Or, les variations génétiques et épigénétiques se produisant à ce stade sont primordiales pour l'obtention de gamètes fonctionnels. Ainsi, bien que les lignées iPS-like produites ne permettent pour l'instant pas d'étudier la différenciation vers la voie germinale, elles restent un outil à fort potentiel pour l'étude des modifications de l'expression génique et de la méthylation de l'ADN dans les premiers stades du développement embryonnaire porcine.

Ce travail de thèse s'inscrit dans une dynamique de recherche portant sur la description des acteurs principaux de l'embryogénèse précoce et de l'établissement de fonctions reproductrices efficaces dans une espèce d'élevage d'intérêt agronomique majeur. A l'avenir, une caractérisation plus fine des mécanismes épigénétiques régulant de nombreux processus biologiques permettra d'identifier les éléments communs ou spécifiques de chaque espèce mammifère jouant un rôle dans le développement embryonnaire, et en particulier dans les voies de régulation de la pluripotence et de la différenciation germinale. Ces études garantiront une meilleure compréhension des voies permettant la transmission efficace de l'information d'une génération à l'autre.

REFERENCES

BIBLIOGRAPHIQUES

- Aasen, T., Raya, A., Barrero, M. J., Garreta, E., Consiglio, A., Gonzalez, F., Vassena, R., Bilić, J., Pekarik, V., Tiscornia, G. et al. (2008) 'Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes', *Nat Biotechnol* 26(11): 1276-84.
- Aladjem, M. I., Spike, B. T., Rodewald, L. W., Hope, T. J., Klemm, M., Jaenisch, R. and Wahl, G. M. (1998) 'ES cells do not activate p53-dependent stress responses and undergo p53-independent apoptosis in response to DNA damage', *Curr Biol* 8(3): 145-55.
- Alberio, R., Croxall, N. and Allegrucci, C. (2010) 'Pig epiblast stem cells depend on activin/nodal signaling for pluripotency and self-renewal', *Stem Cells Dev* 19(10): 1627-36.
- Alfarawati, S., Fragouli, E., Colls, P. and Wells, D. (2012) 'Embryos of robertsonian translocation carriers exhibit a mitotic interchromosomal effect that enhances genetic instability during early development', *PLoS Genet* 8(10): e1003025.
- Althouse, G.C. (1997) 'Evaluating porcine semen for artificial insemination. Part 1. Standard Test.', *Compendium on continuing education for the practicing veterinarian*. 19(3).
- Amarger, V., Nguyen, M., Van Laere, A. S., Braunschweig, M., Nezer, C., Georges, M. and Andersson, L. (2002) 'Comparative sequence analysis of the INS-IGF2-H19 gene cluster in pigs', *Mamm Genome* 13(7): 388-98.
- Ambartsumyan, G. and Clark, A. T. (2008) 'Aneuploidy and early human embryo development', *Hum Mol Genet* 17(R1): R10-5.
- An, J. Y., Kim, E. A., Jiang, Y., Zakrzewska, A., Kim, D. E., Lee, M. J., Mook-Jung, I., Zhang, Y. and Kwon, Y. T. (2010) 'UBR2 mediates transcriptional silencing during spermatogenesis via histone ubiquitination', *Proc Natl Acad Sci U S A* 107(5): 1912-7.
- Arai, Y., Ohgane, J., Fujishiro, S. H., Nakano, K., Matsunari, H., Watanabe, M., Umeyama, K., Azuma, D., Uchida, N., Sakamoto, N. et al. (2013) 'DNA methylation profiles provide a viable index for porcine pluripotent stem cells', *Genesis* 51(11): 763-76.
- Arnaud, P., Monk, D., Hitchins, M., Gordon, E., Dean, W., Beechey, C. V., Peters, J., Craigen, W., Preece, M., Stanier, P. et al. (2003) 'Conserved methylation imprints in the human and mouse GRB10 genes with divergent allelic expression suggests differential reading of the same mark', *Hum Mol Genet* 12(9): 1005-19.
- Bajpai, R., Lesperance, J., Kim, M. and Terskikh, A. V. (2008) 'Efficient propagation of single cells Accutase-dissociated human embryonic stem cells', *Mol Reprod Dev* 75(5): 818-27.
- Ball, S. T., Kelly, M. L., Robson, J. E., Turner, M. D., Harrison, J., Jones, L., Napper, D., Beechey, C. V., Hough, T., Plagge, A. et al. (2013) 'Gene Dosage Effects at the Imprinted Gnas Cluster', *PLoS One* 8(6): e65639.
- Bang, W. Y., Kim, S. W., Kwon, S. G., Hwang, J. H., Kim, T. W., Ko, M. S., Cho, I. C., Joo, Y. K., Cho, K. K., Jeong, J. Y. et al. (2013) 'Swine liver methylomes of Berkshire, Duroc and Landrace breeds by MeDIPS', *Anim Genet* 44(4): 463-6.
- Bao, S., Tang, F., Li, X., Hayashi, K., Gillich, A., Lao, K. and Surani, M. A. (2009) 'Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells', *Nature* 461(7268): 1292-5.
- Bastepe, M. (2007) 'The GNAS Locus: Quintessential Complex Gene Encoding Galpha, XLalphas, and other Imprinted Transcripts', *Curr Genomics* 8(6): 398-414.
- Beaujean, N., Hartshorne, G., Cavilla, J., Taylor, J., Gardner, J., Wilmut, I., Meehan, R. and Young, L. (2004) 'Non-conservation of mammalian preimplantation methylation dynamics', *Curr Biol* 14(7): R266-7.
- Beck, S. (2010) 'Taking the measure of the methylome', *Nat Biotechnol* 28(10): 1026-8.

- Bershteyn, M., Hayashi, Y., Desachy, G., Hsiao, E. C., Sami, S., Tsang, K. M., Weiss, L. A., Kriegstein, A. R., Yamanaka, S. and Wynshaw-Boris, A. (2014) 'Cell-autonomous correction of ring chromosomes in human induced pluripotent stem cells', *Nature* 507(7490): 99-103.
- Bibel, M., Richter, J., Schrenk, K., Tucker, K. L., Staiger, V., Korte, M., Goetz, M. and Barde, Y. A. (2004) 'Differentiation of mouse embryonic stem cells into a defined neuronal lineage', *Nat Neurosci* 7(9): 1003-9.
- Bieberich, Erhard and Wang, Guanghu (2013) Molecular Mechanisms Underlying Pluripotency. in Dr. Deepa Bhartiya (ed.) *Pluripotent Stem Cells*: InTech.
- Bischoff, S. R., Tsai, S., Hardison, N., Motsinger-Reif, A. A., Freking, B. A., Nonneman, D., Rohrer, G. and Piedrahita, J. A. (2009) 'Characterization of conserved and nonconserved imprinted genes in swine', *Biol Reprod* 81(5): 906-20.
- Blomberg, L. A. and Telugu, B. P. (2012) 'Twenty years of embryonic stem cell research in farm animals', *Reprod Domest Anim* 47 Suppl 4: 80-5.
- Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z. D., Ziller, M., Croft, G. F., Amoroso, M. W., Oakley, D. H. et al. (2011) 'Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines', *Cell* 144(3): 439-52.
- Bock, C., Tomazou, E. M., Brinkman, A. B., Müller, F., Simmer, F., Gu, H., Jäger, N., Gnirke, A., Stunnenberg, H. G. and Meissner, A. (2010) 'Quantitative comparison of genome-wide DNA methylation mapping technologies', *Nat Biotechnol* 28(10): 1106-14.
- Boheler, K. R., Czyn, J., Tweedie, D., Yang, H. T., Anisimov, S. V. and Wobus, A. M. (2002) 'Differentiation of pluripotent embryonic stem cells into cardiomyocytes', *Circ Res* 91(3): 189-201.
- Boissonnas, C. C., Abdalaoui, H. E., Haelewyn, V., Fauque, P., Dupont, J. M., Gut, I., Vaiman, D., Jouannet, P., Tost, J. and Jammes, H. (2010) 'Specific epigenetic alterations of IGF2-H19 locus in spermatozoa from infertile men', *Eur J Hum Genet* 18(1): 73-80.
- 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(12): 1093-100.
- Bourc'his, D. and Bestor, T. H. (2004) 'Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L', *Nature* 431(7004): 96-9.
- Boué, S., Paramonov, I., Barrero, M. J. and Izpisua Belmonte, J. C. (2010) 'Analysis of human and mouse reprogramming of somatic cells to induced pluripotent stem cells. What is in the plate?', *PLoS One* 5(9).
- Bowles, J. and Koopman, P. (2007) 'Retinoic acid, meiosis and germ cell fate in mammals', *Development* 134(19): 3401-11.
- Brambrink, T., Foreman, R., Welstead, G. G., Lengner, C. J., Wernig, M., Suh, H. and Jaenisch, R. (2008) 'Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells', *Cell Stem Cell* 2(2): 151-9.
- Braunschweig, M. H. (2012) 'Biallelic transcription of the porcine IGF2R gene', *Gene* 500(2): 181-5.
- Braunschweig, M. H., Owczarek-Lipska, M. and Stahlberger-Saitbekova, N. (2011) 'Relationship of porcine IGF2 imprinting status to DNA methylation at the H19 DMD and the IGF2 DMRs 1 and 2', *BMC Genet* 12: 47.
- Brenin, D., Look, J., Bader, M., Hübner, N., Levan, G. and Iannaccone, P. (1997) 'Rat embryonic stem cells: a progress report', *Transplant Proc* 29(3): 1761-5.
- Brevini, T. A., Antonini, S., Cillo, F., Crestan, M. and Gandolfi, F. (2007a) 'Porcine embryonic stem cells: Facts, challenges and hopes', *Theriogenology* 68 Suppl 1: S206-13.

- Brevini, T. A., Antonini, S., Pennarossa, G. and Gandolfi, F. (2008) 'Recent progress in embryonic stem cell research and its application in domestic species', *Reprod Domest Anim* 43 Suppl 2: 193-9.
- Brevini, T. A., Tosetti, V., Crestan, M., Antonini, S. and Gandolfi, F. (2007b) 'Derivation and characterization of pluripotent cell lines from pig embryos of different origins', *Theriogenology* 67(1): 54-63.
- Brimble, S. N., Zeng, X., Weiler, D. A., Luo, Y., Liu, Y., Lyons, I. G., Freed, W. J., Robins, A. J., Rao, M. S. and Schulz, T. C. (2004) 'Karyotypic stability, genotyping, differentiation, feeder-free maintenance, and gene expression sampling in three human embryonic stem cell lines derived prior to August 9, 2001', *Stem Cells Dev* 13(6): 585-97.
- Brons, I. G., Smithers, L. E., Trotter, M. W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S. M., Howlett, S. K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R. A. et al. (2007) 'Derivation of pluripotent epiblast stem cells from mammalian embryos', *Nature* 448(7150): 191-5.
- Buehr, M., Meek, S., Blair, K., Yang, J., Ure, J., Silva, J., McLay, R., Hall, J., Ying, Q. L. and Smith, A. (2008) 'Capture of authentic embryonic stem cells from rat blastocysts', *Cell* 135(7): 1287-98.
- Buganim, Y., Faddah, D. A., Cheng, A. W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S. L., van Oudenaarden, A. and Jaenisch, R. (2012) 'Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase', *Cell* 150(6): 1209-22.
- Buganim, Y., Faddah, D. A. and Jaenisch, R. (2013) 'Mechanisms and models of somatic cell reprogramming', *Nat Rev Genet* 14(6): 427-39.
- Buiting, K., Gross, S., Lich, C., Gillessen-Kaesbach, G., el-Maarri, O. and Horsthemke, B. (2003) 'Epimutations in Prader-Willi and Angelman syndromes: a molecular study of 136 patients with an imprinting defect', *Am J Hum Genet* 72(3): 571-7.
- Burdon, T., Smith, A. and Savatier, P. (2002) 'Signalling, cell cycle and pluripotency in embryonic stem cells', *Trends Cell Biol* 12(9): 432-8.
- Burgoyne, P. S., Mahadevaiah, S. K. and Turner, J. M. (2009) 'The consequences of asynapsis for mammalian meiosis', *Nat Rev Genet* 10(3): 207-16.
- Cao, S., Han, J., Wu, J., Li, Q., Liu, S., Zhang, W., Pei, Y., Ruan, X., Liu, Z., Wang, X. et al. (2014) 'Specific gene-regulation networks during the pre-implantation development of the pig embryo as revealed by deep sequencing', *BMC Genomics* 15(1): 4.
- Carey, B. W., Markoulaki, S., Hanna, J. H., Faddah, D. A., Buganim, Y., Kim, J., Ganz, K., Steine, E. J., Cassady, J. P., Creighton, M. P. et al. (2011) 'Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells', *Cell Stem Cell* 9(6): 588-98.
- Carey, B. W., Markoulaki, S., Hanna, J., Saha, K., Gao, Q., Mitalipova, M. and Jaenisch, R. (2009) 'Reprogramming of murine and human somatic cells using a single polycistronic vector', *Proc Natl Acad Sci U S A* 106(1): 157-62.
- Carmona-Saez, P., Chagoyen, M., Tirado, F., Carazo, J. M. and Pascual-Montano, A. (2007) 'GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists', *Genome Biol* 8(1): R3.
- Carrell, D. T., Emery, B. R. and Hammoud, S. (2007) 'Altered protamine expression and diminished spermatogenesis: what is the link?', *Hum Reprod Update* 13(3): 313-27.
- Carrell, D. T. and Hammoud, S. S. (2010) 'The human sperm epigenome and its potential role in embryonic development', *Mol Hum Reprod* 16(1): 37-47.
- Cattanach, B. M. and Kirk, M. (1985) 'Differential activity of maternally and paternally derived chromosome regions in mice', *Nature* 315(6019): 496-8.

- Chamberlain, S. J., Chen, P. F., Ng, K. Y., Bourgois-Rocha, F., Lemtiri-Chlieh, F., Levine, E. S. and Lalande, M. (2010) 'Induced pluripotent stem cell models of the genomic imprinting disorders Angelman and Prader-Willi syndromes', *Proc Natl Acad Sci U S A* 107(41): 17668-73.
- Chang, G., Gao, S., Hou, X., Xu, Z., Liu, Y., Kang, L., Tao, Y., Liu, W., Huang, B., Kou, X. et al. (2013) 'High-throughput sequencing reveals the disruption of methylation of imprinted gene in induced pluripotent stem cells', *Cell Res.*
- Chavez, L., Lienhard, M and Dietrich, J (2013) MEDIPS: (MeD)IP-seq data analysis.
- Chen, L. R., Shiue, Y. L., Bertolini, L., Medrano, J. F., BonDurant, R. H. and Anderson, G. B. (1999) 'Establishment of pluripotent cell lines from porcine preimplantation embryos', *Theriogenology* 52(2): 195-212.
- Chen, Y., Guo, L., Chen, J., Zhao, X., Zhou, W., Zhang, C., Wang, J., Jin, L., Pei, D. and Zhang, F. (2014) 'Genome-wide CNV analysis in mouse induced pluripotent stem cells reveals dosage effect of pluripotent factors on genome integrity', *BMC Genomics* 15: 79.
- Cheng, D., Guo, Y., Li, Z., Liu, Y., Gao, X., Gao, Y., Cheng, X., Hu, J. and Wang, H. (2012) 'Porcine induced pluripotent stem cells require LIF and maintain their developmental potential in early stage of embryos', *PLoS One* 7(12): e51778.
- Cheng, H. C., Zhang, F. W., Deng, C. Y., Jiang, C. D., Xiong, Y. Z., Li, F. E. and Lei, M. G. (2007) 'NNAT and DIRAS3 genes are paternally expressed in pigs', *Genet Sel Evol* 39(5): 599-607.
- 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(7262): 415-8.
- Clark, A. T., Bodnar, M. S., Fox, M., Rodriguez, R. T., Abeyta, M. J., Firpo, M. T. and Pera, R. A. (2004) 'Spontaneous differentiation of germ cells from human embryonic stem cells in vitro', *Hum Mol Genet* 13(7): 727-39.
- Coronado, D., Godet, M., Bourillot, P. Y., Taponnier, Y., Bernat, A., Petit, M., Afanassieff, M., Markossian, S., Malashicheva, A., Iacone, R. et al. (2013) 'A short G1 phase is an intrinsic determinant of naïve embryonic stem cell pluripotency', *Stem Cell Res* 10(1): 118-31.
- Court, F., Tayama, C., Romanelli, V., Martin Trujillo, A., Iglesias-Platas, I., Okamura, K., Sugahara, N., Simón, C., Moore, H., Harness, J. V. et al. (2014) 'Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of the human imprintome and suggests a germline methylation independent establishment of imprinting', *Genome Res.*
- Cremer, T. and Cremer, C. (2001) 'Chromosome territories, nuclear architecture and gene regulation in mammalian cells', *Nat Rev Genet* 2(4): 292-301.
- Daxinger, L. and Whitelaw, E. (2012) 'Understanding transgenerational epigenetic inheritance via the gametes in mammals', *Nat Rev Genet* 13(3): 153-62.
- De La Fuente, R., Baumann, C., Fan, T., Schmidtman, A., Dobrinski, I. and Muegge, K. (2006) 'Lsh is required for meiotic chromosome synapsis and retrotransposon silencing in female germ cells', *Nat Cell Biol* 8(12): 1448-54.
- De Los Angeles, A., Loh, Y. H., Tesar, P. J. and Daley, G. Q. (2012) 'Accessing naïve human pluripotency', *Curr Opin Genet Dev* 22(3): 272-82.
- Dean, W., Santos, F., Stojkovic, M., Zakhartchenko, V., Walter, J., Wolf, E. and Reik, W. (2001) 'Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos', *Proc Natl Acad Sci U S A* 98(24): 13734-8.
- Delhanty, J. D., Griffin, D. K., Handyside, A. H., Harper, J., Atkinson, G. H., Pieters, M. H. and Winston, R. M. (1993) 'Detection of aneuploidy and chromosomal mosaicism in human embryos during preimplantation sex determination by fluorescent in situ hybridisation, (FISH)', *Hum Mol Genet* 2(8): 1183-5.

- Ding, Y. Y., Liu, L. Y., Zhou, J., Zhang, X. D., Huang, L., Zhang, S. J. and Yin, Z. J. (2014) 'Expression and genomic imprinting of the porcine Rasgrf1 gene', *Gene* 536(2): 357-61.
- 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(7): 779-85.
- du Puy, Leonie.
- Ducos, A., Berland, H. M., Bonnet, N., Calgario, A., Billoux, S., Mary, N., Garnier-Bonnet, A., Darré, R. and Pinton, A. (2007) 'Chromosomal control of pig populations in France: 2002-2006 survey', *Genet Sel Evol* 39(5): 583-97.
- Ducos, A., Berland, H. M., Pinton, A., Guillemot, E., Seguela, A., Blanc, M. F., Darre, A. and Darre, R. (1998a) 'Nine new cases of reciprocal translocation in the domestic pig (*Sus scrofa domestica* L.)', *J Hered* 89(2): 136-42.
- Ducos, A., Pinton, A., Berland, H. M., Seguela, A., Blanc, M. F., Darre, A. and Darre, R. (1998b) 'Five new cases of reciprocal translocation in the domestic pig', *Hereditas* 128(3): 221-9.
- Dupont, J. M., Tost, J., Jammes, H. and Gut, I. G. (2004) 'De novo quantitative bisulfite sequencing using the pyrosequencing technology', *Anal Biochem* 333(1): 119-27.
- duPuy, Leonie, Lopes, Susana, Haagsman, Henk and Roelen, Bernard (2011) 'Analysis of co-expression of OCT4, NANOG and SOX2 in pluripotent cells of the porcine embryo, *in vivo* and *in vitro*', *Theriogenology* 75: 513-526.
- Eaton, S. A., Hough, T., Fischer-Colbrie, R. and Peters, J. (2013) 'Maternal inheritance of the Gnas cluster mutation Ex1A-T affects size, implicating NESP55 in growth', *Mamm Genome* 24(7-8): 276-85.
- Eden, E., Navon, R., Steinfeld, I., Lipson, D. and Yakhini, Z. (2009) 'GORilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists', *BMC Bioinformatics* 10: 48.
- Edwards, C. A., Mungall, A. J., Matthews, L., Ryder, E., Gray, D. J., Pask, A. J., Shaw, G., Graves, J. A., Rogers, J., Dunham, I. et al. (2008) 'The evolution of the DLK1-DIO3 imprinted domain in mammals', *PLoS Biol* 6(6): e135.
- Eggan, K., Akutsu, H., Loring, J., Jackson-Grusby, L., Klemm, M., Rideout, W. M., Yanagimachi, R. and Jaenisch, R. (2001) 'Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation', *Proc Natl Acad Sci U S A* 98(11): 6209-14.
- Eguizabal, C., Montserrat, N., Vassena, R., Barragan, M., Garreta, E., Garcia-Quevedo, L., Vidal, F., Giorgetti, A., Veiga, A. and Izpisua Belmonte, J. C. (2011) 'Complete meiosis from human induced pluripotent stem cells', *Stem Cells* 29(8): 1186-95.
- Ehrlich, M., Gama-Sosa, M. A., Huang, L. H., Midgett, R. M., Kuo, K. C., McCune, R. A. and Gehrke, C. (1982) 'Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells', *Nucleic Acids Res* 10(8): 2709-21.
- Erkek, S., Hisano, M., Liang, C. Y., Gill, M., Murr, R., Dieker, J., Schübeler, D., van der Vlag, J., Stadler, M. B. and Peters, A. H. (2013) 'Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa', *Nat Struct Mol Biol* 20(7): 868-75.
- Esteban, M. A., Peng, M., Deli, Z., Cai, J., Yang, J., Xu, J., Lai, L. and Pei, D. (2010a) 'Porcine induced pluripotent stem cells may bridge the gap between mouse and human iPS', *IUBMB Life* 62(4): 277-82.
- Esteban, M. A., Wang, T., Qin, B., Yang, J., Qin, D., Cai, J., Li, W., Weng, Z., Chen, J., Ni, S. et al. (2010b) 'Vitamin C enhances the generation of mouse and human induced pluripotent stem cells', *Cell Stem Cell* 6(1): 71-9.

- Esteban, M. A., Xu, J., Yang, J., Peng, M., Qin, D., Li, W., Jiang, Z., Chen, J., Deng, K., Zhong, M. et al. (2009) 'Generation of induced pluripotent stem cell lines from Tibetan miniature pig', *J Biol Chem* 284(26): 17634-40.
- Evans, M. J. and Kaufman, M. H. (1981) 'Establishment in culture of pluripotential cells from mouse embryos', *Nature* 292(5819): 154-6.
- Ezashi, T., Matsuyama, H., Telugu, B. P. and Roberts, R. M. (2011) 'Generation of colonies of induced trophoblast cells during standard reprogramming of porcine fibroblasts to induced pluripotent stem cells', *Biol Reprod* 85(4): 779-87.
- Ezashi, T., Telugu, B. P., Alexenko, A. P., Sachdev, S., Sinha, S. and Roberts, R. M. (2009) 'Derivation of induced pluripotent stem cells from pig somatic cells', *Proc Natl Acad Sci U S A* 106(27): 10993-8.
- Ezashi, T., Telugu, B. P. and Roberts, R. M. (2012) 'Induced pluripotent stem cells from pigs and other ungulate species: an alternative to embryonic stem cells?', *Reprod Domest Anim* 47 Suppl 4: 92-7.
- Feng, J., Liu, T. and Zhang, Y. (2011) 'Using MACS to identify peaks from ChIP-Seq data', *Curr Protoc Bioinformatics* Chapter 2: Unit 2.14.
- Ferchaud, S., Furtoss, V. and Guillouet, P. (2009) Evolution de la durée de carrière et de la production spermatique des verrats de CIA, étude comparative 1991-1997 et 2001-2007. . in INRA IFIP (ed.) *Journées de la Recherche Porcine*. Paris.
- Ficz, G., Branco, M. R., Seisenberger, S., Santos, F., Krueger, F., Hore, T. A., Marques, C. J., Andrews, S. and Reik, W. (2011) 'Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation', *Nature* 473(7347): 398-402.
- Ford, C. E., Pollock, D. L. and Gustavsson, I. (1980) 'Proceedings of the First International Conference for the Standardisation of Banded Karyotypes of Domestic Animals. University of Reading Reading, England. 2nd-6th August 1976', *Hereditas* 92(1): 145-62.
- Fournier, A., McLeer-Florin, A., Lefebvre, C., Duley, S., Barki, L., Ribeyron, J., Alboukadel, K., Hamaidia, S., Granjon, A., Gressin, R. et al. (2010) '1q12 chromosome translocations form aberrant heterochromatic foci associated with changes in nuclear architecture and gene expression in B cell lymphoma', *EMBO Mol Med* 2(5): 159-71.
- Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L. and Paul, C. L. (1992) 'A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands', *Proc Natl Acad Sci U S A* 89(5): 1827-31.
- Fröhlich, L. F., Mrakovcic, M., Steinborn, R., Chung, U. I., Bastepe, M. and Jüppner, H. (2010) 'Targeted deletion of the Nesp55 DMR defines another Gnas imprinting control region and provides a mouse model of autosomal dominant PHP-Ib', *Proc Natl Acad Sci U S A* 107(20): 9275-80.
- Fujishiro, S. H., Nakano, K., Mizukami, Y., Azami, T., Arai, Y., Matsunari, H., Ishino, R., Nishimura, T., Watanabe, M., Abe, T. et al. (2012) 'Generation of Naive-Like Porcine-Induced Pluripotent Stem Cells Capable of Contributing to Embryonic and Fetal Development', *Stem Cells Dev*.
- Fulka, H., Mrazek, M., Tepla, O. and Fulka, J. (2004) 'DNA methylation pattern in human zygotes and developing embryos', *Reproduction* 128(6): 703-8.
- Fulka, J., Fulka, H., Slavik, T. and Okada, K. (2006) 'DNA methylation pattern in pig in vivo produced embryos', *Histochem Cell Biol* 126(2): 213-7.
- Fusaki, N., Ban, H., Nishiyama, A., Saeki, K. and Hasegawa, M. (2009) 'Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome', *Proc Jpn Acad Ser B Phys Biol Sci* 85(8): 348-62.
- Gafni, O., Weinberger, L., Mansour, A. A., Manor, Y. S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A. et al. (2013) 'Derivation of novel human ground state naive pluripotent stem cells', *Nature*.

- Geijsen, N., Horoschak, M., Kim, K., Gribnau, J., Eggan, K. and Daley, G. Q. (2004) 'Derivation of embryonic germ cells and male gametes from embryonic stem cells', *Nature* 427(6970): 148-54.
- Glaser, S., Lubitz, S., Loveland, K. L., Ohbo, K., Robb, L., Schwenk, F., Seibler, J., Roellig, D., Kranz, A., Anastassiadis, K. et al. (2009) 'The histone 3 lysine 4 methyltransferase, Mll2, is only required briefly in development and spermatogenesis', *Epigenetics Chromatin* 2(1): 5.
- Gonzalez, F., Barragan Monasterio, M., Tiscornia, G., Montserrat Pulido, N., Vassena, R., Batlle Morera, L., Rodriguez Piza, I. and Izpisua Belmonte, J. C. (2009) 'Generation of mouse-induced pluripotent stem cells by transient expression of a single nonviral polycistronic vector', *Proc Natl Acad Sci U S A* 106(22): 8918-22.
- 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(5992): 643-8.
- Grootegeod, J. A., Siep, M. and Baarends, W. M. (2000) 'Molecular and cellular mechanisms in spermatogenesis', *Baillieres Best Pract Res Clin Endocrinol Metab* 14(3): 331-43.
- 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(7366): 606-10.
- Gurdon, J. B. (1975) 'Nuclear transplantation and the analysis of gene activity in early amphibian development', *Adv Exp Med Biol* 62: 35-44.
- Gustavsson, I. (1988) 'Standard karyotype of the domestic pig. Committee for the Standardized Karyotype of the Domestic Pig', *Hereditas* 109(2): 151-7.
- Hackett, J. A., Zyllicz, J. J. and Surani, M. A. (2012) 'Parallel mechanisms of epigenetic reprogramming in the germline', *Trends Genet* 28(4): 164-74.
- Hackett, Jamie A., Sengupta, Roopsha, Zyllicz, Jan J., Murakami, Kazuhiro, Lee, Caroline, Down, Thomas A. and Surani, M. Azim (2013) 'Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine', *Science (New York, N.Y.)* 339(6118): 448-52.
- Hajkova, P., Ancelin, K., Waldmann, T., Lacoste, N., Lange, U. C., Cesari, F., Lee, C., Almouzni, G., Schneider, R. and Surani, M. A. (2008) 'Chromatin dynamics during epigenetic reprogramming in the mouse germ line', *Nature* 452(7189): 877-81.
- Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J. and Surani, M. A. (2002) 'Epigenetic reprogramming in mouse primordial germ cells', *Mech Dev* 117(1-2): 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(5987): 78-82.
- Hall, V. J., Christensen, J., Gao, Y., Schmidt, M. H. and Hyttel, P. (2009) 'Porcine pluripotency cell signaling develops from the inner cell mass to the epiblast during early development', *Dev Dyn* 238(8): 2014-24.
- Hall, V. J. and Hyttel, P. (2014) 'Breaking down pluripotency in the porcine embryo reveals both a premature and reticent stem cell state in the inner cell mass and unique expression profiles of the naive and primed stem cell States', *Stem Cells Dev* 23(17): 2030-45.
- Hall, V. J., Jacobsen, J. V., Rasmussen, M. A. and Hyttel, P. (2010) 'Ultrastructural and molecular distinctions between the porcine inner cell mass and epiblast reveal unique pluripotent cell states', *Dev Dyn* 239(11): 2911-20.
- Hall, V. J., Kristensen, M., Rasmussen, M. A., Ujhelly, O., Dinnyés, A. and Hyttel, P. (2012) 'Temporal repression of endogenous pluripotency genes during reprogramming of porcine induced pluripotent stem cells', *Cell Reprogram* 14(3): 204-16.
- Hammoud, S. S., Low, D. H., Yi, C., Carrell, D. T., Guccione, E. and Cairns, B. R. (2014) 'Chromatin and Transcription Transitions of Mammalian Adult Germline Stem Cells and Spermatogenesis', *Cell Stem Cell*.

- Hammoud, S. S., Nix, D. A., Zhang, H., Purwar, J., Carrell, D. T. and Cairns, B. R. (2009) 'Distinctive chromatin in human sperm packages genes for embryo development', *Nature* 460(7254): 473-8.
- Hammoud, S. S., Purwar, J., Pflueger, C., Cairns, B. R. and Carrell, D. T. (2010) 'Alterations in sperm DNA methylation patterns at imprinted loci in two classes of infertility', *Fertil Steril* 94(5): 1728-33.
- Han, J., Yuan, P., Yang, H., Zhang, J., Soh, B. S., Li, P., Lim, S. L., Cao, S., Tay, J., Orlov, Y. L. et al. (2010) 'Tbx3 improves the germ-line competency of induced pluripotent stem cells', *Nature* 463(7284): 1096-100.
- Handel, M. A. and Schimenti, J. C. (2010) 'Genetics of mammalian meiosis: regulation, dynamics and impact on fertility', *Nat Rev Genet* 11(2): 124-36.
- Hanna, J. H. (2010) 'The STATs on naive iPSC reprogramming', *Cell Stem Cell* 7(3): 274-6.
- Hanna, J. H., Saha, K. and Jaenisch, R. (2010) 'Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues', *Cell* 143(4): 508-25.
- Hanna, J., Markoulaki, S., Schorderet, P., Carey, B. W., Beard, C., Wernig, M., Creighton, M. P., Steine, E. J., Cassady, J. P., Foreman, R. et al. (2008) 'Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency', *Cell* 133(2): 250-64.
- Hansis, C., Grifo, J. A. and Krey, L. C. (2000) 'Oct-4 expression in inner cell mass and trophectoderm of human blastocysts', *Mol Hum Reprod* 6(11): 999-1004.
- Harris, R. A., Wang, T., Coarfa, C., Nagarajan, R. P., Hong, C., Downey, S. L., Johnson, B. E., Fouse, S. D., Delaney, A., Zhao, Y. et al. (2010) 'Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications', *Nat Biotechnol* 28(10): 1097-105.
- Hartmann, S., Bergmann, M., Bohle, R. M., Weidner, W. and Steger, K. (2006) 'Genetic imprinting during impaired spermatogenesis', *Mol Hum Reprod* 12(6): 407-11.
- Hasegawa, K., Fujioka, T., Nakamura, Y., Nakatsuji, N. and Suemori, H. (2006) 'A method for the selection of human embryonic stem cell sublines with high replating efficiency after single-cell dissociation', *Stem Cells* 24(12): 2649-60.
- Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. and Saitou, M. (2011) 'Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells', *Cell* 146(4): 519-32.
- Hayashi, K., Yoshida, K. and Matsui, Y. (2005) 'A histone H3 methyltransferase controls epigenetic events required for meiotic prophase', *Nature* 438(7066): 374-8.
- Henikoff, S. (1990) 'Position-effect variegation after 60 years', *Trends Genet* 6(12): 422-6.
- Hotta, A., Cheung, A. Y., Farra, N., Garcha, K., Chang, W. Y., Pasceri, P., Stanford, W. L. and Ellis, J. (2009a) 'EOS lentiviral vector selection system for human induced pluripotent stem cells', *Nat Protoc* 4(12): 1828-44.
- Hotta, A., Cheung, A. Y., Farra, N., Vijayaragavan, K., Séguin, C. A., Draper, J. S., Pasceri, P., Maksakova, I. A., Mager, D. L., Rossant, J. et al. (2009b) 'Isolation of human iPSCs using EOS lentiviral vectors to select for pluripotency', *Nat Methods* 6(5): 370-6.
- Houshdaran, S., Cortessis, V. K., Siegmund, K., Yang, A., Laird, P. W. and Sokol, R. Z. (2007) 'Widespread epigenetic abnormalities suggest a broad DNA methylation erasure defect in abnormal human sperm', *PLoS One* 2(12): e1289.
- Hovatta, O., Jaconi, M., Tökönen, V., Béna, F., Gimelli, S., Bosman, A., Holm, F., Wyder, S., Zdobnov, E. M., Irion, O. et al. (2010) 'A teratocarcinoma-like human embryonic stem cell (hESC) line and four hESC lines reveal potentially oncogenic genomic changes', *PLoS One* 5(4): e10263.
- Hu, Y., Xu, H., Li, Z., Zheng, X., Jia, X., Nie, Q. and Zhang, X. (2013) 'Comparison of the genome-wide DNA methylation profiles between fast-growing and slow-growing broilers', *PLoS One* 8(2): e56411.

- Huangfu, D., Maehr, R., Guo, W., Eijkelenboom, A., Snitow, M., Chen, A. E. and Melton, D. A. (2008a) 'Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds', *Nat Biotechnol* 26(7): 795-7.
- Huangfu, D., Osafune, K., Maehr, R., Guo, W., Eijkelenboom, A., Chen, S., Muhlestein, W. and Melton, D. A. (2008b) 'Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2', *Nat Biotechnol* 26(11): 1269-75.
- Hussein, S. M., Batada, N. N., Vuoristo, S., Ching, R. W., Autio, R., Närvä, E., Ng, S., Sourour, M., Hämmäläinen, R., Olsson, C. et al. (2011) 'Copy number variation and selection during reprogramming to pluripotency', *Nature* 471(7336): 58-62.
- Hyldig, S. M., Croxall, N., Contreras, D. A., Thomsen, P. D. and Alberio, R. (2011a) 'Epigenetic reprogramming in the porcine germ line', *BMC Dev Biol* 11: 11.
- Hyldig, S. M., Ostrup, O., Vejlsted, M. and Thomsen, P. D. (2011b) 'Changes of DNA methylation level and spatial arrangement of primordial germ cells in embryonic day 15 to embryonic day 28 pig embryos', *Biol Reprod* 84(6): 1087-93.
- Ioannou, D. and Griffin, D. K. (2011) 'Male fertility, chromosome abnormalities, and nuclear organization', *Cytogenet Genome Res* 133(2-4): 269-79.
- Iqbal, K., Jin, S. G., Pfeifer, G. P. and Szabó, P. E. (2011) 'Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine', *Proc Natl Acad Sci U S A* 108(9): 3642-7.
- Jena, S. C., Kumar, S., Rajput, S., Roy, B., Verma, A., Kumaresan, A., Mohanty, T. K., De, S., Kumar, R. and Datta, T. K. (2014) 'Differential methylation status of IGF2-H19 locus does not affect the fertility of crossbred bulls but some of the CTCF binding sites could be potentially important', *Mol Reprod Dev*.
- Ji, J., Ng, S. H., Sharma, V., Neculai, D., Hussein, S., Sam, M., Trinh, Q., Church, G. M., McPherson, J. D., Nagy, A. et al. (2012) 'Elevated coding mutation rate during the reprogramming of human somatic cells into induced pluripotent stem cells', *Stem Cells* 30(3): 435-40.
- Kagiwada, Saya, Kurimoto, Kazuki, Hirota, Takayuki, Yamaji, Masashi and Saitou, Mitinori (2013) 'Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice', *The EMBO journal* 32(3): 340-53.
- Kaji, K., Caballero, I. M., MacLeod, R., Nichols, J., Wilson, V. A. and Hendrich, B. (2006) 'The NuRD component Mbd3 is required for pluripotency of embryonic stem cells', *Nat Cell Biol* 8(3): 285-92.
- Kaji, K., Norrby, K., Paca, A., Mileikovsky, M., Mohseni, P. and Woltjen, K. (2009) 'Virus-free induction of pluripotency and subsequent excision of reprogramming factors', *Nature* 458(7239): 771-5.
- Kang, L. and Gao, S. (2012) 'Pluripotency of induced pluripotent stem cells', *J Anim Sci Biotechnol* 3(1): 5.
- Karimi, M., Johansson, S. and Ekström, T. J. (2006a) 'Using LUMA: a Luminometric-based assay for global DNA-methylation', *Epigenetics* 1(1): 45-8.
- Karimi, M., Johansson, S., Stach, D., Corcoran, M., Grandér, D., Schalling, M., Bakalkin, G., Lyko, F., Larsson, C. and Ekström, T. J. (2006b) 'LUMA (LUMinometric Methylation Assay)--a high throughput method to the analysis of genomic DNA methylation', *Exp Cell Res* 312(11): 1989-95.
- Kee, K., Angeles, V. T., Flores, M., Nguyen, H. N. and Reijo Pera, R. A. (2009) 'Human DAZL, DAZ and BOULE genes modulate primordial germ-cell and haploid gamete formation', *Nature* 462(7270): 222-5.
- Kee, K., Gonsalves, J. M., Clark, A. T. and Pera, R. A. (2006) 'Bone morphogenetic proteins induce germ cell differentiation from human embryonic stem cells', *Stem Cells Dev* 15(6): 831-7.

- 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(2): 129-38.
- Kerjean, A., Dupont, J. M., Vasseur, C., Le Tessier, D., Cuisset, L., Paldi, A., Jouannet, P. and Jeanpierre, M. (2000) 'Establishment of the paternal methylation imprint of the human H19 and MEST/PEG1 genes during spermatogenesis', *Hum Mol Genet* 9(14): 2183-7.
- Kerkel, K., Schupf, N., Hatta, K., Pang, D., Salas, M., Kratz, A., Minden, M., Murty, V., Zigman, W. B., Mayeux, R. P. et al. (2010) 'Altered DNA methylation in leukocytes with trisomy 21', *PLoS Genet* 6(11): e1001212.
- Kim, J. B., Sebastiano, V., Wu, G., Araúzo-Bravo, M. J., Sasse, P., Gentile, L., Ko, K., Ruau, D., Ehrlich, M., van den Boom, D. et al. (2009) 'Oct4-induced pluripotency in adult neural stem cells', *Cell* 136(3): 411-9.
- Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M. J., Ji, H., Ehrlich, L. I. et al. (2010) 'Epigenetic memory in induced pluripotent stem cells', *Nature* 467(7313): 285-90.
- King, W. A., Coppola, G., Alexander, B., Mastromonaco, G., Perrault, S., Nino-Soto, M. I., Pinton, A., Joudrey, E. M. and Betts, D. H. (2006) 'The impact of chromosomal alteration on embryo development', *Theriogenology* 65(1): 166-77.
- Knoll, J. H., Nicholls, R. D., Magenis, R. E., Graham, J. M., Lalonde, M. and Latt, S. A. (1989) 'Angelman and Prader-Willi syndromes share a common chromosome 15 deletion but differ in parental origin of the deletion', *Am J Med Genet* 32(2): 285-90.
- Kobayashi, H., Hiura, H., John, R. M., Sato, A., Otsu, E., Kobayashi, N., Suzuki, R., Suzuki, F., Hayashi, C., Utsunomiya, T. et al. (2009) 'DNA methylation errors at imprinted loci after assisted conception originate in the parental sperm', *Eur J Hum Genet* 17(12): 1582-91.
- Kobayashi, H. and Kono, T. (2012) 'DNA methylation analysis of germ cells by using bisulfite-based sequencing methods', *Methods Mol Biol* 825: 223-35.
- Kobayashi, H., Sakurai, T., Imai, M., Takahashi, N., Fukuda, A., Yayoi, O., Sato, S., Nakabayashi, K., Hata, K., Sotomaru, Y. et al. (2012) 'Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks', *PLoS Genet* 8(1): e1002440.
- Kobayashi, H., Sato, A., Otsu, E., Hiura, H., Tomatsu, C., Utsunomiya, T., Sasaki, H., Yaegashi, N. and Arima, T. (2007) 'Aberrant DNA methylation of imprinted loci in sperm from oligospermic patients', *Hum Mol Genet* 16(21): 2542-51.
- Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera, M. el C and Yusa, K. (2014) 'Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library', *Nat Biotechnol* 32(3): 267-73.
- Kota, S. K. and Feil, R. (2010) 'Epigenetic transitions in germ cell development and meiosis', *Dev Cell* 19(5): 675-86.
- Kues, W. A., Herrmann, D., Barg-Kues, B., Haridoss, S., Nowak-Imialek, M., Buchholz, T., Streeck, M., Grebe, A., Grabundzija, I., Merkert, S. et al. (2013) 'Derivation and characterization of sleeping beauty transposon-mediated porcine induced pluripotent stem cells', *Stem Cells Dev* 22(1): 124-35.
- Kwon, D. J., Jeon, H., Oh, K. B., Ock, S. A., Im, G. S., Lee, S. S., Im, S. K., Lee, J. W., Oh, S. J., Park, J. K. et al. (2013) 'Generation of leukemia inhibitory factor-dependent induced pluripotent stem cells from the massachusetts general hospital miniature pig', *Biomed Res Int* 2013: 140639.
- Laird, P. W. (2010) 'Principles and challenges of genomewide DNA methylation analysis', *Nat Rev Genet* 11(3): 191-203.

- Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., Walter, J. and Reik, W. (2003) 'Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse', *Genesis* 35(2): 88-93.
- Laurent, L. C., Ulitsky, I., Slavin, I., Tran, H., Schork, A., Morey, R., Lynch, C., Harness, J. V., Lee, S., Barrero, M. J. et al. (2011) 'Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture', *Cell Stem Cell* 8(1): 106-18.
- Law, J. A. and Jacobsen, S. E. (2010) 'Establishing, maintaining and modifying DNA methylation patterns in plants and animals', *Nat Rev Genet* 11(3): 204-20.
- Lefort, N., Feyeux, M., Bas, C., Féraud, O., Bennaceur-Griscelli, A., Tachdjian, G., Peschanski, M. and Perrier, A. L. (2008) 'Human embryonic stem cells reveal recurrent genomic instability at 20q11.21', *Nat Biotechnol* 26(12): 1364-6.
- Li, B., Li, J. B., Xiao, X. F., Ma, Y. F., Wang, J., Liang, X. X., Zhao, H. X., Jiang, F., Yao, Y. Q. and Wang, X. H. (2013) 'Altered DNA methylation patterns of the H19 differentially methylated region and the DAZL gene promoter are associated with defective human sperm', *PLoS One* 8(8): e71215.
- Li, Q., Li, N., Hu, X., Li, J., Du, Z., Chen, L., Yin, G., Duan, J., Zhang, H., Zhao, Y. et al. (2011) 'Genome-wide mapping of DNA methylation in chicken', *PLoS One* 6(5): e19428.
- Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q. et al. (2010) 'A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts', *Cell Stem Cell* 7(1): 51-63.
- Li, S., Li, J., Tian, J., Dong, R., Wei, J., Qiu, X. and Jiang, C. (2012) 'Characterization, tissue expression, and imprinting analysis of the porcine CDKN1C and NAP1L4 genes', *J Biomed Biotechnol* 2012: 946527.
- Li, W. and Ding, S. (2010) 'Generation of novel rat and human pluripotent stem cells by reprogramming and chemical approaches', *Methods Mol Biol* 636: 293-300.
- Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P. and Ferguson-Smith, A. C. (2008a) 'A maternal-zygotic effect gene, *Zfp57*, maintains both maternal and paternal imprints', *Dev Cell* 15(4): 547-57.
- Li, X. P., Do, K. T., Kim, J. J., Huang, J., Zhao, S. H., Lee, Y., Rothschild, M. F., Lee, C. K. and Kim, K. S. (2008b) 'Molecular characteristics of the porcine DLK1 and MEG3 genes', *Anim Genet* 39(2): 189-92.
- Lienhard, M., Grimm, C., Morkel, M., Herwig, R. and Chavez, L. (2014) 'MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments', *Bioinformatics* 30(2): 284-6.
- Litten-Brown, J. C., Corson, A. M. and Clarke, L. (2010) 'Porcine models for the metabolic syndrome, digestive and bone disorders: a general overview', *Animal* 4(6): 899-920.
- Liu, C., Zhang, R., Gao, J., Li, T., Ren, Z., Zhou, C. and Wen, A. (2014) 'Derivation of human embryonic stem cell lines without any exogenous growth factors', *Mol Reprod Dev*.
- Liu, H., Zhu, F., Yong, J., Zhang, P., Hou, P., Li, H., Jiang, W., Cai, J., Liu, M., Cui, K. et al. (2008) 'Generation of induced pluripotent stem cells from adult rhesus monkey fibroblasts', *Cell Stem Cell* 3(6): 587-90.
- 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', *Proc Natl Acad Sci U S A* 102(15): 5513-8.
- Liu, J., Litman, D., Rosenberg, M. J., Yu, S., Biesecker, L. G. and Weinstein, L. S. (2000a) 'A GNAS1 imprinting defect in pseudohypoparathyroidism type IB', *J Clin Invest* 106(9): 1167-74.
- Liu, J., Yu, S., Litman, D., Chen, W. and Weinstein, L. S. (2000b) 'Identification of a methylation imprint mark within the mouse *Gnas* locus', *Mol Cell Biol* 20(16): 5808-17.

- Liu, Y., Ma, Y., Yang, J. Y., Cheng, D., Liu, X., Ma, X., West, F. D. and Wang, H. (2013) 'Comparative Gene Expression Signature of Pig, Human and Mouse Induced Pluripotent Stem Cell Lines Reveals Insight into Pig Pluripotency Gene Networks', *Stem Cell Rev.*
- Livak, K. J. and Schmittgen, T. D. (2001) 'Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method', *Methods* 25(4): 402-8.
- Loewer, S., Cabili, M. N., Guttman, M., Loh, Y. H., Thomas, K., Park, I. H., Garber, M., Curran, M., Onder, T., Agarwal, S. et al. (2010) 'Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells', *Nat Genet* 42(12): 1113-7.
- Loh, Y. H., Hartung, O., Li, H., Guo, C., Sahalie, J. M., Manos, P. D., Urbach, A., Heffner, G. C., Grskovic, M., Vigneault, F. et al. (2010) 'Reprogramming of T cells from human peripheral blood', *Cell Stem Cell* 7(1): 15-9.
- Lowry, W. E., Richter, L., Yachechko, R., Pyle, A. D., Tchieu, J., Sridharan, R., Clark, A. T. and Plath, K. (2008) 'Generation of human induced pluripotent stem cells from dermal fibroblasts', *Proc Natl Acad Sci U S A* 105(8): 2883-8.
- Lu, L. Y., Wu, J., Ye, L., Gavriline, G. B., Saunders, T. L. and Yu, X. (2010) 'RNF8-dependent histone modifications regulate nucleosome removal during spermatogenesis', *Dev Cell* 18(3): 371-84.
- Lund, R. J., Närvä, E. and Lahesmaa, R. (2012) 'Genetic and epigenetic stability of human pluripotent stem cells', *Nat Rev Genet* 13(10): 732-44.
- Lunney, J. K. (2007) 'Advances in swine biomedical model genomics', *Int J Biol Sci* 3(3): 179-84.
- Luo, H., Zhou, Y., Li, Y. and Li, Q. (2013) 'Splice variants and promoter methylation status of the Bovine Vasa Homology (Bvh) gene may be involved in bull spermatogenesis', *BMC Genet* 14(1): 58.
- Lyons, J., Landis, C. A., Harsh, G., Vallar, L., Grünwald, K., Feichtinger, H., Duh, Q. Y., Clark, O. H., Kawasaki, E. and Bourne, H. R. (1990) 'Two G protein oncogenes in human endocrine tumors', *Science* 249(4969): 655-9.
- Macfarlan, T. S., Gifford, W. D., Driscoll, S., Lettieri, K., Rowe, H. M., Bonanomi, D., Firth, A., Singer, O., Trono, D. and Pfaff, S. L. (2012) 'Embryonic stem cell potency fluctuates with endogenous retrovirus activity', *Nature* 487(7405): 57-63.
- Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R. et al. (2007) 'Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution', *Cell Stem Cell* 1(1): 55-70.
- Mantel, C., Guo, Y., Lee, M. R., Kim, M. K., Han, M. K., Shibayama, H., Fukuda, S., Yoder, M. C., Pelus, L. M., Kim, K. S. et al. (2007) 'Checkpoint-apoptosis uncoupling in human and mouse embryonic stem cells: a source of karyotypic instability', *Blood* 109(10): 4518-27.
- Mantzouratou, A. and Delhanty, J. D. (2011) 'Aneuploidy in the human cleavage stage embryo', *Cytogenet Genome Res* 133(2-4): 141-8.
- Marques, C. J., Carvalho, F., Sousa, M. and Barros, A. (2004) 'Genomic imprinting in disruptive spermatogenesis', *Lancet* 363(9422): 1700-2.
- Marques, C. J., Costa, P., Vaz, B., Carvalho, F., Fernandes, S., Barros, A. and Sousa, M. (2008) 'Abnormal methylation of imprinted genes in human sperm is associated with oligozoospermia', *Mol Hum Reprod* 14(2): 67-74.
- Marques, C. J., Francisco, T., Sousa, S., Carvalho, F., Barros, A. and Sousa, M. (2010) 'Methylation defects of imprinted genes in human testicular spermatozoa', *Fertil Steril* 94(2): 585-94.
- Martinat-Botté, F., Plat, M. and Guillouet, P. (2009) 'Biotechnologies de la reproduction porcine : des techniques de routine aux méthodes en émergence', *Inra Productions Animales* 22(2): 97-116.

- Martins-Taylor, K., Nisler, B. S., Taapken, S. M., Compton, T., Crandall, L., Montgomery, K. D., Lalande, M. and Xu, R. H. (2011) 'Recurrent copy number variations in human induced pluripotent stem cells', *Nat Biotechnol* 29(6): 488-91.
- Martins-Taylor, K. and Xu, R. H. (2012) 'Concise review: Genomic stability of human induced pluripotent stem cells', *Stem Cells* 30(1): 22-7.
- Martínez, D., Pentinat, T., Ribó, S., Daviaud, C., Bloks, V. W., Cebrià, J., Villalmanzo, N., Kalko, S. G., Ramón-Krauel, M., Díaz, R. et al. (2014) 'In utero undernutrition in male mice programs liver lipid metabolism in the second-generation offspring involving altered Lxra DNA methylation', *Cell Metab* 19(6): 941-51.
- Massip, K., Bonnet, N., Calgaro, A., Billoux, S., Baquié, V., Mary, N., Bonnet-Garnier, A., Ducos, A., Yerle, M. and Pinton, A. (2009) 'Male meiotic segregation analyses of peri- and paracentric inversions in the pig species', *Cytogenet Genome Res* 125(2): 117-24.
- Massip, K., Yerle, M., Billon, Y., Ferchaud, S., Bonnet, N., Calgaro, A., Mary, N., Dudez, A. M., Sentenac, C., Plard, C. et al. (2010) 'Studies of male and female meiosis in inv(4)(p1.4;q2.3) pig carriers', *Chromosome Res* 18(8): 925-38.
- Matzuk, M. M. and Lamb, D. J. (2008) 'The biology of infertility: research advances and clinical challenges', *Nat Med* 14(11): 1197-213.
- Mayshar, Y., Ben-David, U., Lavon, N., Biancotti, J. C., Yakir, B., Clark, A. T., Plath, K., Lowry, W. E. and Benvenisty, N. (2010) 'Identification and classification of chromosomal aberrations in human induced pluripotent stem cells', *Cell Stem Cell* 7(4): 521-31.
- McGrath, J. and Solter, D. (1984) 'Completion of mouse embryogenesis requires both the maternal and paternal genomes', *Cell* 37(1): 179-83.
- 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(7205): 766-70.
- Melville, H., Carpiniello, M., Hollis, K., Staffaroni, A. and Golestaneh, N. (2013) 'Stem cells: a new paradigm for disease modeling and developing therapies for age-related macular degeneration', *J Transl Med* 11: 53.
- Mikkelsen, T. S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B. E., Jaenisch, R., Lander, E. S. and Meissner, A. (2008) 'Dissecting direct reprogramming through integrative genomic analysis', *Nature* 454(7200): 49-55.
- Mitalipova, M. M., Rao, R. R., Hoyer, D. M., Johnson, J. A., Meisner, L. F., Jones, K. L., Dalton, S. and Stice, S. L. (2005) 'Preserving the genetic integrity of human embryonic stem cells', *Nat Biotechnol* 23(1): 19-20.
- Miyoshi, N., Ishii, H., Nagano, H., Haraguchi, N., Dewi, D. L., Kano, Y., Nishikawa, S., Tanemura, M., Mimori, K., Tanaka, F. et al. (2011) 'Reprogramming of mouse and human cells to pluripotency using mature microRNAs', *Cell Stem Cell* 8(6): 633-8.
- Miyoshi, N., Kuroiwa, Y., Kohda, T., Shitara, H., Yonekawa, H., Kawabe, T., Hasegawa, H., Barton, S. C., Surani, M. A., Kaneko-Ishino, T. et al. (1998) 'Identification of the Meg1/Grb10 imprinted gene on mouse proximal chromosome 11, a candidate for the Silver-Russell syndrome gene', *Proc Natl Acad Sci U S A* 95(3): 1102-7.
- Molaro, A., Hodges, E., Fang, F., Song, Q., McCombie, W. R., Hannon, G. J. and Smith, A. D. (2011) 'Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates', *Cell* 146(6): 1029-41.
- Monk, D., Arnaud, P., Apostolidou, S., Hills, F. A., Kelsey, G., Stanier, P., Feil, R. and Moore, G. E. (2006) 'Limited evolutionary conservation of imprinting in the human placenta', *Proc Natl Acad Sci U S A* 103(17): 6623-8.

- Montserrat, N., Bahima, E. G., Batlle, L., Häfner, S., Rodrigues, A. M., González, F. and Izpisua Belmonte, J. C. (2011) 'Generation of pig iPS cells: a model for cell therapy', *J Cardiovasc Transl Res* 4(2): 121-30.
- Montserrat, N., de Oñate, L., Garreta, E., González, F., Adamo, A., Eguizábal, C., Häfner, S., Vassena, R. and Izpisua Belmonte, J. C. (2012) 'Generation of feeder-free pig induced pluripotent stem cells without Pou5f1', *Cell Transplant* 21(5): 815-25.
- Moore, T. and Haig, D. (1991) 'Genomic imprinting in mammalian development: a parental tug-of-war', *Trends Genet* 7(2): 45-9.
- Morgan, H. D., Sutherland, H. G., Martin, D. I. and Whitelaw, E. (1999) 'Epigenetic inheritance at the agouti locus in the mouse', *Nat Genet* 23(3): 314-8.
- Morison, I. M., Paton, C. J. and Cleverley, S. D. (2001) 'The imprinted gene and parent-of-origin effect database', *Nucleic Acids Res* 29(1): 275-6.
- Morris, J. K., Alberman, E., Scott, C. and Jacobs, P. (2008) 'Is the prevalence of Klinefelter syndrome increasing?', *Eur J Hum Genet* 16(2): 163-70.
- Nagy, A. (2013) 'Secondary cell reprogramming systems: as years go by', *Curr Opin Genet Dev* 23(5): 534-9.
- Nagy, A., Góczy, E., Diaz, E. M., Prideaux, V. R., Iványi, E., Markkula, M. and Rossant, J. (1990) 'Embryonic stem cells alone are able to support fetal development in the mouse', *Development* 110(3): 815-21.
- Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N. and Yamanaka, S. (2008) 'Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts', *Nat Biotechnol* 26(1): 101-6.
- Nakagawa, M., Takizawa, N., Narita, M., Ichisaka, T. and Yamanaka, S. (2010) 'Promotion of direct reprogramming by transformation-deficient Myc', *Proc Natl Acad Sci U S A* 107(32): 14152-7.
- Navarro-Costa, P., Nogueira, P., Carvalho, M., Leal, F., Cordeiro, I., Calhaz-Jorge, C., Gonçalves, J. and Plancha, C. E. (2010) 'Incorrect DNA methylation of the DAZL promoter CpG island associates with defective human sperm', *Hum Reprod* 25(10): 2647-54.
- Nayernia, K., Nolte, J., Michelmann, H. W., Lee, J. H., Rathsack, K., Drusenheimer, N., Dev, A., Wulf, G., Ehrmann, I. E., Elliott, D. J. et al. (2006) 'In vitro-differentiated embryonic stem cells give rise to male gametes that can generate offspring mice', *Dev Cell* 11(1): 125-32.
- Nichols, J. and Smith, A. (2009) 'Naive and primed pluripotent states', *Cell Stem Cell* 4(6): 487-92.
- Nissan, X., Lemaitre, G., Peschanski, M. and Baldeschi, C. (2011) '[Melanocytes derived from pluripotent stem cells: a new biological resource for pigmentary disorders]', *Med Sci (Paris)* 27(11): 935-8.
- Nogales-Cadenas, R., Carmona-Saez, P., Vazquez, M., Vicente, C., Yang, X., Tirado, F., Carazo, J. M. and Pascual-Montano, A. (2009) 'GeneCodis: interpreting gene lists through enrichment analysis and integration of diverse biological information', *Nucleic Acids Res* 37(Web Server issue): W317-22.
- Notarianni, E., Laurie, S., Moor, R. M. and Evans, M. J. (1990) 'Maintenance and differentiation in culture of pluripotential embryonic cell lines from pig blastocysts', *J Reprod Fertil Suppl* 41: 51-6.
- O'Flynn O'Brien, K., Varghese, A and Agarwal, A (2010) The genetic causes of male factor infertility: A review, vol. 93: Fertility and Sterility.
- Oczkiewicz, M., Piestrzyńska-Kajtoch, A., Ropka-Molik, K., Rejduch, B. and Eckert, R. (2012) 'Expression and imprinting analysis of the NESP55 gene in pigs', *Gene Expr Patterns* 12(1-2): 18-23.
- Oestrup, O., Hall, V., Petkov, S. G., Wolf, X. A., Hyldig, S. and Hyttel, P. (2009) 'From zygote to implantation: morphological and molecular dynamics during embryo development in the pig', *Reprod Domest Anim* 44 Suppl 3: 39-49.

- Okada, Y., Scott, G., Ray, M. K., Mishina, Y. and Zhang, Y. (2007) 'Histone demethylase JHDM2A is critical for Tnp1 and Prm1 transcription and spermatogenesis', *Nature* 450(7166): 119-23.
- Okamura, K., Hagiwara-Takeuchi, Y., Li, T., Vu, T. H., Hirai, M., Hattori, M., Sakaki, Y., Hoffman, A. R. and Ito, T. (2000) 'Comparative genome analysis of the mouse imprinted gene impact and its nonimprinted human homolog IMPACT: toward the structural basis for species-specific imprinting', *Genome Res* 10(12): 1878-89.
- Okamura, K., Sakaki, Y. and Ito, T. (2005) 'Comparative genomics approach toward critical determinants for the imprinting of an evolutionarily conserved gene Impact', *Biochem Biophys Res Commun* 329(3): 824-30.
- 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(3): 247-57.
- Okita, K., Ichisaka, T. and Yamanaka, S. (2007) 'Generation of germline-competent induced pluripotent stem cells', *Nature* 448(7151): 313-7.
- Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T. and Yamanaka, S. (2008) 'Generation of mouse induced pluripotent stem cells without viral vectors', *Science* 322(5903): 949-53.
- Osteil, P., Tapponnier, Y., Markossian, S., Godet, M., Schmaltz-Panneau, B., Jouneau, L., Cabau, C., Joly, T., Blachère, T., Góczy, E. et al. (2013) 'Induced pluripotent stem cells derived from rabbits exhibit some characteristics of naïve pluripotency', *Biol Open* 2(6): 613-28.
- Pacheco, S. E., Houseman, E. A., Christensen, B. C., Marsit, C. J., Kelsey, K. T., Sigman, M. and Boekelheide, K. (2011) 'Integrative DNA methylation and gene expression analyses identify DNA packaging and epigenetic regulatory genes associated with low motility sperm', *PLoS One* 6(6): e20280.
- Panula, S., Medrano, J. V., Kee, K., Bergström, R., Nguyen, H. N., Byers, B., Wilson, K. D., Wu, J. C., Simon, C., Hovatta, O. et al. (2011) 'Human germ cell differentiation from fetal- and adult-derived induced pluripotent stem cells', *Hum Mol Genet* 20(4): 752-62.
- Park, C. H., Kim, H. S., Lee, S. G. and Lee, C. K. (2009) 'Methylation status of differentially methylated regions at Igf2/H19 locus in porcine gametes and preimplantation embryos', *Genomics* 93(2): 179-86.
- Park, C. H., Uh, K. J., Mulligan, B. P., Jeung, E. B., Hyun, S. H., Shin, T., Ka, H. and Lee, C. K. (2011) 'Analysis of imprinted gene expression in normal fertilized and uniparental preimplantation porcine embryos', *PLoS One* 6(7): e22216.
- Park, I. H., Arora, N., Huo, H., Maherali, N., Ahfeldt, T., Shimamura, A., Lensch, M. W., Cowan, C., Hochedlinger, K. and Daley, G. Q. (2008a) 'Disease-specific induced pluripotent stem cells', *Cell* 134(5): 877-86.
- Park, I. H., Zhao, R., West, J. A., Yabuuchi, A., Huo, H., Ince, T. A., Lerou, P. H., Lensch, M. W. and Daley, G. Q. (2008b) 'Reprogramming of human somatic cells to pluripotency with defined factors', *Nature* 451(7175): 141-6.
- Patti, M. E. (2013) 'Intergenerational programming of metabolic disease: evidence from human populations and experimental animal models', *Cell Mol Life Sci* 70(9): 1597-608.
- Pearsall, R. S., Plass, C., Romano, M. A., Garrick, M. D., Shibata, H., Hayashizaki, Y. and Held, W. A. (1999) 'A direct repeat sequence at the Rasgrf1 locus and imprinted expression', *Genomics* 55(2): 194-201.
- Pease, S. and Williams, R. L. (1990) 'Formation of germ-line chimeras from embryonic stem cells maintained with recombinant leukemia inhibitory factor', *Exp Cell Res* 190(2): 209-11.
- Pelizzola, M. and Ecker, J. R. (2011) 'The DNA methylome', *FEBS Lett* 585(13): 1994-2000.
- Pellestor, F. (2014) 'Chromothripsis: how does such a catastrophic event impact human reproduction?', *Hum Reprod* 29(3): 388-93.

- Pera, M. F. and Tam, P. P. (2010) 'Extrinsic regulation of pluripotent stem cells', *Nature* 465(7299): 713-20.
- Peters, A. H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schöfer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A. et al. (2001) 'Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability', *Cell* 107(3): 323-37.
- Peters, J. and Williamson, C. M. (2007) 'Control of imprinting at the Gnas cluster', *Epigenetics* 2(4): 207-13.
- Peters, J., Wroe, S. F., Wells, C. A., Miller, H. J., Bodle, D., Beechey, C. V., Williamson, C. M. and Kelsey, G. (1999) 'A cluster of oppositely imprinted transcripts at the Gnas locus in the distal imprinting region of mouse chromosome 2', *Proc Natl Acad Sci U S A* 96(7): 3830-5.
- Peterson, S. E. and Loring, J. F. (2013) 'Genomic instability in pluripotent stem cells: implications for clinical applications', *J Biol Chem*.
- Peterson, S. E., Westra, J. W., Rehen, S. K., Young, H., Bushman, D. M., Paczkowski, C. M., Yung, Y. C., Lynch, C. L., Tran, H. T., Nickey, K. S. et al. (2011) 'Normal human pluripotent stem cell lines exhibit pervasive mosaic aneuploidy', *PLoS One* 6(8): e23018.
- Petkov, S. G., Reh, W. A. and Anderson, G. B. (2009) 'Methylation changes in porcine primordial germ cells', *Mol Reprod Dev* 76(1): 22-30.
- Piccolo, F. M. and Fisher, A. G. (2014) 'Getting rid of DNA methylation', *Trends Cell Biol* 24(2): 136-43.
- Pinton, A., Barasc, H., Raymond Letron, I., Bordedebat, M., Mary, N., Massip, K., Bonnet, N., Calgario, A., Dudez, A. M., Feve, K. et al. (2011) 'Meiotic studies of a 38,XY/39,XXY mosaic boar', *Cytogenet Genome Res* 133(2-4): 202-8.
- Pinton, A., Raymond Letron, I., Berland, H. M., Bonnet, N., Calgario, A., Garnier-Bonnet, A., Yerle, M. and Ducos, A. (2008) 'Meiotic studies in an azoospermic boar carrying a Y;14 translocation', *Cytogenet Genome Res* 120(1-2): 106-11.
- Pitamber, P. N., Lombard, Z. and Ramsay, M. (2012) 'No Evidence for a Parent-of-Origin Specific Differentially Methylated Region Linked to RASGRF1', *Front Genet* 3: 41.
- Plagge, A., Gordon, E., Dean, W., Boiani, R., Cinti, S., Peters, J. and Kelsey, G. (2004) 'The imprinted signaling protein XL alpha s is required for postnatal adaptation to feeding', *Nat Genet* 36(8): 818-26.
- Polo, J. M., Liu, S., Figueroa, M. E., Kulalert, W., Eminli, S., Tan, K. Y., Apostolou, E., Stadtfeld, M., Li, Y., Shioda, T. et al. (2010) 'Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells', *Nat Biotechnol* 28(8): 848-55.
- Popescu, C., Boscher, J., Le Henaff, G. and Jouet, H. (1986) 'A new reciprocal translocation in a hypoproliferic boar', *Genet Sel Evol* 18(2): 123-30.
- Poplinski, A., Tüttelmann, F., Kanber, D., Horsthemke, B. and Gromoll, J. (2010) 'Idiopathic male infertility is strongly associated with aberrant methylation of MEST and IGF2/H19 ICR1', *Int J Androl* 33(4): 642-9.
- 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(7284): 1101-5.
- Proudhon, C. and Bourc'his, D. (2010) 'Identification and resolution of artifacts in the interpretation of imprinted gene expression', *Brief Funct Genomics* 9(5-6): 374-84.
- Proudhon, C., Duffié, R., Ajjan, S., Cowley, M., Iranzo, J., Carbajosa, G., Saadeh, H., Holland, M. L., Oakey, R. J., Rakyant, V. K. et al. (2012) 'Protection against de novo methylation is instrumental in maintaining parent-of-origin methylation inherited from the gametes', *Mol Cell* 47(6): 909-20.

- Qin, D., Gan, Y., Shao, K., Wang, H., Li, W., Wang, T., He, W., Xu, J., Zhang, Y., Kou, Z. et al. (2008) 'Mouse meningiocytes express Sox2 and yield high efficiency of chimeras after nuclear reprogramming with exogenous factors', *J Biol Chem* 283(48): 33730-5.
- Rais, Y., Zviran, A., Geula, S., Gafni, O., Chomsky, E., Viukov, S., Mansour, A. A., Caspi, I., Krupalnik, V., Zerbib, M. et al. (2013) 'Deterministic direct reprogramming of somatic cells to pluripotency', *Nature*.
- Rajarajan, K., Engels, M. C. and Wu, S. M. (2012) 'Reprogramming of mouse, rat, pig, and human fibroblasts into iPS cells', *Curr Protoc Mol Biol* Chapter 23: Unit 23.15.
- Ramos, L., van der Heijden, G. W., Derijck, A., Berden, J. H., Kremer, J. A., van der Vlag, J. and de Boer, P. (2008) 'Incomplete nuclear transformation of human spermatozoa in oligo-astheno-teratospermia: characterization by indirect immunofluorescence of chromatin and thiol status', *Hum Reprod* 23(2): 259-70.
- Rashid, S. T., Corbineau, S., Hannan, N., Marciniak, S. J., Miranda, E., Alexander, G., Huang-Doran, I., Griffin, J., Ahrlund-Richter, L., Skepper, J. et al. (2010) 'Modeling inherited metabolic disorders of the liver using human induced pluripotent stem cells', *J Clin Invest* 120(9): 3127-36.
- Reik, W. and Murrell, A. (2000) 'Genomic imprinting. Silence across the border', *Nature* 405(6785): 408-9.
- Reik, W. and Walter, J. (2001) 'Genomic imprinting: parental influence on the genome', *Nat Rev Genet* 2(1): 21-32.
- Reya, T. and Clevers, H. (2005) 'Wnt signalling in stem cells and cancer', *Nature* 434(7035): 843-50.
- Rice, P., Longden, I. and Bleasby, A. (2000) 'EMBOSS: the European Molecular Biology Open Software Suite', *Trends Genet* 16(6): 276-7.
- Richard, N., Molin, A., Coudray, N., Rault-Guillaume, P., Jüppner, H. and Kottler, M. L. (2013) 'Paternal GNAS mutations lead to severe intrauterine growth retardation (IUGR) and provide evidence for a role of XLAs in fetal development', *J Clin Endocrinol Metab*.
- Roach, M., Wang, L., Yang, X. and Tian, X. C. (2006) 'Bovine embryonic stem cells', *Methods Enzymol* 418: 21-37.
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G. and Mesirov, J. P. (2011) 'Integrative genomics viewer', *Nat Biotechnol* 29(1): 24-6.
- Rodríguez, A., Allegrucci, C. and Alberio, R. (2012) 'Modulation of pluripotency in the porcine embryo and iPS cells', *PLoS One* 7(11): e49079.
- Roest, H. P., van Klaveren, J., de Wit, J., van Gurp, C. G., Koken, M. H., Vermey, M., van Rooijen, J. H., Hoogerbrugge, J. W., Vreeburg, J. T., Baarends, W. M. et al. (1996) 'Inactivation of the HR23B ubiquitin-conjugating DNA repair enzyme in mice causes male sterility associated with chromatin modification', *Cell* 86(5): 799-810.
- Ronaghi, M. (2001) 'Pyrosequencing sheds light on DNA sequencing', *Genome Res* 11(1): 3-11.
- Ronen, D. and Benvenisty, N. (2012) 'Genomic stability in reprogramming', *Curr Opin Genet Dev* 22(5): 444-9.
- Rotondo, J. C., Selvatici, R., Di Domenico, M., Marci, R., Vesce, F., Tognon, M. and Martini, F. (2013) 'Methylation loss at H19 imprinted gene correlates with methylenetetrahydrofolate reductase gene promoter hypermethylation in semen samples from infertile males', *Epigenetics* 8(9).
- Ruike, Y., Imanaka, Y., Sato, F., Shimizu, K. and Tsujimoto, G. (2010) 'Genome-wide analysis of aberrant methylation in human breast cancer cells using methyl-DNA immunoprecipitation combined with high-throughput sequencing', *BMC Genomics* 11: 137.

- Ruiz, S., Panopoulos, A. D., Herreras, A., Bissig, K. D., Lutz, M., Berggren, W. T., Verma, I. M. and Izpisua Belmonte, J. C. (2011) 'A high proliferation rate is required for cell reprogramming and maintenance of human embryonic stem cell identity', *Curr Biol* 21(1): 45-52.
- Saha, K. and Jaenisch, R. (2009) 'Technical challenges in using human induced pluripotent stem cells to model disease', *Cell Stem Cell* 5(6): 584-95.
- Saitou, M., Kagiwada, S. and Kurimoto, K. (2012) 'Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells', *Development* 139(1): 15-31.
- Sasaki, H. and Matsui, Y. (2008) 'Epigenetic events in mammalian germ-cell development: reprogramming and beyond', *Nat Rev Genet* 9(2): 129-40.
- Sati, S., Tanwar, V. S., Kumar, K. A., Patowary, A., Jain, V., Ghosh, S., Ahmad, S., Singh, M., Reddy, S. U., Chandak, G. R. et al. (2012) 'High resolution methylome map of rat indicates role of intragenic DNA methylation in identification of coding region', *PLoS One* 7(2): e31621.
- Sato, A., Hiura, H., Okae, H., Miyauchi, N., Abe, Y., Utsunomiya, T., Yaegashi, N. and Arima, T. (2011) 'Assessing loss of imprint methylation in sperm from subfertile men using novel methylation polymerase chain reaction Luminex analysis', *Fertil Steril* 95(1): 129-34, 134.e1-4.
- Sato, N., Meijer, L., Skaltsounis, L., Greengard, P. and Brivanlou, A. H. (2004) 'Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor', *Nat Med* 10(1): 55-63.
- Savatier, P., Huang, S., Szekely, L., Wiman, K. G. and Samarut, J. (1994) 'Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts', *Oncogene* 9(3): 809-18.
- Scheuermann, M. O., Tajbakhsh, J., Kurz, A., Saracoglu, K., Eils, R. and Lichter, P. (2004) 'Topology of genes and nontranscribed sequences in human interphase nuclei', *Exp Cell Res* 301(2): 266-79.
- Schmitt, R. M., Bruyns, E. and Snodgrass, H. R. (1991) 'Hematopoietic development of embryonic stem cells in vitro: cytokine and receptor gene expression', *Genes Dev* 5(5): 728-40.
- Schuettengruber, B. and Cavalli, G. (2009) 'Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice', *Development* 136(21): 3531-42.
- Schulze, M., Ungefroren, H., Bader, M. and Fändrich, F. (2006) 'Derivation, maintenance, and characterization of rat embryonic stem cells in vitro', *Methods Mol Biol* 329: 45-58.
- Seki, Y., Hayashi, K., Itoh, K., Mizugaki, M., Saitou, M. and Matsui, Y. (2005) 'Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice', *Dev Biol* 278(2): 440-58.
- Shah, K., Sivapalan, G., Gibbons, N., Tempest, H. and Griffin, D. K. (2003) 'The genetic basis of infertility', *Reproduction* 126(1): 13-25.
- Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelsen, T. S., Heckl, D., Ebert, B. L., Root, D. E., Doench, J. G. et al. (2014) 'Genome-scale CRISPR-Cas9 knockout screening in human cells', *Science* 343(6166): 84-7.
- Shang, E., Nickerson, H. D., Wen, D., Wang, X. and Wolgemuth, D. J. (2007) 'The first bromodomain of Brdt, a testis-specific member of the BET sub-family of double-bromodomain-containing proteins, is essential for male germ cell differentiation', *Development* 134(19): 3507-15.
- Sikora, K. M., Magee, D. A., Berkowicz, E. W., Berry, D. P., Howard, D. J., Mullen, M. P., Evans, R. D., Machugh, D. E. and Spillane, C. (2011) 'DNA sequence polymorphisms within the bovine guanine nucleotide-binding protein Gs subunit alpha (Gsa)-encoding (GNAS) genomic imprinting domain are associated with performance traits', *BMC Genet* 12: 4.

- Silva, J., Barrandon, O., Nichols, J., Kawaguchi, J., Theunissen, T. W. and Smith, A. (2008) 'Promotion of reprogramming to ground state pluripotency by signal inhibition', *PLoS Biol* 6(10): e253.
- 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(8): 811-4.
- Smith, Z. D., Nachman, I., Regev, A. and Meissner, A. (2010) 'Dynamic single-cell imaging of direct reprogramming reveals an early specifying event', *Nat Biotechnol* 28(5): 521-6.
- Smyth, G. K. (2004) 'Linear models and empirical Bayes methods for assessing differential expression in microarrays experiments.', *Statistical Applications in Genetics and Molecular Biology* 3(1).
- Soejima, H. and Higashimoto, K. (2013) 'Epigenetic and genetic alterations of the imprinting disorder Beckwith-Wiedemann syndrome and related disorders', *J Hum Genet* 58(7): 402-9.
- Soldner, F., Hockemeyer, D., Beard, C., Gao, Q., Bell, G. W., Cook, E. G., Hargus, G., Blak, A., Cooper, O., Mitalipova, M. et al. (2009) 'Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors', *Cell* 136(5): 964-77.
- Sommer, C. A. and Mostoslavsky, G. (2010) 'Experimental approaches for the generation of induced pluripotent stem cells', *Stem Cell Res Ther* 1(3): 26.
- Spira, Alfred, Persoz, Charles, Barouki, Robert, Christin-Maitre, Sophie, Etiemble, Jeanne, Labro, Marie-Thérèse, Le Moal, Joëlle, Léridon, Henri, Levy, Rachel, Royère, Dominique et al. (2012) Les troubles de la fertilité. Etat des connaissances et pistes pour la recherche: INSERM.
- Stadtfeld, M., Apostolou, E., Akutsu, H., Fukuda, A., Follett, P., Natesan, S., Kono, T., Shioda, T. and Hochedlinger, K. (2010) 'Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells', *Nature* 465(7295): 175-81.
- Stadtfeld, M., Apostolou, E., Ferrari, F., Choi, J., Walsh, R. M., Chen, T., Ooi, S. S., Kim, S. Y., Bestor, T. H., Shioda, T. et al. (2012) 'Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and facilitates generation of all-iPS cell mice from terminally differentiated B cells', *Nat Genet* 44(4): 398-405, S1-2.
- Stadtfeld, M. and Hochedlinger, K. (2010) 'Induced pluripotency: history, mechanisms, and applications', *Genes Dev* 24(20): 2239-63.
- Stadtfeld, M., Maherali, N., Breault, D. T. and Hochedlinger, K. (2008a) 'Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse', *Cell Stem Cell* 2(3): 230-40.
- Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G. and Hochedlinger, K. (2008b) 'Induced pluripotent stem cells generated without viral integration', *Science* 322(5903): 945-9.
- Stark, R and Brown, G (2011) DiffBind: differential binding analysis of ChIP-Seq peak data.
- Stephens, P. J., Greenman, C. D., Fu, B., Yang, F., Bignell, G. R., Mudie, L. J., Pleasance, E. D., Lau, K. W., Beare, D., Stebbings, L. A. et al. (2011) 'Massive genomic rearrangement acquired in a single catastrophic event during cancer development', *Cell* 144(1): 27-40.
- Su, J., Wang, Y., Xing, X., Liu, J. and Zhang, Y. (2014) 'Genome-wide analysis of DNA methylation in bovine placentas', *BMC Genomics* 15(1): 12.
- Sugimoto, K., Koh, E., Sin, H. S., Maeda, Y., Narimoto, K., Izumi, K., Kobori, Y., Kitamura, E., Nagase, H., Yoshida, A. et al. (2009) 'Tissue-specific differentially methylated regions of the human VASA gene are potentially associated with maturation arrest phenotype in the testis', *J Hum Genet* 54(8): 450-6.
- 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(5959): 548-50.

- Surani, M. A., Hayashi, K. and Hajkova, P. (2007) 'Genetic and epigenetic regulators of pluripotency', *Cell* 128(4): 747-62.
- Suzuki, M. M. and Bird, A. (2008) 'DNA methylation landscapes: provocative insights from epigenomics', *Nat Rev Genet* 9(6): 465-76.
- Taapken, S. M., Nisler, B. S., Newton, M. A., Sampsel-Barron, T. L., Leonhard, K. A., McIntire, E. M. and Montgomery, K. D. (2011) 'Karyotypic abnormalities in human induced pluripotent stem cells and embryonic stem cells', *Nat Biotechnol* 29(4): 313-4.
- Tabas-Madrid, D., Nogales-Cadenas, R. and Pascual-Montano, A. (2012) 'GeneCodis3: a non-redundant and modular enrichment analysis tool for functional genomics', *Nucleic Acids Res* 40(Web Server issue): W478-83.
- Tachibana, M., Nozaki, M., Takeda, N. and Shinkai, Y. (2007) 'Functional dynamics of H3K9 methylation during meiotic prophase progression', *EMBO J* 26(14): 3346-59.
- Takagi, Y., Talbot, N. C., Rexroad, C. E. and Pursel, V. G. (1997) 'Identification of pig primordial germ cells by immunocytochemistry and lectin binding', *Mol Reprod Dev* 46(4): 567-80.
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S. (2007) 'Induction of pluripotent stem cells from adult human fibroblasts by defined factors', *Cell* 131(5): 861-72.
- Takahashi, K. and Yamanaka, S. (2006) 'Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors', *Cell* 126(4): 663-76.
- Tancos, Z., Nemes, C., Polgar, Z., Gocza, E., Daniel, N., Stout, T. A., Maraghechi, P., Pirity, M. K., Osteil, P., Taponnier, Y. et al. (2012) 'Generation of rabbit pluripotent stem cell lines', *Theriogenology* 78(8): 1774-86.
- Telugu, B. P., Ezashi, T. and Roberts, R. M. (2010) 'Porcine induced pluripotent stem cells analogous to naïve and primed embryonic stem cells of the mouse', *Int J Dev Biol* 54(11-12): 1703-11.
- Telugu, B. P., Ezashi, T., Sinha, S., Alexenko, A. P., Spate, L., Prather, R. S. and Roberts, R. M. (2011) 'Leukemia inhibitory factor (LIF)-dependent, pluripotent stem cells established from inner cell mass of porcine embryos', *J Biol Chem* 286(33): 28948-53.
- Teng, Fei and Zhou, Qi (2013) 'Epigenetic Re-Programming during Mammalian Preimplantation Embryogenesis and PGC Development', *Journal of Fertilization: In Vitro - IVF-Worldwide, Reproductive Medicine, Genetics & Stem Cell Biology* 1:114.
- Tesar, P. J., Chenoweth, J. G., Brook, F. A., Davies, T. J., Evans, E. P., Mack, D. L., Gardner, R. L. and McKay, R. D. (2007) 'New cell lines from mouse epiblast share defining features with human embryonic stem cells', *Nature* 448(7150): 196-9.
- Thomsen, H., Lee, H. K., Rothschild, M. F., Malek, M. and Dekkers, J. C. (2004) 'Characterization of quantitative trait loci for growth and meat quality in a cross between commercial breeds of swine', *J Anim Sci* 82(8): 2213-28.
- Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S. and Jones, J. M. (1998) 'Embryonic stem cell lines derived from human blastocysts', *Science* 282(5391): 1145-7.
- Thomson, J. A., Kalishman, J., Golos, T. G., Durning, M., Harris, C. P., Becker, R. A. and Hearn, J. P. (1995) 'Isolation of a primate embryonic stem cell line', *Proc Natl Acad Sci U S A* 92(17): 7844-8.
- Thomson, J. A. and Marshall, V. S. (1998) 'Primate embryonic stem cells', *Curr Top Dev Biol* 38: 133-65.
- Thornton, J. E. and Gregory, R. I. (2012) 'How does Lin28 let-7 control development and disease?', *Trends Cell Biol* 22(9): 474-82.
- Thuan, N. V., Kishigami, S. and Wakayama, T. (2010) 'How to improve the success rate of mouse cloning technology', *J Reprod Dev* 56(1): 20-30.

- Tilgner, K., Atkinson, S. P., Yung, S., Golebiewska, A., Stojkovic, M., Moreno, R., Lako, M. and Armstrong, L. (2010) 'Expression of GFP under the control of the RNA helicase VASA permits fluorescence-activated cell sorting isolation of human primordial germ cells', *Stem Cells* 28(1): 84-92.
- Tollefsbold, T (2011) *Handbook of Epigenetics : The New Molecular and Medical Genetics*: Academic Press.
- 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(5): 811-20.
- Toyooka, Y., Tsunekawa, N., Akasu, R. and Noce, T. (2003) 'Embryonic stem cells can form germ cells in vitro', *Proc Natl Acad Sci U S A* 100(20): 11457-62.
- Tsukiyama, T., Asano, R., Kawaguchi, T., Kim, N., Yamada, M., Minami, N., Ohinata, Y. and Imai, H. (2011) 'Simple and efficient method for generation of induced pluripotent stem cells using piggyBac transposition of doxycycline-inducible factors and an EOS reporter system', *Genes Cells* 16(7): 815-25.
- Villagómez, D. A. and Pinton, A. (2008) 'Chromosomal abnormalities, meiotic behavior and fertility in domestic animals', *Cytogenet Genome Res* 120(1-2): 69-80.
- Viswanathan, S. R., Daley, G. Q. and Gregory, R. I. (2008) 'Selective blockade of microRNA processing by Lin28', *Science* 320(5872): 97-100.
- Wakayama, T., Perry, A. C., Zuccotti, M., Johnson, K. R. and Yanagimachi, R. (1998) 'Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei', *Nature* 394(6691): 369-74.
- Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Li, G., Ci, W., Li, W. et al. (2014) 'Programming and inheritance of parental DNA methylomes in mammals', *Cell* 157(4): 979-91.
- Wang, M., Zhang, X., Kang, L., Jiang, C. and Jiang, Y. (2012) 'Molecular characterization of porcine NECD, SNRPN and UBE3A genes and imprinting status in the skeletal muscle of neonate pigs', *Mol Biol Rep* 39(10): 9415-22.
- Wang, S., Tang, X., Niu, Y., Chen, H., Li, B., Li, T., Zhang, X., Hu, Z., Zhou, Q. and Ji, W. (2007) 'Generation and characterization of rabbit embryonic stem cells', *Stem Cells* 25(2): 481-9.
- Ward, W. S. and Coffey, D. S. (1991) 'DNA packaging and organization in mammalian spermatozoa: comparison with somatic cells', *Biol Reprod* 44(4): 569-74.
- Warren, L., Manos, P. D., Ahfeldt, T., Loh, Y. H., Li, H., Lau, F., Ebina, W., Mandal, P. K., Smith, Z. D., Meissner, A. et al. (2010) 'Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA', *Cell Stem Cell* 7(5): 618-30.
- Weber, M., Davies, J. J., Wittig, D., Oakeley, E. J., Haase, M., Lam, W. L. and Schübeler, D. (2005) 'Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells', *Nat Genet* 37(8): 853-62.
- Weksberg, R., Shuman, C. and Beckwith, J. B. (2010) 'Beckwith-Wiedemann syndrome', *Eur J Hum Genet* 18(1): 8-14.
- Wernig, M., Lengner, C. J., Hanna, J., Lodato, M. A., Steine, E., Foreman, R., Staerk, J., Markoulaki, S. and Jaenisch, R. (2008a) 'A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types', *Nat Biotechnol* 26(8): 916-24.
- Wernig, M., Meissner, A., Cassady, J. P. and Jaenisch, R. (2008b) 'c-Myc is dispensable for direct reprogramming of mouse fibroblasts', *Cell Stem Cell* 2(1): 10-2.
- West, F. D., Terlouw, S. L., Kwon, D. J., Mumaw, J. L., Dhara, S. K., Hasneen, K., Dobrinsky, J. R. and Stice, S. L. (2010) 'Porcine induced pluripotent stem cells produce chimeric offspring', *Stem Cells Dev* 19(8): 1211-20.

- WHO (2010) *WHO laboratory manual for the examination and processing of human sperm - Fifth edition*: World Health Organisation.
- Williamson, C. M., Ball, S. T., Nottingham, W. T., Skinner, J. A., Plagge, A., Turner, M. D., Powles, N., Hough, T., Papworth, D., Fraser, W. D. et al. (2004) 'A cis-acting control region is required exclusively for the tissue-specific imprinting of Gnas', *Nat Genet* 36(8): 894-9.
- 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', *Nat Genet* 38(3): 350-5.
- Williamson, CM, Blake, A, Thomas, S, Beechey, CV, Hancock, J, Cattanach, BM and Peters, J (2013) World Wide Web Site - Mouse Imprinting Data and References. Oxfordshire: MRC Harwell.
- Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. and Campbell, K. H. (1997) 'Viable offspring derived from fetal and adult mammalian cells', *Nature* 385(6619): 810-3.
- Woodsdalek, J. E. (1913) 'ACCESSORY CHROMOSOMES IN THE PIG', *Science* 38(966): 30a-31.
- Wolf, X. A., Serup, P. and Hyttel, P. (2011) 'Three-dimensional localisation of NANOG, OCT4, and E-CADHERIN in porcine pre- and peri-implantation embryos', *Dev Dyn* 240(1): 204-10.
- Woltjen, K., Michael, I. P., Mohseni, P., Desai, R., Mileikovsky, M., Hämmäläinen, R., Cowling, R., Wang, W., Liu, P., Gertsenstein, M. et al. (2009) 'piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells', *Nature* 458(7239): 766-70.
- Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C. J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W. and Walter, J. (2011) '5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming', *Nat Commun* 2: 241.
- Wu, T., Liu, Y., Wen, D., Tseng, Z., Tahmasian, M., Zhong, M., Rafii, S., Stadtfeld, M., Hochedlinger, K. and Xiao, A. (2014) 'Histone Variant H2A.X Deposition Pattern Serves as a Functional Epigenetic Mark for Distinguishing the Developmental Potentials of iPSCs', *Cell Stem Cell* 15(3): 281-94.
- Wu, W., Shen, O., Qin, Y., Niu, X., Lu, C., Xia, Y., Song, L., Wang, S. and Wang, X. (2010) 'Idiopathic male infertility is strongly associated with aberrant promoter methylation of methylenetetrahydrofolate reductase (MTHFR)', *PLoS One* 5(11): e13884.
- Wu, Z., Chen, J., Ren, J., Bao, L., Liao, J., Cui, C., Rao, L., Li, H., Gu, Y., Dai, H. et al. (2009) 'Generation of pig induced pluripotent stem cells with a drug-inducible system', *J Mol Cell Biol* 1(1): 46-54.
- Yamanaka, S. (2009) 'Elite and stochastic models for induced pluripotent stem cell generation', *Nature* 460(7251): 49-52.
- Yamanaka, S. and Blau, H. M. (2010) 'Nuclear reprogramming to a pluripotent state by three approaches', *Nature* 465(7299): 704-12.
- Yamashita, J., Itoh, H., Hirashima, M., Ogawa, M., Nishikawa, S., Yurugi, T., Naito, M. and Nakao, K. (2000) 'Flk1-positive cells derived from embryonic stem cells serve as vascular progenitors', *Nature* 408(6808): 92-6.
- Yang, A. S., Estéicio, M. R., Doshi, K., Kondo, Y., Tajara, E. H. and Issa, J. P. (2004) 'A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements', *Nucleic Acids Res* 32(3): e38.
- Ye, L., Chang, J. C., Lin, C., Qi, Z., Yu, J. and Kan, Y. W. (2010) 'Generation of induced pluripotent stem cells using site-specific integration with phage integrase', *Proc Natl Acad Sci U S A* 107(45): 19467-72.
- Yerle, M., Schmitz, A., Milan, D., Chaput, B., Monteagudo, L., Vaiman, M., Frelat, G. and Gellin, J. (1993) 'Accurate characterization of porcine bivariate flow karyotype by PCR and fluorescence in situ hybridization', *Genomics* 16(1): 97-103.

- Ying, Q. L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P. and Smith, A. (2008) 'The ground state of embryonic stem cell self-renewal', *Nature* 453(7194): 519-23.
- Yoon, B. J., Herman, H., Sikora, A., Smith, L. T., Plass, C. and Soloway, P. D. (2002) 'Regulation of DNA methylation of Rasgrf1', *Nat Genet* 30(1): 92-6.
- Young, M. A., Larson, D. E., Sun, C. W., George, D. R., Ding, L., Miller, C. A., Lin, L., Pawlik, K. M., Chen, K., Fan, X. et al. (2012) 'Background mutations in parental cells account for most of the genetic heterogeneity of induced pluripotent stem cells', *Cell Stem Cell* 10(5): 570-82.
- Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, I. I. and Thomson, J. A. (2009a) 'Human induced pluripotent stem cells free of vector and transgene sequences', *Science* 324(5928): 797-801.
- Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R. et al. (2007) 'Induced pluripotent stem cell lines derived from human somatic cells', *Science* 318(5858): 1917-20.
- Yu, M., Hon, G. C., Szulwach, K. E., Song, C. X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B. et al. (2012) 'Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome', *Cell* 149(6): 1368-80.
- Yu, S., Yu, D., Lee, E., Eckhaus, M., Lee, R., Corria, Z., Accili, D., Westphal, H. and Weinstein, L. S. (1998) 'Variable and tissue-specific hormone resistance in heterotrimeric Gs protein alpha-subunit (Gsalph) knockout mice is due to tissue-specific imprinting of the gsalph gene', *Proc Natl Acad Sci U S A* 95(15): 8715-20.
- Yu, Z., Ji, P., Cao, J., Zhu, S., Li, Y., Zheng, L., Chen, X. and Feng, L. (2009b) 'Dazl promotes germ cell differentiation from embryonic stem cells', *J Mol Cell Biol* 1(2): 93-103.
- Yusa, K., Rad, R., Takeda, J. and Bradley, A. (2009) 'Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon', *Nat Methods* 6(5): 363-9.
- Yusa, K., Rashid, S. T., Strick-Marchand, H., Varela, I., Liu, P. Q., Paschon, D. E., Miranda, E., Ordóñez, A., Hannan, N. R., Rouhani, F. J. et al. (2011) 'Targeted gene correction of $\alpha 1$ -antitrypsin deficiency in induced pluripotent stem cells', *Nature* 478(7369): 391-4.
- Zalzman, M., Falco, G., Sharova, L. V., Nishiyama, A., Thomas, M., Lee, S. L., Stagg, C. A., Hoang, H. G., Yang, H. T., Indig, F. E. et al. (2010) 'Zscan4 regulates telomere elongation and genomic stability in ES cells', *Nature* 464(7290): 858-63.
- Zhang, C. Z., Leibowitz, M. L. and Pellman, D. (2013a) 'Chromothripsis and beyond: rapid genome evolution from complex chromosomal rearrangements', *Genes Dev* 27(23): 2513-30.
- Zhang, F. W., Cheng, H. C., Jiang, C. D., Deng, C. Y., Xiong, Y. Z., Li, F. E. and Lei, M. G. (2007) 'Imprinted status of pleomorphic adenoma gene-like I and paternal expression gene 10 genes in pigs', *J Anim Sci* 85(4): 886-90.
- Zhang, F. W., Han, Z. B., Deng, C. Y., He, H. J. and Wu, Q. (2012) 'Conservation of genomic imprinting at the NDN, MAGEL2 and MEST loci in pigs', *Genes Genet Syst* 87(1): 53-8.
- Zhang, P., Chang, W. H., Fong, B., Gao, F., Liu, C., Al Alam, D., Bellusci, S. and Lu, W. (2014) 'Regulation of induced pluripotent stem (iPS) cell induction by Wnt/ β -catenin signaling', *J Biol Chem* 289(13): 9221-32.
- Zhang, P., Zhao, M., Liang, G., Yin, G., Huang, D., Su, F., Zhai, H., Wang, L., Su, Y. and Lu, Q. (2013b) 'Whole-genome DNA methylation in skin lesions from patients with psoriasis vulgaris', *J Autoimmun* 41: 17-24.
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoutte, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W. et al. (2008) 'Model-based analysis of ChIP-Seq (MACS)', *Genome Biol* 9(9): R137.
- Zhao, M. T., Rivera, R. M. and Prather, R. S. (2013) 'Locus-Specific DNA Methylation Reprogramming During Early Porcine Embryogenesis', *Biol Reprod*.

Zhao, X. Y., Li, W., Lv, Z., Liu, L., Tong, M., Hai, T., Hao, J., Guo, C. L., Ma, Q. W., Wang, L. et al. (2009) 'iPS cells produce viable mice through tetraploid complementation', *Nature* 461(7260): 86-90.

Zhou, H., Wu, S., Joo, J. Y., Zhu, S., Han, D. W., Lin, T., Trauger, S., Bien, G., Yao, S., Zhu, Y. et al. (2009) 'Generation of induced pluripotent stem cells using recombinant proteins', *Cell Stem Cell* 4(5): 381-4.

Zwaka, T. P. (2010) 'Stem cells: Troublesome memories', *Nature* 467(7313): 280-1.

ANNEXES

ANNEXE 1 : Article publié dans *Biology of Reproduction*

Publication en ligne le 15 octobre 2014, doi: 10.1095/biolreprod.114.119610)

Title: Sperm DNA methylation analysis in swine reveals conserved and species-specific methylation patterns and highlights an altered methylation at the *GNAS* locus in infertile boars

Running Title: Sperm DNA methylation analysis in infertile boars

Summary sentence: Sperm DNA methylation reflects the paternal status of imprinted genes in swine and is increased at the *GNAS* locus of infertile boars.

Authors and affiliations:

Annabelle Congras¹, Martine Yerle-Bouissou¹, Alain Pinton², Florence Vignoles¹, Laurence Liaubet¹, Stéphane Ferchaud³ and Hervé Acloque^{1*}

¹ INRA, UMR1388 Génétique, Physiologie et Systèmes d'Élevage, *GenPhySE*, F-31326 Castanet-Tolosan, France

² Université de Toulouse INPT ENVT, UMR1388 Génétique Physiologie et Systèmes d'Élevage *GenPhySE*, F-31076 Toulouse, France

³ UE1372 GenESI Génétique, Expérimentation et Système Innovants, F-17700 Surgères, France

* **corresponding author** : Herve Acloque, herve.acloque@toulouse.inra.fr, UMR1388 INRA-INPT *GenPhySE*, 24 chemin de Borde Rouge, CS52627, 31326 Castanet-Tolosan, France.

Keywords

Sperm, DNA methylation, imprinted gene, infertility, swine, pig, *GNAS*

Funding

This work was funded by the Animal Genetics Division of the Institut National de la Recherche Agronomique (National Institute of Agronomic Research, INRA/DGA). Annabelle Congras's PhD was granted by the Department of Animal Genetics of the Institut National de la Recherche

Agronomique (National Institute of Agronomic Research, INRA/DGA) and by the Direction Générale de l'Enseignement et de la Recherche (General Direction of Education and Research, DGER).

Abbreviations

A : Asthenospermia

CGI : CpG island

DMR : Differentially Methylated Region

ICR : Imprinted Control Region

LUMA : Luminometric Methylation Assay

MeDIP : Methylation DNA Immunoprecipitation

O : Oligospermia

PGC : Primordial germ cell

qPCR : quantitative Polymerase Chain Reaction

SNP : Single Nucleotide Polymorphism

T : Teratospermia

TSS : Transcription Start Site

QTL : Quantitative Trait Locus

Abstract

Male infertility is an increasing health issue in today's society, for both human or livestock populations. In livestock, male infertility slows the improvement of animal selection programs and agricultural productivity. There are increasing evidences that epigenetic marks play an important role in the production of good quality sperm. We therefore screened for specific or common epigenetic signatures of livestock infertility. To do so, we compared DNA methylation level in sperm DNA from fertile and infertile boars. We evaluated first the global level of sperm DNA methylation and found no difference between the two groups of boars. We then selected 42 loci of interest, most of them known to be imprinted in human or mice and assessed the imprinting status of 5 of them not previously described in swine tissues: WT1, CNTN3, IMPACT, QPCT and GRB10. DNA methylation level was then quantified in fertile and infertile boars, at these 42 loci. Results from fertile boars indicated that the methylation level of the selected loci is highly conserved between pig, human and mice, with a few exceptions, including the POU5F1 (OCT4) promoter and RTL1. Comparison between fertile and infertile boars revealed that one imprinted region, the GNAS locus, shows an increase in sperm DNA methylation in 4 out of 8 infertile boars with low semen quality. This increase in DNA methylation is associated with an altered expression of the genes belonging to the GNAS locus, suggesting a new role for GNAS in the proper formation of functional gametes.

Introduction

Infertility is a growing issue in human health and in livestock production. In humans, one out of six couples have trouble conceiving a child [1]. In swine industry, more than 40% of boars selected for reproduction based on agronomic criteria are culled and excluded from artificial insemination centers due to low semen quality. The main parameters for assessing semen quality are sperm count, mobility and morphology. Causes of infertility or low fertility are numerous (genetic, hormone disorders, life style, age, etc.), but many cases are still defined as idiopathic. In the last decade, epigenetic

mechanisms have been described, with more and more evidence of impaired DNA methylation patterns in spermatozoa of patients having an altered spermogram. Most studies focus on imprinted genes, supposing a flaw in the formation of epigenetic marks at the primordial germ cell stage [2]. One of these genes, MEST, has been shown to be hypermethylated in sperm from oligozoospermic patients [2, 3] whereas another imprinted gene, H19, is hypomethylated [3-5]. Studies on a larger panel of genes confirm a trend of hypermethylation in paternally expressed genes such as PEG1, LIT1, PEG3, SNRPN, HRAS, NTF3, MT1A, PAX8, DIRAS3, PLAGL1 (also known as ZAC1) or SFN and hypomethylation in genes expressed on the maternal allele such as GTL2 (also known as MEG1) and ZDBF2 [2, 6, 7]. Hyper- and hypomethylation can occur separately or, in the most severe cases, together [6, 7]. Flaws in DNA methylation patterns are not only associated with oligospermia, but also with other types of sperm defects: teratospermia with methylation loss at IGF2 DMR2 and the sixth CTCF-binding site (CTCF6) of H19 and methylation gain on MEST [8, 9], azoospermic patients with H19 CTCF6 hypomethylation in testicular spermatozoa [10], altered methylation level of several CpG islands in asthenospermic patients [11]. Abnormal protamination is also correlated with the hypermethylation of LIT1 and SNRPN in sperm DNA [12]. Furthermore, the pool of impacted genes may differ depending on the type of reproductive issue involved and its associated mechanism [12]. This variation in methylation patterns may explain some discrepancies, such as persistent methylation at H19 locus in spermatocytes from tubules with spermatogenic arrest at the spermatocyte stage [5] and the absence of MEST hypermethylation in testicular spermatozoa from azoospermic patients [10]. Sperm abnormalities may also be linked to a global flaw in DNA methylation patterns in sequences other than the promoters of imprinted genes [2, 11]. For example, idiopathic infertility has been found to be associated with hypermethylation of the MTHFR promoter, which plays a role in the regulation of the pool of available methyl groups [13, 14]. The silencing of this gene is a possible explanation for the low methylation level of some loci such as the CTCF-binding sites in oligozoospermic patients, but can also be explained by a more global flaw in DNA methylation.

To our knowledge, there has been no report on the correlation between infertility and abnormal methylation of gene promoters and gametic Differentially Methylated Regions (DMRs) of imprinted

genes in swine. To address this question, we added new data on the imprinted status of genes in swine and compared their methylation level in sperm DNA from fertile and infertile boars.

Using reciprocal crosses between Chinese Meishan and European Large White breeds we showed that RASGRF1 is paternally expressed in boars while WT1, IMPACT, CNTN3, QPCT and GRB10 are biallelically expressed. Then we compared global DNA methylation level in sperm from normal and infertile boars. We next focused our analysis on 29 target genes. Among them, we selected 3 genes implicated in early embryonic development (POU5F1) or in germline establishment and gametogenesis (DAZL and DDX4) as suitable candidates for the characterization of the epigenetic status of their promoters in relation to sperm defects [15-18]. We also analyzed DNA methylation at 11 genes IGF2, DLK1, MEST (also known as PEG1), PEG3, PEG10, NNAT, GNASXL, RTL1, RASGRF1, XIST and IMPACT that are known to be paternally expressed in humans and/or mice and 11 genes H19, IGF2R, NESP55 (GNAS locus), GRB10, OSBPL1A, MEG3, HM13, UBE3A, WT1, CNTN3 and QPCT that are known to be maternally expressed in humans and/or mice. Our data highlight an increase in DNA methylation level at the GNAS locus in infertile boars with low semen quality. Expression analysis confirms that this increase in DNA methylation is associated with an altered expression of GNAS genes and suggest for the first time that altered DNA methylation at this imprinted locus could be associated with infertility.

Materials and methods

Ethics Statement

Our research was conducted in accordance with European Directive 2010/63/EU on the protection of animals used for scientific purposes, and validated by the Animal Experimentation Ethics Committee for the Poitou-Charentes region (France) (no. CE2012-2). Moreover, the technical and scientist staff obtained individual accreditation from the ethic committee to experiment on living animals.

Allele-specific expression analysis

Blood DNA of two Meishan (MS, male and female) and two Large White (LW, male and female) adult animals were genotyped at previously described exonic Single Nucleotide Polymorphisms (SNPs) (Ensembl Database) in RASGRF1, WT1, CNTN3, IMPACT, QPCT and GRB10. These regions were amplified by conventional PCR and PCR products were sequenced using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystem) and an ABI 3730 DNA analyser. SNPs were confirmed or infirmed by aligning the sequences for the 4 animals with ChromasPro software.

Total RNA from liver, muscle, brain and lung tissues from 12 fetuses at 90 days of gestation (6 of each reciprocal crosses LWxMS and MSxLW) were extracted using Trizol reagent and the Nucleospin RNA kit (Macherey-Nagel, Cat. No.740955.50). Quality and quantity of total RNA was assessed using NanoDrop™ 1000 (Thermo scientific) dosage and gel electrophoresis. RNA was converted to cDNA with Superscript II Reverse Transcriptase (Invitrogen, Cat. No.18064-014). Complementary DNA were amplified using the PyroMark PCR kit (Qiagen, Cat. No. 978703) with the primers listed in Supplementary Table S1 and finally sequenced with the PyroMark Q24 and associated reagents to assess the parental origin of the expressed alleles in the four tissues from the 12 fetuses.

Sperm collection and DNA extraction

Boars were raised at the Artificial Insemination Station of INRA GenESI (authorization no. A-17-661), individually lodged on straw litter and fed twice a day. Semen was collected using the gloved hand technique and boars were accustomed to the procedure, which was carried out once a week by the same person. After sampling, sperm was immediately assessed qualitatively and quantitatively and diluted in Beltsville Thawing Solution (Landata) at a final concentration of 3.107 cell/ml. Swine were sampled at least at two different times to confirm semen parameters. Swine sperm samples were obtained from 13 boars, including 5 fertile boars and 8 boars with low sperm quality, characterized by oligospermia, asthenospermia or teratospermia (see Table 1). Boars were karyotyped using classical cytogenetic analyses (GTG banding karyotyping) according to Ducos et al.(1998) [19] and revealed that one fertile and two infertile boars carried chromosomal abnormalities (Supplementary Figure S1). Sperm cells were lysed in Tris-EDTA buffer (10 mM Tris pH 8.0, 1 mM EDTA) with 100 mM NaCl, 2% SDS and 10 mM DTT. After incubation in 10 mg/mL proteinase K, DNA was extracted using

phenol:chloroform:isoamyl alcohol (25:24:1), purified with 70% ethanol and diluted in Tris-EDTA buffer. DNA concentrations were measured using a NanoDrop™ 1000 spectrophotometer (Thermo Scientific).

Global methylation level: luminometric methylation assay

The global methylation level of each DNA sample was measured using LUMA, a pyrosequencing-based method [20, 21] in two independent experiments. First, 500 ng of genomic DNA was digested by EcoRI and either by HpaII or MspI. Enzymatic digestion of DNA was performed using excess of restriction enzymes and a long period of incubation (4 hours) to guaranty the efficiency of the reaction. Digestion efficiencies were checked on an agarose gel. Digested DNA was then diluted in Pyromark Annealing Buffer (Qiagen) and then pyrosequenced on PyroMark Q24 sequencer (Qiagen, product no. 9001514) using PyroMark Gold Q24 Reagents (Qiagen, product no. 970802). The isoschizomers HpaII and MspI target the same DNA CCGG sequence, but HpaII is methylation sensitive and does not cleave methylated sites while MspI is methylation insensitive. Pyrosequencing is used to sequence the overhangs left by both enzymes. During pyrosequencing, the proportion of incorporated C and G nucleotides at 5'-CG overhangs is directly related to the number of digested sites in the sample. The nucleotide dispensation order is defined as A;C+G;T;C+G;G;A where the adenosine incorporation reflects the EcoRI digestion efficiency and the (C+G) simultaneous incorporation reflects both HpaII and MspI digestions (Figure 1A). We first normalized the peak height of C+G incorporation by the peak height of A incorporation to normalize for digestion efficiencies. We then calculated the peak height ratio of simultaneous C+G incorporation in HpaII and MspI digests which is therefore representative of the DNA methylation level in the DNA sample and is close to 1 when the sample is unmethylated. Differences in global methylation level between the two groups of boars were assessed by a t-test.

Methylated DNA immunoprecipitation

Methylated DNA immunoprecipitation (MeDIP) was performed as previously described [22]. For each reaction, DNA was sheared by sonication using a Bioruptor sonication system (Diagenode) to obtain 300-500 bp DNA fragments. Then, 1 µg of sheared DNA was diluted in 450 µL of TE buffer (10 mM Tris pH 8.0, 1 mM EDTA), heat denatured in boiling water for 10 min and immediately chilled on ice for 10 min. Next, 50 µL of 10X IP buffer (100 mM Na-Phosphate pH 7.0, 1.4 M NaCl, 0.5% Triton-X100) was added to the mix together with 1 µg of anti-5-methylcytosine (mouse monoclonal antibody, clone 33D3, Millipore product no. MABE146). The reaction mix was incubated overnight at 4°C with overhead shaking and DNA-antibody complexes were purified using 40 µL Dynabeads-ProteinG (Life Technologies, product no. 10009D) previously washed in PBS buffer (Gibco, product no. 10010-015). DNA-antibody complexes were washed twice in IP buffer (10 mM Na-Phosphate pH 7.0, 140 mM NaCl, 0.05% Triton-X100) for 10 min. Immunoprecipitated DNA and 200ng of input DNA were then purified using the iPure kit (Diagenode, product no. C03010012) following the manufacturer's instructions. For each sperm sample, MeDIP reaction was made in triplicate.

Real-time PCR

Enriched methylated DNA and input DNA were amplified in triplicate using real-time PCR on specific genomic regions to quantify the enrichment rate of methylated DNA. Reactions were run on a Roche LC480 thermal cycler. Loci of interest were selected based on their sequence homology for well-defined, differentially methylated regions (DMRs) in imprinted genes and for promoter regions or in CpG islands in imprinted and other genes. CpG islands were determined using CpGPlot [23]. Primers were designed using Primer3 [24] and tested for their amplification efficiencies on serial dilutions of swine genomic DNA. Primer sequences used in this study are shown in Supplemental Table S2. The enrichment rate of methylated DNA for each locus was calculated using the formula which normalizes the Cp value obtained from methylation-enriched DNA by the ones of non-immunoprecipitated input DNA and takes into account the difference in the starting concentration of the two samples. Results for each locus were then normalized with respect to the enrichment rate of the repeated element LINE1. Differences between control and case groups were assessed by statistical non parametric Wilcoxon test.

Pyrosequencing

Sperm DNA was bisulfite-converted in duplicate using the MethylEdge bisulfite conversion system (Promega, product no. N1301) following the manufacturer's instructions. Primer pairs designed by the PyroMark Assay Design 2.0 software (Qiagen, product no. 9019062) on our sequences of interest were used to amplify specifically converted DNA using the PyroMark PCR kit (Qiagen, product no. 978703). Amplicons were then pyrosequenced on a PyroMark Q24 pyrosequencer (Qiagen, product no. 9001514) with the appropriate buffers and solutions (Qiagen, product nos. 979009, 979006, 979007, 979008 and 970802). Bisulfite conversion efficiency was evaluated during pyrosequencing through the quantification of Cytosine and Thymidine incorporation at some cytosines in a non CpG context. One of the two amplification primers must be biotinylated to allow proper fixation on Sepharose beads and sequencing with the sequencing primer. Primers used for pyrosequencing are listed in Supplementary Table S3. Differences in CpG methylation percentages between boars were assessed by statistical t-test.

Gene expression analysis in the testis

Total RNA from testis of 2 control boars (N1 and N5), 2 infertile boars with normal methylation status at GNAS (A1 and AT2) and 2 infertile boars with increased methylation at GNAS (AT3, OAT1) was extracted using the Nucleospin RNA kit (Macherey-Nagel, Cat. No. 740955.50) following manufacturer's instructions. Quality and quantity of total RNA were assessed using NanoDrop™ 1000 dosage and gel electrophoresis. RNA was converted to cDNA with Superscript II Reverse Transcriptase (Invitrogen, Cat. No.18064-014). Expression analysis of the 4 main transcripts composing the GNAS cluster and POU5F1 (specific marker of spermatogonia used as a positive control) was realized by real-time qPCR using the primers listed in Supplementary Table S4 and the LightCycler 480 Real-Time PCR System from Roche. Results were analyzed by the $2\Delta\Delta C_t$ method and statistical t-test. TATA Box binding protein (TBP) was used as the reference gene for relative quantification [25].

Results

Assessment of the parent-of-origin specific expression of RASGRF1, CNTN3, WT1, IMPACT, GRB10 and QPCT

To maximize the probability of identifying the parental origin of an allele in F1 individuals from two reciprocal crosses, we selected two swine breeds, european Large White (LW) and chinese Meishan (MS) known to be genetically distant. We genotyped two pairs of LW and MS animals at SNPs located in exonic DNA of RASGRF1, CNTN3, WT1, IMPACT, GRB10 and QPCT for which imprinting as not yet been assessed in swine. The identified SNPs were subsequently used to assign the parental origin of transcripts coding for these genes. For that, we extracted total RNA from muscle, brain, liver and lung tissues in fetuses issued from the two reciprocal crosses (male LW x female MS or male MS x female LW) and quantified the ratio between maternally and paternally expressed alleles by pyrosequencing. It allowed us to detect the biallelic or monoallelic expression of the genes and to assign the parental origin of the expressed alleles in fetal tissues in which expression was detected. Results in Table 2 revealed the biallelic expression of CNTN3 in muscle, lung and brain, WT1 in muscle, GRB10 in muscle, lung and brain, and IMPACT and QPCT in the four fetal tissues tested. RASGRF1 expression detected in brain and lung revealed the paternal expression of this gene in both tissues.

High level of DNA methylation in swine sperm

Based on spermogram data we separated the boars into two groups: the control group included animals with normal sperm parameters whereas the case group was composed of boars with different types of semen defects, from simple defects (oligospermia (O), teratospermia (T) or asthenospermia (A), see Table 1) or complex phenotypes combining oligo-, astheno- and/or teratospermia (AT and OAT boars; see Table 1). One fertile and two infertile boars were carrier of reciprocal translocations, respectively t(3;4) for N5, t(13;17) for AT3 and t(1;14) for OAT1 (supplementary Figure 1). Global methylation level in sperm DNA was measured using the LUMA method ([24,25] and Material and methods section). Calculating C+G incorporation ratios between MspI and HpaII digestion products (Figure

1A) allowed to quantify the percentage of global methylation at CCGG sites in the swine genome. It ranged from 68% to 82% depending on the sample (Figure 1B). The mean methylation level was 77% (+/-6%) and there were no significant differences between control and case groups (t-test $p=0.611$). To confirm this result, we then quantified the absolute methylation level of the Long Interspersed Nucleotide Element 1 (LINE1), which is a simple method to assess for global DNA methylation defects between animals [26]. LINE1 is a repeated element distributed all over the swine genome and is normally heavily methylated. After bisulfite conversion and pyrosequencing of a LINE1 region containing 10 CpGs, we observed that the methylation level was variable among each CpG, ranging from 31% to 100%, with a highly conserved pattern in all tested boars (Figure 1C). The mean methylation level in this region of LINE1 is about 61.5% (+/-3.2%), corroborating the high global methylation level in the sperm genome observed by LUMA analysis.

Differential methylation level between loci in sperm DNA from control boars

First, we designed 42 primer pairs to study local methylation by MeDIP-qPCR at several imprinted genes and genes of early development (Table 3). When possible, we used sequence homologies between the swine and mouse genomes to focus on the swine genomic regions corresponding to the gametic differentially methylated regions (gDMRs) defined in mice [27], because gDMRs are slightly different from embryonic DMRs. When data on the mouse gDMRs were not available, we selected regions in the promoters or in CpG islands spanning the locus. We added two loci in the promoter region of non-imprinted genes (C10ORF67 and ACAD8) as a control for hypomethylated regions in sperm and ACTB as a control for moderately methylated regions. Given to its large and homogenous representation in the genome and to its highly conserved methylation level in all tested boars, the LINE1 transposon was used as a reference locus to normalize gene methylation levels between genes and between animals.

The methylation level at selected loci appeared variable compared to the methylation level of the reference locus LINE1 in the five control animals (N1 to N5) (Figure 2A). Some loci harbored a higher methylation level compared to LINE1 while others showed a lower level. Weakly methylated loci were located near or within RASGRF1, C10ORF67, ACAD8, DDX4, DAZL, MEST, GNASXL,

POU5F1, WT1, OSBPL1A, QPCT, HM13, MEG3 and IGF2R DMR. These loci showed a relative methylation enrichment at least two-fold lower compared to the repeated LINE1 sequence (Figure 2A). PEG3 methylation enrichment was not high enough to be detected by MeDIP-qPCR. The relative methylation level of RASGRF1 DMR, IMPACT, XIST, IGF2R, NESP55, PEG10, NNAT, ACTB, CNTN3 and UBE3A was close to 1, indicating that it does not differ strongly from the methylation level of the reference LINE1 sequence (Figure 2A). In contrast, methylation marks were at least two-fold enriched on IGF2 DMR1 and DMR2, H19 imprinted control region (ICR), DLK1, DLK1/MEG3, RTL1 and GRB10 regions (Figure 2A). As expected, we confirmed the methylation of regions known to be methylated in swine sperm DNA like IGF2 DMR and H19 ICR.

Comparative analysis of DNA methylation levels at 32 loci in swine, mice and human sperm .

We then compared our data to data available in the genome-wide sperm methylome for humans and mice [28, 29]. The methylation level of homologous regions in the three species was compared (supplementary Table S5) and classified into three categories: poorly methylated (from 0 to 33% of CpG methylation in human and mice and less than 0.5 relative methylation enrichment for swine, in blue on Figure 2B), moderately methylated (from 33% to 66% of CpG methylation in human and mice and from 0.5 to 2 relative methylation score in swine, in light brown on Figure 2B) and highly methylated (from 66% to 100% of CpG methylation in human and mice and more than 2 relative methylation score in swine, in red on Figure 2B) . As shown in Figure 2B, our results are consistent with whole-genome bisulfite sequencing data for human and mouse sperm, suggesting that the methylation level on almost all studied DMRs are conserved between these three species. Interesting points consist in the few observed differences between species: surprisingly, the region spanning the POU5F1 (also known as OCT4) promoter and transcription start site (TSS) is highly methylated in human sperm DNA but not in swine or mice. Similarly, RTL1 is hypomethylated in humans but highly methylated in swine and mice. The second difference consists of loci with intermediate methylation levels in swine and high methylation scores in human and mice (such as IGF2R, XIST PEG10 gene body or RASGRF1 DMR). These differences may be due to the very low density of CpGs in these

regions that affects the MeDIP efficiency and also to the lower resolution of MeDIP-qPCR compared to whole genome bisulfite sequencing.

Comparison of the methylation level between control and infertile boars

Using MeDIP followed by real-time PCR, we quantified variation in methylation levels at different genomic regions in sperm DNA from controls and infertile boars (Table 1). We did not observe any variation between control and case groups for 21 of the 38 tested loci, independently of their basal methylation level (Supplementary Table S6). This includes the MEST promoter and H19 ICR (Figure 3A). In contrast, 17 loci revealed significant differences between control and case groups (Supplementary Table S6). In 15 of them, methylation was higher in boars with poor semen quality, whereas in 2 others, the level of methylation was lower (Supplementary Table S6 and Supplementary Figure S2). Increases in methylation were observed for RTL1, MEG3 DMR, DLK1/MEG3, NESP55, GNASXL ICR, GRB10, RASGRF1, PEG10, WT1, IMPACT B, DAZL promoter and IGF2 DMRs. Decreases in methylation level were observed for RASGRF1 DMR and IMPACT A.

Due to experimental variability and experimental bias, MeDIP-qPCR is not sufficiently sensitive to discriminate for small variations between boars and we observed ambiguous results for IGF2 DMRs, RASGRF1 and IMPACT. Moreover, most of statistically significant differences observed in our data remained low and not sufficient to conclude in favor of a major change in the methylation level between the two groups.

Thus, to overcome the MeDIP-qPCR limitations, we decided to confirm our results by performing single base methylation analysis in all the regions (insofar as possible) where we observed differences between the two groups.

Pyrosequencing increased the resolution of the analysis to the base level. This technique quantifies the methylation level of each CpG dinucleotide in a bisulfite-converted sequence [30]. As shown in Figure 3B, results confirmed the low basal methylation level of MEST, MEG3 DMR, DAZL, IGF2R DMR and NESP55 loci in sperm of the control group (mean values of 11.8%, 6.4%, 2.7%, 4.3% and 11.5% calculated on 13, 7, 7, 10 and 15 CpGs respectively) and the high basal methylation level of IGF2 DMR1 and DMR2, H19 ICR, DLK1 B and RTL1 (mean values of 86.1%, 80.2% 88.3%, 90% and

89.9% calculated on 8, 8, 17, 7 and 6 CpGs respectively). RASGRF1 DMR and IMPACT A regions, which were classified as moderately methylated by MeDIP-qPCR analysis, appeared highly methylated in sperm DNA by using the pyrosequencing technology (mean values of 97.2% and 87.9% calculated on 4 and 3 CpGs respectively) in accordance with the methylation scores found in mouse and human homologous region for RASGRF1 DMR but in discrepancy for IMPACT, hypomethylated in human and mouse. (Figure 2B).

We then compared the methylation level between fertile and infertile boars. Pyrosequencing of MEG3 DMR, DAZL, MEST, IGF2 DMR1, IGF2 DMR2, H19 ICR, DLK1, IMPACT, RASGRF1, RTL1 and IGF2R did not confirm the differences observed using the MeDIP-qPCR methodology (Figure 3B and supplementary Figure S3). Parallel analyses in blood and sperm DNA confirmed the absence of DNA methylation variations between the two groups of boars either in somatic or germ cells and interestingly highlighted the specific paternal imprinting at MEG3 DMR and H19 ICR with a methylation level in blood around 50% (Figure 3B). On the contrary, the methylation level of IGF2 and RASGRF1 DMRs in blood was unexpectedly high and the one of NESP55 and MEST was only around 30%. The variation of DNA methylation level between sperm and blood was specific to imprinted regions as sperm and blood DNA share the same global level of DNA methylation in the genome as well as in repeated element LINE1 (Supplementary Figure S4). This can also be observed at the not imprinted regions NESP55 A, which is located upstream the GNAS locus, and IMPACT A which is in the gene body, both harboring a high methylation level in sperm and blood samples (Figure 3C). Finally, in accordance with its germ cell specificity, the promoter of DAZL appeared demethylated in sperm while highly methylated in somatic cells like blood samples [31].

NESP55 was the only gene with significant differences between fertile and infertile boars using either MeDIP-qPCR or pyrosequencing of bisulfite-converted DNA methods (Supplementary Figure S2 and Figure 3B). The significant increase detected by MeDIP-qPCR in two CpG islands at the GNAS locus corresponding to the NESP55 promoter in three of the seven infertile boars (Figure 4A), was confirmed with a high fold change (1.9 to 3.3) for three of the eight low semen quality boars at two different locations in the CpG island (T1, AT3 and OAT1, Figure 4B) leading to a slight but

significant increase in the mean methylation level in the case group (17.8% compared to 11.5% in the control group for NESP55 B and C, $p=0,0012$, Welch two sample t-test).

This hypermethylation was sperm-specific because the methylation level in blood DNA remained constant in all boars (mean value 31.7%). Conversely, the variation in methylation level observed between sperm and blood samples in the NESP55 locus, like for MEST B, was representative of the imprinted status of the gene because the NESP55 promoter was unmethylated in sperm but partially methylated in blood (Figure 3B).

Methylation status of the GNAS complex locus varies between the two groups of boars

We then decided to precise whether the increase in DNA methylation observed in three infertile boars was restricted to the NESP55 promoter or expand to a larger genomic region. NESP55 belongs to the GNAS locus that covers more than 100kb on swine chromosome 17 and encodes for four different main transcripts (Figure 5A). Ten pyrosequencing primers were designed in this region to study the different CpG islands. Two primer pairs (G4 and G5) were designed for small CpG islands upstream from the locus, three primer pairs (G6-G7-G8) were located in the CpG island covering the promoter and the first exon of the NESP55 transcript, three others (G9-G10-G12) covered the promoter region and the first exon of GNASXL transcript and two primers (G13 and G14) were designed for the CpG island spanning the promoter of 1A (also known as A/B) and GSA transcripts (Figure 5A and Supplementary Table S3). The regions covered by G6 to G14 correspond to the three DMRs identified in the *Gnas* locus in mice [32].

By pyrosequencing of bisulfite-converted DNA, we observed an increased level of methylation in three boars for the NESP55 promoter and the first GNASXL exon (Figure 5B). These boars include the teratospermic boar T1, the astheno-/teratospermic boar AT3 and the oligo-/astheno-/teratospermic boar OAT1, the two last carrying chromosomal abnormalities. The significant increase detected by MeDIP-qPCR on one oligospermic boar (O1) (Figure 4A) could not be confirmed by pyrosequencing, as the moderate increase detected at the G6 region was not significant. Non-promoter and non-coding CpG islands upstream from the GNAS complex locus were highly methylated in all boars unlike the

promoter region and first exon of transcript 1A and GSA that were poorly methylated and showed no variation with respect to the fertility status of the boars.

Gene expression at the GNAS locus varies between the two groups of boars

Genes with high levels of methylation at 5-methylcytosine in their promoter are generally transcriptionally silent and we hypothesized that the increase in methylation at the NESP55 promoter may decrease NESP55 expression in male germ cells.

In order to see whether the altered methylation profile is specifically associated to a change in gene expression, we extracted total RNA from testis of 2 fertile boars (N1 and N5, control), 2 infertile boars with normal methylation levels at the NESP55 promoter assessed by pyrosequencing (A1 and AT2, group1) and 2 infertile boars with increased methylation level at the NESP55 promoter (AT3 and OAT1, group2). We quantified gene expression level by real-time PCR for the four genes of the GNAS locus and for POU5F1, used as a positive control for male germ cells.

We observed a decrease in NESP55 expression in group2 compared to controls, suggesting that higher methylation level of NESP55 promoter decreases the transcriptional activity for this gene. Surprisingly we also observed a strong increase in the expression level of the three other genes (GNASXL, GSA and 1A) in infertile boars from groups 1 and 2, with a stronger and significant increase in boars with increase methylation at the NESP55 promoter (Figure 6).

Discussion

Methylation pattern of swine sperm DNA is consistent with data from other mammals

Our data provided further evidence that the global pattern of the sperm cell methylome is conserved in mammals, with a high level of methylation and specific hypomethylation of promoters [28, 29]. Our LUMA analysis revealed a global high level of methylation in the sperm genome, and local analysis of several promoters (DDX4, DAZL, C10ORF67, ACAD8, NESP55, WT1, MEST, RASGRF1 and POU5F1) indicated low levels of methylation characteristic of the TSS region, mainly conserved in humans and mice with the exception of the POU5F1 promoter, which is strongly methylated in

humans and in bovine sperm [33] and hypomethylated in mice and swine. This discrepancy has already been observed [34, 35] and may possibly indicate that this gene is differentially regulated during early embryogenesis in these four species, depending on the activation of demethylation mechanisms [36]. We did not see any broad change in global methylation level between control boars and boars with low semen quality, as observed in a study on low-quality human sperm [3].

Imprinting status of QPCT, CNTN3, IMPACT, WT1, GRB10 and RASGRF1

We first investigated the imprinted status of 5 genes that were not reported in swine but known to be imprinted either in mice or humans. Two of them, QPCT and CNTN3, are described as maternally expressed in mouse placenta [37], but their status is unknown in human. In pigs, we found a biallelic expression for both genes in several fetal tissues (lung, muscle, liver and brain). Our data also indicate a biallelic expression for GRB10 and WT1 which are described as maternally expressed in several mouse and human tissues. We found that the biallelic expression of IMPACT in humans is conserved in the pig species [38]. Finally, the paternal expression of RASGRF1 in mice is conserved in pig lungs, as already described by Ding et al. [39] and according to our study in brain as well. We only looked at four embryonic tissues and can not exclude that allele-specific expression would be observed in placental or other embryonic tissues. However these data confirmed that the imprinting status of many genes is species-dependent in mammals, mostly for those located outside of the well-described large imprinted clusters like IGF2-H19 or DLK1/MEG3.

Methylation analysis in sperm DNA reflects the conserved imprinted status of most of the selected genes

Our data provide further support that the methylation pattern of most of the selected genes is conserved between mice, swine and humans in sperm DNA. Thus, we observed a weak methylation of the MEST and PEG3 promoters, corresponding to MEST and PEG3 DMR in mouse [40, 41] and a higher methylation in the blood DNA. Our data also support that MEST and PEG3 are paternally expressed in swine, as previously described [42].

We also observed a moderate methylation of NESP55 promoter (12% of CpG methylation) that is quite similar to that observed in mouse and human and is known to be a DMR and Imprinting controlling region of the GNAS locus. NESP55 is maternally expressed and its promoter is paternally methylated but the imprinting appeared after fertilization [31].

As expected, we also found that the DMR of IGF2R, which is maternally expressed and methylated on the maternal chromosomes, was poorly methylated in swine sperm, in agreement with human and mouse sperm data. The two IGF2 DMRs and the H19 ICR showed high methylation in swine sperm DNA, corresponding to the paternal expression of IGF2 and the maternal expression of H19. In accordance with human and mouse data, the gene body of GRB10 was also highly methylated (compared to LINE1) in swine sperm DNA. Pyrosequencing confirmed the moderate methylation of H19 ICR in blood DNA (39% of CpG methylation compared to 89% in sperm DNA) but both DMRs of IGF2 appeared highly methylated. This surprising data either suggest that IGF2 DMR1 and DMR2 are highly methylated in swine blood DNA or that the methylation level of the 8 selected CpGs is not representative of the global methylation levels of the full DMRs. Pyrosequencing of sperm DNA also confirmed the high methylation of RASGRF1 DMR on the paternal allele which is conserved between species and is associated to its paternal expression in mice and swine. In blood DNA, we observed a high methylation level of this region that is also reported in human blood [42] and may question whether it can be considered as a DMR in these two species.

The DLK1/MEG3 cluster is a complex locus in which MEG3 is expressed on the maternal allele whereas DLK1 is expressed on the paternal allele. This imprinting is controlled by several DMRs, which are all methylated on the paternal allele [43, 44]. Our data suggest strong methylation of DLK1 on the paternal allele, with all studied loci being highly methylated, except MEG3 DMR which appeared demethylated. The surprising hypomethylation of MEG3 DMR is conserved in human and mouse sperm DNA, suggesting that specific methylation of the MEG3 DMR on the paternal genome occurs after fertilization in mammals, during embryonic development. Moreover, pyrosequencing of bisulfite-converted sperm and blood DNA at the MEG3 DMR confirm our MeDIP-qPCR data, the CpG methylation in blood DNA being around 59% against 9% in sperm DNA. We failed to detect homologous sequences for either the human or mouse IG DMR in the swine genome, but the

intergenic CpG island, which we named DLK1/MEG3, may correspond to the mouse DLK1/MEG3 IG DMR in terms of its genomic localization. Moreover, like the human or mouse IG DMR, this region is methylated on the paternal allele.

OSBPL1A has been reported as maternally expressed in pigs [42]. Our results demonstrated a low methylation level in the analyzed region of this gene, also observed in human and mouse data. The selected regions was located in the first intron, within a CpG island spanning the TSS, which are known to often escape DNA methylation and might not be the most informative sites of general methylation patterns [29, 45-47]. However, one explanation may be that the differential methylation of this locus occurs later after fertilization, like for MEG3 DMR.

Species specific methylation patterns in sperm DNA

Finally, our study highlights species-specific methylation patterns as observed for RTL1, IMPACT and POU5F1. RTL1 has been shown to be paternally methylated in mice and we observed the same pattern in swine sperm DNA, but surprisingly it is specifically hypomethylated in humans. RTL1 is a poorly characterized gene that is derived from the domestication of a sushi-ichi-related retrotransposon and is unique to placental mammals [48]. RTL1 may have been conserved in mammals by gaining a vital function in growth and development. Thus, its specific hypomethylation in human sperm may reflect the acquisition of new or different functions in human tissues. For IMPACT, pyrosequencing revealed the high methylation of the gene body, confirming the discrepancy between the species. In mouse, only the paternal allele is hypomethylated and mouse Impact is maternally expressed while in human, both alleles are hypomethylated and bi-allelically expressed [38], suggesting that different molecular mechanisms exist in those species to regulate Impact expression. Differences regarding POU5F1 were previously discussed.

Differences in methylation levels between normal and low semen quality boars

We used first MeDIP-qPCR to quantify relative methylation of a large panel of candidate genes including many imprinted loci. However, the low sensitivity of the MeDIP-qPCR technique [49], primarily due to the size of the immunoprecipitated DNA fragments (around 400 bp), to an enrichment bias for CpG-rich regions and to the difficulty to amplify low amount of DNA corresponding to low methylated regions, complicated data interpretation and was inadequate to highlight moderate variations between animals. Nevertheless, using this technique, we detected a strong and significant enrichment of methylated DNA at two different sites of the NESP55 promoter in three boars with low semen quality. To overcome MeDIP-qPCR limitations, we used another method with higher resolution, combining bisulfite conversion and pyrosequencing. By using this approach, we were able to discard the false positive cases detected using MeDIP-qPCR technologies and to confirm methylation differences at the NESP55 promoter in infertile boars.

Surprisingly, we did not observe any variation in methylation level between control and case boars at most of the selected loci, including at the IGF2/H19 reciprocally imprinted region, nor at the MEST promoter, both of which are widely described in the literature as hyper- or hypomethylated loci in sperm of infertile patients [9]. This result is not related to poor identification and localization in swine of IGF2 and H19 DMRs, which are highly homologous to the corresponding human gene cluster and well described in the literature [50, 51]. Moreover, our results are in accordance with a study on bovine sperm that did not find any significant difference in the methylation level of IGF2/H19 DMRs of high- or low-fertility bulls [52].

Pyrosequencing analysis confirmed that the pattern of methylation in the GNAS complex locus is conserved in mouse and swine sperm, with promoter regions being hypomethylated and non-promoter regions being highly methylated [29]. It also confirmed that methylation at paternally methylated DMR of NESP55, occurs after fertilization also in swine [32].

More importantly, this analysis confirmed the enrichment in methylation marks found using MeDIP-qPCR in the GNAS complex locus in sperm DNA of several infertile boars. Imprinting of the GNAS complex locus, which includes several transcripts, has been described in mice and humans, in which they are conserved. The NESP55 coding transcript is maternally expressed, whereas the different forms of GNASXL (also known as XLas) protein-coding transcripts and A/B non-coding transcripts

are paternally expressed, as well as the non-coding antisense AS transcript [53]. The maternal imprint of NESP55 is set up after implantation, explaining the low methylation level in the male gametes at the promoter of this maternally expressed transcript [54]. Expression of the GSA transcript coding for the alpha-G-protein subunit is mainly biallelic and maternally expressed in some tissues such as the thyroid, the pituitary gland and in ovaries [53, 55]. Non-coding transcripts are key regulating elements for the maintenance of GNAS imprinting. The quantification of transcripts has hindered the assessment of the imprinting status of the GNAS locus in swine using the parthenogenetic model [42]. Nevertheless, the maternal expression of NESP55 is conserved in swine [56]. Mutations or imprinting defaults in GNAS are recurrent in the McCune-Albright syndrome, progressive osseous heteroplasia, osteodystrophy, hormonal resistance to parathyroid hormone, with high variability of phenotypes that can be attributed to the complex regulation of GNAS imprinting [53, 57-59]. Gene dosage and imprinting of the GNAS complex locus is also associated with postnatal growth and metabolism [60-62] and intrauterine growth retardation [58]. In swine, the GNAS locus is located close to quantitative trait loci (QTLs) for body mass and piglet growth [63]. Our study is, to our knowledge, the first to correlate infertility with epigenetic modification in the GNAS region. In 3 out of 8 infertile boars, we detected an hypermethylation of the NESP55 promoter and of the putative first exon of swine GNASXL, both described as Nesp55 DMRs in mice. This variation is sperm-specific because no increase was detected in blood DNA. In the same animals, we also observed an altered expression of transcripts from the GNAS complex locus, with a decreased in NESP55 expression and an increase in GNASXL, 1A and GSA expression. This is consistent with the phenotype of mutant mice in which the maternal Nesp55 DMR is deleted. These mice lose Nesp55 expression and overexpressed Gnasxl and 1A compared to normal mice [32]. To date, it is not known whether GNAS genes are bi-allelically or mono-allelically expressed in testis and it would be of interest to check whether NESP55 and GNASXL DMRs hypermethylation may modulate a parent-of-origin expression of GNAS genes in the testis of adult boars.

Increase methylation within the GNAS complex locus did not appear to be linked to the presence of chromosomal abnormalities, because two of the affected boars had a normal karyotype and the fertile boar with the reciprocal translocation was not affected. However, the affected AT boar carried a

reciprocal translocation implicating chromosome 17 close to the GNAS locus. Moreover, modification in the methylation pattern at this imprinted locus may indicate severe sperm defects, which are more frequent in the presence of chromosomal abnormalities. Our results did not directly link a specific etiology with our methylation defects with only 1 out of 3 AT boars being impacted. However it is interesting to note that the 3 concerned boars are teratospermic. Here again, we hypothesize that the alteration in methylation pattern could be related to the severity of the case, the AT3 boar being the one with the most degraded spermogram among others AT boars. These results suggest a possible role of NESP55 or GNASXL in the development of efficient gametes. One of the three DMRs of GNAS described in mice includes the GNASXL promoting regions as well as the promoter of the anti-sense NESPAS transcript. NESPAS has not been described in swine, but modification of its transcription may affect proper imprinting in the GNAS complex locus and therefore the expression of its protein-coding transcripts.

To conclude, our results confirm that the global methylation level is high in swine sperm DNA, as observed in other mammals including humans and mice. We did not detect any significant difference in this level of DNA methylation in animals with sperm defects, whatever the defect type. By examining several gene regions, we also confirmed that the pattern of methylation of sperm DNA is generally conserved in mammals, with some exceptions such as in the POU5F1 promoter region or RTL1 gene body. Finally, using bisulfite conversion and pyrosequencing techniques, we revealed an increase in the methylation level of NESP55 and GNASXL promoters, in sperm DNA from some boars (3 out of 8) with low semen quality. These hypermethylated regions belong to the imprinted GNAS complex locus and are localized in homologous regions of two Gnas DMRs in mice. This hypermethylation, which has never been reported, is sperm-specific and associated with changes in expression level, suggesting a new role for the GNAS complex locus in gametogenesis.

Acknowledgments

We are very grateful to other members of the Cytogene team for helpful discussions and staff at the Plateforme Genomique (Get-PlaGe, Toulouse) and INRA Experimental Unit GenESI (Rouillé) for technical assistance. We also wish to thank David Robelin for his statistical advice, Melissa Gabanou for her contribution to data production, Juliette Riquet for imprinting studies and Frédérique Pitel and Sophie Leroux for setting up the pyrosequencing facility. We thank french breeders, artificial insemination centers and IFIP-Institut du Porc for their help in detecting and providing us interesting cases.

References

1. Shah K, Sivapalan G, Gibbons N, Tempest H, Griffin DK. The genetic basis of infertility. *Reproduction* 2003; 126:13-25.
2. Houshdaran S, Cortessis VK, Siegmund K, Yang A, Laird PW, Sokol RZ. Widespread epigenetic abnormalities suggest a broad DNA methylation erasure defect in abnormal human sperm. *PLoS One* 2007; 2:e1289.
3. Marques CJ, Costa P, Vaz B, Carvalho F, Fernandes S, Barros A, Sousa M. Abnormal methylation of imprinted genes in human sperm is associated with oligozoospermia. *Mol Hum Reprod* 2008; 14:67-74.
4. Marques CJ, Carvalho F, Sousa M, Barros A. Genomic imprinting in disruptive spermatogenesis. *Lancet* 2004; 363:1700-1702.
5. Hartmann S, Bergmann M, Bohle RM, Weidner W, Steger K. Genetic imprinting during impaired spermatogenesis. *Mol Hum Reprod* 2006; 12:407-411.
6. Sato A, Hiura H, Okae H, Miyauchi N, Abe Y, Utsunomiya T, Yaegashi N, Arima T. Assessing loss of imprint methylation in sperm from subfertile men using novel methylation polymerase chain reaction Luminex analysis. *Fertil Steril* 2011; 95:129-134, 134.e121-124.
7. Kobayashi H, Sato A, Otsu E, Hiura H, Tomatsu C, Utsunomiya T, Sasaki H, Yaegashi N, Arima T. Aberrant DNA methylation of imprinted loci in sperm from oligospermic patients. *Hum Mol Genet* 2007; 16:2542-2551.
8. Boissonnas CC, Abdalaoui HE, Haelewyn V, Fauque P, Dupont JM, Gut I, Vaiman D, Jouannet P, Tost J, Jammes H. Specific epigenetic alterations of IGF2-H19 locus in spermatozoa from infertile men. *Eur J Hum Genet* 2010; 18:73-80.
9. Poplinski A, Tüttelmann F, Kanber D, Horsthemke B, Gromoll J. Idiopathic male infertility is strongly associated with aberrant methylation of MEST and IGF2/H19 ICR1. *Int J Androl* 2010; 33:642-649.
10. Marques CJ, Francisco T, Sousa S, Carvalho F, Barros A, Sousa M. Methylation defects of imprinted genes in human testicular spermatozoa. *Fertil Steril* 2010; 94:585-594.
11. Pacheco SE, Houseman EA, Christensen BC, Marsit CJ, Kelsey KT, Sigman M, Boekelheide K. Integrative DNA methylation and gene expression analyses identify DNA packaging and epigenetic regulatory genes associated with low motility sperm. *PLoS One* 2011; 6:e20280.
12. Hammoud SS, Purwar J, Pflueger C, Cairns BR, Carrell DT. Alterations in sperm DNA methylation patterns at imprinted loci in two classes of infertility. *Fertil Steril* 2010; 94:1728-1733.
13. Wu W, Shen O, Qin Y, Niu X, Lu C, Xia Y, Song L, Wang S, Wang X. Idiopathic male infertility is strongly associated with aberrant promoter methylation of methylenetetrahydrofolate reductase (MTHFR). *PLoS One* 2010; 5:e13884.
14. Rotondo JC, Selvatici R, Di Domenico M, Marci R, Vesce F, Tognon M, Martini F. Methylation loss at H19 imprinted gene correlates with methylenetetrahydrofolate reductase gene promoter hypermethylation in semen samples from infertile males. *Epigenetics* 2013; 8.
15. Sugimoto K, Koh E, Sin HS, Maeda Y, Narimoto K, Izumi K, Kobori Y, Kitamura E, Nagase H, Yoshida A, Namiki M. Tissue-specific differentially methylated regions of the human VASA gene are potentially associated with maturation arrest phenotype in the testis. *J Hum Genet* 2009; 54:450-456.
16. Luo H, Zhou Y, Li Y, Li Q. Splice variants and promoter methylation status of the Bovine Vasa Homology (Bvh) gene may be involved in bull spermatogenesis. *BMC Genet* 2013; 14:58.
17. Navarro-Costa P, Nogueira P, Carvalho M, Leal F, Cordeiro I, Calhaz-Jorge C, Gonçalves J, Plancha CE. Incorrect DNA methylation of the DAZL promoter CpG island associates with defective human sperm. *Hum Reprod* 2010; 25:2647-2654.
18. Li B, Li JB, Xiao XF, Ma YF, Wang J, Liang XX, Zhao HX, Jiang F, Yao YQ, Wang XH. Altered DNA methylation patterns of the H19 differentially methylated region and the DAZL gene promoter are associated with defective human sperm. *PLoS One* 2013; 8:e71215.
19. Ducos A, Pinton A, Berland HM, Seguela A, Blanc MF, Darre A, Darre R. Five new cases of reciprocal translocation in the domestic pig. *Hereditas* 1998; 128:221-229.
20. Karimi M, Johansson S, Stach D, Corcoran M, Grandér D, Schalling M, Bakalkin G, Lyko F, Larsson C, Ekström TJ. LUMA (LUMinometric Methylation Assay)--a high throughput method to the analysis of genomic DNA methylation. *Exp Cell Res* 2006; 312:1989-1995.
21. Karimi M, Johansson S, Ekström TJ. Using LUMA: a Luminometric-based assay for global DNA-methylation. *Epigenetics* 2006; 1:45-48.
22. Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schübeler D. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* 2005; 37:853-862.

23. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 2000; 16:276-277.
24. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* 2007; 35:W71-74.
25. Nygard AB, Jørgensen CB, Cirera S, Fredholm M. Selection of reference genes for gene expression studies in pig tissues using SYBR green qPCR. *BMC Mol Biol* 2007; 8:67.
26. Yang AS, Estéicio MR, Doshi K, Kondo Y, Tajara EH, Issa JP. A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res* 2004; 32:e38.
27. Tomizawa S, Kobayashi H, Watanabe T, Andrews S, Hata K, Kelsey G, Sasaki H. Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development* 2011; 138:811-820.
28. Molaro A, Hodges E, Fang F, Song Q, McCombie WR, Hannon GJ, Smith AD. Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* 2011; 146:1029-1041.
29. Kobayashi H, Sakurai T, Imai M, Takahashi N, Fukuda A, Yayoi O, Sato S, Nakabayashi K, Hata K, Sotomaru Y, Suzuki Y, Kono T. Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet* 2012; 8:e1002440.
30. Dupont JM, Tost J, Jammes H, Gut IG. De novo quantitative bisulfite sequencing using the pyrosequencing technology. *Anal Biochem* 2004; 333:119-127.
31. Hackett JA, Reddington JP, Nestor CE, Dunican DS, Branco MR, Reichmann J, Reik W, Surani MA, Adams IR, Meehan RR. Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development* 2012; 139:3623-3632.
32. Fröhlich LF, Mrakovcic M, Steinborn R, Chung UI, Bastepe M, Jüppner H. Targeted deletion of the Nesp55 DMR defines another Gnas imprinting control region and provides a mouse model of autosomal dominant PHP-Ib. *Proc Natl Acad Sci U S A* 2010; 107:9275-9280.
33. Lan J, Hua S, Yuan Y, Zhan L, He X, Zhang Y. Methylation patterns in 5' terminal regions of pluripotency-related genes in mature bovine gametes. *Zygote* 2011; 19:165-169.
34. Zhao MT, Rivera RM, Prather RS. Locus-Specific DNA Methylation Reprogramming During Early Porcine Embryogenesis. *Biol Reprod* 2013.
35. Al-Khtib M, Blachère T, Guérin JF, Lefèvre A. Methylation profile of the promoters of Nanog and Oct4 in ICSI human embryos. *Hum Reprod* 2012; 27:2948-2954.
36. Gu TP, Guo F, Yang H, Wu HP, Xu GF, Liu W, Xie ZG, Shi L, He X, Jin SG, Iqbal K, Shi YG, et al. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* 2011; 477:606-610.
37. Brideau CM, Eilertson KE, Hagarman JA, Bustamante CD, Soloway PD. Successful computational prediction of novel imprinted genes from epigenomic features. *Mol Cell Biol* 2010; 30:3357-3370.
38. Okamura K, Hagiwara-Takeuchi Y, Li T, Vu TH, Hirai M, Hattori M, Sakaki Y, Hoffman AR, Ito T. Comparative genome analysis of the mouse imprinted gene impact and its nonimprinted human homolog IMPACT: toward the structural basis for species-specific imprinting. *Genome Res* 2000; 10:1878-1889.
39. Ding YY, Liu LY, Zhou J, Zhang XD, Huang L, Zhang SJ, Yin ZJ. Expression and genomic imprinting of the porcine Rasgrf1 gene. *Gene* 2014; 536:357-361.
40. Imamura T, Kerjean A, Heams T, Kupiec JJ, Thenevin C, Paldi A. Dynamic CpG and non-CpG methylation of the Peg1/Mest gene in the mouse oocyte and preimplantation embryo. *J Biol Chem* 2005; 280:20171-20175.
41. Kim J, Ekram MB, Kim H, Faisal M, Frey WD, Huang JM, Tran K, Kim MM, Yu S. Imprinting control region (ICR) of the Peg3 domain. *Hum Mol Genet* 2012; 21:2677-2687.
42. Bischoff SR, Tsai S, Hardison N, Motsinger-Reif AA, Freking BA, Nonneman D, Rohrer G, Piedrahita JA. Characterization of conserved and nonconserved imprinted genes in swine. *Biol Reprod* 2009; 81:906-920.
43. Zhou Y, Zhang X, Klibanski A. MEG3 noncoding RNA: a tumor suppressor. *J Mol Endocrinol* 2012; 48:R45-53.
44. Takada S, Tevendale M, Baker J, Georgiades P, Campbell E, Freeman T, Johnson MH, Paulsen M, Ferguson-Smith AC. Delta-like and gtl2 are reciprocally expressed, differentially methylated linked imprinted genes on mouse chromosome 12. *Curr Biol* 2000; 10:1135-1138.
45. Down TA, Rakyen VK, Turner DJ, Flicek P, Li H, Kulesha E, Gräf S, Johnson N, Herrero J, Tomazou EM, Thorne NP, Bäckdahl L, et al. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat Biotechnol* 2008; 26:779-785.
46. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 2010; 11:204-220.
47. Pelizzola M, Ecker JR. The DNA methylome. *FEBS Lett* 2011; 585:1994-2000.

48. Edwards CA, Mungall AJ, Matthews L, Ryder E, Gray DJ, Pask AJ, Shaw G, Graves JA, Rogers J, Dunham I, Renfree MB, Ferguson-Smith AC, et al. The evolution of the DLK1-DIO3 imprinted domain in mammals. *PLoS Biol* 2008; 6:e135.
49. Bock C, Tomazou EM, Brinkman AB, Müller F, Simmer F, Gu H, Jäger N, Gnirke A, Stunnenberg HG, Meissner A. Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat Biotechnol* 2010; 28:1106-1114.
50. Braunschweig MH, Owczarek-Lipska M, Stahlberger-Saitbekova N. Relationship of porcine IGF2 imprinting status to DNA methylation at the H19 DMD and the IGF2 DMRs 1 and 2. *BMC Genet* 2011; 12:47.
51. Amarger V, Nguyen M, Van Laere AS, Braunschweig M, Nezer C, Georges M, Andersson L. Comparative sequence analysis of the INS-IGF2-H19 gene cluster in pigs. *Mamm Genome* 2002; 13:388-398.
52. Jena SC, Kumar S, Rajput S, Roy B, Verma A, Kumaresan A, Mohanty TK, De S, Kumar R, Datta TK. Differential methylation status of IGF2-H19 locus does not affect the fertility of crossbred bulls but some of the CTCF binding sites could be potentially important. *Mol Reprod Dev* 2014.
53. Bastepe M. The GNAS Locus: Quintessential Complex Gene Encoding Gsalpha, XLalphas, and other Imprinted Transcripts. *Curr Genomics* 2007; 8:398-414.
54. Liu J, Yu S, Litman D, Chen W, Weinstein LS. Identification of a methylation imprint mark within the mouse Gnas locus. *Mol Cell Biol* 2000; 20:5808-5817.
55. Crane JL, Shambloott MJ, Axelman J, Hsu S, Levine MA, Germain-Lee EL. Imprinting status of Galpha(s), NESP55, and XLalphas in cell cultures derived from human embryonic germ cells: GNAS imprinting in human embryonic germ cells. *Clin Transl Sci* 2009; 2:355-360.
56. Oczkiewicz M, Piestrzyńska-Kajtoch A, Ropka-Molik K, Rejduch B, Eckert R. Expression and imprinting analysis of the NESP55 gene in pigs. *Gene Expr Patterns* 2012; 12:18-23.
57. Williamson CM, Ball ST, Nottingham WT, Skinner JA, Plagge A, Turner MD, Powles N, Hough T, Papworth D, Fraser WD, Maconochie M, Peters J. A cis-acting control region is required exclusively for the tissue-specific imprinting of Gnas. *Nat Genet* 2004; 36:894-899.
58. Richard N, Molin A, Coudray N, Rault-Guillaume P, Jüppner H, Kottler ML. Paternal GNAS mutations lead to severe intrauterine growth retardation (IUGR) and provide evidence for a role of XL α s in fetal development. *J Clin Endocrinol Metab* 2013.
59. Linglart A, Maupetit-Méhouas S, Silve C. GNAS -Related Loss-of-Function Disorders and the Role of Imprinting. *Horm Res Paediatr* 2013:119-129.
60. Ball ST, Kelly ML, Robson JE, Turner MD, Harrison J, Jones L, Napper D, Beechey CV, Hough T, Plagge A, Cattanaach BM, Cox RD, et al. Gene Dosage Effects at the Imprinted Gnas Cluster. *PLoS One* 2013; 8:e65639.
61. Plagge A, Gordon E, Dean W, Boiani R, Cinti S, Peters J, Kelsey G. The imprinted signaling protein XL alpha s is required for postnatal adaptation to feeding. *Nat Genet* 2004; 36:818-826.
62. Eaton SA, Hough T, Fischer-Colbrie R, Peters J. Maternal inheritance of the Gnas cluster mutation Ex1A-T affects size, implicating NESP55 in growth. *Mamm Genome* 2013; 24:276-285.
63. Thomsen H, Lee HK, Rothschild MF, Malek M, Dekkers JC. Characterization of quantitative trait loci for growth and meat quality in a cross between commercial breeds of swine. *J Anim Sci* 2004; 82:2213-2228.
64. Li XP, Do KT, Kim JJ, Huang J, Zhao SH, Lee Y, Rothschild MF, Lee CK, Kim KS. Molecular characteristics of the porcine DLK1 and MEG3 genes. *Anim Genet* 2008; 39:189-192.
65. Morison IM, Paton CJ, Cleverley SD. The imprinted gene and parent-of-origin effect database. *Nucleic Acids Res* 2001; 29:275-276.
66. McMurray EN, Schmidt JV. Identification of imprinting regulators at the Meg3 differentially methylated region. *Genomics* 2012; 100:184-194.
67. Williamson CM, Turner MD, Ball ST, Nottingham WT, Glenister P, Fray M, Tymowska-Lalanne Z, Plagge A, Powles-Glover N, Kelsey G, Maconochie M, Peters J. Identification of an imprinting control region affecting the expression of all transcripts in the Gnas cluster. *Nat Genet* 2006; 38:350-355.
68. Park CH, Kim HS, Lee SG, Lee CK. Methylation status of differentially methylated regions at Igf2/H19 locus in porcine gametes and preimplantation embryos. *Genomics* 2009; 93:179-186.
69. Nezer C, Moreau L, Brouwers B, Coppeters W, Detilleux J, Hanset R, Karim L, Kvasz A, Leroy P, Georges M. An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs. *Nat Genet* 1999; 21:155-156.
70. Wutz A, Smrzka OW, Schweifer N, Schellander K, Wagner EF, Barlow DP. Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature* 1997; 389:745-749.
71. Park CH, Uh KJ, Mulligan BP, Jeung EB, Hyun SH, Shin T, Ka H, Lee CK. Analysis of imprinted gene expression in normal fertilized and uniparental preimplantation porcine embryos. *PLoS One* 2011; 6:e22216.

72. Zhang FW, Han ZB, Deng CY, He HJ, Wu Q. Conservation of genomic imprinting at the NDN, MAGEL2 and MEST loci in pigs. *Genes Genet Syst* 2012; 87:53-58.
73. XU C, SU L, ZHOU Q, LI C, ZHAO S. Imprinting analysis of the porcine MEST gene in 75 and 90 day placentas and prenatal tissues. *Acta Biochim Biophys Sin (Shanghai)* 2007; 39:633-639.
74. Wang M, Zhang X, Kang L, Jiang C, Jiang Y. Molecular characterization of porcine NECD, SNRPN and UBE3A genes and imprinting status in the skeletal muscle of neonate pigs. *Mol Biol Rep* 2012; 39:9415-9422.
75. Williamson C, Blake A, Thomas S, Beechey C, Hancock J, Cattnach B, Peters J. World Wide Web Site - Mouse Imprinting Data and References. In; 2014.
76. Cheng HC, Zhang FW, Deng CY, Jiang CD, Xiong YZ, Li FE, Lei MG. NNAT and DIRAS3 genes are paternally expressed in pigs. *Genet Sel Evol* 2007; 39:599-607.
77. Zhang FW, Cheng HC, Jiang CD, Deng CY, Xiong YZ, Li FE, Lei MG. Imprinted status of pleomorphic adenoma gene-like I and paternal expression gene 10 genes in pigs. *J Anim Sci* 2007; 85:886-890.
78. Zhou QY, Huang JN, Xiong YZ, Zhao SH. Imprinting analyses of the porcine GATM and PEG10 genes in placentas on days 75 and 90 of gestation. *Genes Genet Syst* 2007; 82:265-269.

Figure legends

Figure 1: Global methylation level in swine sperm DNA. **A.** Example of a pyrogram showing a higher peak at the simultaneous C+G incorporation step in a MspI digest than in a HpaII digest (sample N2). E and S peaks represent the incorporation of enzyme and substrate, respectively, which catalyze the pyrosequencing reaction while the A peak represents the A incorporation at the EcoRI digest, used for normalization of the C+G peak. **B.** Global DNA methylation level at CCGG sites in sperm DNA is higher than 70% and is similar between fertile and infertile boars. Results are the means of two independent experiments +/- SD. **C.** Methylation level at *LINE1* as determined by bisulfite conversion and pyrosequencing in sperm DNA from fertile (control group) and infertile (case group) boars. Results are the means of two independent experiments.

Figure 2: Relative methylation level at several loci in sperm DNA from fertile boars. **A.** Methylation level varied with locus identity compared to the methylation level of *LINE1* in fertile boars (N1 to N5). Colors indicate the expected expressed allele (orange: maternal, green: paternal, gray: biallelic). Results are the mean values of three independent experiments +/-SD **B.** Comparison with human and mouse sperm methylome data sets (Molaro et al., 2011; Kobayashi et al., 2012). Mean values for CpG methylation at each locus is precised for human and mice. Colors indicate the methylation levels in the different loci. Blue: poorly methylated (0 to 33% of CpG methylation in human and mice and less than 0.5 relative methylation enrichment for swine); light brown, moderately methylated (33% to 66% of of CpG methylation in human and mice and from 0.5 to 2 relative methylation score in swine); red, highly methylated (> 66% of CpG methylation in human and mice and more than 2 relative methylation score in swine)).

Figure 3: Variation in sperm DNA methylation level from boars with normal or low semen quality. **A.** Methylation level of the paternally expressed *MEST* gene and the paternally methylated *H19* ICR remained constant in all animals. Results are the means of the relative methylation scores from three independent experiments +/- SD. **B.** Absolute methylation level in sperm and blood DNA

of control and case groups in several loci assessed by pyrosequencing on bisulfite-converted DNA. Significant differences between groups were only detected at the *NESP55* promoter (*NESP55* B and C). Results are the means of two independent experiments +/- SD (Welch *t*-test ** $p < 0,01$)

Figure 4: Increase in the methylation level at the *NESP55* locus in infertile boars with low semen quality. **A.** Relative enrichment of methylated DNA at the *NESP55* promoter (*NESP55* B) and first exon (*NESP55* C) determined by MeDIP-qPCR analysis in sperm DNA of infertile boars compared to fertile ones (N1 to N5). Results are the means of three independent experiments +/- SD (*t*-test * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$). **B.** Pyrosequencing of bisulfite-converted DNA of the *NESP55* promoter (*NESP55* B) and first exon (*NESP55* C) in fertile boars (Control group, black line) and infertile boars (colored lines). Results are the means of two independent experiments.

Figure 5: Methylation pattern in the swine *GNAS* complex locus in fertile boars and boars with low semen quality. **A.** Diagram of the main alternative first exons in the *GNAS* complex locus with the defined pyrosequencing primers (purple dashes) and CpG islands (green boxes). *NESP55* is expressed on the maternal allele, whereas all the variants of *GNASXL* and *IA* (also known as *A/B*) transcripts are expressed on the paternal allele. *GSA* coding transcripts are generally biallelic (blue arrows: coding transcripts; red arrow: non-coding transcript). **B.** Methylation level in fertile and infertile boars with low semen quality in the *GNAS* complex locus determined by pyrosequencing of bisulfite-converted DNA. Significant increase in DNA methylation of the CpG islands covering *NESP55* promoter and the first exon of *GNASXL* were observed in three boars with low semen quality. Results are the mean of two independent experiments +/-SD (for G6, G8 and G12, *t*-test * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$)

Figure 6: Expression level of the main *GNAS* transcripts in testis of fertile and infertile boars. Fertile boars (Control group, N1 and N5), low semen quality boars with constant methylation level in the *GNAS* locus (Group 1, A1 and AT2) and low semen quality animals harboring methylation changes in the *GNAS* locus (Group 2, AT3 and OAT1) were assessed for the expression of the four

main *GNAS* transcripts (*NESP55*, *GNASXL*, *IA* and *GSA*) and the common transcribed region (*GNAS*) by real-time PCR. *POU5F1* was used as a positive control and *TBP* as a reference gene for relative quantification. *NESP55* expression decreases in boars with methylated *NESP55* promoter (group 2) while *GNAS*, *GNASXL*, *IA* and *GSA* expression significantly increases in these animals. Results are the mean of three independent experiments. (*t*-test * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$)

Supplemental data legends

Supplementary Figure S1: G-banding karyotype of the three animal carriers of reciprocal translocation. Arrowheads highlight translocation positions. A. Animal N5 t(3;4) **B.** Animal AT3 t(13;17) **C.** Animal OAT1 t(1;14) (Alain Pinton, personal communication)

Supplementary Figure S2: MeDIP-qPCR analysis revealed significant variation in methylation level between the control group and the case group at several loci (Wilcoxon rank sum test. * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$)

Supplementary Figure S3: Mean methylation profile assessed by pyrosequencing of bisulfite-converted DNA at *MEG3*, *DAZL*, *MEST*, *IGF2*, *H19*, *DLK1*, *IGF2R*, *IMPACT*, *RASGRF1* and *RTL1* loci in fertile boars (black line) and infertile boars (colored lines).

Supplementary Figure S4: Conservation of methylation pattern in blood and sperm DNA in fertile and infertile boars. A. Similar global methylation level in blood and sperm of fertile (N1) and infertile animals (T1, O1, O2, AT2, AT3) as determined using LUMA analysis. **B.** Similar pattern and level of methylation in sperm and blood DNA samples at the *LINE1* locus.

Supplementary Figure S5 : Methylations profiles assessed by pyrosequencing of bisulfite converted DNA at all selected loci for each replicate.

Supplementary Table S1: Amplification and sequencing primers used to detect parental imprint by pyrosequencing.

Supplementary Table S2: Primer pairs used for MeDIP-qPCR.

Supplementary Table S3: Amplification and sequencing primers used for pyrosequencing of bisulfite-converted DNA.

Supplementary Table S4 : Primer pairs used for expression analysis by real-time PCR in testicular tissues.

Supplementary Table S5: Genomic coordinates and CpG methylation score for human and mice genes.

Supplementary Table S6 : Variation in methylation level between control and case groups by MeDIP-qPCR (Wilcoxon rank sum test; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$).

Tables

Table 1: Panel of the 13 studied boars, including 5 control boars and 8 boars with low semen quality. One boar in the first group and two boars in the second group are carriers of a reciprocal translocation. Mean values are calculated on 12000 boars (Stéphane Ferchaud, personal communication). *: commercial synthetic lines

	Name	Chromosomal status	Semen parameters	Genotype	Total number of spermatozoa (10 ⁹ /ejaculate)	Concentration (10 ⁶ /mL)	Mobility (%)	Motility (1 to 5)	Alive normal (%)
Mean values					95	402	88	3.3	
Limit values					>15		>70	>2.5	
Control group	N1	38, XY	Normal	Large White	22	303	90	4	90
	N2	38, XY	Normal	Large White	144	660	90	4	85
	N3	38, XY	Normal	Pietrain	91	585	75	2.5	80
	N4	38, XY	Normal	Pietrain	76	475	85	3.5	90
	N5	38, XY t(3;4)	Normal	Large White	50	200	90	3.7	NA
Experimental group	A1	38, XY	Asthenospermia	Large White	NA	490	30	1	NA
	T1	38, XY	Teratospermia	Pietrain	46	342	50	3	40
	O1	38, XY	Oligospermia	Pietrain	7	81	85	4	85
	O2	38, XY	Oligospermia	Pietrain	11	59	91	3.4	91
	AT1	38, XY	Asthenospermia Teratospermia	P76*	99	340	5	0.5	43
	AT2	38, XY	Asthenospermia, Teratospermia	Pietrain	62	677	20	2	0
	AT3	38, XY t(13;17)	Asthenospermia Teratospermia	GP1062*	31	282	0.5	5	0
	OAT1	38, XY t(1;14)	Oligospermia Asthenospermia Teratospermia	Pietrain	NA	19	38	1.5	40

Table 2: Genotypes and allele-specific expression of *WT1*, *CNTN3*, *IMPACT*, *RASGRF1*, *GRB10* and *QPCT* in fetal tissues issued from reciprocal crosses between Large White and Meishan breeds.

Region	Genomic position on <i>Sus scrofa</i> assembly 10.2	SNP reference	Genotype of the parents				Expressed nucleotide in fetal tissue		Expression
			Male LW	Femelle MS	Male MS	Femelle LW	LWxMS	MSxLW	
WT1 exon 9	chr2:30945103	rs55619464	G/A	A/A	A/A	G/G	G+A	G+A	Biallelic
CNTN3 exon 31	chr13:61558099	rs339003919	G/G	A/A	A/A	A/A	G+A	A	Biallelic
IMPACT exon 11	chr6:102037866	rs329889650	T/T	G/G	G/G	T/T	T+G	T+G	Biallelic
RASGRF1 exon 15	chr7:53697957	rs337793313	G/A	G/G	G/G	A/A	A or G	G or G+A	Paternal
GRB10 exon 7	chr9 :150298033	rs330552874	C/C	C/T	T/T	C/C	C or C+T	C+T	Biallelic
QPCT exon 8	chr3 :109334945	rs81214310	A/A	G/G	G/G	A/A	A+G	A+G	Biallelic

Table 3: Overview of the 42 selected loci for MeDIP-qPCR analysis of methylation level (M : Maternal, P : Paternal, B : Biallelic, NA: not assessed)

Name	Genomic Location on Sus Scrofa assembly 10.2	Description	Expressed allele in swine	Expressed allele in mouse	Expressed allele in human	ICR/DMR Methylation	References
ACTB	chr3:4733279-4733511 233	ACTB promoter	B	B	B		
RTL1	chr7:132099126-132099220	RTL1 gene body ; CGI	NA	P	NA	P	[42][64][65]
DLK1/MEG3	chr7:132176228-132176330	Intergenic between <i>MEG3</i> and <i>DLK1</i> ; CGI					
DLK1 A	chr7:132195144-132195248	2.5kb upstream <i>DLK1</i>					
DLK1 B	chr7:132448071-132448177	100kb downstream <i>DLK1</i>	P	P	P	P	[42][64][65]
MEG3 DMR	chr7:132163735-132163884	<i>MEG3</i> gene body	M	M	M	P	[27][65][66]
NESP55 A	chr17:66295668-66295785	8kb upstream <i>NESP55</i> ; CGI					
NESP55 B	chr17:66304289-66304459	<i>NESP55</i> promoter ; CGI					
NESP55 C	chr17:66306603-66306705	<i>NESP55</i> gene body ; CGI	M	M	M		[67][56][65]
GNASXL ICR	chr17:66314745-66314957	<i>NESP55</i> gene body ; <i>GNASXL</i> promoter ; CGI	NA	P	P		[67][56][65]
GRB10	chr9:150301782-150301877	<i>GRB10</i> gene body ; CGI	B	M (P in brain)	M (P in brain)	M	[27] present study
H19 ICR A	31841-32007 AY044827						
H19 ICR B	33075-33249 AY044827		M	M	M	P	[27][42][68][50][51][65]
IGF2 DMR1 A	16820-16976 AY044828						
IGF2 DMR1 B	17423-17667 AY044828		P	P	P	P	[42][68][50][51][69][65]
IGF2 DMR2 A	32905-33025 AY044827						
IGF2 DMR2 B	29112-29235 AY044828		P	P	P	P	[42][68][50][51][69][65]
IGF2R DMR	chr1:9245108-9245270	<i>IGF2R</i> gene body					
IGF2R A	chr1:9248709-9248851	<i>IGF2R</i> gene body					
IGF2R B	chr1:9039118-9039232	AF342812, <i>IGF2R</i> 3'UTR	M	M	M	M	[27][42][70][65][71]
LINE1	repeated sequence	many locations in the genome	B	B	B		
MEST A	chr18:19341929-19342047	<i>MEST</i> promoter ; CGI					
MEST B	chr18:19347216-19347307	8kb upstream <i>MEST</i> ; CGI	P	P	P	M	[27][42][72][73][65]
POU5F1	chr7:27263237-27263354	<i>POU5F1</i> gene body ; CGI	B	B	B		

RASGRF1 DMR	chr7:53820144-53820295	35kb upstream <i>RASGRF1</i>											
RASGRF1	chr7:53784911-53785053	<i>RASGRF1</i> 5'UTR											[27][65][39] present study
DDX4	chr16:37047974-37048093	<i>DDX4</i> gene body ; CGI	B	B	B	B	B	B	B	NA	P		
UBE3A	chr1:157794010-157794107	<i>UBE3A</i> gene body	B	B	M	M				M			[42][74]
IMPACT A	chr6:102003692-102003835	<i>IMPACT</i> gene body											
IMPACT B	chr6:102002040-102002238	<i>IMPACT</i> promoter ; CGI	B	B	B	P	B	B	B	B			[27][65] present study
DAZL	chr13:4036859-4037100	<i>DAZL</i> promoter; CGI	B	B	B	B	B	B	B	B			
OSBPL1A	chr6:101984829-101984928	<i>OSBPL1A</i> intron1 ; CGI	M	M						M			[42]
ACAD8	chr9:67366614-67366731	<i>ACAD8</i> promoter ; CGI	B	B	B	B	B	B	B	B			
QPCT	chr3:109359903-109360002	<i>QPCT</i> gene body intron1	B	B	B	M	B	B	B	NA			[75] present study
WT1	chr2:30902723-30902901	<i>WT1</i> promoter ; CGI	B	B	B	M	B	B	B	B			[75][65] present study
HM13	chr17:40074655-40074837	4kb downstream HM13	B	B	B	M	B	B	B	B			[42][65]
CNTN3	chr13:6107550-61606791	<i>CNTN3</i>	B	B	B	M	B	B	B	B			[75] present study
C10ORF67	chr10:57337222-57337382	<i>C10ORF67</i> promoter/exon1	B	B	B	B	B	B	B	B			
XIST	chrX:67157193-67157325	AJ429140, <i>XIST</i> gene body	NA	NA	NA	P	NA	NA	NA	NA			[65][71]
NNAT	chr17:46045193-46045324	DQ666422, <i>NNAT</i> 3'UTR	P	P	P	P	P	P	P	P			[42][76][65][71]
PEG10	chr9:81652380-81652519	DQ323403, <i>PEG10</i> 3'UTR	P	P	P	P	P	P	P	P			[27][42][77][78][65][71]
PEG3	chr6:56642050-56642263	12kb upstream <i>PEG3</i>	P	P	P	P	P	P	P	P			[27][42][65]

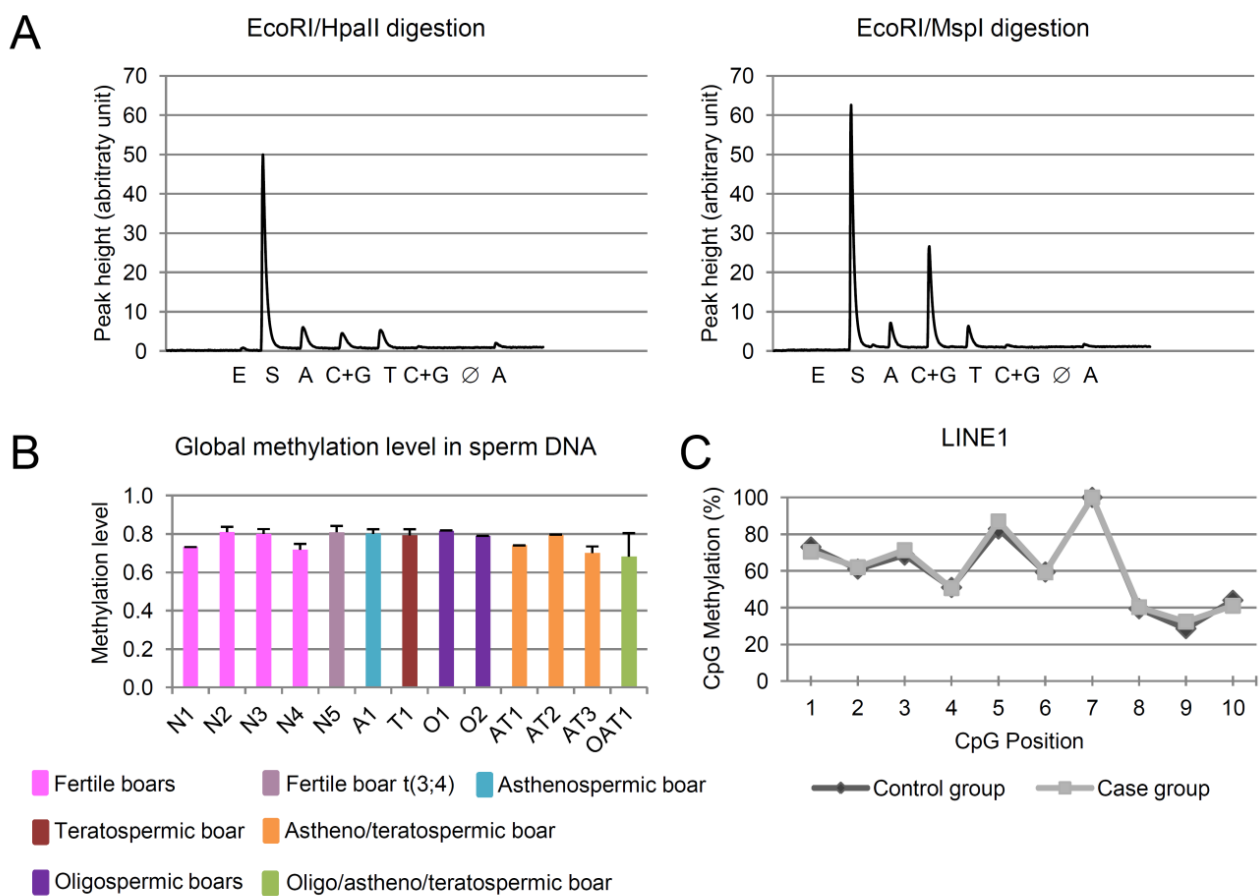


Figure 1

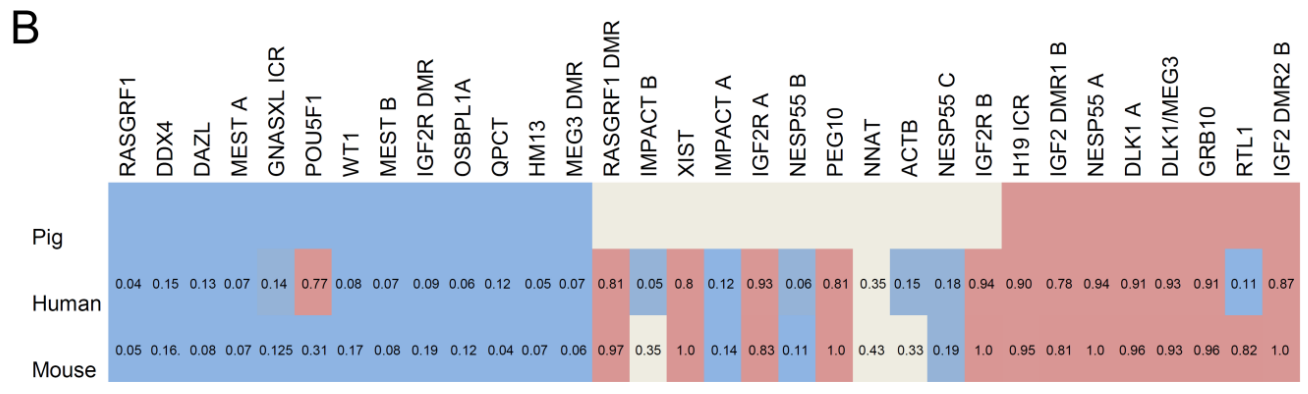
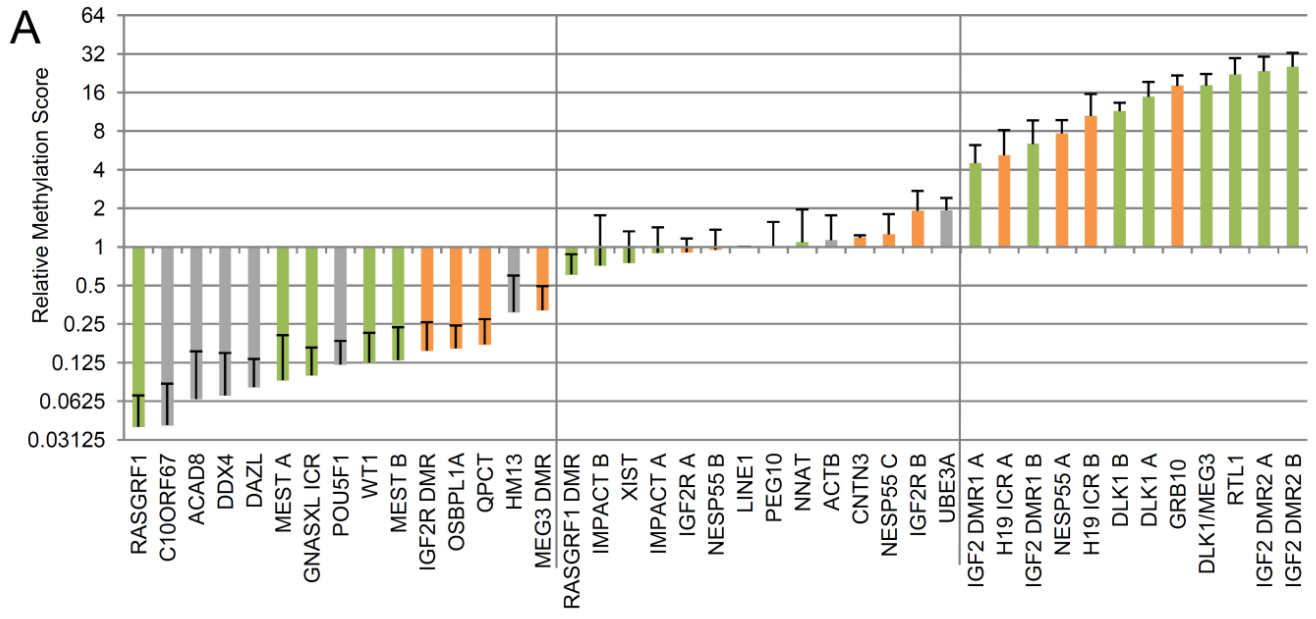


Figure 2

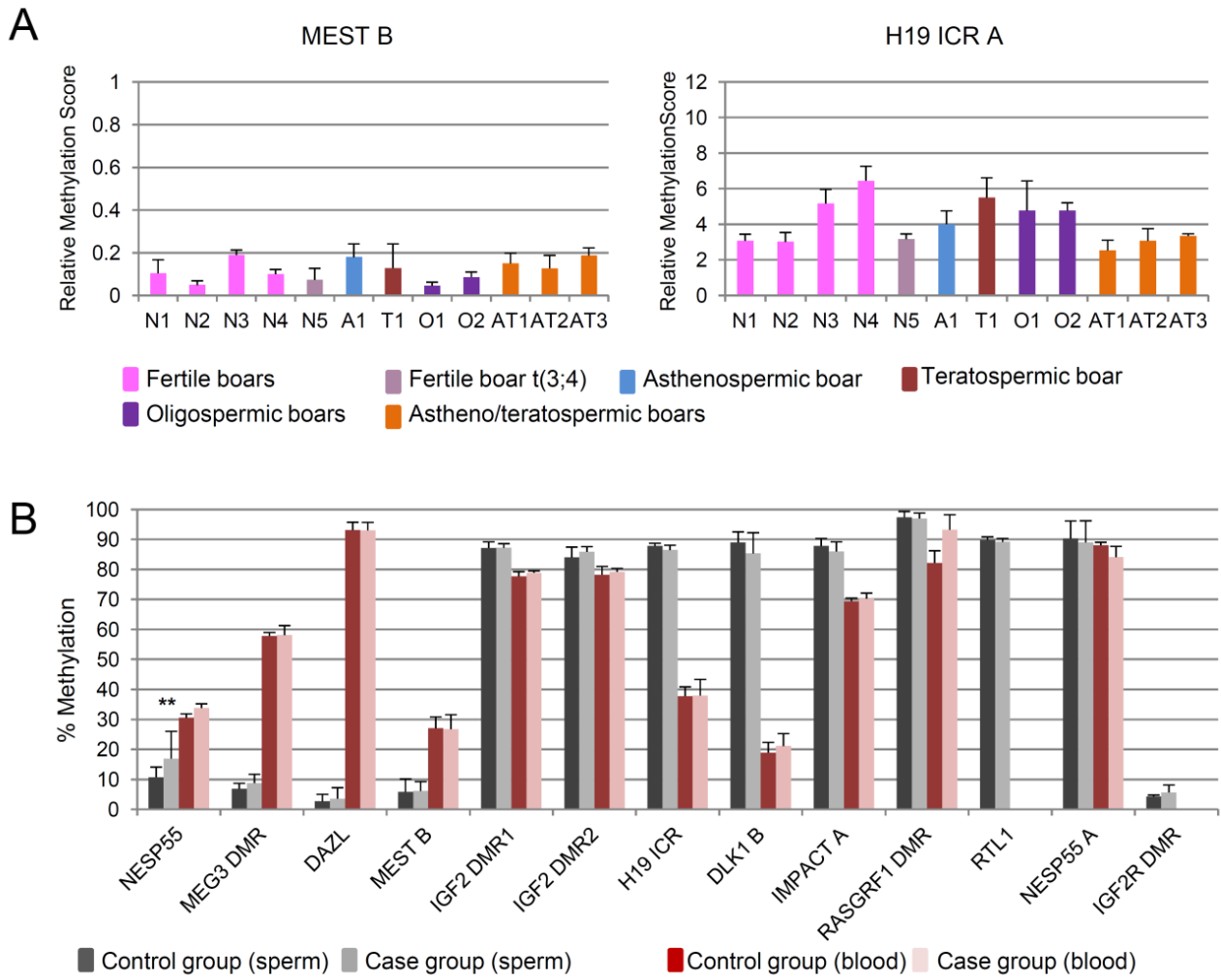


Figure 3

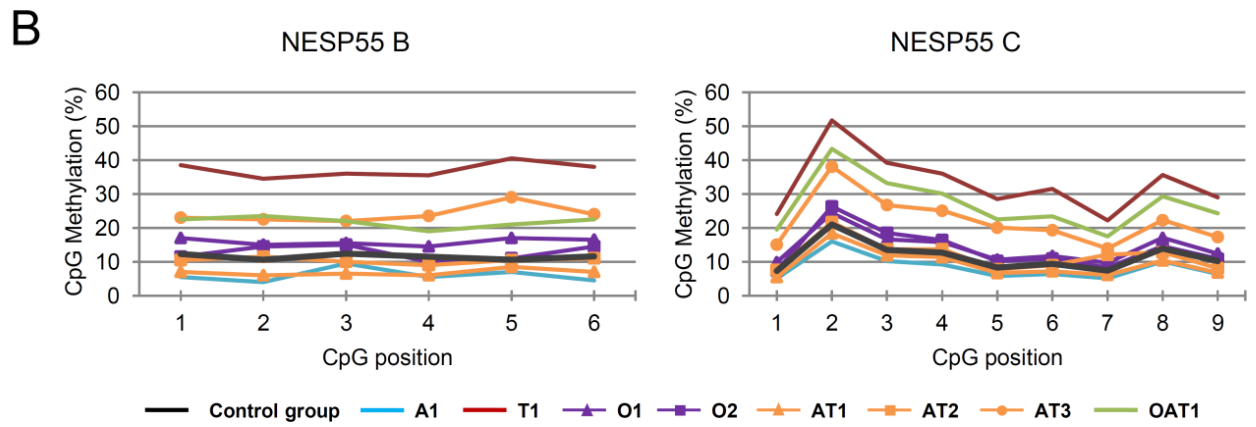
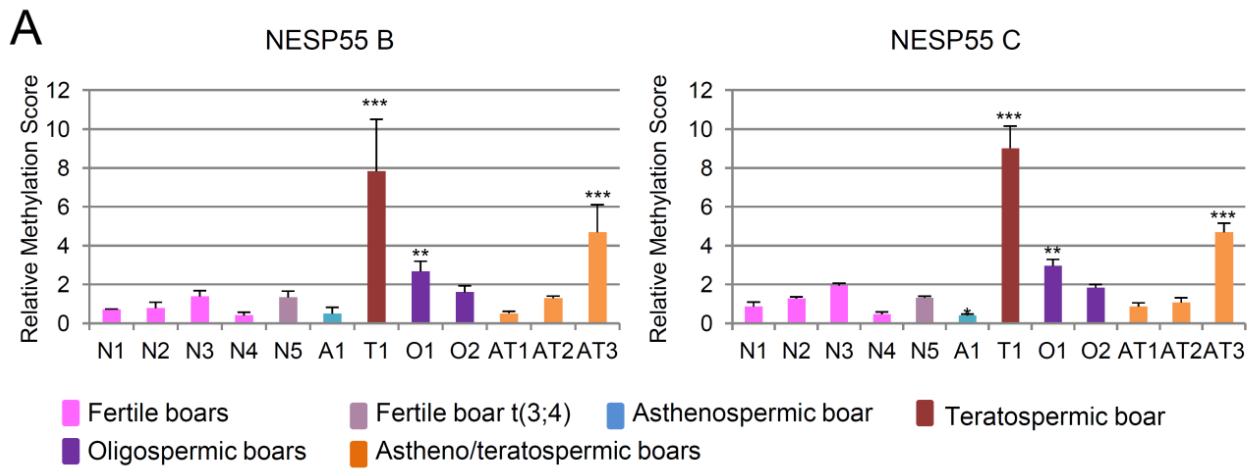


Figure 4

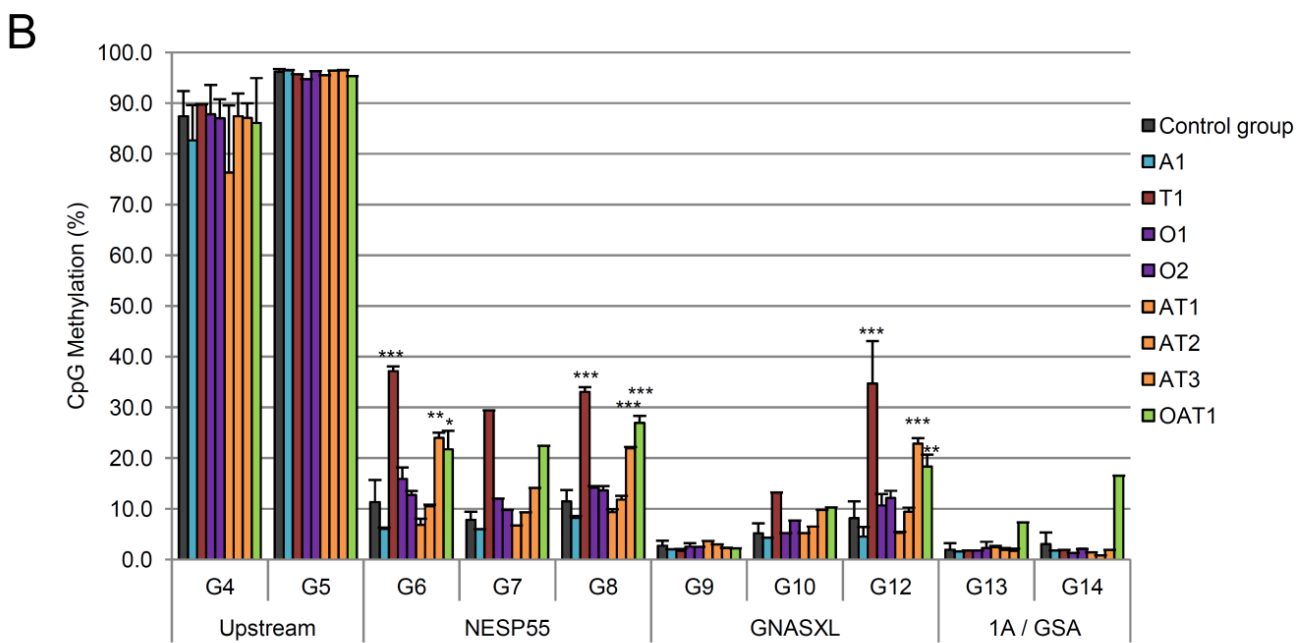
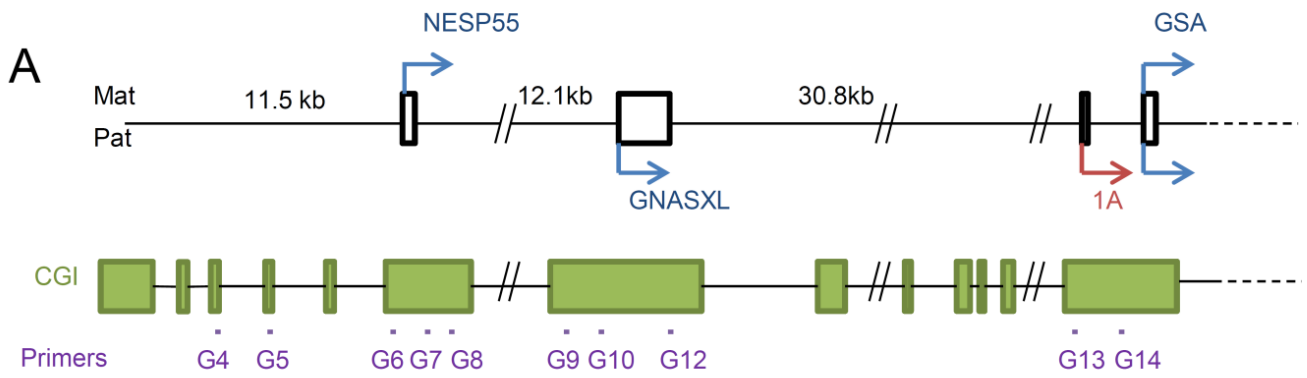


Figure 5

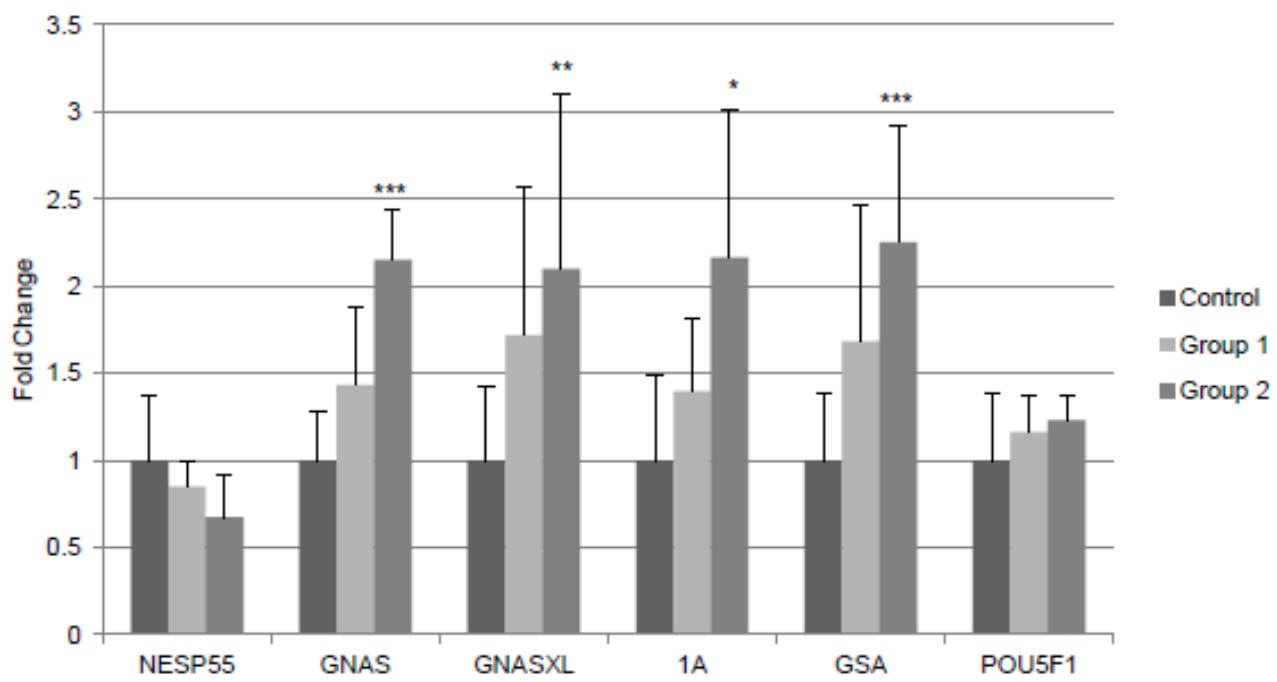
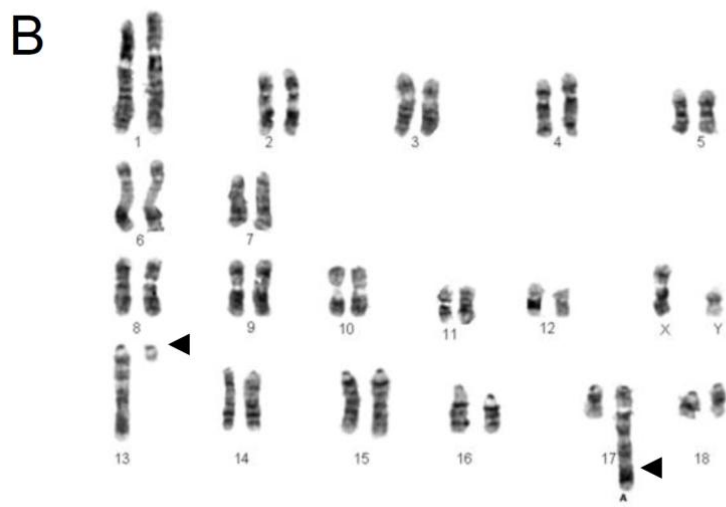
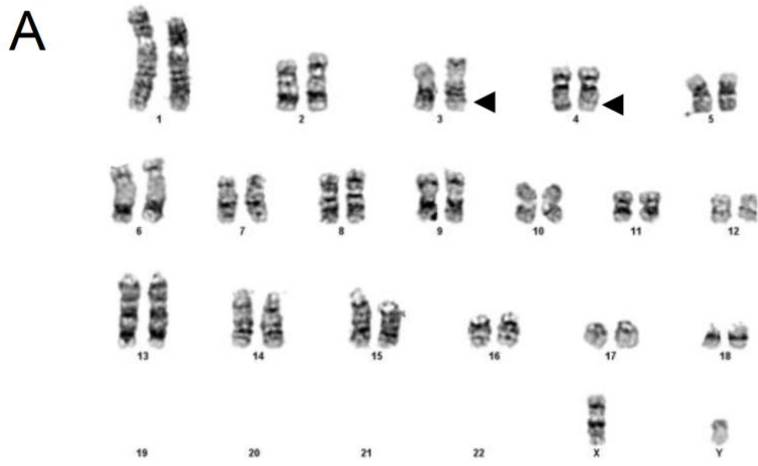
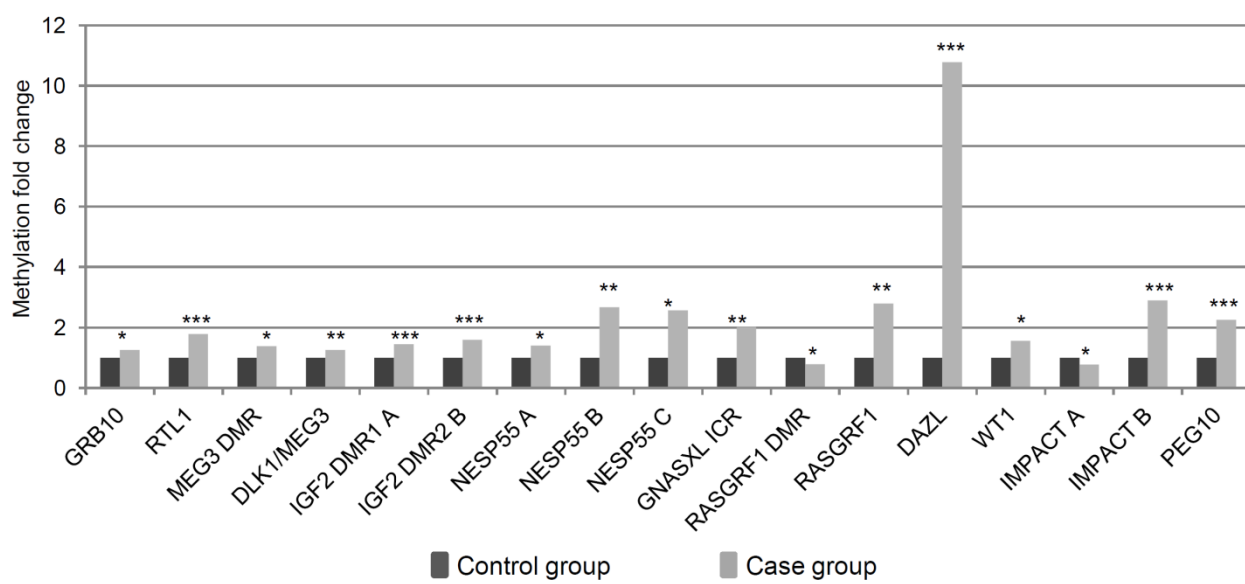


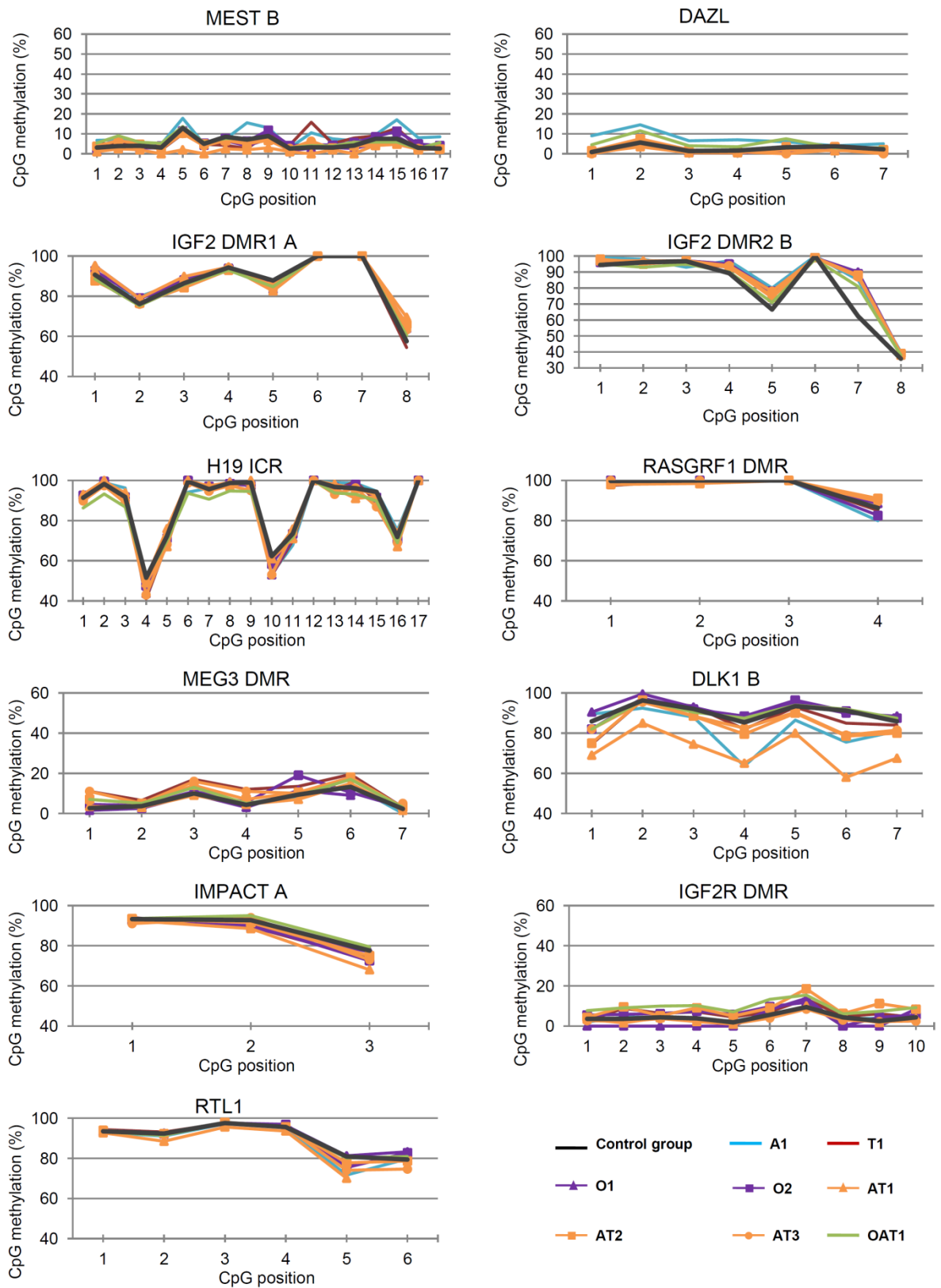
Figure 6



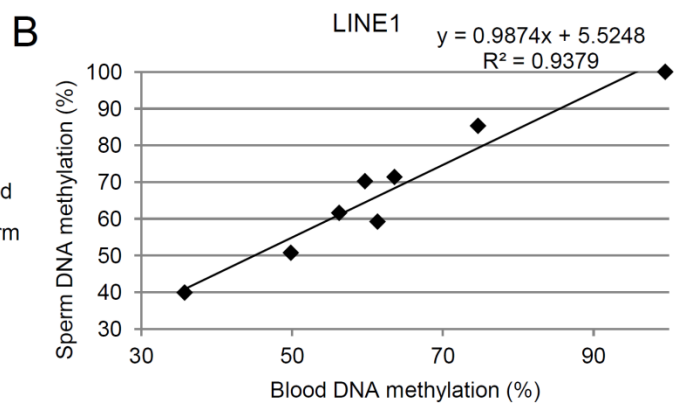
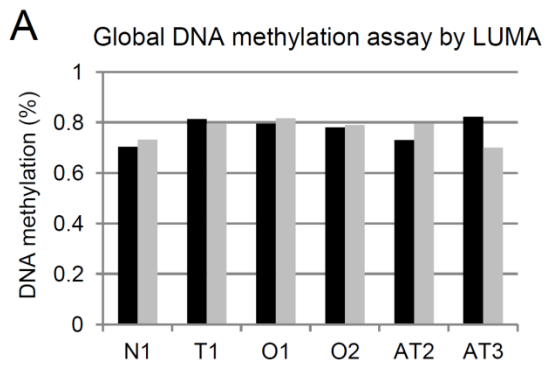
Supplementary Figure S1



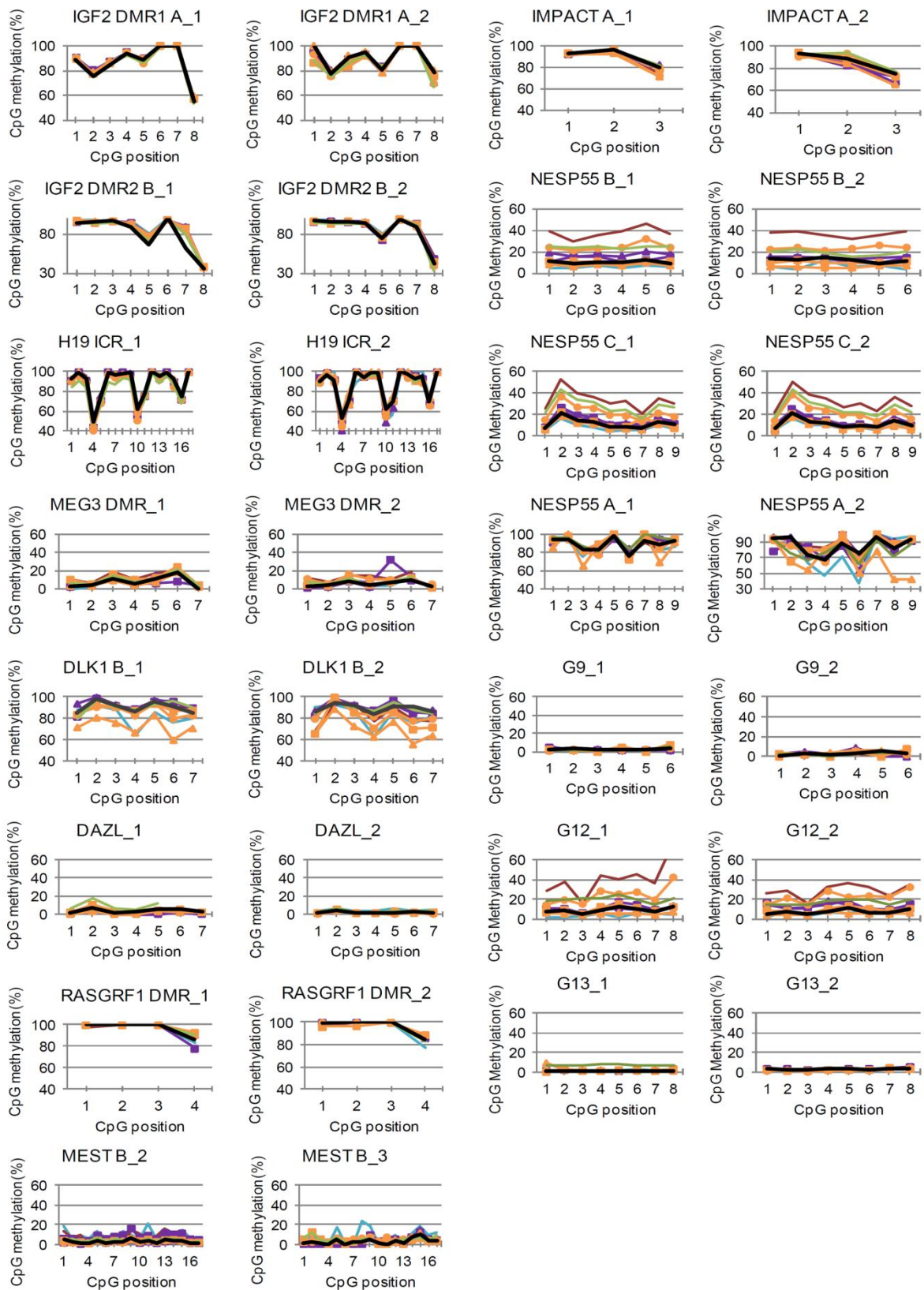
Supplementary Figure S2



Supplementary Figure S3



Supplementary Figure S4



Supplementary Figure S5

Supplementary table S1 : Amplification and sequencing primers used to detect parental imprint by pyrosequencing.

Name		Sequence 5' to 3'
CNTN3 exon 31	F	[Btm]CAATGGTCATATGAAATCTACCTG
	R	AACTTTCTCTGCAAGGTCTGTAAA
	S	CTCTGCAAGGTCTGTAAA
WT1 exon 9	F	AACGACTTTGCAAGCTGACTT
	R	[Btm]CAAACCTGGTAGGATCCAGTG
	S	CTTCATCCAGCGTCC
RASGRF1 exon 15	F	[Btm]TTGCTGAAAGGCGACATG
	R	AATATCCCCATCATCACAGGC
	S	TCAGCTGCAACTCAA
IMPACT exon 11	F	GCACAACCTGTAGATGAAAGAAGA
	R	[Btm]GTTATGTAATTGTGGTGCCTGTTT
	S	AGGAAGATTTGATAAGAATG
QPCT exon 8	F	[Btm]GCCATACTTCAGGGAAAGGAGAT
	R	ATTTGGTTTGCTCAAGGATCACTC
	S	TGTGATTCAGGATGACC
GRB10 exon 7	F	GTGGGATTTTTGAAAACTCATAT
	R	[Btm]GCGAGCGAAAGCAAATTTTTAT
	S	ATTTTTGAAAACTCATATT

Supplementary Table S2: Primer pairs used for MeDIP-qPCR

Name		Sequence 5'-3'
ACTB	F	TCCATGGGAACAGCTTCTCT
	R	TTGCTAAGGGGGCACATTAC
RTL1	F	CCTCCTGGCTCATGGAGTAG
	R	CCAGAGCGTGATCACTTCA
DLK1/MEG3	F	CTGCATCTTCGTGCACTTGT
	R	AACGTTTAAATGCGCAGGAG
DLK1 A	F	TGTCCCGATGACAAAGAAT
	R	TGGAGATGCTTCTGATTGA
DLK1 B	F	GCGTGACGTGTGAAAATCAA
	R	CCGTTGACCTGAGAAGGAAA
MEG3 DMR	F	GTTAGTGCAAAGCCGACTCC
	R	CCTAGCAGCCCTTCTCACAC
NESP55 A	F	ACTCCTTCCCTTGAGCTTCC
	R	GCTGCGGTAAGAGCAAGTCT
NESP55 B	F	TCTTACCTCGCAAGGAGCAC
	R	GGGCACAGATCGCTGTAAGT
NESP55 C	F	CACAAAACCTCACCCGACTT
	R	TTTACCCCGCAAACATTCTC
GNASXL ICR	F	CCCTGTGTTAGGTGGGTTTG
	R	AACGCCCTAAGAGGTTGTGA
GRB10	F	GTGGCAGAAGGTGAGGAAGA
	R	GGATCAAATTTCCGATCCTGT
H19 ICR A	F	GTGCAGGCTTACACATCTCG
	R	CTAAGGTGGCTGGAGAGAGC
H19 ICR B	F	GCGGCTTTAGGGGGATAC
	R	CCGCTTCTACCTCCACACC
IGF2 DMR1 A	F	TGACAACCCTGCCCTGTTAG

	R	GGACCTTGGAGGTAGGAAGG
IGF2 DMR1 B	F	GGGAGACAGGCCATATGTTT
	R	GGTCCCCTTAAAGCCTGACT
IGF2 DMR2 A	F	CAGCGGGATACCGTAATGAT
	R	TTCTCCTCTCCCCACTTCT
IGF2 DMR2 B	F	GGAAGAGTGCTTCCGTA
	R	AGAGAGGGGCTGCCTTAGG
IGF2R I2 DMR	F	TTAAGATGCCAACGGAGAGG
	R	CGCCTATTTGGTCGATTTTC
IGF2R A	F	GACCAGCAACTGCAATCAGA
	R	AAAGCTATGCGTCCGCTAAA
IGF2R B	F	AGGTCTCACCTCTTCAGGTT
	R	CTGTGCAAATTAAGGCTTCT
LINE1	F	CACTGTTAGCCCCATCAACA
	R	GCAGTGTTCCTTGTCACT
MEST A	F	ACTCTAGTCCGCACCCATT
	R	ATGGGATAAGACGCCACCAT
MEST B	F	TGCGTGATTCAACTTCGTTT
	R	CTCCCAGAACCTCGCTGTAG
POU5F1	F	GCCCTAGGAAGTGGAAGGAC
	R	CTAGGCGCCGAATTTAGAGA
RASGRF1 DMR	F	TCCACACAGAACATGGCTTC
	R	ACGGTTTGCGCATTTTTAAG
RASGRF1	F	ACTCCACGCGTTACATCTT
	R	CCCGGAGTGTAGAGATTTGC
DDX4	F	CTGAAGCCTCAAAAAGCCAAC
	R	TCAACAACAAATGGCACCAG
UBE3A	F	GTCTGCAACCTCCACCACA
	R	TCCGACTAGGAACCATGAGG
IMPACT A	F	GTCAGGGATCCTTCATCCAA
	R	AGCTGCCATTGCTTCAATTT
IMPACT B	F	AATGATGAGTCTCGGGCTGT
	R	CCGCTTGTCTTGTCTCTTC
DAZL	F	GAACACTGGGTGTCGCAGTA
	R	CACTCCCAGCCTCTCGAC
OSBPL1A	F	GCTCCTAGCCCAGCATCTC
	R	TTCCCCTCTTGCCTTTGTAG
ACAD8	F	CGGGACTACAACCTCCAGAC
	R	AGGCTCTAGCCCTCACACCT
QPCT	F	AAAGCTGGTCTCCATTTACGC
	R	GCTGAAGTTTCCAGGCAGTC
WT1	F	CCTTACCCTGGTTGCGAATA
	R	CGCCTTCACTGTCCATTCT
HM13	F	AGCAACACAGGCAGGAAAAG
	R	AAGCGAATCGTTCAAAGGAA
CNTN3	F	GTTGAATCGGAGCTGTAGCC
	R	TCCGACTAGGAACCATGAGG
C10ORF67	F	TCATTACGTACAGGGCATCG
	R	CGTGTGGATTCTGGACTTT
XIST	F	ATTCCTGAGGTTTGGGTAAT
	R	AGTGCAGTTTGCCAAATTAT
NNAT	F	CGACAATACCAGATTCTTC
	R	CTTGGTCCAGATCAGAATGT
PEG10	F	GTTGTTAATGGCTGGAAGAG
	R	AGTCACTCCCCTTCTAAG
PEG3	F	AACTGCCCAAGCAGTACCAC
	R	TGGAGATGGAGCCTTACCAC

Supplementary Table S3: Amplification and sequencing primers used for pyrosequencing of bisulfite-converted DNA.

Name		Sequence 5' to 3'
IGF2 DMR1 A	F	TTGGAAGTTGGGGAGGTT
	R	[Btm]CACCAACCCTAAAAAACTTCTACTAATAC
	S	GTAGTGATTGGATT
MEST B	F	GAGTTTTTGGAGGGGTGGTTAA
	R	[Btm]ACCAACCCCAAACCAAACTT
	S	AGTTAATTGAGTGATTTAAT
H19 ICR	F	TTTTTGGTAGTGGGTTGATTTTTATAGG
	R	[Btm]AACACCCCATCCCCTAAATA
	S	GAGGAGGTAGTGTAGG
IGF2 DMR2 B	F	[Btm]GTTGAGATTTGTTTTGTTGGAGATTTAT
	R	AAACCCCTCTAACAACACTTAA
	S	CTACCCAAAAAAAACCTACCTT
IGF2R DMR	F	TGGGGTGAAGGGTGAAA
	R	[Btm]CCCCCTTCTAACCTCAAC
	S	GTGAAGGGGTGAAAA
IMPACT A	F	TGAGGAAATTGAAGTAATGGTAGT
	R	[Btm]ATATACCTACAAACAAAATATCCACTTA
	S	GAAATTGAAGTAATGGTAGT
RASGRF1 DMR	F	GGGAAAATGGATATAGGGATGTAT
	R	[Btm]CAACCCTAAATCAAAAATCAAAAAACCTT
	S	TGTTGAGTTATTGGAGG
MEG3 DMR	F	[Btm]GGTTAGAAATTAGAGGGAAAAT
	R	CACTAAAATCCAAACCCCAAATT
	S	CCAAACCCCAAATTTT
DLK1 B	F	TTTGGGGAGAGAAGAGGAAGTT
	R	[Btm]ATCCTAAAATCACACCTCCTAATATTC
	S	AAGAGGAAGTTAGTGAT
RTL1	F	[Btm]GGTTTTTGTTTAGGGAGTAGTAG
	R	AACTACCAACCCTTTCTAAT
	S	AAACCCCATCATCCC
DAZL	F	AGGAGAGGGTGAGAGTTG
	R	[Btm]CTCTCTACCCCTCAAACCCCTATA
	S	TTTTTGTGTTTTTTTTGGGA
LINE1	F	GGGAGAAAAGAGGAGAGTGAGAGA
	R	[Btm]ACTCCCACTTACTACTACTCTTCTCTAC
	S	AGAGTGAGAGAAGGT
G4 (GNAS locus)	F	AGTTTTTTAGTTAAATGGGAAAGTTTAT
	R	[Btm]AACCCATCCCTAAAAAACCAATAAAAT
	S	AGTAGAGATTTTAGGAAGAAATAG
G5 (GNAS locus)	F	AAGGGTTTGAGTTGAAAGGA
	R	[Btm]AACCTCAAATTTTTTCTCTACTC
	S	AGTTGAAAGGATGGT
G6 (GNAS locus)	F	[Btm]AGGTGGGGTTGGGATTTT
	R	ACCCACCCTCTAACTCTACAAA
	S	CTAACTCTACAAAAACAACCT
G7 (GNAS locus)	F	GTAATGAGAGAAAAGTAATTAGATTGGTAT
	R	[Btm]CCCAAATTAACAACCACTAAACCTA
	S	AAAAGTAATTAGATTGGTATT
G8 (GNAS locus)	F	TGTGTTTTTTTAGTTAATTTTTGGGGGAGAA
	R	[Btm]AAACCCTTTAACTATATTCCTTCTA
	S	GGGAGAAGATAGATTAGTGAAAT

G9 (GNAS locus)	F	[Btn]GTTATTAGAGAGAGTAAGGGTTGGTT
	R	CCCTAAAAAATTATAAAACTAAAATAAATT
	S	ATAACCACCAAACCTAAAATAC
G10 (GNAS locus)	F	[Btn]GTTTGGGAGGGAGGGTTTTAGT
	R	CTCCTTCCAATCCTCTCCCTAACTAATTT
	S	TCTCCCCTAACCCAAA
G12 (GNAS locus)	F	AGAAAGAGAAGATGGGTTATATGTG
	R	[Btn]AAACCCTTCACTAAAACCTTCTAAC
	S	GATTTTGTAGTGGAGAGTA
G13 (GNAS locus)	F	TTAATAGGGAAGTTGGAGGTAAGG
	R	[Btn]CCAAACCTACCTCCCAATTTTC
	S	AGTTGGAGGTAAGGA
G14 (GNAS locus)	F	GGTATAGTGGGTATGAGTAGGATT
	R	[Btn]AACCCCAACCAAATACCCCTAAA
	S	ATTTGGGTAGAGTGGGAA

Supplemental table S4 : Primers pairs used for expression analysis by RT-qPCR in testicular tissues.

Name		Séquence 5'-3'
NESP55	F	TCTTACCTCGCAAGGAGCAC
	R	GGGCACAGATCGCTGTAAGT
GNAS	F	TGCTCGCTACACTACTCCTG
	R	CAGGTGAAGTGAGGGTAGCA
GNASXL	F	CAACTTCTCGTCCAGGCTT
	R	CAGGATCCTCATCTGCTTCACA
1A	F	GGACACTCAGTCGCGTTCG
	R	CAGGATCCTCATCTGCTTCACA
GSA	F	CAGTAAGACCGAGGACCAGC
	R	CAGGATCCTCATCTGCTTCACA
POU5F1	F	AGTGCCCAAAGCCCCTCT
	R	TTCTGGCGACGGTTGCA
TBP	F	AACAGTTCAGTAGTTATGAGCCAGA
	R	AGATGTTCTCAAACGCTTCG

Supplementary Table S5: Genomic coordinates and CpG methylation score of human and mice homologous regions

Gene names	Pig genome sscrofa10.2	Human genome hg18	Mouse genome mm10	Relative Methylation score pig	Absolute Methylation score at CpG human (in %)	Absolute Methylation score at CpG mouse (in %)
RASGRF1	chr7:53,784,911-53,785,053	chr15:77,169,916-77,170,0430	chr9:89,909,830-89,909,960	0.04	0.04	0.05
DDX4	chr16:37,047,974-37,048,093	chr5:55,069,454-55,069,600	chr13:112,652,400-112,652,850	0.07	0.14	0.16
DAZL	chr13:4,036,859-4,037,100	chr3:16,622,050-16,622,200	chr17:50,293,500-50,293,800	0.08	0.13	0.08
MEST A	chr18:19,341,929-19,342,047	chr7:129,919,400-129,919,700	chr6:30,738,300-30,738,450	0.09	0.07	0.07
GNASXL ICR	chr17:66,314,745-66,314,957	chr20:56,859,100-56,859,300	chr2:174,295,240-174,295,498	0.10	0.14	0.125
POU5F1	chr7:27,263,237-27,263,354	chr6:31,245,700-31,245,900	chr17:35,506,550-35,506,700	0.12	0.68	0.31
WT1	chr2:30,902,723-30,902,901	chr11:32,412,935-32,413,135	chr2:105,127,270-105,127,480	0.12	0.08	0.17
MEST B	chr18:19,347,216-19,347,307	chr7:129,913,972-129,914,200	chr6:30,732,400-30,732,700	0.13	0.07	0.08
IGF2R DMR	chr1:9,245,108-9,245,270	chr6:160,347,300-160,347,500	chr17:12,742,500-12,743,000	0.15	0.09	0.19
OSBPL1A	chr6:101,984,829-101,984,928	chr18:20,231,420-20,231,530	chr18:12,941,400-12,941,620	0.16	0.06	0.12
QPCT	chr3:109,359,903-109,360,002	chr2:37,425,800-37,426,300	chr17:79,052,700-79,052,900	0.17	0.12	0.04
HM13	chr17:40,074,655-40,074,837	chr20:29,624,700-29,624,900	chr2:152,707,700-152,707,950	0.31	0.05	0.07
MEG3 DMR	chr7:132,163,735-132,163,884	chr14:100,362,700-100,362,900	chr12:109,541,600-109,541,800	0.32	0.07	0.06
RASGRF1 DMR	chr7:53,820,144-53,820,295	chr15:77,213,300-77,213,500	chr9:89,872,350-89,872,550	0.61	0.81	0.86
IMPACT B	chr6:102,002,040-102,002,238	chr18:20,260,270-20,260,450	chr18:12,971,900-12,972,150	0.72	0.05	0.35
XIST	chrX:67,157,193-67,157,325	chrX:72,962,540-72,962,740	chrX:103,465,600-103,465,800	0.75	0.8	1
IMPACT A	chr6:102,003,692-102,003,835	chr18:20,261,760-20,261,960	chr18:12,974,550-12,974,750	0.90	0.12	0.14
IGF2R A	chr1:9,248,709-9,248,851	chr6:160,343,950-160,344,150	chr17:12,743,500-12,744,000	0.91	0.93	0.83
NESP55B	chr17:66,304,289-66,304,459	chr20:56,848,400-56,848,600	chr2:174,284,400-174,284,700	0.95	0.06	0.11
PEG10	chr9:81,652,380-81,652,519	chr7:94,133,930-94,134,150	chr6:4,757,900-4,758,100	1.01	0.79	1
NNAT	chr17:46,045,193-46,045,324	chr20:35,584,700-35,584,900	chr2:157,561,800-157,562,100	1.09	0.35	0.43
ACTB	chr3:4,733,279-4,733,511	chr7:5,537,500-5,538,000	chr5:142,907,100-142,907,600	1.13	0.15	0.33
NESP55 C	chr17:66,306,603-66,306,705	chr20:56,850,700-56,850,900	chr2:174,285,600-174,285,800	1.25	0.18	0.19
IGF2R B	chr1:9,039,118-9,039,232	chr6:160,447,000-160,447,200	chr17:12,682,800-12,683,000	1.91	0.94	1
H19 ICR A	31841-32007 AY044827	chr11:1,978,450-1,978,600	chr7:142,580,400-142,580,700	5.18	0.9	0.95
IGF2 DMR1 B	17423-17667 AY044828	chr11:2,122,900-2,123,150	chr7:142,665,000-142,665,350	6.34	0.78	0.81
NESP55 A	chr17:66,295,668-66,295,785	chr20:56,834,307-56,834,500	chr2:174,273,200-174,273,400	7.65	0.94	1
DLK1 A	chr7:132,195,144-132,195,248	chr14:100,273,500-100,273,700	chr12:109,462,600-109,462,800	14.84	0.91	0.96
GRB10	chr9:150,301,782-150,301,877	chr7:50,658,500-50,658,800	chr11:11,948,480-11,948,836	18.03	0.91	0.96
DLK1/MEG3	chr7:132,176,228-132,176,330	chr14:100,350,399-100,350,600	chr12:109,530,600-109,530,900	18.14	0.93	0.93
RTL1	chr7:132,099,126-132,099,220	chr14:100,418,600-100,418,800	chr12:109,592,350-109,592,550	22.16	0.11	0.82
IGF2 DMR2 B	29112-29235 AY044828	chr11:2,111,150-2,111,400	chr7:142,654,150-142,654,450	25.38	0.87	1

Supplementary Table S6: Variation in methylation level between control and case groups by MedIP-qPCR (Wilcoxon rank sum test; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$).

Name	Variation	Significance	P-value	Fold Change
ACTB	No		0.0670	0.73
RTL1	Increase	***	2.61e-5	1.78
DLK1/MEG3	Increase	**	0.0075	1.25
DLK1 A	No		0.7187	1.05
DLK1 B	No		0.1354	0.89
MEG3 DMR	Increase	*	0.0168	1.38
NESP55 A	Increase	*	0.0436	1.40
NESP55 B	Increase	**	0.0087	2.67
NESP55 C	Increase	*	0.0253	2.56
GNASXL ICR	Increase	**	0.0074	2.02
GRB10	Increase	*	0.0101	1.25
H19 ICR A	No		0.4722	1.29
H19 ICR B	No		0.5744	1.04
IGF2 DMR1 A	Increase	***	0.0005	1.45
IGF2 DMR1 B	No		0.7467	1.52
IGF2 DMR2 A	No		0.117	0.86
IGF2 DMR2 B	Increase	***	2.44e-5	1.59
IGF2R DMR	No		0.4811	1.18
IGF2R A	No		0.8651	1.10
IGF2R B	No		0.7467	1.01
MEST A	No		0.1944	1.67
MEST B	No		1.8734	1.01
POU5F1	No		0.3135	1.51
RASGRF1 DMR	Decrease	*	0.0477	0.78
RASGRF1	Increase	**	0.0043	2.79
DDX4	No		0.7376	1.17
UBE3A	No		0.1549	1.17
IMPACT A	Decrease	*	0.0253	0.78
IMPACT B	Increase	***	0.0006	2.89
DAZL	Increase	***	4.82e-6	10.78
OSBPL1A	No		0.9000	1.22
ACAD8	No		0.1922	2.83
QPCT	No		0.1216	0.73
WT1	Increase	*	0.0460	1.56
HM13	No		0.9753	1.28
C10ORF67	No		0.5052	2.87
XIST	No		0.8716	0.96
PEG10	Increase	***	0.0008	2.26

ANNEXE 2 : Liste des réactifs et matériels utilisés

Nom	Fournisseur	Référence
Dulbecco's Modified Eagle's Medium - high glucose	Sigma-Aldrich	D0819
Dulbecco's Modified Eagle's Medium/Nutrient Mixture F-12 Ham	Sigma-Aldrich	D6421
Opti-MEM® Reduced Serum Medium, GlutaMAX™ Supplement	Gibco	51985-034
KnockOut Serum Replacement	Gibco	10828-028
Foetal Bovine Serum	Gibco	10270-106
PANSera ES	Pan Biotech	P30-2602
100mM Sodium Pyruvate	Gibco	11360-039
100X MEM Non Essential Amino Acids Solution	Gibco	11140-035
200mM L-Glutamine	Gibco	25030081
2-mercaptoethanol	Gibco	31350-010
Pénicilline	Sigma-Aldrich	6501
Streptomycine	Sigma-Aldrich	PEN-NA
PBS, pH 7.4	Gibco	10010-015
StemPro® Accutase® Cell Dissociation Reagent	Life Technologies	A1110501
0.25% Trypsine-EDTA (1X), phenol red	Gibco	25200-056
Fibroblast Growth Factor basic Protein, Human recombinant	Millipore	GF003
ESGRO Mouse LIF Medium Supplement	Millipore	ESG1106
N-2 Supplement (100X)	Gibco	17502-048
B-27 Supplement (50X), serum free	Gibco	17504-044
PD 0325901	Sigma-Aldrich	PZ0162
1-Azakenpauillone	Sigma-Aldrich	A3734
Plasmocin™ Treatment	Invivogen	ant-mpt
BD Matrigel	BD Biosciences	354277
Gélatine	Sigma-Aldrich	G9391-100G
Puromycin	Invivogen	ant-pr-1
Colcemide	Sigma-Aldrich	D7385
Boîtes de cultures	Corning	3548, 3526, 3513, 3516, 430166, 430720
Paraformaldehyde	Subra	10440051000
NBT/BCIP	Roche	11681451001
Anti-human Nanog	Peptotech	500-P236
Anti-Oct4 antibody - ChIP Grade	Abcam	ab19857
Human/Mouse Sox2 MAb (Clone 245610), Mouse IgG2A	R&D Systems	MAB2018
Anti-Lin28 antibody	Abcam	ab63740
E-cadherin	BD Biosciences	610181
Anti-SSEA4 [MC813] antibody	Abcam	ab16287
Anti-SSEA1 [MC480] antibody	Hybridoma Bank	MC-480
Anti SSEA1-AF647 (480)	Santa Cruz	SC21702
Anti-SSEA4 (813-70)	Santa Cruz	SC21704
Anti-Mouse IgG F(ab') ₂ -APC	Santa Cruz	SC3818
Anti-Rabbit 568	Life Technologies	A11011
Alexa-Fluor Goat Anti-Mouse 568 (IgG)	Life Technologies	A11004
VECTASHIELD Mounting medium with DAPI	Eurobio	H-1200
PL-SIN-EOS-C(3+)-EiP	SBRI, Marielle Afanassieff	
pMXs-hOct4 ; pMXs-hSox2 ; pMXs-hKLF4, pMXs-hc-Myc	SBRI, Marielle Afanassieff	
CytoTune™ iPS Reprogramming Kit	Life Technologies	A13780
Propidium Iodide	Invitrogen	P3566
Doxorubicin hydrochloride	Sigma-Aldrich	44583
RNase A	Sigma-Aldrich	R5503-1G
BioPrime DNA Labeling SYStem	Invitrogen	18094-011
ChromaTide® Alexa Fluor® 488-5-dUTP	Invitrogen	C-11397
ChromaTide® Alexa Fluor® 568-5-dUTP	Invitrogen	C11399
Biotin-16-dUTP	Roche	11093070910
Streptavidin, Alexa Fluor® 488 conjugate	Invitrogen	S11223
Hybridization Solution	Sigma	H7782
Giemsa stain	Sigma	11700
Porcine Hybloc DNA	Applied Genetics Laboratories	PHB-0.5
Illustra Microspin G-50 columns	GE Health Care Life	27-5330-01

	Sciences	
Formamide	Sigma-Aldrich	47671-1L-F
SSC 20X	Fisher Scientific	11598976
Rubber Cement	Euromedex	72172
Proteinase K	Eurobio	GEXPRK010D
Oligonucléotides	Sigma-Aldrich	
GenElute™ Mammalian Genomic DNA Miniprep Kit	Sigma-Aldrich	G1N70
Nucleospin RNA	Macherey-Nagel	740955.50
Superscript® II Reverse Transcriptase	Life Technologies	18064-014
RNasin® RNase Inhibitor	Promega	N2511
5X Green GoTaq® Reaction Buffer	Promega	M7911
GoTaq® DNA Polymerase	Promega	M3178
FastStart SYBR Green Master mix	Roche	04673484001
SYBR Green PCR Master Mix	Applied Biosystem	4309155
BigDye Terminator v3.1 cycle sequencing kit	Applied Biosystem	4337455
Ethanol absolu AnalaR Normapur	VWR	20821.296
Phenol TRIS	Eurobio	GEXPHE0301
Chloroforme AnalaR Normapur	VWR	22709.292
Alcool isoamylique	Merck Millipore	818969
TriZol Reagent	Invitrogen	15596-018
Phase Lock Gel Light 2mL	5Prime	2302820
MethylEdge™ Bisulfite Conversion System	Promega	N1301
PyroMark Gold Q24 Reagents	Qiagen	970802
PyroMark PCR kit	Qiagen	978703
PyroMark Binding Buffer	Qiagen	979006
PyroMark Denaturation Solution	Qiagen	979007
PyroMark Wash Buffer	Qiagen	979008
PyroMark Annealing Buffer	Qiagen	979009
Streptavidin Sepharose High Performance	GE Healthcare	17-5113-01
Anti-5-methylcytosine, clone 33D3	Millipore	MABE146
5-hmC monoclonal antibody (mouse)	Diagenode	C12500200 Mab-31HMC
Dynabeads-ProteinG	Life Technologies	10009D
iPure kit	Diagenode	C03010012
NEXTflex Methyl Sequencing Kit	Bioo Scientific	5118
Exonuclease I	New England Biolabs	M0293S
TSAP Thermosensitive Alkaline Phosphatase	Promega	M9910
HpaII	New England Biolabs	R0171S
MspI	New England Biolabs	R0106S
EcoRI	New England Biolabs	R3101S
PyroMark Q24	Qiagen	9001514
Light Cycler ® 480	Roche	
BioMark™ HD system	Fluidigm	
Bioruptor	Diagenode	
BD FACScalibur™	BD Biosciences	
TissueLyser	Retsch	
ABI3730	Applied Biosystem	
Zeiss Imager ZI	Zeiss	
Leica DM IRBE statif inversé	Leica	
CYTOVISION Genus	Applied Imaging	
Volocity	Perkin Elmer	
Hybridizer	DakoCytomation	S2450
Pyromark Assay Design 2.0	Qiagen	
Fluidigm Real-Time PCR Analysis	Fluidigm	
Seqmonk	Babraham Bioinformatics	
Integrative Genomics Viewer (IGV)	Broad Institute	
Ingenuity Pathway Analysis	Qiagen	
ChromasPro	Technelysium	

ANNEXE 3 : Tableau bilan des clones reprogrammés analysés

Le tableau ci-dessous présente les résultats obtenus pour l'ensemble des clones sélectionnés et analysés. Pour la majorité de ces clones, la morphologie, l'expression de la phosphatase alcaline et l'expression des gènes NANOG, OCT4, SOX2 et LIN28 ont été analysés. Le choix de conserver ou non ces clones a été réalisé sur plusieurs critères (morphologique, expressionnel, caryotypique, stabilité en culture) et seul un faible nombre a été analysé de manière plus approfondie. De plus, certains clones issus de fibroblastes normaux présentant des caractéristiques moléculaire de pluripotence ont du être éliminés car les cultures se sont avérées être des mélanges avec les lignées issues de fibroblastes t(Y;14) qui possèdent un fort avantage sélectif. Les caryotypes indiqués sont ceux obtenus à passage 10-20 et la mention anomalies indique l'accumulation d'anomalies dans des passages plus tardifs.

Infection	Type	Nom	GFP+	AP+	Morphologie iPS-like	Nanog+	Lin28B+	Oct4+	Sox2+	hOSKM+	Formation EB	Maintenance en LIF+2i	Phase G1 raccourcie	Point de contrôle G1/S	Caryotype
V1	N1 EOS	A	oui	oui	oui	oui	non	oui	non	oui					
V1	N1 EOS	B	non	non	oui	oui	non	non	non	oui					
V1	N1 EOS	C	oui	oui	oui	faible	oui	faible	oui	oui					
V1	N1 EOS	D	non	oui	oui	faible									
V1	N1 EOS	E	oui	oui	non	oui	non	oui	non	oui					
V1	N1 EOS	F	oui	oui	non	oui	non	oui	non	oui		non			
V1	N1 EOS	G	oui	oui	oui	faible	non	oui	non	oui					
V1	N1 EOS	H	oui	oui	oui	oui	non	oui	faible	oui					
V1	N1 EOS	I	oui	oui	oui	faible	non	oui	non	oui					
V1	N1 EOS	J	oui	oui	oui	oui	non	oui	non	oui					
V1	N1 EOS	K	non	oui	oui	non				oui					
V1	N1 EOS	L	non	oui	oui	faible	non	faible	non	oui					
V1	N1 EOS	M	oui	oui	oui	oui	non	oui	non	oui					
V1	N1 EOS	N	oui	oui	oui	oui	non	oui	non	oui					
V1	N1 EOS	O	oui	oui	oui	oui	non	w	non	oui					
V1	N1 EOS	P	oui	oui	oui	oui	non	oui	non	oui					
V1	N1 EOS	Q	oui	oui	oui	non									
V1	N1 EOS	R	oui	oui	oui	non									
V1	N1 EOS	S	oui	oui	oui	oui	non	oui	non	oui					
V1	N1 EOS	T	non	oui	oui	faible	non	non	non	oui					
V1	N1 EOS	U	non	oui	oui	faible									
V1	N1 EOS	V	non	oui	oui										
V1	N1 EOS	W	non	oui	oui										
V1	N1 EOS	A'	oui	oui	oui	oui	non	oui	non	oui					2n=38,XY
V1	N1 EOS	B'	oui	oui	oui	non									

Infection	Type	Nom	GFP+	AP+	Morphologie iPS-like	Nanog+	Lin28B+	Oct4+	Sox2+	hOSKM+	Formation EB	Maintenance en LIF+2i	Phase G1 raccourcie	Point de contrôle G1/S	Caryotype
V1	N1 EOS	C'	oui	oui	oui	non									
V1	N1 EOS	D'	oui	oui	oui	faible									
V1	N1 EOS	E'	oui	oui	oui	faible									
V1	N1 EOS	F'	oui	oui	oui	faible									
V1	N1 EOS	G'	oui	oui	oui	oui	non	oui	oui	oui					
V1	N1 EOS	H'	oui	oui	oui	oui	oui	oui	oui	oui					
V1	N1 EOS	I'	non	oui	oui	faible									
V1	N1 EOS G	A	non	oui	oui	oui	non	oui	non	oui					
V1	N1 EOS G	B	non	oui	oui	faible									
V1	N1 EOS G	C	oui	oui	oui	non	non	faible	non	oui					
V1	N1 EOS G	D	non	oui	oui	oui	non	oui	non	oui					
V1	N1 EOS G	E	non	oui	oui	non									
V1	N1 EOS G	F	non	oui	oui	non									
V1	N1 EOS G	G	non	oui	oui	oui	non	oui	non	oui					
V1	N1 EOS G	H	oui	oui	oui	non									
V1	N1 EOS G	I	non	oui	oui	oui	non	oui	non	oui					
V1	N1 EOS G	J	oui	oui	oui	non									
V1	N1 EOS G	K	non	oui	oui										
V1	N1 EOS G	L	oui	oui	oui	faible									
V1	N1 EOS G	M	non	oui	oui	non									
V1	N1 EOS G	N	oui	oui	oui	non									
V1	N1 EOS G	O	non	oui	oui										
V1	N1 EOS G	P	non	oui	oui										
V1	N1 EOS G	Q	oui	oui	oui	oui	non	oui	faible	oui					
V1	N1 EOS G	R	oui	oui	oui										

Infection	Type	Nom	GFP+	AP+	Morphologie iPS-like	Nanog+	Lin28B+	Oct4+	Sox2+	hOSKM+	Formation EB	Maintenance en LIF+2i	Phase G1 raccourcie	Point de contrôle G1/S	Caryotype
V1	N1 EOS G	S	non	oui	oui										
V1	N1 EOS G	T	non	oui	oui										
V1	N1 EOS G	U	oui	oui	oui	faible	non	oui	non	oui					
V1	N1 EOS G	V	oui	oui	oui										
V1	N1 EOS G	W	oui	oui	oui										
V1	N1 EOS G	X	oui	oui	oui										
V1	N1 EOS G	A'	non	oui	oui										
V1	N1 EOS G	B'	non	oui	oui										
V1	N1 G	A			oui										
V1	N1 G	B		oui	oui										
V1	N1 G	C		oui	oui										
V1	N1 G	D		oui	oui										
V1	N1 G	E		oui	oui										
V1	MEF G	A		oui	oui										
V1	MEF G	B		oui	oui										
V1	MEF G	C		oui	oui										
V1	MEF G	D		oui	oui										
V1	MEF G	E		oui	oui										
V2	AT3 EOS	B	oui	non	oui										
V2	AT3 EOS	C	oui	oui	oui										
V2	AT3 EOS	D	non	oui	oui										
V2	AT3 EOS	E	non	non	oui										
V2	AT3 EOS	F	non	non	oui										
V2	AT3 EOS	G	non	non	oui										
V2	AT3 EOS	H	non	non	oui	oui	oui	oui	non	oui					mélange cellulaire

Infection	Type	Nom	GFP+	AP+	Morphologie iP5-like	Nanog+	Lin28B+	Oct4+	Sox2+	hOSKM+	Formation EB	Maintenance en LIF+2i	Phase G1 raccourcie	Point de contrôle G1/S	Caryotype
V2	AT3 EOS (LIF)	1	oui	oui	oui	oui	faible	faible	non	oui					2n=39
V2	AT3 EOS (LIF)	2	oui	non	oui										
V2	AT3 EOS (LIF)	3	non	non	oui										
V2	AT3 EOS (LIF)	4	oui	non	oui										
V2	MEF	1		oui	oui	faible		non	oui						
V2	MEF	3		oui	oui	faible		non	oui						
V3	Y14 EOS	A		oui	non										
V3	Y14 EOS	B		oui	non										
V3	Y14 EOS	C		oui	non										
V3	Y14 EOS	1	oui	faible	non										
V3	Y14 EOS	2	oui	oui	oui	oui		oui	oui						2n=38, XY t(Y ;14)
V3	Y14 EOS	3	oui	oui	oui	oui	faible	non	oui	oui	faible	oui	oui	absent	2n=38, XY t(Y ;14) anomalies
V3	Y14 EOS	4	oui	oui	oui	oui	oui	non	oui	oui	faible	oui	oui	absent	2n=38, XY t(Y ;14) anomalies
V3	Y14 EOS	5	oui	oui	non										
V3	Y14 EOS	6	oui		non										
V3	Y14 EOS	7	oui	oui	non										
V3	Y14 EOS	8	oui	oui	non										
V3	Y14 EOS	9	oui	faible	non										
V3	Y14 EOS	10	oui	oui	oui	oui	oui	oui	oui	oui					2n=38, XY t(Y ;14)
V3	Y14 EOS	11	oui	oui	oui	oui	faible	non	oui	oui					
V3	Y14 EOS	12	oui		non										
V3	Y14 EOS	13			oui										
V3	Y14 EOS	14			oui										
V3	Y14 EOS	15	oui		oui										
V3	Y14 EOS	16	oui		oui	oui	oui	non	oui	oui	faible	oui	oui	absent	2n=38, XY t(Y ;14) anomalies

Infection	Type	Nom	GFP+	AP+	Morphologie iPS-like	Nanog+	Lin28B+	Oct4+	Sox2+	hOSKM+	Formation EB	Maintenance en LIF+2i	Phase G1 raccourcie	Point de contrôle G1/S	Caryotype
V3	XXY EOS	A	oui	oui	non										
V3	XXY EOS	B	oui												
V3	XXY EOS	1	oui	faible	oui	faible	non	oui	non	oui					2n=39
V3	XXY EOS	2	oui	faible	oui	oui	non	oui	faible	oui					2n=39
V3	XXY EOS	3	oui	non											
V3	XXY EOS	4	oui	faible	oui	faible	oui	faible	non						
V3	XXY EOS	5	oui	faible	oui	oui	oui	non	faible						
V3	XXY EOS	6	oui	non	oui	faible	faible	oui	non	oui					2n=39
V3	XXY EOS	7	oui		oui	oui	oui	faible	non						2n=39
V3	XXY EOS	8	oui	faible											
V3	XXY EOS	11	faible												
V3	XXY EOS	12	oui												
V3	XXY EOS	13	oui	faible											
V3	XXY EOS	14	oui	non	oui	oui	oui	oui	faible						
V3	XXY EOS	15	faible	faible	oui	oui	non	oui	oui			non			2n=39,XXY
V3	XXY EOS	16	oui	oui	oui	oui	non	oui	faible	oui					
V3	XXY EOS	17	faible	oui											
V3	XXY EOS	18	faible	faible											
V3	XXY EOS	19	oui	non	oui	oui	non	faible	faible	oui					mélange cellulaire
V3	XXY EOS	20	oui	faible	oui										
V3	XXY EOS	21	oui	faible											
V3	XXY EOS	22	oui	oui	oui	oui	oui	oui	oui			non			2n=39,XXY
V3	XXY EOS	23			oui	faible	faible	non	oui						
V3	XXY EOS	24			oui	oui	non	non	non						
V5	N2 EOS	2	non	oui	non										

Infection	Type	Nom	GFP+	AP+	Morphologie iPS-like	Nanog+	Lin28B+	Oct4+	Sox2+	hOSKM+	Formation EB	Maintenance en LIF+2i	Phase G1 raccourcie	Point de contrôle G1/S	Caryotype
V5	N2 EOS	3	non	non	hétérogène	oui	oui	faible	oui	oui		non			2n=38
V5	N2 EOS	5	non	non	hétérogène	oui	oui	non	oui	oui					2n=39
V5	N2 EOS	6	non	non	hétérogène	oui	oui	oui	oui	oui					2n=39
V5	N2 EOS	11	non	non	oui	oui	non	non	non						
V5	N2 EOS	12	non	non	hétérogène	oui	oui	oui	oui	oui		non	oui	présent	2n=38
V5	N2 EOS	15	non	non	hétérogène	oui	non	non	faible						
V5	N2 EOS	16	non	non	hétérogène	oui	oui	faible	oui	oui					
V5	OAT1 EOS	1	non	non	oui	oui	oui	non	oui	oui	oui	oui	oui	absent	mélange cellulaire
V5	OAT1 EOS	2	non	non	oui	oui	oui	non	oui	oui					
V5	OAT1 EOS	4	non	oui	oui	oui	oui	non	oui	oui					
V5	OAT1 EOS	5	non	non	oui										
V5	OAT1 EOS	7	non	moyen	oui	faible	oui	non	oui	oui					
V5	OAT1 EOS	9	non	non	oui	faible	oui	non	oui	oui					
V6	N2 EOS	1			non										
V6	N2 EOS	2			oui	oui	non	non	non						
V6	N2 EOS	3			non	oui	non	faible	non						
V6	N2 EOS	4			hétérogène	oui	oui	oui	oui	oui		non			
V6	N2 EOS	5			oui	oui	non	non	non						
V6	N1 EOS	1			oui	non	non	non	non						
V6	N1 EOS	2			non	oui	oui	oui	oui	oui					
V6	N1 EOS	3			oui	faible	faible	non	non						
V6	N1 EOS	4			oui	oui	oui	faible	oui	oui		non			2n=38, XY anomalies
V6	N1 EOS	5			hétérogène	oui	faible	non	non						
V6	N1 EOS	6			non	oui	non	oui	oui	oui					
V6	N1 EOS	7			hétérogène	non	non	non	non						

Infection	Type	Nom	GFP+	AP+	Morphologie iP5-like	Nanog+	Lin28B+	Oct4+	Sox2+	hOSKM+	Formation EB	Maintenance en LIF+2i	Phase G1 raccourcie	Point de contrôle G1/S	Caryotype
V6	N1 EOS	8			hétérogène	oui	oui	faible	oui	oui		non			2n=38, XY anomalies
V6	PEF2 EOS	1			oui	oui	non	non	non						
V6	PEF2 EOS	2			oui	oui	non	non	non						
V6	PEF2 EOS	3			oui	oui	non	non	non						
V6	PEF2 EOS	4			oui	faible	non	non	non						
V6	PEF EOS	1			oui	oui	non	non	non						
V6	PEF EOS	2			oui	oui	oui	non	non	oui					
V6	PEF EOS	3			oui	oui	oui	faible	oui	oui					
V6	PEF EOS	8			hétérogène	oui	oui	faible	oui	oui		non			mélange cellulaire
V6	PEF EOS	9			oui	oui	faible	non	non						
V6	PEF EOS	10			oui	oui	oui	non	non						
V6	PEF EOS	12			oui	oui	non	faible	faible	oui					
V6	PEF EOS	13			oui	oui	oui	faible	non						
V6	PEF EOS	14			hétérogène	oui	oui	oui	oui	oui		oui	absent		mélange cellulaire
V6	PEF EOS	15			oui	oui	oui	faible	oui	oui					
V6	PEF EOS	18			oui	oui	oui	non	oui	oui		non			
V6	PEF EOS	19			oui	oui	oui	non	non	oui					
V6	PEF EOS	20			hétérogène	oui	oui	faible	faible	oui					
V6	PEF EOS	21			oui	oui	oui	non	non						
V6	PEF EOS	22			hétérogène	oui	oui	non	faible	oui					
V6	PEF EOS	24			oui	oui	oui	faible	non	oui					
V6	PEF EOS	25			oui	oui	oui	oui	non	oui					
V6	PEF EOS	26			oui	oui	oui	non	oui						
V6	PEF EOS	27			oui	oui	faible	non	non						
V6	PEF EOS	28	oui		oui	oui	oui	non	non	oui		non	présent		2n=38

Infection	Type	Nom	GFP+	AP+	Morphologie iPS-like	Nanog+	Lin28B+	Oct4+	Sox2+	hOSKM+	Formation EB	Maintenance en LIF+2i	Phase G1 raccourcie	Point de contrôle G1/S	Caryotype
V7	Y14	1	non	oui	oui	oui	faible	oui	oui	oui	oui				
V7	Y14	2	non	oui	hétérogène	oui	faible	oui	oui	oui	oui				
V7	Y14	3	non	oui	oui	oui	oui	oui	oui	oui	oui				
V7	Y14	4L	non	oui	oui	oui	faible	faible	oui	oui	oui				
V7	Y14	6	non	oui	oui	oui	faible	oui	oui	oui	non				
V7	Y14	9	non	oui	oui	oui	oui	oui	oui	oui	oui				
V7	Y14	10	non	oui	oui	oui	oui	oui	oui	oui	oui				
V7	Y14	11	non	oui	oui	oui	oui	oui	oui	oui	oui				
V7	Y14	12	non	oui	oui	oui	oui	oui	oui	oui	oui				
V7	Y14	13	non	oui	oui	oui	oui	oui	oui	oui	oui				2n=38, XY t(Y;14)
V7	Y14	14	non	oui	oui	oui	oui	oui	oui	oui	oui				
V7	Y14	15	non	oui	oui	oui	oui	oui	oui	oui	oui				
V7	Y14	19L	non	oui	oui	oui	oui	oui	oui	oui	oui				
V7	Y14	20L	non	oui	oui	oui	oui	oui	oui	oui	oui				
V7	AT3	2	non	non	non	non	non	non	non	oui	oui				
V7	AT3	4	non	non	hétérogène	non	non	non	non	oui	oui				
V7	AT3	5	non	non	non	non	non	non	non	oui	oui				
V7	AT3	6L	non	moyen	hétérogène	non	non	non	non	oui	oui				
V7	AT3	16	non	non	oui	non	oui	non	non	oui	oui				

Analyse de la méthylation de l'ADN spermatique et développement de cellules pluripotentes induites chez des verrats infertiles porteurs ou non de remaniements chromosomiques.

Annabelle Congras -- Vendredi 12 décembre 2014, Toulouse -- Biologie cellulaire

UMR1388 GenPhySE, INRA, 24 chemin de Borderouge - Auzeville, CS 52627, 31326 Castanet-Tolosan

Directeurs de thèse : Martine YERLE-BOUISSOU et Hervé ACLOQUE.

Les problèmes de fertilité s'appliquent à la fois à l'homme, touchant une part croissante de la population, et aux espèces d'élevage comme le porc, freinant la diffusion de caractères d'intérêt agronomique. Dans ce projet, nous avons étudié deux mécanismes liés à l'infertilité : la dérégulation du profil de méthylation de l'ADN gamétique et la présence de remaniements chromosomiques.

Nous avons tout d'abord concentré notre étude sur le profil de méthylation de l'ADN spermatique de verrats fertiles et infertiles. Nous avons ainsi constaté que ce niveau de méthylation est conservé entre trois espèces de mammifères, au niveau global et local, ainsi qu'entre verrats fertiles et infertiles. Une augmentation spécifique de la méthylation dans le locus GNAS a été mise en évidence, ainsi qu'une dérégulation de son expression chez certains animaux infertiles, reliant pour la première fois hyperméthylation de cette région et infertilité chez les mammifères.

Pour étudier ensuite l'impact des remaniements chromosomiques sur la fertilité des animaux, nous avons choisi de produire des lignées de cellules souches pluripotentes induites (iPSCs) par reprogrammation cellulaire de fibroblastes issus de ces animaux, afin de pouvoir ensuite étudier leur différenciation dans la voie germinale. Les lignées obtenues pour un verrat azoospermique porteur de la translocation réciproque t(Y;14) possèdent de nombreuses caractéristiques des cellules pluripotentes telles que l'expression de gènes spécifiques, un cycle cellulaire proche de celui des cellules souches embryonnaires, et une capacité à évoluer vers l'état naif de pluripotence dans un milieu de culture adapté. Toutefois, leur capacité de différenciation est faible et elles présentent une instabilité génomique augmentant avec les passages, comme les lignées issues d'animaux fertiles à caryotype normal, phénomène pouvant être associé au choix d'une stratégie de reprogrammation intégrative. En effet, l'utilisation d'une méthode non intégrative a permis de démontrer que les cellules iPS ainsi obtenues ne présentent pas cette instabilité. Leur caractérisation partielle laisse présager l'obtention de lignées plus stables rassemblant plus de caractéristiques de pluripotence.

Mots-clés : infertilité, cellules souches pluripotentes induites, méthylation de l'ADN, épigénétique, remaniements chromosomiques, porc.

Spermatic DNA methylation analysis and development of induced pluripotent stem cells from infertile boars carriers or not of chromosomal rearrangements.

Fertility issues concern both humans, affecting a growing part of the population, and farm animals including pigs in which they slow down the diffusion of agronomical traits of interest. In this project, we focused on two mechanisms linked to infertility: alterations in gametic DNA methylation and chromosomal rearrangements.

We first studied DNA methylation in sperm cells of fertile or infertile boars. We observed that the methylation level of spermatic DNA is conserved between three mammalian species, both at the global and local level as well as between fertile and infertile boars. A specific increase in DNA methylation in the GNAS locus was nevertheless identified, as well as a deregulation of its expression in some boars with low quality semen, linking for the first time hypermethylation of this region and male infertility in mammals.

To study the effect of chromosomal rearrangements on fertility, we chose to produce induced pluripotent stem cell lines (iPSCs) derived from fibroblasts of these animals, as a tool for studying their differentiation towards the germ cell lineage. Cell lines derived from an azoospermic boar carrier of a reciprocal translocation t(Y;14) harbor several characteristics of pluripotency including the expression of specific genes, a cell cycle resembling the one of embryonic stem cells, and an ability to evolve into the naïve state in an adapted culture medium. However they revealed a poor differentiation potential and a genomic instability increasing with passaging as control lines derived from fertile and karyotypically normal boars, phenomena that might be associated with the integrative reprogramming strategy. The use of a non-integrative technique allowed us to demonstrate that the cell lines obtained with this method did not harbor this instability. Their preliminary characterization may be predictive of production of more stable cell lines gathering more characters of pluripotency.

Keywords : infertility, induced pluripotent stem cells, DNA methylation, epigenetics, chromosomal rearrangements, pig.