



HAL
open science

Epigénétique de la semence bovine : analyse moléculaire de la qualité de la semence et impact potentiel sur le développement embryonnaire

Jean-Philippe Perrier

► To cite this version:

Jean-Philippe Perrier. Epigénétique de la semence bovine : analyse moléculaire de la qualité de la semence et impact potentiel sur le développement embryonnaire. Biologie animale. Université Paris Saclay (COMUE), 2017. Français. NNT : 2017SACLA037 . tel-02098254

HAL Id: tel-02098254

<https://theses.hal.science/tel-02098254>

Submitted on 12 Apr 2019

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

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

Epigénétique de la semence bovine : analyse moléculaire de la qualité de la semence et impact potentiel sur le développement embryonnaire

Thèse de doctorat de l'Université Paris-Saclay
préparée à AgroParisTech (l'institut des sciences et industries du
vivant et de l'environnement)

École doctorale n°581 Agriculture, Alimentation, Biologie,
Environnement et Santé (ABIES)
Spécialité de doctorat : Biologie de la reproduction

Thèse présentée et soutenue à Jouy-en-Josas, le 14 décembre 2017, par

Jean-Philippe Perrier

Composition du Jury :

Pr. Etienne Verrier

Professeur, AgroParisTech

Président

Dr. Carole Charlier

Principal investigator, GIGA (Unité de Génomique Animale)

Rapporteur

Pr. Romain Barrès

Professeur associé, CBMR (Section for Integrative Physiology)

Rapporteur

Dr. Christophe Arnoult

Directeur de recherche, CNRS (Institut de Biosciences Avancées)

Examineur

Dr. Hélène Jammes

Directeur de recherche, INRA (Biologie du Développement et Reproduction)

Directeur de thèse

Dr. Hélène Kiefer

Chargé de recherche, INRA (Biologie du Développement et Reproduction)

Co-encadrant de thèse

Dr. Laurent Schibler

Responsable Développement et Innovation, Allice

Invité

Remerciements

Près de 4 ans se sont écoulées depuis mon arrivée sur le centre INRA de Jouy-en-Josas, et en écrivant ces remerciements un grand nombre de souvenirs me viennent à l'esprit. J'ai pu y rencontrer tant de gens, j'ai l'impression d'avoir tellement appris et grandi pendant cette période que je suis sûr de conserver ces moments en moi toute la vie.

Tout naturellement, mes premiers remerciements vont à mes encadrantes, Hélène Kiefer et Hélène Jammes. Je vous suis extrêmement reconnaissant, tout d'abord d'avoir accepté ma venue après vous avoir contacté pour mon stage de M2, puis de m'avoir fait confiance pour intégrer l'équipe en tant que doctorant sur le projet SeQuaMol. Vous Toutes 2 êtes toujours restées l'écoute, m'avez toujours guidé dans la bonne direction quand j'avais des hésitations, et m'avez laissé de l'autonomie pour mener ce projet. La thèse m'aura apporté énormément de choses, autant sur le plan personnel que professionnel, et c'est à vous que je le dois. Hélène Kiefer, j'ai eu beaucoup de chance de t'avoir comme encadrante principale, notamment pour la rigueur scientifique dont tu fais toujours preuve et que j'espère pouvoir appliquer maintenant à mon tour. Félicitation pour la naissance de Joseph, qui sera certainement un petit scientifique en herbe. Hélène Jammes, malgré ton emploi du temps toujours très chargé tu as pu me consacrer de ton temps, et notamment pendant la rédaction et alors que les mauvaises nouvelles arrivaient l'une après l'autre, et que je commençais à me demander si j'allais vraiment y arriver. Merci beaucoup pour ton aide et tes avis précieux.

Je souhaite remercier l'ensemble de l'équipe ER5 avec qui j'ai eu la chance de travailler durant ces quelques années : les titulaires (les Hélénes, Anne, Audrey, Aurélie, Mélanie et Charline), ceux qui sont parti à la retraite (Evelyne, Michel), les thésards (Maxime, Sofiane, Polina, Sarah), et les étudiants (Chaneze, Laura, Mélanie, Justin). Je souhaite remercier en particulier Audrey, ma camarade de bureau. Les pauses tricot, les pauses thé, et tous les moments de rigolade qu'on a eu vont me manquer ! Avant de partir, je te laisserai mon pistolet Nerf, je suis sûr que tu en trouveras une nouvelle utilité. Je pense avoir beaucoup appris de toi également, surtout concernant la rigueur à avoir pour les manip, et pour ton sens de l'organisation particulièrement développé. Je te souhaite de t'épanouir dans ta nouvelle équipe.

Maxime et Sofiane, camarades de pauses cafés – clopes. Désolé de vous en avoir taxé autant, dorénavant c'est terminé car après une 112^{ème} tentative, j'ai finalement réussi à arrêter ! Maxime, le pro de la bioinfo, je te souhaite une bonne continuation dans ta nouvelle voie. Bon courage à toi Sofiane, tu es l'un des derniers thésards de la BDR, soit endurant et représente bien l'unité.

Eli, depuis mon arrivée tu as toujours été là et pour moi tu fais partie de l'équipe : un grand merci à toi aussi. C'était toujours sympa de venir dans ton bureau pour discuter de tout et de rien, et pour tous ces moments de rigolade. J'ai beaucoup appris grâce à toi, mise à part peut-être niveau blague, qui a dû baisser à force d'entendre les tiennes. Bon courage pour ta thèse, et pense à moi si tu finis par monter une boîte !

Aurélie, bienvenu à nouveau dans l'équipe. J'espère que je n'ai pas été un trop mauvais professeur de RRBS et de bioinfo, et surtout prend bien soin du robot !

Merci à la partie « souris et microcèbe » de l'ER5, Anne Mélanie et Charline, pour votre bonne humeur constante, et pour nous avoir laissé manger tous les gâteaux d'anniversaire de votre côté du labo. Charline, bienvenu à toi, j'espère que tu vas t'épanouir en BDR.

Merci à toutes les personnes d'Alice avec qui j'ai pu travailler, en particulier Laurent Schibler et Chrystelle le Danvic. Ça a été vraiment enrichissant de pouvoir participer aux réunions du projet, d'avoir un aperçu des autres aspects du projet SeQuaMol. Merci particulièrement à Laurent d'avoir pris sur son temps précieux pour me relire lors de la rédaction.

Un grand merci aux bio informaticiens de l'unité, Luc Jouneau et François Piumi. Luc, merci pour ta patience et pour avoir toujours été là pour répondre à mes questions et lorsque j'ai eu un pépin avec les analyses des données de RRBS. François, je te souhaite une bonne continuation dans ton nouveau labo.

Une pensée pour tous les étudiants que j'ai croisé au labo : Laura, Mélanie, Justin, Mikela, et particulièrement à Chaneze que j'ai encadré. Tu faisais de super pyro ! J'espère avoir été un bon encadrant et je te souhaite le meilleur pour la suite.

Je souhaite remercier L'équipe Doc'J, et en particulier, Tatiana sans qui le congrès annuel de 2016 n'aurait pas pu avoir lieu. Faire partie de Doc'J a été une très bonne expérience, et on a

passé de bonnes soirées, fait plein d'activité (comme les cafés entre étudiants, les ateliers pendant la journée portes-ouvertes...) et j'en garderai un bon souvenir.

Merci à tous les autres thésards que j'ai eu la chance de côtoyer : Matteo, Clara, Audrey, Sarah Valentino, Sarah Voisin, Polina et Mouna. Ça fait longtemps qu'on n'a pas pu se voir pour la plupart, chacun d'entre vous étant parti poursuivre sa carrière aux 4 coins du monde, mais ça a été un plaisir de vous rencontrer. Je ne désespère pas de vous croiser un jour au détour d'un congrès ! J'ai une pensée particulière pour Clara, avec qui je n'ai pas toujours été sympa, t'ayant posé beaucoup de lapins pour les pauses déjeuné...mais je suis certain que tu ne m'en veux pas car tu sais bien que j'ai une petite tête. Pour ta rédaction c'est la dernière ligne droite, mais n'ayant jamais vu quelqu'un d'aussi solide que toi je ne doute pas que tu vas t'en sortir brillamment.

Merci à Hala et Sara Fneich, vous avez apporté de la bonne humeur au labo et j'espère également avoir la chance de vous recroiser un jour.

Je souhaite également remercier toutes les autres personnes de la BDR que j'ai eu la chance de croiser, avec qui j'ai pu discuter et échanger. Chacun d'entre vous m'a apporté une petite chose, et c'est chacun d'entre vous qui faisait que j'étais de bonne humeur le matin en allant travailler. Merci en particulier à Céline, qui passait souvent me faire un coucou au bureau, et à Olivier Dubois pour avoir bien rigolé pendant les pauses clopes.

Merci aux membres du jury (Carole Charlier, Romain Barrès, Etienne Verrier et Christophe Arnoult), qui ont gentiment accepté de prendre de leur temps pour évaluer mon travail de thèse, et merci également aux membres de mon comité de thèse (Joël Drevet et Thomas Heams), qui m'ont suivi au cours de ces 3 années et qui m'ont toujours donné des conseils avisés.

Enfin, je souhaite remercier ma famille et surtout mes parents, qui ont toujours été là pour moi, qui m'ont toujours soutenu et à qui je dois absolument tout. Une pensée pour ma Sixtine, qui a su me pousser à donner le meilleur de moi-même, qui m'a supporté aussi alors que j'étais au bout du rouleau et que je ne devais pas être particulièrement agréable à vivre. A mes amis les Binious, ma seconde famille qui m'ont également soutenu et avec qui j'ai grandi, même pendant les moments difficiles qu'on a traversé ensemble. Je suis heureux d'avoir des amis comme vous.

Table des matières

Liste des abréviations	5
Table des illustrations	8
Introduction	12
Préambule	13
Chapitre I : La semence bovine, vecteur du patrimoine génétique paternel et produit à forte valeur ajoutée	15
1. Aspects biologiques : la voie mâle dans la reproduction	15
1.1. La structuration du testicule bovin	15
1.1.1. La différenciation des gonades fœtales	15
1.1.2. De la naissance à la puberté	17
1.2. La production des spermatozoïdes	18
1.2.1. La spermatocytogenèse	18
1.2.2. La spermiogenèse	19
1.2.3. La spermiation et la maturation des spermatozoïdes	20
1.3. La rencontre avec le gamète femelle et la génération d'un descendant	21
1.3.1. La fécondation	21
1.3.2. Le développement embryonnaire	22
2. Aspects biotechnologiques et économiques : production et diffusion de la semence bovine... ..	24
2.1. L'organisation de la filière	24
2.1.1. Trois siècles de développement de l'IA	24
2.1.2. Les avantages apportés par l'IA	25
2.1.3. La situation de la filière en France	26
2.2. La sélection des reproducteurs	27
2.2.1. Le testage sur descendance	27
2.2.2. La révolution génomique	28
2.2.3. L'évolution des pratiques	30
2.3. Les étapes de production de semence	31
2.3.1. La collecte et le conditionnement des éjaculats	31
2.3.2. La mise en place de l'IA	32
2.3.3. Aspects économiques	32
Chapitre II : Les enjeux de la fertilité mâle pour les filières bovines	33
1. Causes et conséquences de la baisse de fertilité des élevages	34
1.1. La baisse de fertilité est délétère pour les filières bovines	34
1.1.1. Observations	34
1.1.2. Conséquences organisationnelles et économiques	35
1.1.3. Estimation de la contribution des mâles à la baisse de la fertilité	36
1.2. Les causes de la baisse de fertilité mâle	38
1.2.1. Effets environnementaux sur la baisse de la fertilité mâle	38

1.2.2.	Effets des traitements technologiques de la semence	39
1.2.3.	Effets génétiques	40
1.2.4.	Autres effets	41
2.	Méthodes d'évaluation de la fertilité mâle	42
2.1.	Analyses phénotypiques	42
2.1.1.	Le spermogramme	42
2.1.2.	Bull Breeding Soundness Evaluation	43
2.1.3.	Analyses morphologiques des spermatozoïdes	44
2.1.4.	Méthodes exploratoires	45
2.2.	Données de terrain après IA	47
2.2.1.	Taux de fécondation	47
2.2.2.	Taux de non-retour	47
2.2.3.	Taux de conception.....	48
2.2.4.	Fertilité et sub-fertilité.....	48
3.	Les limites des méthodes d'évaluation de la fertilité mâle	49
3.1.	Les analyses phénotypiques de la semence ne permettent pas de prédire la sub-fertilité .	49
3.2.	Effets compensables et non-compensables	51
3.3.	L'évaluation génétique ne permet pas de prédire la sub-fertilité	51
Chapitre III : L'épigénétique de la fertilité mâle.....		53
1.	Les différentes marques épigénétiques	54
1.1.	Les modifications post-traductionnelles des histones	54
1.2.	Les ARN non-codants	55
1.3.	La méthylation de l'ADN.....	57
1.3.1.	Répartition génomique des sites CpG	58
1.3.2.	Régulation de l'expression génique par la méthylation.....	59
1.3.3.	La machinerie de la méthylation de l'ADN.....	61
1.3.4.	Dialogue avec les autres marques épigénétiques.....	64
1.4.	Rôles biologiques de la méthylation de l'ADN.....	65
1.4.1.	Rôle structural et stabilité du génome	65
1.4.2.	L'empreinte parentale.....	67
1.4.3.	L'inactivation du chromosome X.....	67
1.4.4.	Différenciation cellulaire et expression tissu-spécifique des gènes	68
1.4.5.	Adaptation à des conditions environnementales	69
2.	La reprogrammation épigénétique, un processus biologique fondamental pour la fertilité mâle	71
2.1.	La reprogrammation dans la lignée germinale mâle	72
2.1.1.	Déméthylation puis méthylation de novo dans les gonades fœtales.....	72
2.1.2.	Remaniements épigénétiques après la naissance et dans la gonade adulte.....	75
2.2.	La reprogrammation de l'épigénome embryonnaire	77
2.3.	Effets multigénérationnels et transgénérationnels.....	80

3. Méthylation de l'ADN et infertilité mâle	81
3.1. Manipulations pharmacologiques et génétiques chez la souris	82
3.2. Troubles de la fertilité et de la spermatogenèse chez l'Homme en lien avec la méthylation de l'ADN.....	83
3.3. Altération épigénétique et fertilité bovine	86
Objectifs de la thèse.....	88
Méthodologie.....	91
1. Méthylation de l'ADN : quelle stratégie pour quel objectif ?	92
1.1. Les stratégies d'extraction de l'information de méthylation	94
1.1.1. L'utilisation d'enzymes de restriction	94
1.1.2. Les stratégies par affinité.....	94
1.1.3. La conversion de l'ADN au bisulfite de sodium	95
1.2. Les stratégies de révélation	96
1.2.1. Les méthodes basées sur la PCR	96
1.2.2. L'hybridation sur micropuce	98
1.2.3. Le séquençage à haut-débit	99
2. Choix des méthodes retenues pour notre étude	99
2.1. Le bisulfite-pyroséquençage.....	101
2.2. Le LUMinometric Methylation Assay (LUMA)	102
2.3. Le Reduced Representation Bisulfite Sequencing (RRBS).....	103
2.3.1. Optimisation du protocole par analyses in silico.....	103
2.3.2. La préparation des banques :	104
2.3.3. Le séquençage des banques	106
2.3.4. Le traitement bio-informatique des données	106
Résultats	111
Bull spermatozoa exhibit genome-wide DNA hypomethylation with both species peculiarities and conserved tissue-specific features	112
Identification de nouveaux biomarqueurs de la fertilité mâle dans le méthylome de la semence bovine : une analyse par RRBS.....	152
Introduction	153
Matériels et méthodes.....	155
A. Dispositif expérimental et collecte des échantillons.....	155
B. Extraction d'ADN génomique.....	156
C. Analyse de la méthylation de l'ADN par RRBS	157
D. Evaluation des apports de la double digestion enzymatique dans la description du méthylome.....	159
E. Validations de DMR par pyroséquençage	161
F. Intégration des données de génotypage aux données de méthylation	163
Résultats	164
A. Constitution des groupes de taureaux en fonction de leurs caractéristiques de fertilité	164
B. Bonne qualité des séquences de RRBS	165

C.	La race influence de manière significative le taux de méthylation	166
D.	Des différentiels de méthylation entre les catégories de fertilité sont mis en évidence pour les 2 races.....	167
E.	Les DMC représentent de bons biomarqueurs de la fertilité	170
F.	Répartition génomique des DMC identifiées	170
G.	Identification des fonctions biologiques enrichies en gènes associés aux DMC/DMR identifiées	172
H.	Validation des DMR par bisulfite-pyroséquençage.....	175
I.	Intégration des données de SNP et de DMC	177
J.	Intérêts d'utiliser une double digestion enzymatique HaeIII + MspI, pour construire une banque RRBS chez le bovin.....	178
a.	Par analyse descriptive, il apparaît que la double digestion permet d'identifier es CpG ₁₀ discriminant mieux les catégories Fertiles+ et Déviants.	179
b.	La double digestion permet d'identifier un nombre plus important de DMC entre catégories de fertilité.	180
c.	Annotation génomique des DMC _[HaeIII + MspI]	180
K.	Mise en évidence d'un effet « âge » sur le méthylome spermatique - Correction des biomarqueurs de fertilité	183
	Discussion.....	186
A.	Le méthylome spermatique présente des variations importantes entre races	186
B.	Des modifications du méthylome spermatique en fonction de la fertilité ciblent des processus biologiques variés	187
C.	Impact fonctionnel des altérations du méthylome spermatique : conséquences pour la descendance.....	189
D.	Le catalogue de DMC identifiées peut-il servir de base pour le développement d'un outil d'épigénotypage de la semence bovine ?.....	190
E.	Effet de l'âge à la production de semence sur le méthylome spermatique.....	192
F.	Nouvelle analyse du méthylome spermatique en fonction de la catégorie de fertilité en excluant les effets liés à l'âge à la production de la semence.....	194
	Discussion générale	197
	Points technologiques critiques pour une fiabilité des biomarqueurs.....	198
	Une limite de l'approche par RRBS : l'absence de données d'hydroxyméthylation	203
	Hypométhylation de l'ADN spermatique bovin et compaction nucléaire.....	205
	Des biomarqueurs révélateurs de la fertilité à l'éjaculat.....	207
	Quelles sont les causes et les conséquences des différentiels de méthylation ?	209
	Les causes.....	209
	Les conséquences	212
	Prise en compte de l'âge à la production de semence pour l'identification de biomarqueurs pertinents	215
	Intégration des données épigénétiques et génétiques pour la compréhension du phénotype spermatique.....	216
	Considérations pour la mise au point d'un outil d'épigénotypage	217
	Bibliographie.....	220
	Annexes	239

Liste des abréviations

5caC : 5-carboxylcytosine
5fC : 5-formylcytosine
5hmC : 5-hydroxyméthylcytosine
5mC : 5-méthylcytosine
A : Adénine
ACP : Analyses en Composantes Principales
ADN : Acide DésoxyriboNucléique
ADNsat : ADN satellite
AID/APOBEC : Activation-induced cytidine deaminase / Apolipoprotein B mRNA Editing Enzyme Catalytic
ARN : Acide RiboNucléique
AMH : Anti-Müllerian Hormone
AMP : Aide Médicale à la Procréation
AMPc : Adénosine MonoPhosphate cyclique
ANR : Agence National de la Recherche
BBSE : Bull Breeding Soundness Evaluation
BCS : Body Condition Score
BER : Base Excision Repair
Bsp: *Bisulfite Sequencing PCR*
C : Cytosine
CASA : Computer Assisted Semen Analyzer
CASMA : Computer Assisted Sperm Morphology Analyzer
CGI : CpG Island
ChIP : Chromatin ImmunoPrecipitation
CNV : Copy Number Variation
CS : Circonférence Scrotale
CTCF: *CCCTC-Binding factor*
CTIG : Centre de Traitement de l'Information Génétique
CVM : *Complex Vertebral Malformation*
DAPI : 4,6-DiAmidino-2-PhenylIndole
Dazl: Deleted in AZoospermia like
Ddx4 : DEAD-Box Helicase 4
DMC : Differentially Methylated Cytosine
DMR : Differentially Methylated Region
DNMT: ADN MéthylTransférase
DOHaD: Developmental Origins of Health and Disease
DSB: Double-Strand Break
EGA: Embryonic Genome Activation
ELISA: *Enzyme-Linked ImmunoSorbent Assay*
EMP: Entreprise de Mise en Place
Emx2 : Empty Spiracles Homeobox 2
ERK: Extracellular signal-Regulated Kinase
ES: Entreprise de Sélection
FGF: Fibroblast Growth Factor
FIV: Fécondation *In Vitro*
FSH: Follicle-Stimulating Hormone
G : Guanine
GnRH: Gonadotropin-Releasing Hormone
GO: Gene Ontology
GTF: *General Transfer Format*
GWAS: Genome-Wide Association Studies
HDAC: Histone Déacétylase
HDM: Histone DéMéthylase
HH: Holstein Haplotype
HMT: Histone MéthylTransférase
HRM : *High Resolution Melting*
IA : Insémination Artificielle
IAP : Insémination Artificielle Première

IAP : Intra-cisternal A Particle
ICF : Immunodeficiency with Centromeric instability and Facial anomalies
ICM : Inner Cell Mass
ICR: Imprinting Control Region
ICSI: Intracytoplasmic Sperm Injection
IDELE : Institut de l'Élevage
IGF : Insulin-like Growth Factor
INRA : Institut National de la Recherche Agronomique
INSL3 : Insulin-like factor 3
IPE : Insémination Par l'Éleveur
jpc : jours post-coïtum
KAT : Lysine AcétylTransférase
KRuO4 : perruntenate de potassium
QTL : Quantitative Trait Loci
LAMB2 : *Laminin Subunit Beta 2*
LH : Luteinizing hormone
LIF : Leukemia Inhibitory Factor
LIM1 : LIM Homeobox 1
LINE : Long Interspersed Nuclear Element
LNCR: Laboratoire National de Contrôle des Reproducteurs
lncRNA : long non-coding RNA
LTR : Long Terminal Repeat
LUMA : Luminometric Methylation Assay
MAR: Matrix Attachment Region
MBD: Methyl-Binding-Domain
MDA: Malondialdehyde
MeCP2: Methyl-CpG-binding Protein 2
MeDIP: Methylated DNA ImmunoPrecipitation
MH: Montbeliard Haplotype
MPT: Modification Post-Traductionnelle
MSRE : *Methylation-Sensitive Restriction Enzymes*
N6mA: N6-méthyl-adénine
NF2 : *Neurofibromin 2*
NGS: Next Generation Sequencing
OAT: oligoasthénotératozoospermique
OS: Oligospermique
pb : paire de base
PLCζ: PhosphoLipase C zeta
PCR : Polymerase Chain Reaction
PGC: Primordial Germ Cell
piARN: PIWI-interacting RNA
PFN1: Profiline 1
PLS-DA: *Partial Least Squares - Discriminant Analysis*
RARb: Retinoic Acid Receptor b
RISC: RNA-Induced Silencing Complex
ROS: Reactive Oxygen Species
RRBS : Reduced Representation Bisulfite Sequencing
SAM: S-adénosylméthionine
SBH: Sperm Bos Halomax
SCF: *Stem Cell Factor*
SCNT: Somatic Cell Nuclear Transfer
SCR: Sire Conception Rate
SCSA: Sperm Chromatin Structure Assay
SF1: Steroidogenic Factor 1
SHBG: Sex Hormone Binding Globulin
SHGC: Syndrome d'Hypoplasie Généralisé Capréoliforme
SINE: Short Interspersed Nuclear Element
SMRT: Single Molecule Real-Time Sequencing
SMARCD3: SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily D, Member 3
SMUG1 : Single-Strand Selective Monofunctional Uracil-DNA glycosylase 1
SNIG: Système National d'Information Génétique

SNIP1 : *Smad Nuclear Interacting Protein 1*
SNP : Single-Nucleotide Polymorphism
Sox9: SRY-related HMG-box gene 9
SPLS: *Sparse Partial Least Squares*
SRA: SET and RING-finger Associated domain
SRY: Sex determining region of the Y-chromosome
SSB: Single-Strand Break
Sycp3: SYnaptonemal Complex Protein 3
T : Thymine
TAB-seq : *TET-Assisted Bisulphite sequencing*
TDG: Thymine DNA Glycosylase
TE: Transposable Element
TEAD1-4: *TEA Domain Transcription Factor 1-4*
TERRA: Telomeric Repeat Containing RNA
Tet: Ten-Eleven Translocation
TF : Transcription Factor
TGF β : Transforming Growth Factor β
TLE3 : *Transducin Like Enhancer Of Split 3*
TNR : Taux de Non-Retour
TRD: Transcriptional Repression Domain
TSS : Transcription Start Site
TTS : Transcription Termination Site
TUNEL : Terminal deoxynucleotidyl transferase mediated dUTP Nick end labelling
U : Uracile
UCSC: *University of California Santa Cruz*
UHRF: Ubiquitin-like PHD and RING Finger
WGBS: *Whole Genome Bisulfite Sequencing*
WT1 : Wilms Tumor 1
WWC1-2: *WW And C2 Domain Containing*
YAP: Yes Association Protein
ZP: Zone Pellucide

Table des illustrations

Introduction

Figure 1. Chronologie comparative du développement de la fonction reproductrice chez le taureau, l'Homme et la souris	16
Figure 2. Chronologie des grandes étapes de la spermatogenèse	16
Figure 3. Migration des PGC au cours du développement embryonnaire chez la souris.....	16
Figure 4. Différenciation du tractus génital pendant la période embryonnaire chez le bovin.....	17
Figure 5. Représentation de l'axe hypothalamo-hypophysio-testiculaire	18
Figure 6. Cinétique de la maturation sexuelle chez le taureau.....	18
Figure 7. Organisation du tractus génital mâle bovin	19
Figure 8. La méiose.....	20
Figure 9. Les étapes de la spermiogenèse chez le bovin.....	20
Figure 10 Spermatozoïde bovin.....	21
Figure 11. La fécondation	22
Figure 12. Ecllosion et élongation du blastocyste chez le bovin	24
Figure 13. Principe du testage sur descendance	28
Figure 14. Principe de la sélection génomique.....	29
Figure 15. La mise en place de l'IA	32
Figure 16. L'IA en France.....	33
Figure 17. Le taux de conception bovin est en diminution, et est principalement lié à des pertes embryonnaires précoces	35
Figure 18. Estimation de la fertilité des cheptels et des mâles uniquement	37
Figure 19. Le spermogramme	43
Figure 20. Exemple d'anomalies morphologiques des spermatozoïdes retrouvées chez le bovin.....	45
Figure 21. Relation entre la valeur de fertilité prédite et le taux de non-retour à 56 jours.....	50
Figure 22. Relation entre le nombre de spermatozoïdes par paille et la fertilité des taureaux	52
Figure 23. Représentation schématique des QTL associés à la fertilité mâle bovine, chromosome par chromosome.....	52
Figure 24. Le paysage développemental selon Conrad Waddington	54
Figure 25. Principales marques épigénétiques chez les eucaryotes	54
Figure 26. Etats chromatinien et régulation de la transcription	54
Figure 27. Les modifications post-traductionnelles des histones.....	54
Figure 28. Modes d'action des microARN.....	57
Figure 29. Mécanismes de régulation des transposons par les piARN	57
Figure 30. La méthylation de l'ADN est la marque épigénétique la plus étudiée.....	58
Figure 31. La méthylation de l'ADN.....	58
Figure 32. Les diverses formes principales de méthylation de l'ADN	59
Figure 33. Concepts d'îlots CpG, de rivages CpG, de plateaux CpG et de mer ouverte	60
Figure 34. Répartition génomique de la méthylation de l'ADN	61
Figure 35. Degré de méthylation du promoteur et transcription génique	62
Figure 36. Représentation schématique de la structure des protéines de la famille des DNMT	62
Figure 37. Modes d'action de la DNMT de maintenance (DNMT1), et des DNMT <i>de novo</i> (DNMT3a et DNMT3b)	62
Figure 38. Représentation schématique du cycle du folate et de la méthionine.....	62
Figure 39. Représentation de la voie passive de la déméthylation de l'ADN	63
Figure 40. Voie active de la déméthylation de l'ADN	63

Figure 41. Les trois familles de protéines de liaison aux sites CpG méthylés chez les mammifères.....	64
Figure 42. Interactions entre la méthylation de l'ADN et les modifications post-traductionnelles de histones.....	65
Figure 43. Maintenance de la méthylation de l'ADN par DNMT1 et Uhrfl	65
Figure 44. Les éléments de type transposables dans le génome bovin	66
Figure 45. Le phénomène de l'empreinte parentale.....	68
Figure 46. Représentation de l'inactivation des gènes et de la réorganisation spatiale du chromosome X	69
Figure 47. Les différentes phases de la reprogrammation épigénétique chez la souris.....	72
Figure 48. La reprogrammation épigénétique dans les PGC	73
Figure 49. Etat chromatinien de type bivalent.....	75
Figure 50. Remplacement des histones par les protamines	77
Figure 51. Modifications des marques épigénétiques au cours de la spermatogenèse.....	78
Figure 52. La reprogrammation épigénétique dans le zygote	79
Figure 53. Principe du clonage par transfert nucléaire (<i>Somatic Cell Nuclear Transfer</i>).....	80
Figure 54. Effets multi et transgénérationnels	81
Figure 55. Les phases du projet Sequamol	82
Tableau 1. Liste des haplotypes létaux au stade embryonnaire chez le bovin	42
Tableau 2. Corrélations génétiques entre paramètres spermatiques et phénotypiques liés à la fertilité chez le bovin.....	50
Tableau 3. Héritabilité des paramètres spermatiques et de fertilité bovine.....	53
Tableau 4. Principaux algorithmes de prédiction des îlots CpG du génome Humain.....	59
Tableau 5. Listes des gènes soumis à l'empreinte connus chez le bovin	68
Tableau 6. Liste des études faisant le lien entre nutrition et méthylation de l'ADN chez les espèces d'intérêt agronomique...71	

Méthodologie

Figure 56. Arbre décisionnel pour l'analyse de la méthylation de l'ADN	93
Figure 57. Schéma représentatif de l'analyse de la méthylation de l'ADN à différentes échelles génomiques	93
Figure 58. Schéma représentatif de la stratégie d'analyse de la méthylation de l'ADN par affinité.....	95
Figure 59. La conversion de l'ADN par bisulfite de sodium.....	96
Figure 60. Automatisation du protocole de RRBS sur un automate de pipetage	101
Figure 61. Principe de la mesure de la méthylation de l'ADN par pyroséquençage.....	102
Figure 62. Principe de la mesure de la méthylation de l'ADN par LUMA	103
Figure 63. Etapes de préparation des banques RRBS.....	104
Figure 64. Sélection de taille des fragments et densité en sites CpG.....	105
Figure 65. Contrôles qualité de production de production des banques RRBS	106
Figure 66. Vue d'ensemble de la procédure de traitement bio-informatique des données de RRBS.....	107
Figure 67. Rapports de contrôles qualité fournis par le logiciel FastQC	108
Figure 68. Contrôle qualité, nettoyage des séquences et alignement sur le génome de référence	108
Figure 69. Identification et annotation des cytosines et régions différenciellement méthylées	109
Tableau 7. Liste des principales techniques développées pour l'analyse de la méthylation de l'ADN	94

Résultats

Bull spermatozoa display a genome-wide DNA hypomethylation with both species peculiarities and conserved tissue-specific features

Figure 1. Global methylation measured by LUMA is low in bovine sperm.....	147
Figure 2. Cell type is a massive determinant of DNA methylation landscapes in cattle.....	148
Figure 3. Two distinct sets of differentially methylated regions and differentially methylated CpGs display hypomethylation in sperm	149
Figure 4. Hypo-DMRs identified by MeDIP target genes involved in mRNA processing and spermatogenesis	150
Figure 5. Hypo-DMCs identified by RRBS target specific genomic features and functions.....	151
Figure 6. Validation by pyrosequencing.....	152

Identification de nouveaux biomarqueurs de la fertilité mâle dans le méthylome de la semence bovine : une analyse par RRBS

Figure 1. Dispositif expérimental	156
Figure 2. Analyses phénotypiques.....	164
Figure 3. Analyses fonctionnelles	165
Figure 4. Contrôles qualité des séquences de RRBS	166
Figure 5. Comparaisons de la méthylation entre les races Holstein et Montbéliarde	167
Figure 6. Analyse descriptive comparant les races Holstein et Montbéliarde	167
Figure 7. Méthylation globale des taureaux Holstein et Montbéliard, avec distinction entre les catégories de fertilité	168
Figure 8. Analyse descriptive comparant les catégories de fertilité pour les races Holstein et Montbéliarde	168
Figure 9. Analyse différentielle entre les catégories de fertilité, pour les races Holstein et Montbéliarde	169
Figure 10. Analyse discriminante pour les DMC identifiées.....	171
Figure 11. Annotation des DMC et DMR	172
Figure 12. Processus biologiques enrichis en DMC – enrichissements pour les comparaisons Fertiles+ contre Fertiles-, Déviants/Fertiles+, Déviants/Fertiles- – taureaux Holstein	173/174
Figure 13. Processus biologiques enrichis en DMC – enrichissements pour les comparaisons Fertiles+ contre Fertiles-, Déviants/Fertiles+ et Déviants/Fertiles- – taureaux Montbéliard	175
Figure 14. Validations par bisulfite-pyroséquençage sur 3 des gènes pour les taureaux Holstein.....	176
Figure 15. Validations par bisulfite-pyroséquençage sur 1 autre gène pour les taureaux Holstein, et le gène LAMB2 pour les taureaux Holstein et Montbéliard	176
Figure 16. Corrélation entre les résultats de RRBS et de bisulfite-pyroséquençage.....	177
Figure 17 Répartition des polymorphismes de séquence sur la DMC du gène SMARCD3	178
Figure 18. Intégration des données de génotypage aux DMC	178
Figure 19. Relation spatiale entre les DMC et les SNP pour la race Holstein	178
Figure 20. Résultats obtenus par la double digestion enzymatique HaeIII + MspI	180
Figure 21. Identification et caractérisation des DMC et DMR pour la double digestion HaeIII + MspI.....	181
Figure 22. Processus biologiques enrichis en DMC – enrichissements pour les comparaisons Fertiles+ contre Fertiles-, pour les taureaux Holstein analysés par double digestion enzymatique	182
Figure 23. Processus biologiques enrichis en DMC – enrichissements pour les comparaisons Déviants/Fertiles+, pour les taureaux Holstein analysés par double digestion enzymatique.....	182
Figure 24. Processus biologiques enrichis en DMC – enrichissements pour les comparaisons Déviants/Fertiles-, pour les taureaux Holstein analysés par double digestion enzymatique.....	182
Figure 25. Différences d'âge entre les catégories de fertilité	183
Figure 26. Prise en compte de l'âge à la production.....	184
Figure 27. Processus biologiques enrichis en DMC – enrichissements pour les comparaisons Fertiles+ contre Fertiles-, Déviants/Fertiles+, Déviants/Fertiles- – taureaux Holstein analysés après retrait des DMC liés à l'âge.....	184

Figure 28. Régulation du lignage Trophectoderme-ICM par la voie Hippo dans l'embryon préimplantatoire	196
Tableau 1. Comparaisons de la méthylation entre les races Holstein et Montbéliarde	168
Tableau 2. Implications biologiques des gènes présentant une DMR et validés par bisulfite-pyroséquenceage	176
Tableau 3. Liste des gènes de la voie Hippo impactés par un différentiel de méthylation	186

Discussion générale

Figure 1. Mesure de la 5mC et 5hmC globale par ELISA dans la semence fraîche et congelée de taureaux	204
Figure 2. Représentation schématique des méthodes de RRBS, de oxRRBS et de TAB-seq	205
Figure 3. Analyse descriptive prenant en compte toutes les races du dispositif	217

Introduction

Préambule

Au cours de la période Néolithique, les premières traces de domestication datent de 14 000 à 12 000 ans. Elles représentent l'un des tournants majeurs ayant contribué au développement des civilisations Humaines. Parmi les premiers animaux domestiqués, les bovins tiennent une place essentielle. Leur domestication apparaît au Proche-Orient vers 6500 avant J.C (Zeder 2009). Les bovins ont fourni l'Homme en aliments (lait, viande), en matériaux (os et cuir) et l'ont soutenu dans les travaux agricoles. La domestication s'est progressivement accompagnée d'une sélection par utilisation, comme reproducteurs, des individus au sein des troupeaux ayant les performances ou les aptitudes recherchées par les éleveurs : force physique, richesse et quantité de lait produit par les vaches, docilité... Au gré des adaptations à des conditions environnementales variées et des processus de sélection, il est dénombré aujourd'hui près de 900 races bovines à travers le monde (*Commission on genetic resources for food and agriculture, Food and Agriculture Organization of the United Nations, Rome 2007*).

Basée initialement sur l'observation de la conformation des animaux, la sélection s'est rationalisée à partir du XIX^{ème} siècle et surtout après la Seconde Guerre Mondiale, avec la création de l'Institut National de la Recherche Agronomique (INRA) et le développement de la génétique quantitative sous l'impulsion de chercheurs tels B. Vissac et J. Poly. Délaissant les performances propres de l'animal, la sélection se fonde alors progressivement sur l'évaluation des performances de la descendance des taureaux (« testage »), profitant pleinement de l'essor de l'Insémination Animale (IA). L'IA consiste en un dépôt intra-utérin de la semence préalablement prélevée sur un taureau et conditionnée en paillettes. Outre les avantages sur le plan de la sécurité sanitaire, cette technique permet une large diffusion du progrès génétique et, ce faisant, facilite l'évaluation des reproducteurs sur descendance. Avec les progrès de la connaissance sur le génome, les méthodes d'évaluation se sont aussi perfectionnées, pour aboutir aujourd'hui à une sélection « génomique », où les animaux sont évalués directement par une analyse ADN (génotypage) grâce à des équations de prédiction établies sur des populations de référence phénotypées et génotypées. En quelques années, le génotypage s'est imposé en France comme la méthode de choix pour les races à fort effectif (grandes races laitières et allaitantes, principales races laitières régionales). Plus de 95% des taureaux mis en marché sont issus de la sélection génomique et plus de 100.000 génisses sont testées chaque année afin de mieux gérer le renouvellement des troupeaux.

La sélection génomique introduit un nouveau paradigme pour l'évaluation du potentiel reproducteur des mâles. En effet, lors du testage sur descendance, une centaine de filles étaient procréées par taureau, permettant par conséquent de mesurer la fertilité de la semence avant diffusion à grande échelle des paillettes. Avec la sélection génomique, les taureaux sont aujourd'hui sélectionnés par l'analyse de leur génome dans les premiers mois de vie et leurs paillettes sont commercialisées sans information sur leur fertilité effective, puisque ces taureaux n'ont alors pas encore de descendants. Or, l'efficacité reproductive est un facteur clef de l'efficacité globale de l'élevage bovin. L'utilisation de taureaux sub-fertiles se traduit par des pertes de production et une augmentation du taux de réforme des vaches pour infertilité, induisant des pertes économiques importantes et un gaspillage des ressources biologiques.

Garantir le progrès génétique et la durabilité des élevages suppose de faire évoluer les procédés de contrôle qualité de la semence afin de prendre en compte ce nouveau contexte. En particulier, il est essentiel de pointer des défauts épigénétiques pouvant avoir un retentissement sur le développement embryonnaire ou le phénotype de la descendance, et d'en identifier les éventuelles causes environnementales. L'identification de marques épigénétiques pourrait ainsi permettre d'affiner l'évaluation génomique, et contribuer à mettre en place des indicateurs de gestion individuelle des animaux dans une logique d'élevage de précision.

L'UMR 1198 Biologie du Développement et Reproduction (BDR) est une unité mixte de recherche du Centre INRA de Jouy-en-Josas dont les thématiques de recherche portent sur les mécanismes développementaux qui contribuent à l'établissement des phénotypes, en particulier via la programmation précoce et la régulation de l'expression du génome par l'environnement. L'équipe d'Hélène Jammes, en collaboration avec l'équipe R&D d'ALLICE, la fédération des Entreprises de sélection et des coopératives d'IA, pilote le laboratoire commun (Labcom ANR) SeQuaMol « Qualité Moléculaire de la Semence ». L'objectif du projet SeQuaMol, dans lequel s'inscrit mon travail de thèse, est le développement du contrôle qualité épigénétique de la semence, afin de mieux prédire et d'optimiser la fertilité chez le taureau.

Le manuscrit comporte en introduction des données sur la reproduction du mâle en se concentrant sur le modèle bovin, puis des données sur l'organisation de la filière professionnelle concernant la semence bovine mettant en œuvre la sélection génomique et l'IA. Nous consacrons aussi un chapitre afin de mieux cerner l'apport de la recherche en épigénétique et la contribution de l'identification de marqueurs épigénétiques de la fertilité afin d'écarter les taureaux sub-fertiles des élevages.

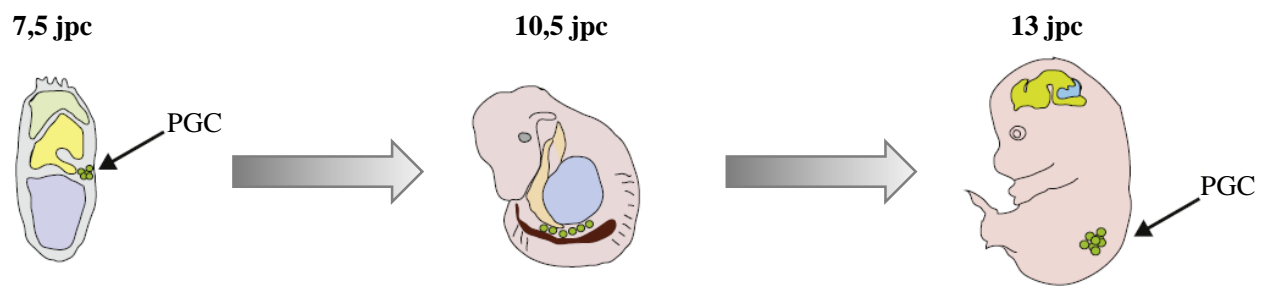


Figure 3. Migration des PGC au cours du développement embryonnaire chez la souris

Les précurseurs des cellules germinales se différencient dans les tissus extra-embryonnaires à la base de l'allantoïde. Ils vont migrer, se multiplier et coloniser les crêtes génitales pour former la réserve de cellules germinales primordiales (PGC).

jpc, jours *post-coïtum*

(adapté de Bogdanović et Lister, 2017)

Chapitre I : La semence bovine, vecteur du patrimoine génétique paternel et produit à forte valeur ajoutée

Les schémas d'organisation de la reproduction sont assez semblables entre les espèces de mammifères. Pour simplifier la description et rester dans le contexte de la filière de la semence bovine, dans ce document ne sera décrit dans la mesure du possible que celle du taureau. Une cinétique comparative des principaux événements dans la mise en place du processus de reproduction chez le mâle et des principaux événements de la spermatogenèse sont donnés respectivement Figure 1 et Figure 2.

Dans le but de brosser un schéma complet du processus de reproduction, cette partie débutera par la différenciation et le développement des organes génitaux mâles au stade embryonnaire, puis décrira la puberté et la production des gamètes mâles, et enfin évoquera la fécondation et le développement embryonnaire.

1. Aspects biologiques : la voie mâle dans la reproduction

1.1. La structuration du testicule bovin

1.1.1. La différenciation des gonades fœtales

Après la fécondation et durant les premières semaines de développement, les gonades embryonnaires se développent de manière indifférenciée. La crête génitale, originaire du mésonéphros, apparaît entre le 27^{ème} et le 31^{ème} jour *post-coitum* (jpc). Elle constitue alors l'ébauche des gonades et en formera la composante somatique. Chez la souris, il a été montré que la formation de la gonade indifférenciée est sous le contrôle de gènes codant pour des facteurs de transcription comme le *Steroidogenic factor-1 (Sf1)*, *Wilms tumor 1 (Wt1)*, *Empty spiracles homeobox 2 (Emx2)*, *Lim homeobox 1 (Lim1)* ou *M33* (Capel 2000).

Les cellules germinales primordiales (*Primordial Germ Cells*, PGC) sont d'origine extra-embryonnaire, en provenance du sac vitellin (*yolk sac*) chez le bovin. Les PGC sont les premières cellules de la lignée germinale. Elles migrent et colonisent les crêtes génitales entre le 25^{ème} et le 31^{ème} jpc (Wrobel et Süß 1998)(Figure 3). La crête génitale croît alors rapidement en taille pour atteindre une forme globulaire au 38^{ème} jpc. Chez la souris, la

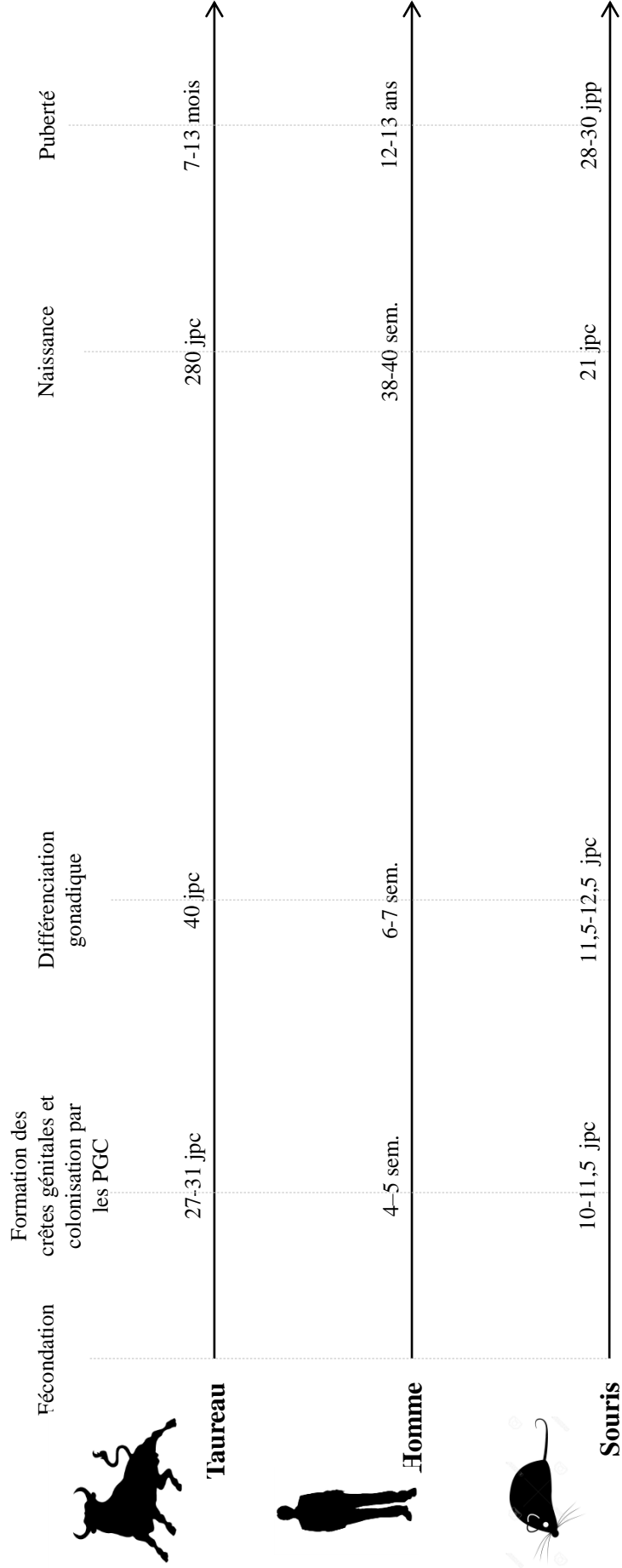


Figure 1. Chronologie comparative du développement de la fonction reproductrice chez le taureau, l'Homme et la souris

La chronologie débute à partir de la fécondation.

jpc, jours post-coïtum ; jpp, jours post-partum

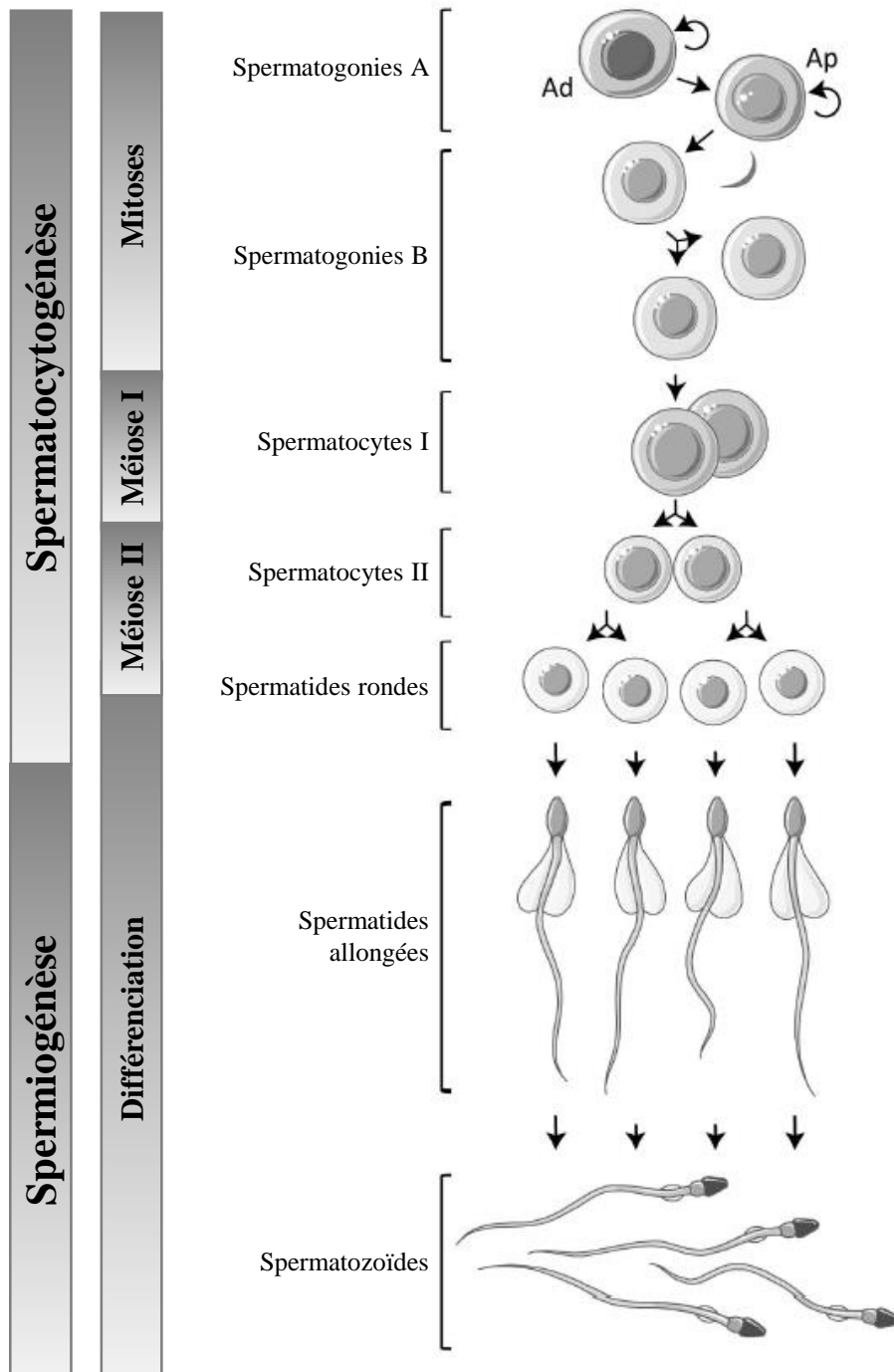


Figure 2. Chronologie des grandes étapes de la spermatogénèse

La spermatogénèse est un processus continu qui peut être subdivisé en 2 grandes étapes : la spermatocytogénèse et la spermiogénèse. La spermatocytogénèse correspond aux différentes phases permettant à partir de cellules souches, les spermatogonies ($2n$) de générer des spermatides rondes (n). Le renouvellement du stock de cellules souches est assuré par de nombreuses mitoses des spermatogonies A. Certaines spermatogonies A enclenchent une première différenciation qui donne naissance aux spermatogonies B puis aux spermatocytes de type I. Ces cellules entrent alors en méiose afin d'aboutir à la formation de spermatocytes de type II puis à des spermatides rondes et de passer d'un statut diploïde à un statut haploïde : pour ce faire, 2 phases méiotiques se mettent en place. Les spermatocytes I à $2n$ chromosomes subissent une division réductionnelle et donne ainsi 2 spermatocytes II à n chromosomes. Chaque spermatocyte II subit la deuxième division de méiose et donne 2 spermatides à n chromosomes. La dernière phase est la spermiogénèse, phase pendant laquelle les spermatocytes vont progressivement se différencier en spermatides allongées puis en spermatozoïdes.

(adapté de Jégou, Rolland et Albert, La reproduction animale et humaine, Synthèse, 2014)

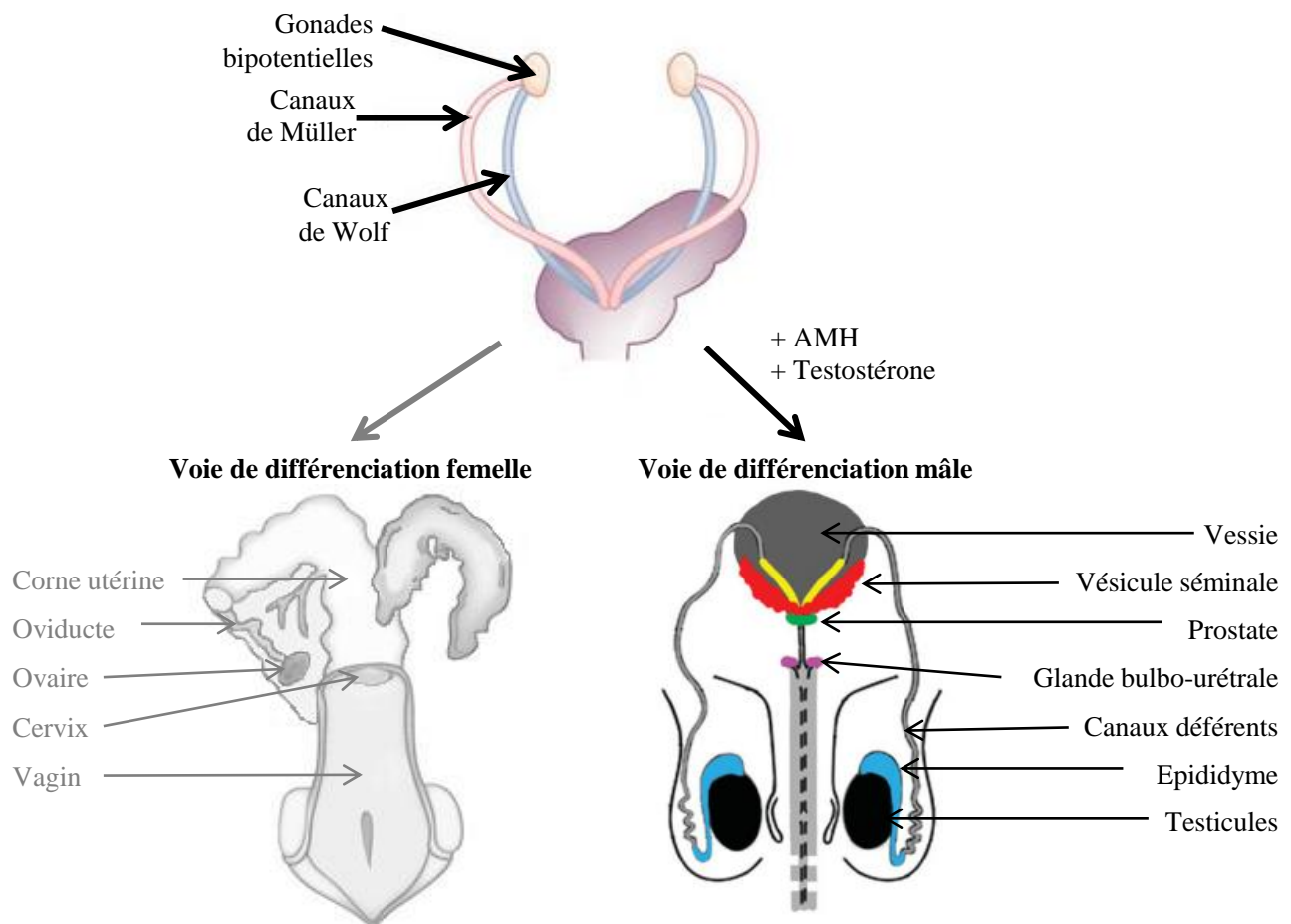


Figure 4. Différenciation du tractus génital pendant la période embryonnaire chez le bovin

La sécrétion de l'hormone anti-müllérienne (AMH) par le testicule fœtal provoque la régression des canaux de Müller. Dans la voie de différenciation femelle, les canaux de Müller donneront les oviductes, l'utérus et la partie supérieure du vagin. La sécrétion de testostérone est responsable de la transformation des canaux de Wolf en les canaux déférents, l'épididyme et les vésicules séminales.

(adapté de Matzuk et Lamb, 2008)

migration et la prolifération des PGC est sous la dépendance des facteurs de croissance *Stem cell factor (Scf)*, *Leukemia inhibitory factor (Lif)* ou *Fibroblast growth factor (FGF)* (De Felici 2000).

Si le génome de l'embryon est porteur d'un chromosome sexuel Y, les gonades embryonnaires se dirigent vers une voie de différenciation de type testiculaire. Cette voie de différenciation est enclenchée par l'activité d'un gène présent sur le chromosome Y, le *Sex determining Region of the Y-chromosome (SRY)*. L'expression de *SRY* est détectée dès les stades 4 à 8 cellules chez le bovin, et tout au long du stade blastocyste (Gutiérrez-Adán et al. 1997). Il s'agit d'un facteur de transcription activant ou réprimant l'expression de gènes favorisant la voie de différenciation mâle. Par exemple, il active en partenariat avec *SFI* le gène *SRY-related HMG-box gene 9 (Sox9)*. *Sox9* est impliqué dans la différenciation et l'organisation des cellules somatiques de soutien de la lignée germinale, les précurseurs des cellules de Sertoli (Sekido et Lovell-Badge 2008).

Le processus de morphogénèse testiculaire prend place au 40^{ème} jpc chez le bovin et est principalement orchestré par les cellules de Sertoli. A cette période débute la formation des cordons testiculaires (qui donneront les futurs tubes séminifères) et l'apparition de la vascularisation du testicule. Ces événements sont sous l'action des hormones testiculaires (Figure 4). Les cellules de Sertoli, sécrètent l'hormone anti-Müllérienne (*Anti-Müllerian Hormone, AMH*) qui induit la dégénérescence des canaux de Müller (Jost et al. 1973). Les cellules de Leydig de type fœtal sont présentes dans l'espace interstitiel et sécrètent les androgènes, parmi lesquels la testostérone est la composante la plus importante. La sécrétion de testostérone provoquera la transformation des canaux de Wolf pour masculiniser le tractus génital et ainsi donner l'épididyme et les canaux déférents. Les cellules de Leydig sécrètent aussi l'*Insulin-like factor 3 (INSL3)*, dont un des rôles est la descente testiculaire dans le scrotum (Adham 2000).

Une fois formé, le testicule croît rapidement en période fœtale (Orth 1993). Les gènes impliqués dans la régulation de la morphogénèse testiculaire sont notamment le *Transforming Growth Factor Beta (TGFB)*, les *Insulin-like Growth Factors (IGF)* ou encore le *Fibroblast Growth Factor 9 (FGF9)*. Au stade fœtal démarre la différenciation précoce des PGC en cellules germinales spermatogoniales. C'est la phase de pré-spermatogénèse. Les cellules germinales continuent à proliférer par vague de mitose, notamment entre 50 et 80 jpc. Cette prolifération décroît ensuite en intensité jusqu'à s'arrêter à 200 jpc (Wrobel 2000). Elles n'entrent cependant pas en méiose à ce stade, et restent à un stade de différenciation appelé « pré-spermatogonie ».

1.1.2. De la naissance à la puberté

A la naissance, les pré-spermatogonies sont en état de quiescence. La structure générale des testicules est cependant bien formée. Les testicules sont composés de centaines de tubes séminifères dans lesquels les cellules germinales sont étroitement associées aux cellules de Sertoli. Les tubes séminifères sont entourés du tissu interstitiel contenant les cellules de Leydig. Les premières semaines après la naissance, une fraction importante des pré-spermatogonies entre en phase dégénérative (Wrobel 2000). A ce stade, les sécrétions en hormones sexuelles sont encore très faibles. A partir de la fin du premier mois postnatal, on observe une augmentation du diamètre des tubes séminifères et de la concentration plasmatique des gonadotropines originaires de la glande pituitaire antérieure : l'hormone lutéinisante (*Luteinizing hormone*, LH) et l'hormone folliculostimulante (*Follicle-Stimulating Hormone*, FSH). Leurs sécrétions sont régulées par la sécrétion pulsatile de l'hormone de libération des gonadotrophines hypophysaires (*Gonadotropin-Releasing Hormone*, GnRH) par l'hypothalamus. LH et FSH agissent sur le testicule pour stimuler respectivement la production de testostérone et d'inhibine (Rawlings et al. 2008). Ces sécrétions hormonales correspondent à la mise en place de l'axe hypothalamo-hypophyso-testiculaire (Figure 5). A ce stade, les pré-spermatogonies entrent alors en phase proliférative.

La période prépubère s'étend de la 9^{ème} à la 22^{ème} semaine. Un pic de concentration en LH est atteint entre la 12^{ème} et la 16^{ème} semaine, et la sécrétion de testostérone augmente également (Figure 6). Les premières différenciations cellulaires des pré-spermatogonies en spermatogonies apparaissent. A la 25^{ème} semaine, le développement testiculaire s'accélère : les tubes séminifères croissent en diamètre et en longueur. Les impulsions de GnRH sont plus fréquentes, de 4 à 8 fois par jour. Les concentrations plasmatiques en FSH et LH diminuent progressivement, et la concentration en testostérone augmente rapidement. Certaines cellules germinales commencent à proliférer et à entrer en méiose.

Chez le bovin, la puberté est généralement atteinte entre le 7^{ème} et le 13^{ème} mois. La grande variabilité dans l'âge à la puberté dépend de facteurs génétiques et notamment de la race, mais aussi de facteurs environnementaux (Lunstra, Ford, et Echternkamp 1978). On considère généralement qu'un taureau a définitivement atteint sa puberté quand il peut produire un éjaculat contenant $50 \cdot 10^6$ spermatozoïdes, dont plus de 10% sont motiles (la motilité étant entendue comme capacité de se déplacer de manière autonome) (Wolf, Almquist, et Hale 1965). Il faut également que le développement pénien soit entièrement réalisé, et que la circonférence scrotale atteigne au moins 28cm. On considère alors que le taureau est apte à la reproduction.

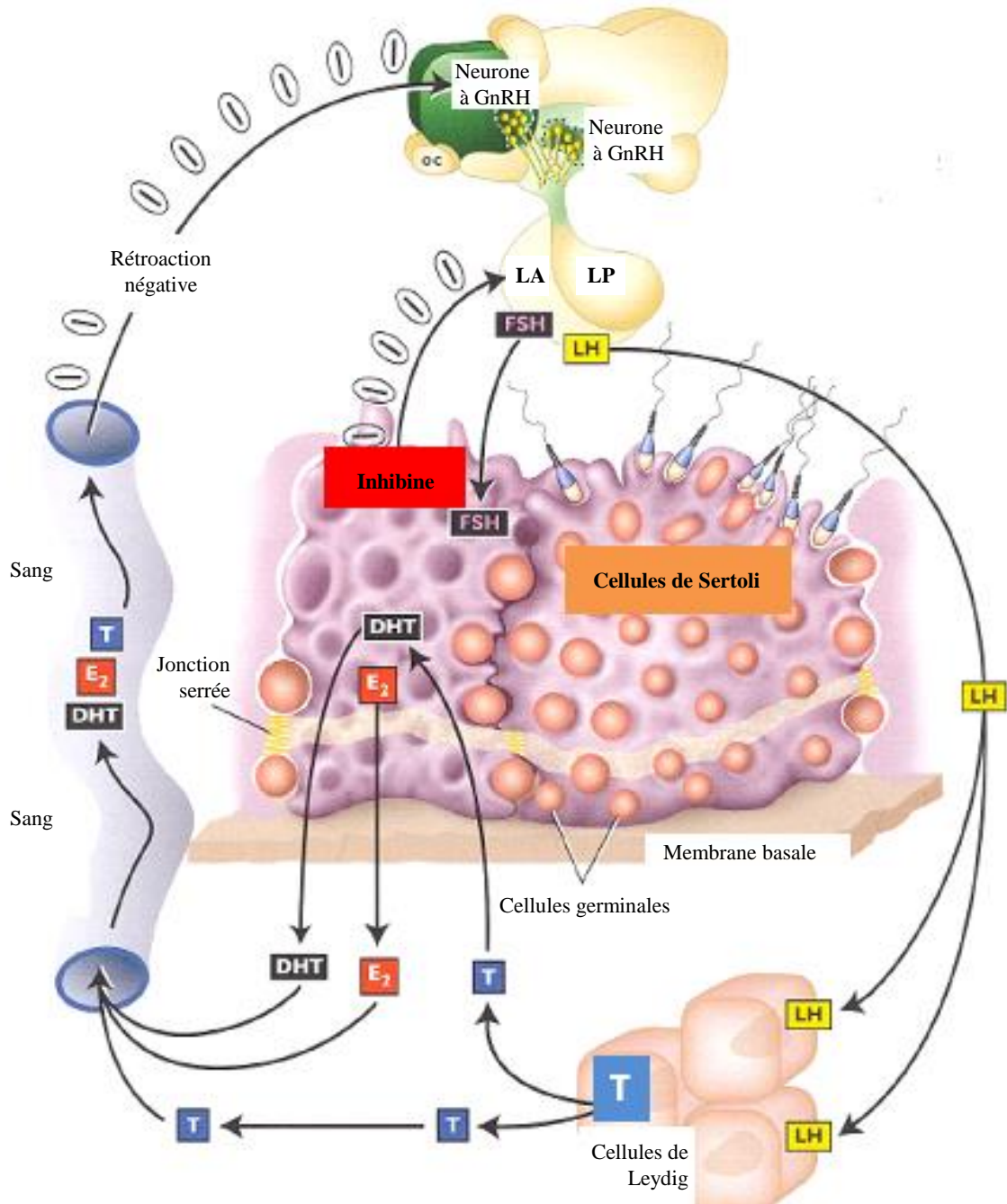


Figure 5. Représentation de l'axe hypothalamo-hypophysio-testiculaire

Les neurones à GnRH de l'hypothalamus sécrètent le GnRH de façon pulsatile. Ces sécrétions déclenchent la sécrétion de LH et de FSH. La LH agit sur les cellules de Leydig qui en possèdent les récepteurs. En réponse, elles sécrètent la testostérone, qui dans les cellules de Sertoli sera convertie en DHT. Les cellules de Sertoli sont responsables de la conversion de testostérone en estradiol et autres estrogènes, ainsi que de la production d'inhibine sous l'action de la FSH. Alors que l'estradiol joue, conjointement avec la testostérone et la DHT, le rôle de régulateur négatif de la production de LH et FSH en agissant sur la production de GnRH, l'inhibine agit directement sur la sécrétion de FSH dans le lobe antérieur de la glande pituitaire.

DHT, dihydrotestostérone; E₂, estrogène; FSH, Hormone folliculostimulante; LA, lobe antérieur de la glande pituitaire; LH, Hormone Lutéinisante LP, lobe postérieur de la glande pituitaire; T, testostérone (adapté de Senger, 2004).

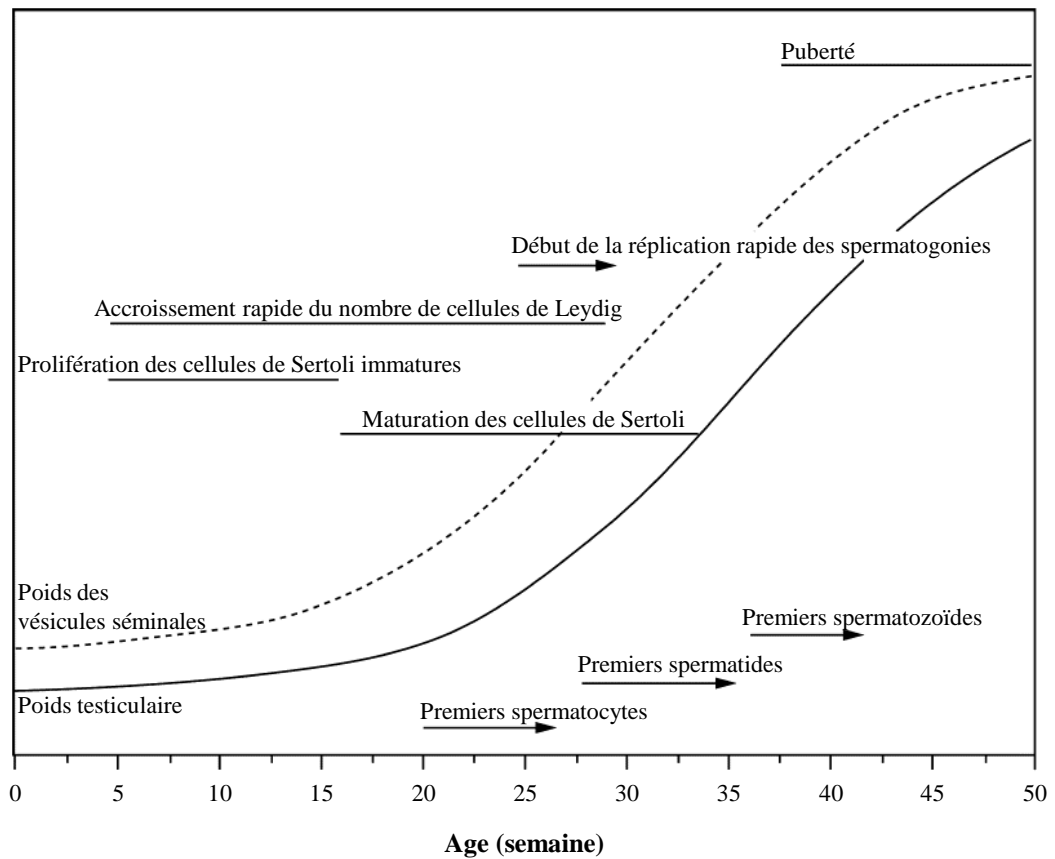
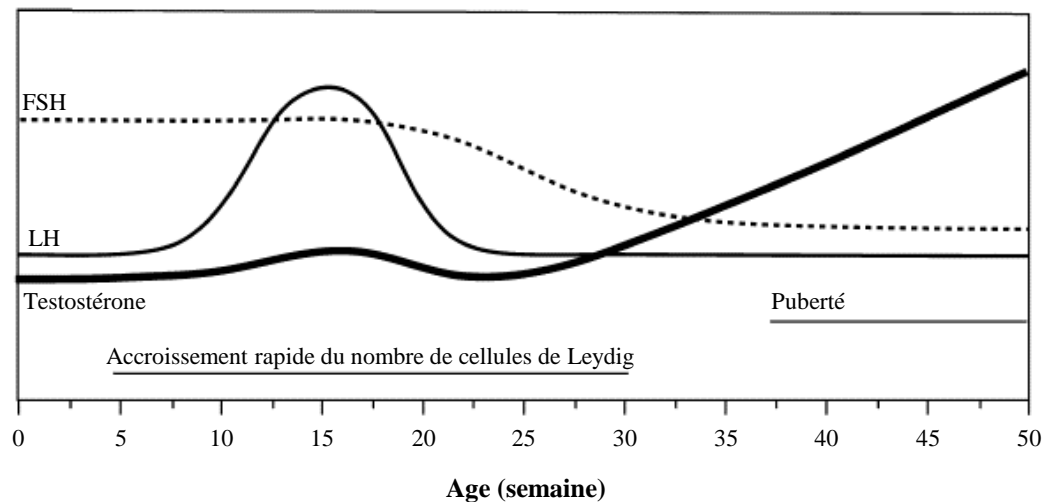
A**B**

Figure 6. Cinétique de la maturation sexuelle chez le taureau

A. Croissance et développement des éléments du système reproducteur mâle bovin, de la naissance à la puberté. Les lignes horizontales indiquent les périodes approximatives pendant lesquelles les activités indiquées prennent place. Les flèches indiquent le point de départ d'une activité continue.

B. Profil temporel de la concentration plasmatique des hormones sexuelles de la naissance à la puberté. Les lignes horizontales indiquent les périodes approximatives pendant lesquelles les activités indiquées prennent place.

(adapté de Rawlings *et al.*, 2008)

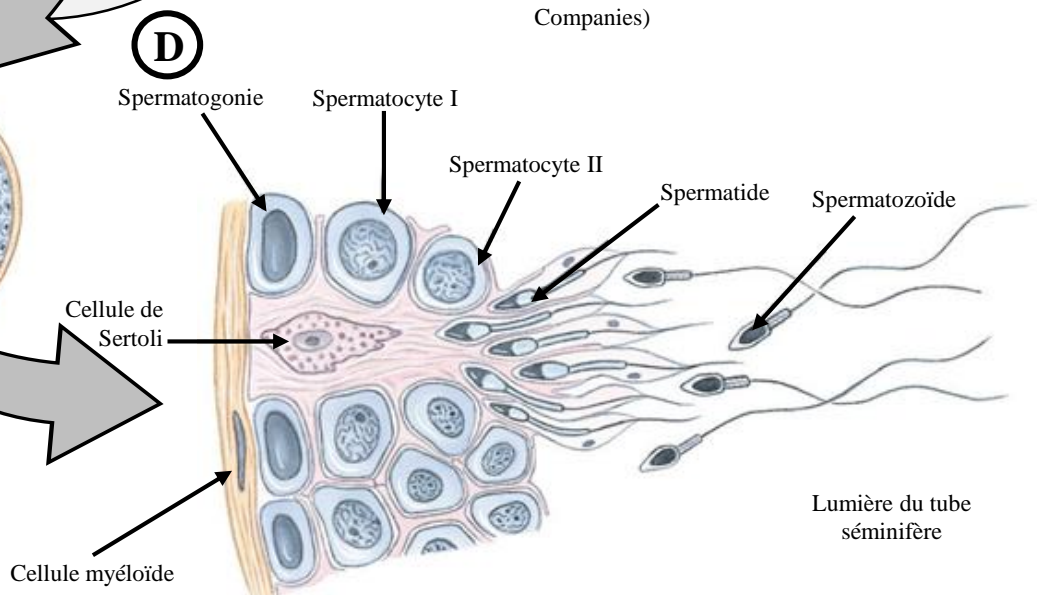
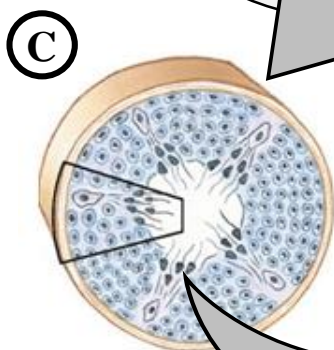
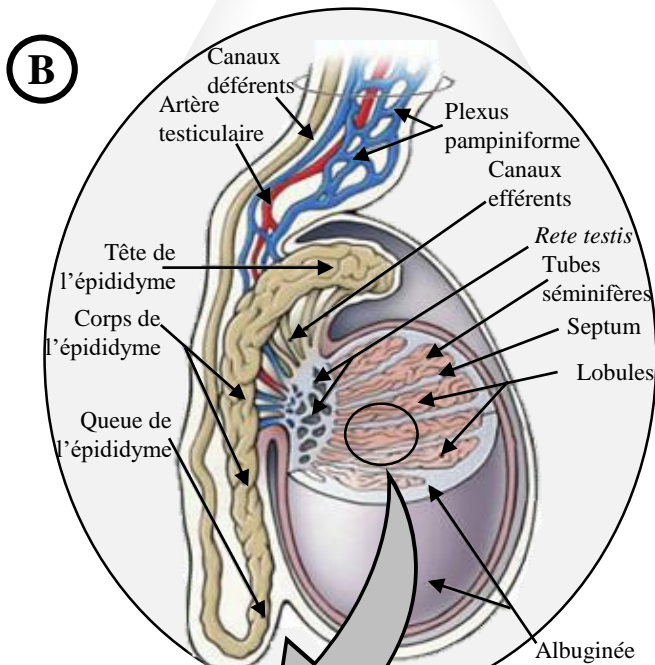
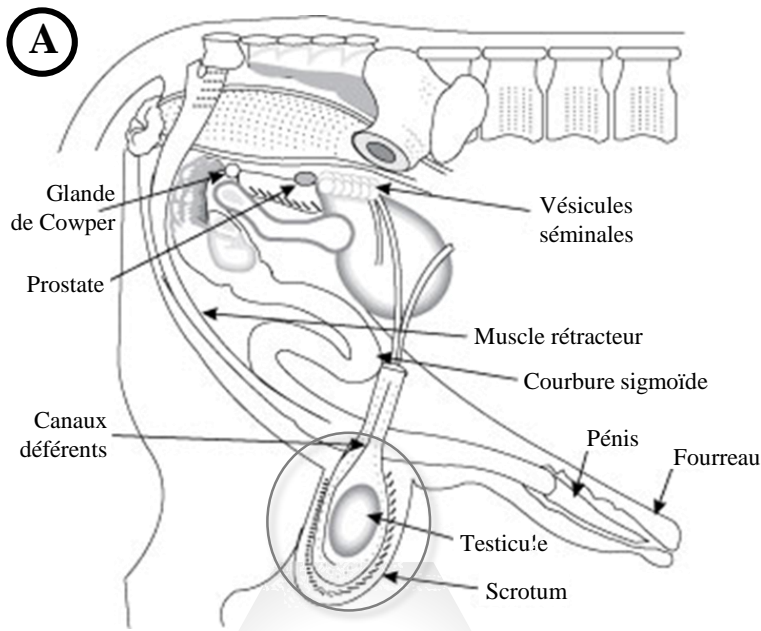


Figure 7. Organisation du tractus génital mâle bovin

A. Ultrastructure de l'appareil génital mâle bovin. Le tractus est composé des testicules, des organes sexuels secondaires (épididymes, canaux déférents et pénis), ainsi que de 3 glandes accessoires (vésicules séminales, prostate et glande bulbo-urétral (ou glande de Cowper). **B.** Structure du testicule. Les testicules sont entourés de l'albuginée, dont des expansions (septum) cloisonnent le testicule en centaines de lobules. Chaque lobule comporte 1 à 4 tubes séminifères. Ces derniers se raccordent au *rete testis*, sur lequel s'abouche les canaux efférents, puis la tête de l'épididyme. L'épididyme est composée de 3 segments (tête, corps et queue), qui donnent sur les canaux déférents. L'irrigation des testicules est réalisée par le biais de l'artère testiculaire et du plexus pampiniforme, issu de l'anastomose des veines testiculaires. **C.** Coupe transversale d'un tube séminifère. **D.** Section d'un tube séminifère. Les spermatogonies sont situées à la membrane basale. Alors qu'une partie d'entre elles se divisent par mitose pour en conserver leur réserve, une autre partie entre en différenciation pour former les spermatocytes I. Au cours des phases de méiose et de différenciation des précurseurs des spermatozoïdes, les cellules s'avancent vers la lumière des tubes séminifères. Les spermatozoïdes seront relargués dans la lumière lors de la spermiation. Tout au long du processus, les précurseurs des spermatozoïdes sont en étroite relation avec les cellules de Sertoli.

(adapté de Duke University School of Medicine ; The McGraw-Hill Companies)

1.2. La production des spermatozoïdes

La spermatogenèse peut être définie comme l'ensemble des processus de division et de différenciation cellulaires nécessaires à la formation de spermatozoïdes fonctionnels. Chez l'adulte, un cycle complet de spermatogenèse dure 54 jours chez le bovin et comprend 14 stades de différenciation cellulaire (Berndston et Desjardins 1974). L'ensemble du cycle requiert le contrôle de l'expression de plusieurs milliers de gènes, qui ont essentiellement été décrits chez l'Homme et la souris. Leurs rôles dans le contrôle épigénétique et transcriptionnel de la spermatogenèse sont abordés dans le chapitre III.

La spermatogenèse peut être décomposée en 2 phases principales : la spermatocytogenèse et la spermiogenèse. Ces étapes ont lieu dans les tubes séminifères des testicules (Figure 7). La libération des spermatozoïdes dans l'épididyme est appelée spermiation. Elle est suivie d'un processus de maturation continu des spermatozoïdes pendant leur transit dans l'épididyme et les voies femelles, jusqu'au moment de la fécondation.

1.2.1. La spermatocytogenèse

La première phase de la spermatogenèse, appelée spermatocytogenèse, recouvre l'ensemble des étapes de différenciation des spermatogonies en spermatocytes. L'ensemble du processus dure 13,5 jours chez le bovin (Hochereau 1964). La spermatocytogenèse est elle-même subdivisée en 2 phases, l'une mitotique et l'autre méiotique. Les spermatogonies, situées à la membrane basale des tubes séminifères et en étroite relation avec les cellules de Sertoli, se multiplient par mitose et constituent ainsi une réserve constante de cellules souches. A chaque division, certaines d'entre elles entrent en voie de différenciation pour former progressivement la réserve de spermatocytes. Les stades de différenciation ont été définis sur des critères morphologiques des cellules (spermatogonies A₀, puis A₁, A₂, A₃, intermédiaires, B₁ et B₂ chez le bovin). A la fin de la phase mitotique, les cellules atteignent le stade de spermatocytes de type I et sont prêtes à entrer en méiose.

La méiose est le processus de division cellulaire permettant de passer d'une cellule diploïde à 2n chromosomes (dans notre cas, les spermatocytes de type I) à une cellule haploïde à n chromosomes (les spermatides). La phase méiotique de la production des gamètes est composée d'une phase de réplication de l'ADN suivie immédiatement de 2 cycles de division. La première division méiotique est dite réductionnelle car elle permet de séparer les chromosomes homologues. Les chromosomes homologues s'apparient et forment des structures appelées bivalents. Ces structures vont permettre les recombinaisons

chromosomiques, à la base de la diversité génétique qui sera propre au nouvel individu formé après fécondation par rapport à ses géniteurs. La prophase du premier cycle de méiose est très spécifique et est divisée en plusieurs sous-étapes (Figure 8). La seconde division méiotique intervient après une courte interphase et est qualifiée d'équationnelle car elle se passe de réplication d'ADN. Les chromatides sœurs se séparent à ce stade. Tout au long de ces étapes de multiplication et de différenciation, les spermatogonies et les spermatocytes migrent de la membrane basale en direction de la lumière des tubes séminifères. Une méiose altérée peut entraîner un arrêt de la spermatogenèse ou produire des spermatozoïdes non-fonctionnels. Une mauvaise ségrégation des chromosomes (notamment par l'inactivation de certains gènes impliqués dans la recombinaison) peut conduire à une aneuploïdie (un ou plusieurs chromosomes manquants) ou une polyploïdie (un ou plusieurs chromosomes surnuméraires) dans les spermatides. Il est également possible d'observer une mauvaise réparation des cassures double brins de l'ADN nécessaires à la recombinaison. La réparation des cassures double brins est sous le contrôle du complexe protéique RAD50-MRE11 et des gènes *SPO11*, *RAD51* et *DMC1* (Keeney 2008).

1.2.2. La spermiogenèse

Les spermatides nouvellement formées entrent ensuite dans des phases de profondes modifications structurelles regroupées sous le nom de spermiogenèse. Les spermatides passent par 14 stades morphologiques distincts, établis sur des critères d'observation morphologique par microscopie (Leblond et Clermont 1952) (Figure 9). La première modification est celle de la formation de l'acrosome à la tête du futur spermatozoïde. L'acrosome est un réservoir enzymatique formé à partir de l'appareil de Golgi, nécessaire à la fécondation. En parallèle on observe la formation de la pièce intermédiaire (partie riche en mitochondries fournissant l'énergie nécessaire au fonctionnement de la cellule) et du flagelle (partie permettant la motilité cellulaire). Jusqu'à la spermiogenèse, la chromatine nucléaire était organisée de façon similaire aux autres types cellulaires, sous forme de nucléosomes. Pendant la spermiogenèse a lieu la condensation de l'ADN spermatique notamment par le remplacement graduel d'une large majorité des histones par des protéines dites de transition, puis par des protamines (détaillé dans le chapitre III). La condensation nucléaire concourt à la protection de l'intégrité génomique, tout en réduisant la dimension de la tête du spermatozoïde, pour lui conférer une forme hydrodynamique (Malo et al. 2006). Le cytoplasme est progressivement éliminé pour former une gouttelette cytoplasmique résiduelle. La transcription est arrêtée (Grunewald et al. 2005). A la fin de la spermiogenèse, les spermatozoïdes sont évacués dans la lumière des tubes séminifères.

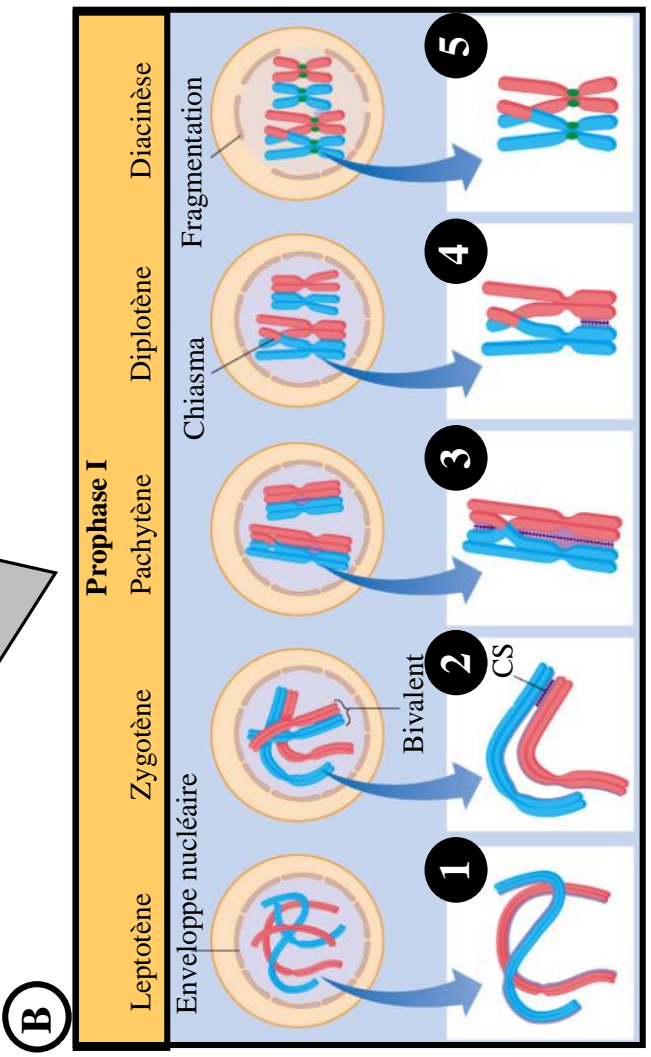
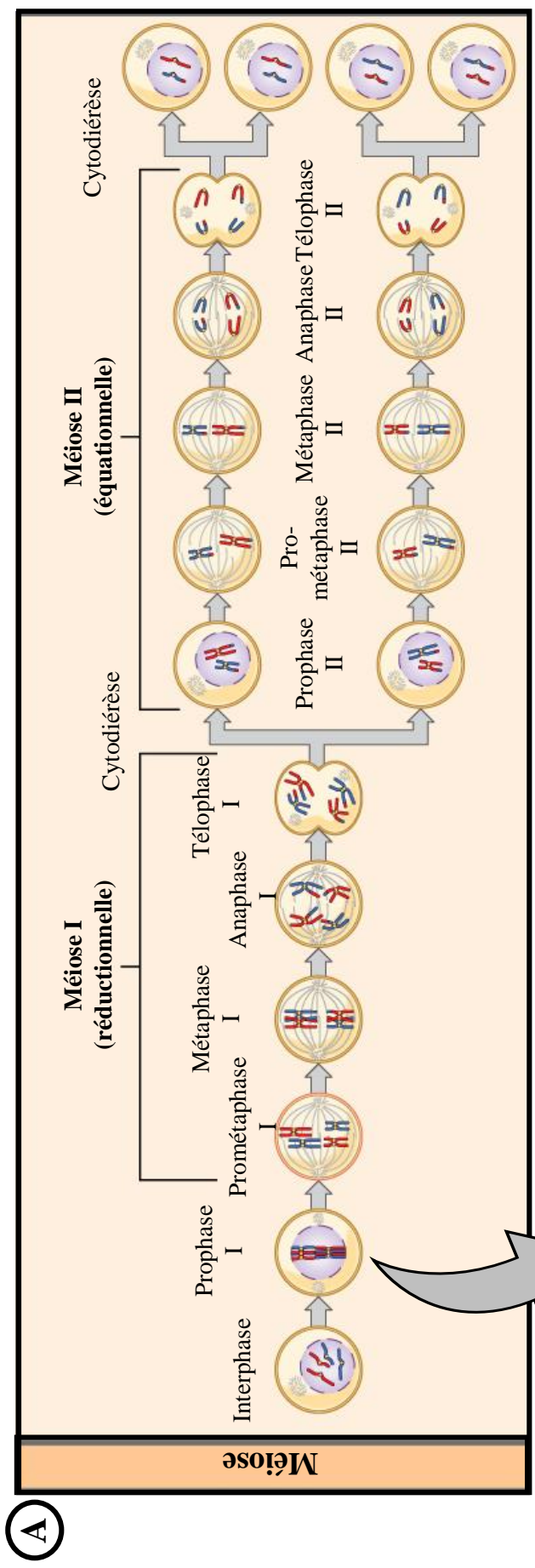


Figure 8. La méiose

A. La méiose débute par une réplication de l'ADN. Chaque chromosome est alors constitué de deux chromatides sœurs. Les deux phases méiotiques sont composées de l'enchaînement des mêmes phases: prophase, prométaphase, métaphase, anaphase et télaphase. Les cellules se séparent définitivement avec la cytotiérèse. La méiose I est dite réductionnelle, car elle permet de passer de $2n$ chromosomes doubles à n chromosomes doubles. La méiose II est dite équationnelle car elle conserve le nombre de chromosomes: on passe de n chromosomes doubles à n chromosomes simples. Au final, à partir d'une cellule diploïde la méiose donne naissance à 4 cellules haploïdes. **B.** La prophase I méiotique est la phase la plus longue, et est elle-même subdivisée en 5 sous-phases: leptotène, zygotène, pachytène, diplotène et diacynèse. **1.** En phase leptotène, la chromatine commence à se condenser et à s'attacher par les télomères à l'enveloppe nucléaire. Les chromatides s'associent par chromatides sœurs. **2.** En phase zygotène, les chromosomes homologues commencent à s'apparier par le complexe synaptonémal (CS). L'ensemble des 2 chromosomes homologues s'appelle un bivalent. **3.** Au stade pachytène, les chromosomes homologues s'apparient et des nodules de recombinaisons apparaissent permettant les échanges entre chromatides. **4.** Au stade diplotène, les chromosomes homologues se séparent, mais restent attachés au niveau des chiasmats. **5.** Au stade diacynèse, l'enveloppe nucléaire commence à disparaître.

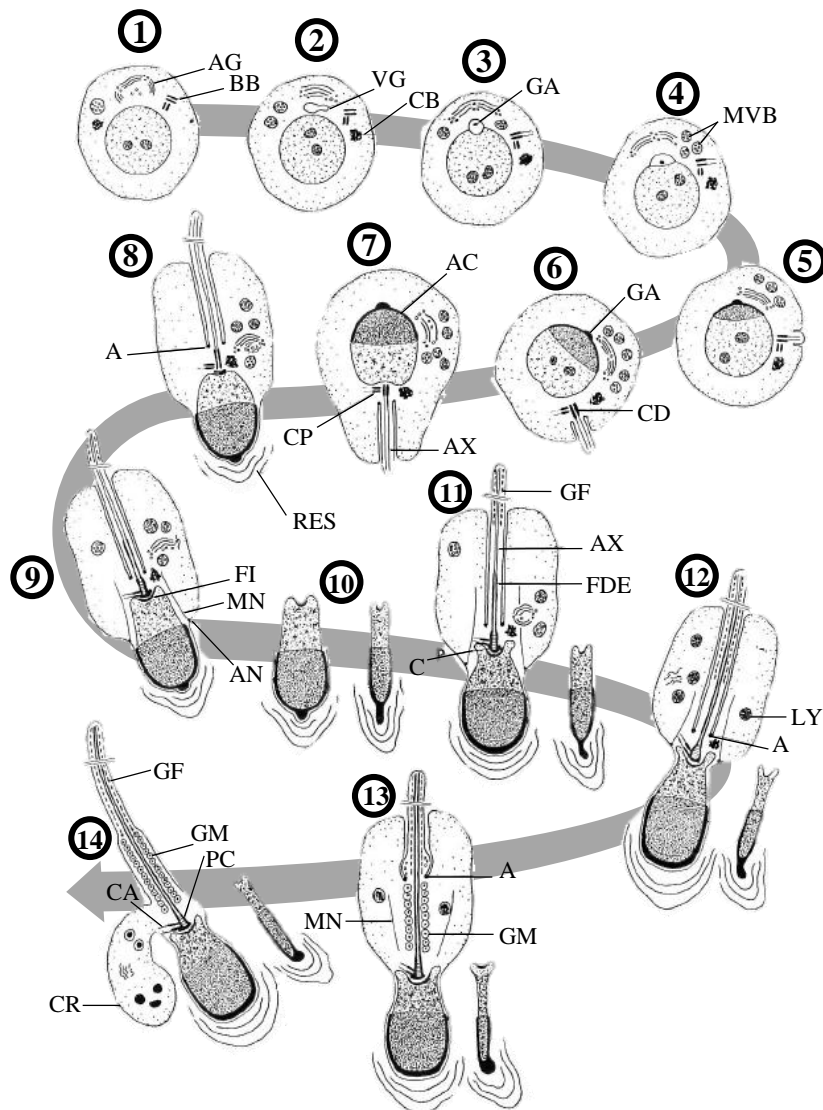


Figure 9. Les étapes de la spermiogénèse chez le bovin

Représentation schématique d'images en microscopie électronique des 14 étapes de la spermiogénèse chez le bovin. Ces étapes peuvent être regroupées en 4 grandes phases chez le bovin. La phase dite de Golgi va de la formation des spermatides rondes (étape 1) à la formation du granule acrosomique associé à la membrane nucléaire (étape 3). Pendant les phases 4 à 7 le capuchon acrosomique des spermatides est en croissance. La phase acrosomique recouvre les étapes 8 à 12 et consiste en des modifications morphologiques majeures de l'acrosome (condensation et élongation). Les étapes 13 et 14 sont des étapes de maturation au cours desquelles les spermatides terminent leur différenciation, et incluent la formation du flagelle et la condensation finale du noyau. AG, Appareil de Golgi ; BB, *Basal Body* (corpuscule basal) ; VG, Vésicule de Golgi ; CB, *Chromatoid Body* (corpuscule chromatoïde) ; GA, Granule Acrosomique ; MVB, *Multivesicular Body* (corpuscule multivésiculaire) ; A, Annulus ; RES, Reticulum Endoplasmique de Sertoli ; CP, Centriole Proximale ; AX, Axonème ; AC, Acrosome ; CD, Centriole Distale ; FI, Fosse Implantatoire ; MN, Manchette ; AN, Anneau Nucléaire ; GF, Gaine Fibreuse ; FDE, Fibres Denses Externes ; C, Capitulum ; LY, Lysosome ; CA, *Centriolar Adjunct* (Accessoire centriolaire) ; GM, Gaine Mitochondriale ; CR, Corps Résiduel ; PC, Pièce Connectrice.

(adapté de Barth et Oko, 1989)

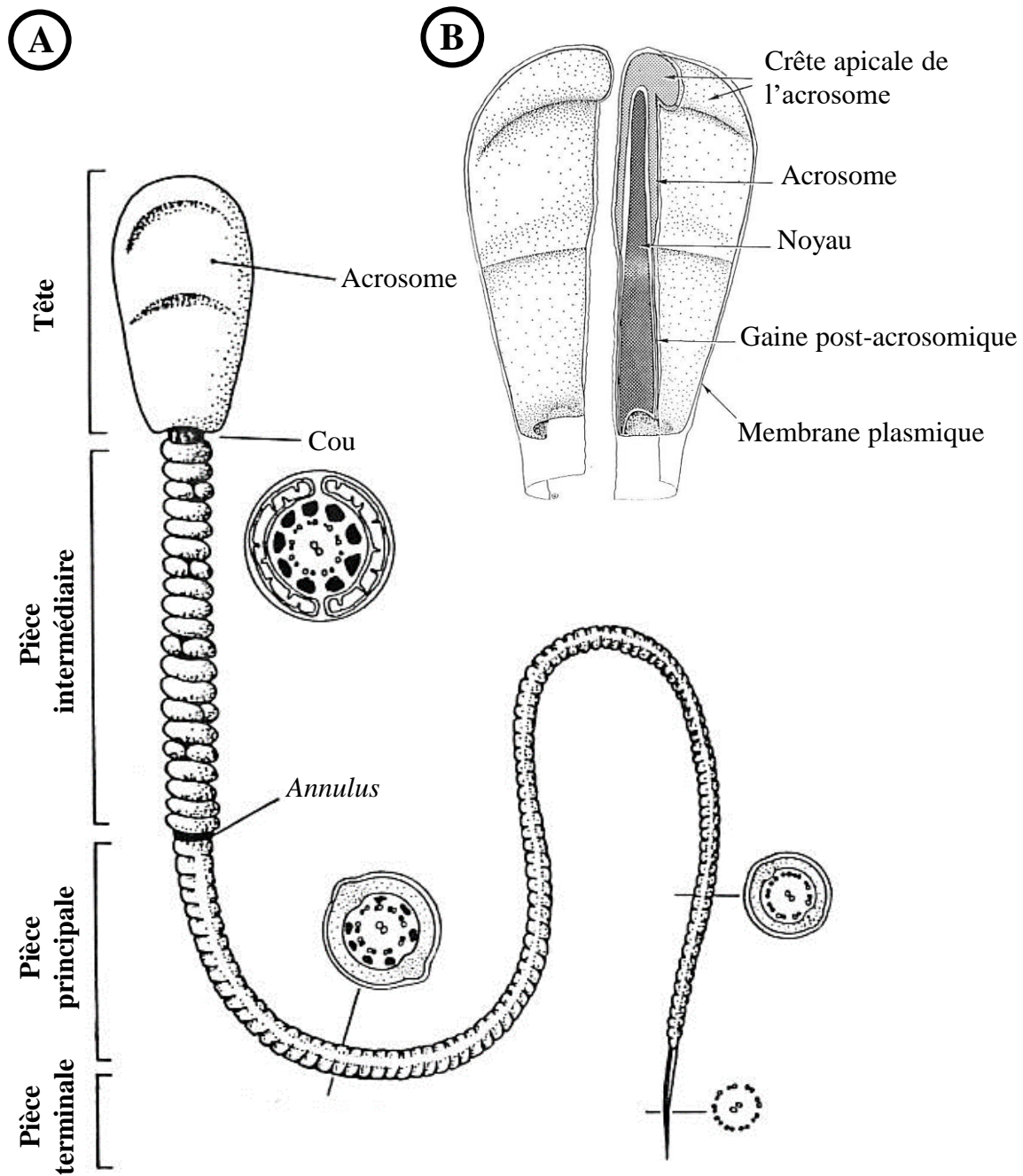


Figure 10 Spermatozoïde bovin

A. Représentation de la structure d'un spermatozoïde bovin. Les sections transversales de la pièce intermédiaire, de la pièce principale et la pièce terminale montrent la structure de l'axonème. Il s'agit de la partie motrice du flagelle, constituée d'un cylindre de 9 doublets de microtubules associés entre eux par des bras de dynéine et des liens de nexine. L'axonème est entouré de la gaine mitochondriale dans la pièce intermédiaire, de la gaine fibreuse dans la pièce principale. Il est libre dans la pièce terminale. **B.** Représentation détaillée des structures de la tête d'un spermatozoïde bovin. (adapté de Barth et Oko, 1989)

1.2.3. La spermiation et la maturation des spermatozoïdes

La sortie des spermatozoïdes des testicules est appelée la spermiation. Les spermatozoïdes ont acquis à ce stade leurs caractéristiques morphologiques mais ne sont pas encore féconds. L'acquisition du pouvoir fécondant nécessite des étapes de maturation dans l'épididyme (Figure 10). Chez le taureau le transit à travers l'épididyme dure entre 8 à 11 jours (Orgebin 1961). L'épididyme est subdivisé en 3 segments (tête, corps et queue). Sa lumière contient le fluide épидидymaire dont la composition évolue en fonction du segment (Cornwall 2008). Au cours du transit dans l'épididyme, les spermatozoïdes vont progressivement acquérir leur motilité. Leur volume sera encore réduit par la condensation finale du noyau et la perte de la gouttelette cytoplasmique. Les spermatozoïdes sont concentrés par la réabsorption d'une partie du fluide d'origine testiculaire. La composition du liquide séminal sera aussi modifiée par la dégradation de certaines protéines solubles et la sécrétion d'autres protéines par l'épithélium épидидymaire. Ces protéines assurent la protection des spermatozoïdes et un apport énergétique nécessaire à leur survie au cours du transit dans les voies femelles (actions antioxydante, antimicrobienne, etc.). Les phospholipides de la membrane plasmique des spermatozoïdes sont modifiés, et des protéines membranaires sont dégradées, chimiquement modifiées ou changent de localisation membranaire. Ces protéines participeront notamment aux cascades de signalisation enclenchées à la fécondation. Avant l'éjaculation, les spermatozoïdes matures sont conservés jusqu'à plusieurs semaines dans la partie distale de l'épididyme.

La motilité des spermatozoïdes est encore accrue à l'éjaculation, au contact du fluide riche en HCO_3^- sécrété par les vésicules séminales (Okamura et al. 1985) (Hereng et al. 2014). Le liquide séminal représente alors jusqu'à 80% du volume final de l'éjaculat. Il contient du fructose et des protéines protégeant les spermatozoïdes lors du transit dans le tractus génital femelle (spermadhésine pour l'immunosuppression, séménogéline pour la coagulation, facteurs décapacitants, etc.). Le liquide séminal est complété par les sécrétions prostatiques, sources principales de cholestérol pour les spermatozoïdes. L'insertion de cholestérol dans la bicouche phospholipidique assure la stabilité de la membrane plasmique des spermatozoïdes. Les spermatozoïdes sont finalement évacués par l'urètre lors de l'éjaculation. Un schéma représentatif de la structure d'un spermatozoïde mature est donné Figure 11.A.

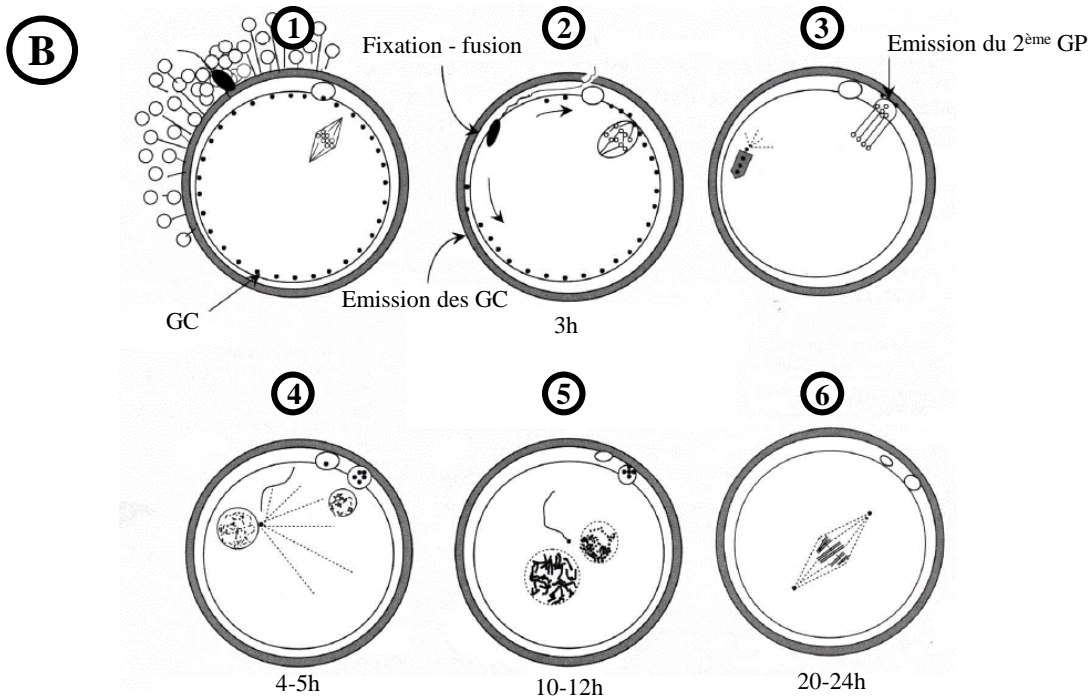
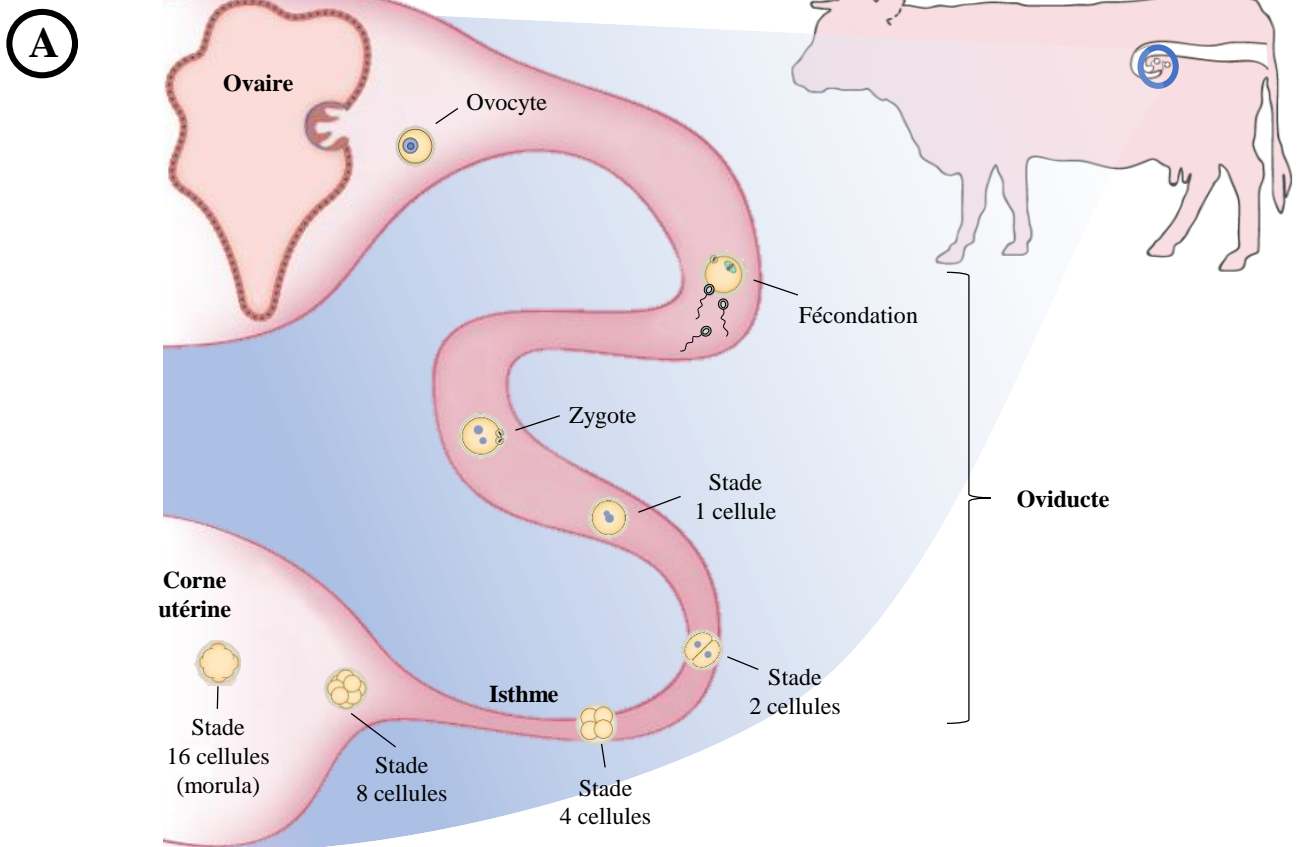


Figure 11. La fécondation

A. Rencontre entre le gamète mâle et le gamète femelle. Elle se déroule dans l’oviducte. L’embryon atteint la corne utérine au stade morula. **B.** Événements cellulaires lors de la fécondation. **1.** Arrivée du spermatozoïde fécondant. **2.** Début de la pénétration et émission des granules corticaux (GC). Anaphase de seconde division méiotique pour le noyau ovocytaire. **3.** Décondensation du noyau du spermatozoïde, émission du second globule polaire (GP). **4.** Développement du spermaster. **5.** Les *pronuclei* paternel et maternel occupent une position subcentrale ; réplication de l’ADN. **6.** Première division de segmentation. Les chromosomes homologues sur la plaque métaphasique sont décondensés. Sous chaque figure est indiqué le temps écoulé depuis le début de la fécondation, pour l’espèce bovine. (adapté de Clift et Schuh 2013; Goudet, 2014)

1.3. La rencontre avec le gamète femelle et la génération d'un descendant

1.3.1. La fécondation

A l'accouplement, la semence est déposée dans le tractus génital femelle. Les spermatozoïdes vont alors tenter de traverser l'utérus, et s'établir en réservoir à la jonction utéro-tubaire. Au moment de l'ovulation, des sécrétions endocriniennes femelles vont enclencher une modification du métabolisme spermatique, provoquant une augmentation de la concentration en espèces réactives de l'oxygène (*Reactive Oxygen Species*, ROS) et de l'adénosine monophosphate cyclique intracellulaire (AMPc). Le phénomène de capacitation définit l'ensemble des modifications des spermatozoïdes qui les rendent compétents à pénétrer la zone pellucide (ZP) puis l'ovocyte. Elle consiste en la perte du cholestérol membranaire, la phosphorylation de protéines flagellaires comme des canaux ioniques (notamment par la voie *Extracellular signal-Regulated Kinases*, ERK), et l'externalisation de récepteurs protéiques responsables de la fusion avec la membrane de l'ovocyte (comme IZUMO). En réponse aux signaux de capacitation, le relargage des réservoirs calciques du spermatozoïde induit aussi son hyperactivation. L'hyperactivation se traduit par une augmentation de l'amplitude et de la fréquence des battements flagellaires, se traduisant par l'acquisition d'une mobilité non-progressive.

La fécondation représente l'étape de fusion entre un spermatozoïde et l'ovocyte, et a lieu dans l'oviducte dans lequel l'ovocyte a migré après l'ovulation (Figure 11.B). Le processus de fécondation est enclenché lorsque le spermatozoïde entre en contact avec le complexe cumulo-ovocytaire. Les hyaluronidases membranaires du spermatozoïde dégradent la matrice du cumulus et permettent sa traversée. Le spermatozoïde peut alors se fixer à la zone pellucide grâce aux glycoprotéines membranaires de la famille des ZP, ce qui déclenche la réaction acrosomique. Le contenu en enzyme de l'acrosome est alors libéré, ce qui permet au spermatozoïde de digérer localement la ZP et de la traverser. Une fois dans l'espace périvitellin, le spermatozoïde peut alors entrer en contact avec la membrane ovocytaire et y adhérer. Cette adhésion est rendue possible par la présence des protéines IZUMO1, ADAM15, ou encore des cadhérines (Evans 2012). Les membranes cellulaires spermatiques et ovocytaires fusionnent, et le noyau contenu dans la tête spermatique est internalisé dans le cytoplasme ovocytaire. L'entrée du spermatozoïde dans l'ovocyte déclenche un blocage membranaire ainsi qu'au niveau de la ZP pour empêcher la polyspermie.

1.3.2. Le développement embryonnaire

Le développement préimplantatoire débute après la fécondation (qu'on considérera ici comme le jour 0). Suite à l'entrée de la tête du spermatozoïde dans l'ovocyte, la phospholipase C zeta (PLCZ) d'origine spermatique induit le relargage d'ions Ca^{2+} par le réticulum endoplasmique ovocytaire (Ito et Kashiwazaki 2012). En réaction, l'ovocyte achève sa seconde division méiotique et expulse le second globule polaire. Une enveloppe nucléaire se forme autour de la chromatine ovocytaire pour former le pronoyau maternel. De la même façon, une enveloppe se forme autour du génome paternel après qu'il ait perdu son enveloppe d'origine spermatique, pour former le pronoyau mâle. Les pronoyaux migrent au centre de l'œuf, puis perdent leur enveloppe. Une fois la chromatine condensée sous forme de chromosomes, ces derniers s'alignent sur le fuseau de division mitotique. Ce n'est qu'au stade 2 cellules que les génomes paternel et maternel se retrouvent au sein d'un même noyau, processus de fusion appelée la syngamie.

S'en suit une période d'intenses clivages cellulaires, sous contrôle des transcrits et des protéines cytoplasmiques ovocytaires car la transcription est encore inactive. Une activation du génome embryonnaire (*Embryonic Genome Activation*, EGA) a lieu entre les stades 8 et 16 cellules chez le bovin (Frei, Schultz, et Church 1989). Le développement embryonnaire passe à ce stade sous le contrôle transcriptionnel de son propre génome. L'embryon migre progressivement le long de l'oviducte maternel pour accéder à la corne utérine entre le 4^{ème} et 5^{ème} jour. A ce stade, l'embryon possède entre 16 et 32 cellules et est appelé morula. Les premières jonctions serrées apparaissent entre les cellules et induisent une compaction de l'embryon.

Au 6^{ème} jour post-fécondation s'en suit la formation du blastocyste. La blastulation consiste en la formation d'une cavité au sein du blastocyste, appelé le blastocœle (Holm et al. 1998). Entre le 7^{ème} et le 8^{ème} jour, l'embryon possède une centaine de cellules. Le blastocyste augmente en diamètre et le premier lignage cellulaire se forme. Cela conduit à la formation de la masse cellulaire interne (*Inner Cell Mass*, ICM) et du trophoctoderme, respectivement à l'origine de l'embryon et des annexes extra-embryonnaires (Berg et al. 2011). La spécification des lignées est en partie contrôlée par la voie de signalisation Hippo. Dans les cellules qui donneront l'ICM, le facteur de transcription *Yes Association Protein* (YAP) est phosphorylé par des protéines de la voie Hippo et est séquestré dans le cytoplasme. Dans les cellules du futur trophoctoderme, YAP n'est pas phosphorylé et migre dans le noyau. Il peut alors activer l'expression de gènes cibles dont *Cdx2*. Celui-ci réprime l'expression du facteur

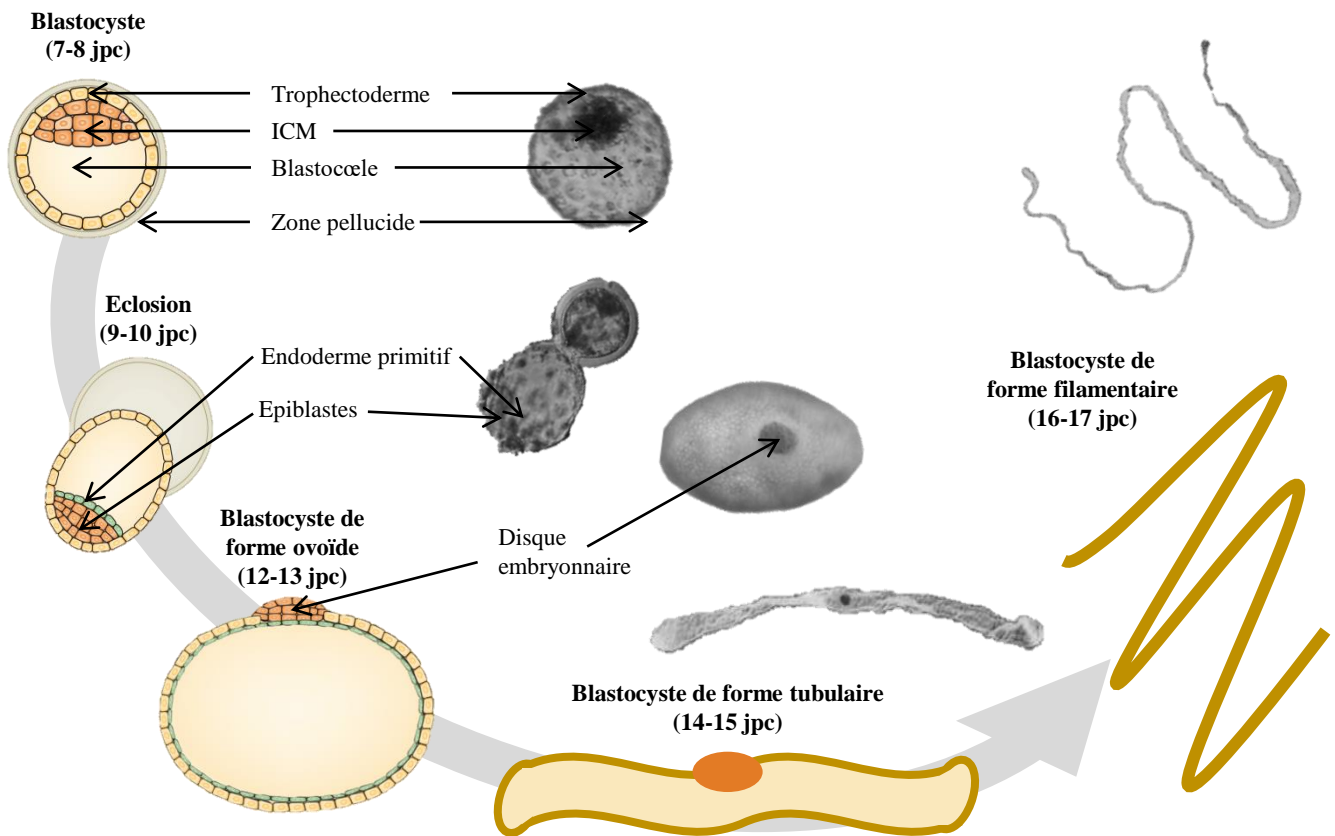


Figure 12. Eclosion et élongation du blastocyste chez le bovin

Chez les ruminants, une fois éclos le blastocyste est libre dans la corne utérine et passe progressivement d'une forme ovoïde à une forme tubulaire, puis filamenteire. Le conceptus reste libre dans les voies génitales femelles. L'implantation n'a lieu qu'à partir du 21^{ème} jour.

jpc, jour-post-coïtum

(adapté de Degrelle *et al*, 2005 ; Clift and Schuh, 2013)

de pluripotence *Oct4*, ce qui dirige les cellules vers la spécification en la lignée trophoblastique (Sasaki 2017).

Entre le 9^{ème} et le 10^{ème} jour, le blastocyste éclot de la ZP. Le processus de formation des lignages se poursuit. L'ICM se différencie en épiblaste et en hypoblaste (ou endoderme primitif). L'épiblaste donne naissance au disque embryonnaire qui participera à la formation de tous les tissus embryonnaires, et participe aussi à la formation de certains tissus extra-embryonnaires. L'hypoblaste se dépose quant à lui en une couche de cellules à la face interne du trophoblaste. Pendant la gastrulation, le lignage continue pour donner naissance aux trois feuilletts primordiaux : le mésoderme, l'endoderme et l'ectoderme.

Chez les ruminants, le *conceptus* entre en phase d'élongation (Figure 12). Il est alors libre dans la corne utérine et passe progressivement d'une forme ovoïde (12^{ème}-13^{ème} jour), à une forme tubulaire (14^{ème}-15^{ème} jour), et une forme filamenteuse (16^{ème}-17^{ème} jour) (Degrelle et al. 2005). A cette étape, l'embryon est déjà 1000 fois plus grand qu'au départ. Entre le passage de la forme ovoïdale à la forme tubulaire, plus de 500 gènes sont différentiellement exprimés dans le *conceptus* (Ushizawa et al. 2004) (Mamo et al. 2011) (Hue, Degrelle, et Turenne 2012). La composition en protéines cytoplasmiques est alors profondément modifiée. Il existe aussi un dialogue entre le *conceptus* et l'endomètre utérin, nécessaire à la stimulation des sécrétions endométriales et à la reconnaissance du *conceptus* par l'utérus. Au 21^{ème} jour, le *conceptus* occupe la totalité de la corne utérine.

Au stade où le *conceptus* est en neurulation débute l'implantation, c'est-à-dire la mise en place des tous premiers contacts cellulaires entre le trophoblaste et l'endomètre de la paroi utérine (Wooding 1992) (Yamada et al. 2002). L'implantation reste superficielle chez le bovin, contrairement à l'espèce humaine chez laquelle les trophoblastes migrent profondément dans l'endomètre maternel (Chavatte-Palmer et Tarrade 2016). Le développement embryonnaire passe progressivement sous la dépendance des hormones maternelles et notamment de la progestérone, alors sécrétée par le corps jaune. Le processus implantatoire est définitivement terminé au 25^{ème} jour. La placentation, c'est-à-dire la mise en place des structures placentaires permettant les échanges entre le *conceptus* et l'endomètre utérin, est achevée entre le 50 et le 60^{ème} jpc.

Au 45^{ème} jour, l'organogenèse est terminée et c'est le début de la vie fœtale. Les gonades sont formées, et la croissance des organes se poursuit jusqu'à la naissance de l'individu. Le vêlage pour les races bovines élevées en France intervient en moyenne entre 277 à 295 jours après la fécondation.

2. Aspects biotechnologiques et économiques : production et diffusion de la semence bovine

A l'issue de la seconde guerre mondiale, l'exode rural et la croissance démographique ont conduit à une modernisation et une intensification de l'élevage, de façon à garantir l'autonomie alimentaire et la sécurité sanitaire. Il a fallu renouveler un cheptel peu productif (1500kg de lait par an, contre 4000kg aux Etats-Unis) et décimé (92 000 vaches en 1938 contre 45 000 en 1945). La sélection a été rationalisée, profitant de l'avènement des premiers ordinateurs et de l'adoption rapide d'un nouveau mode de reproduction : l'Insémination Animale (IA), en particulier en élevage laitier (*De la paillette à l'ère du génome, 70 ans d'aventure humaine*, Alice, 2017). Aujourd'hui, près de 80% des vaches laitières sont inséminées. En élevage allaitant, la « monte naturelle » reste importante (85%), pour des raisons pratiques essentiellement (difficulté de détection des chaleurs et intervention plus complexe sur des vaches en pâture). Une telle révolution des pratiques suppose une organisation collective performante, mise en place depuis la loi sur l'élevage de 1966.

2.1. L'organisation de la filière

2.1.1. Trois siècles de développement de l'IA

Les premières utilisations de l'IA remonteraient au XIV^{ème} siècle. La première source mentionne Henri IV roi de Castille, dit « l'impotent » (1425-1474), qui aurait le premier recouru à l'IA avec sa femme la reine Juana. Une seconde relate l'utilisation de l'IA par un chef de tribu d'Arabie, qui aurait volé la semence de l'étalon de son rival pour inséminer ses juments. La première IA attestée fut celle pratiquée par le scientifique et prêtre italien Lazzaro Spallanzani en 1784, qui la réalisa chez une chienne qui donna 3 petits. Il faudra cependant attendre le XX^{ème} siècle pour en voir les premières applications. Ivanov développa cette pratique pour diverses espèces d'intérêt agricole, développa des diluants, et entraîna des techniciens à réaliser des IA en utilisant de la semence d'étalons. En URSS, et faisant suite aux travaux d'Ivanov, Milovanov développa les premiers programmes d'IA chez les bovins et ovins, ainsi que les premiers vagins artificiels pour la collecte de la semence. A la même période, les premières coopératives d'IA furent organisées au Danemark par Sørensen et Gylling-Holm. Il fallut attendre 1946 pour en voir apparaître la première en France, suivie par de nombreuses autres sous l'impulsion de Martial Laplaud et Robert Cassou.

2.1.2. Les avantages apportés par l'IA

L'IA permet de fractionner un éjaculat, qui contient des milliards de spermatozoïdes, en un grand nombre de paillettes et ainsi d'optimiser la diffusion du potentiel génétique des reproducteurs. En effet, chez le bovin, un reproducteur peut en théorie féconder des milliers de femelles par IA au cours de sa carrière, contre entre 100 à 150 par monte naturelle. Par l'IA, l'éleveur a accès à un vaste choix de taureaux d'élite de très haute performance. Il peut donc faire varier le reproducteur utilisé pour son cheptel, et ainsi limiter la consanguinité. Il existe cependant un risque de réduction du potentiel génétique des cheptels si les accouplements ne sont pas conduits de manière raisonnée. Par exemple, dans le système des « taureaux *stars* », taureaux de très haute valeur génétique, le risque était une utilisation intensive de la semence. Il est ainsi bien établi qu'une conséquence de cette pratique est un rythme particulièrement soutenu d'accroissement de la consanguinité au sein de races laitières ayant pourtant de très larges effectifs de femelles (Danchin-Burge et al. 2012).(Yue, Dechow, et Liu 2015).

En plus de réduire les coûts d'entretien en supprimant les taureaux, l'amélioration génétique du cheptel permet un gain financier pour les exploitations. A chaque génération, l'utilisation de taureaux de haute valeur génétique engendre une amélioration de la production globale du cheptel. Des études ont montré que l'usage de la sélection et de l'IA sur un cheptel de 60 vaches laitières permet une augmentation du bénéfice de 3610 euros par an (étude Osiris, Ministère de l'agriculture).

Avec le recours à des gaines d'insémination à usage unique, la quarantaine imposée pour les paillettes en attendant le contrôle sanitaire opéré par le Laboratoire National de Contrôle des Reproducteurs (LNCR), et l'absence de contact entre les animaux, le risque de transmission d'éléments pathogènes a chuté drastiquement. C'est d'ailleurs grâce à cela qu'une dizaine de maladies a quasi disparues depuis les années 1950, comme la gourme.

L'IA a transformé le travail de l'éleveur. Elle l'a affranchi des contraintes traditionnelles de la reproduction. Néanmoins, l'IA requiert une meilleure observation du troupeau pour la détection des chaleurs et la planification de la venue de l'inséminateur. Le fractionnement de l'éjaculat en paillettes et leur congélation permet aussi de faciliter le transport et de découpler la production de semence de son utilisation. La traçabilité des croisements est améliorée, avec un système de code-barre permettant de gérer de manière informatisée les réserves et les transports de paillettes.

Associée à la technologie de la cryoconservation, l'IA permet de conserver le patrimoine génétique de races qui ne sont plus exploitées ou des individus de haute valeur génétique (Locatelli 2012). Les races Bretonne, Flamande ou Vosgienne ont par exemple été préservées par ce biais (Lauvie et Couix 2012). Ces banques peuvent aussi servir à limiter la consanguinité, en réintroduisant de manière contrôlée d'anciens génotypes dans la population. Des réserves de semence peuvent être conservées dans des banques de sperme dans le cadre de la Cryobanque Nationale en France. L'IA permet aussi l'utilisation de la semence sexée, c'est-à-dire de la semence dont les spermatozoïdes porteurs d'un chromosome X et les spermatozoïdes porteurs d'un chromosome Y ont été séparés. Ainsi, un éleveur de race allaitante souhaitera donner naissance à un maximum de mâles, alors qu'un éleveur de bovins laitiers souhaitera obtenir un maximum de femelles. L'efficacité de l'IA en semence sexée est en revanche moindre qu'en semence conventionnelle (DeJarnette et al. 2010).

2.1.3. La situation de la filière en France

L'organisation de la filière de la reproduction bovine, articulée autour de la sélection des reproducteurs et de l'IA, a permis à la France d'avoir une place prépondérante dans la génétique mondiale. Cela a favorisé le développement de son élevage tout en limitant la compétition entre les entreprises, et en préservant les petites races.

L'INRA, au travers de son département de génétique animale, a un rôle moteur dans les recherches relatives aux méthodes d'évaluation génétique et à l'optimisation de la sélection. La fédération des centres d'insémination, Alice, conduit également des projets de recherche et de développement, notamment en partenariat avec l'INRA ou l'Institut de l'Élevage (IDELE). Les projets de recherche sont financés par l'Agence Nationale de la Recherche (ANR), ainsi que par organismes de financement comme Apis-Gene, mais aussi directement par les fédérations.

Les entreprises de sélection (ES) sont les sociétés qui élèvent les reproducteurs dans les stations d'élevage et produisent les paillettes de semence. En France, ce sont ensuite les Entreprises de Mise en Place (EMP) qui organisent et pratiquent l'IA directement chez les éleveurs. En 2017, 75 EMP bovines sont déclarées auprès de l'IDELE. Depuis 2007 a été instauré l'Insémination Par l'Éleveur (IPE) pour permettre aux éleveurs de pratiquer directement l'IA sur leurs reproductrices sans passer par les EMP. Cette pratique reste encore marginale (5,1% en 2013) bien qu'en constante progression depuis plusieurs années. Les performances enregistrées par les animaux sont ensuite envoyées vers le Système National d'Information Génétique (SNIG) pour assurer l'amélioration génétique des populations.

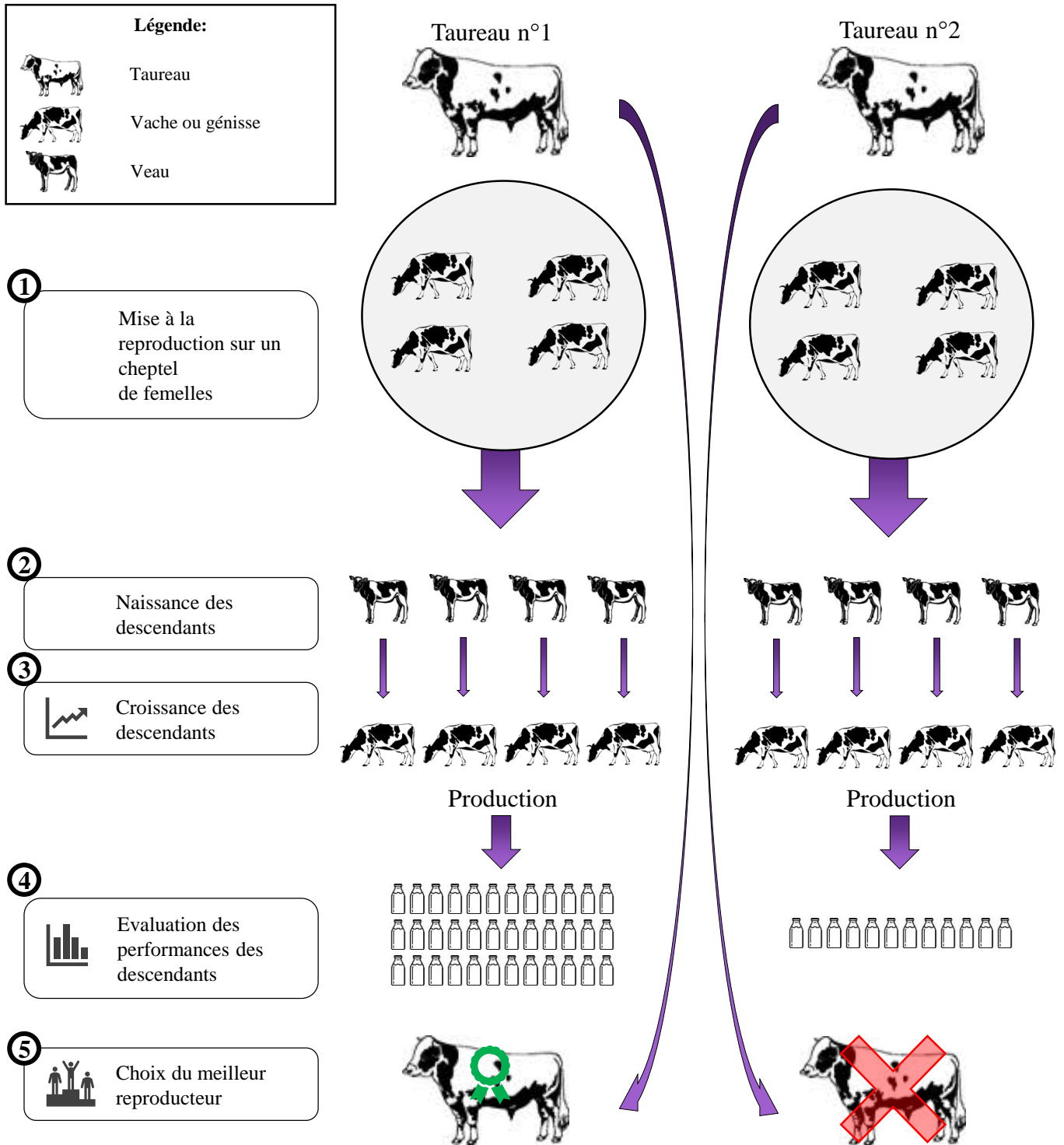


Figure 13. Principe du testage sur descendance

Les taureaux n°1 et n°2 sont évalués. Les 2 individus sont mis à la reproduction sur un cheptel de femelles (1). Il faut alors attendre la naissance des descendants (2), leur croissance (3) pour pouvoir évaluer leurs performances (4, ici représentées par la production laitière). Il est alors possible de savoir quel taureau à donné naissance aux descendants qui sont globalement les plus performants. Ce taureau sera sélectionné et rejoindra la cohorte des nouveaux reproducteurs (5).

2.2. La sélection des reproducteurs

2.2.1. Le testage sur descendance

Avant l'emploi systématique du testage sur descendance, la sélection des reproducteurs était basée sur leur ascendance, c'est-à-dire sur la base des performances enregistrées de leurs parents et grands-parents. Tous les taureaux étaient enregistrés dans un Livre Généalogique (aussi appelé *Herd Book*), qui retraçait leur pedigree et leurs performances en production.

L'idée du testage sur descendance est en fait très ancienne. Un agronome romain de la région de Pompéi, Varron, écrivait déjà en 37 avant J.C: « Surtout, il faut veiller à avoir du bétail bon reproducteur. Il y a en général 2 moyens d'en juger : l'aspect extérieur et la descendance » (Marcus Terentius Varro, *De re rustica*). C'est grâce au développement de l'IA que cette idée a pu être efficacement mise en application. Les premières expériences de testage sur descendance seront enclenchées en France dès 1951 sous l'impulsion de Marc Chevaldonné (Denis Lefevre, *Eleveurs, passion solidaire*), et finiront par être rendues obligatoires après la circulaire du 31 août 1964. Jusqu'à présent, elles sont effectuées par les ES.

Le principe est le suivant (Figure 13). A la puberté, les candidats reproducteurs sont utilisés pour produire une réserve de doses de semence qui sera utilisée en testage (de 600 à 900 doses en fonction de la race et de l'ES). Des IA sont réalisées à partir de ces doses sur un échantillon d'entre 100 à 200 femelles. Plus le nombre de femelles inséminées est important, plus fiables sont les résultats du testage, car cela permet d'écarter l'impact de la fertilité femelle. Ces femelles sont de plus sélectionnées parmi différents élevages répartis dans diverses régions pour limiter les potentiels biais environnementaux. Ces IA permettent de donner naissance à au moins une centaine de descendants pour chacun des taureaux candidats. On peut alors évaluer de manière empirique la valeur génétique du père au travers des performances mesurées chez ses descendants. Seront mesurées, entre autres, leur vitesse de croissance, leur morphologie, et une fois atteint l'âge pour entrer en production, leurs performances de production de lait pour les femelles. Les reproducteurs ayant donné naissance aux descendants qui présentent de manière globale les meilleures performances pour un caractère donné sont alors sélectionnés pour former la cohorte des futurs reproducteurs. Pour obtenir des informations sur la qualité du taureau candidat, il faut donc avoir un retour sur les premiers résultats de performance de ses descendants. Dans le cadre de l'industrie laitière par exemple, cela nécessite d'attendre la fin de la gestation, l'élevage des génisses, leur première lactation, la collecte rigoureuse de tous les résultats de production

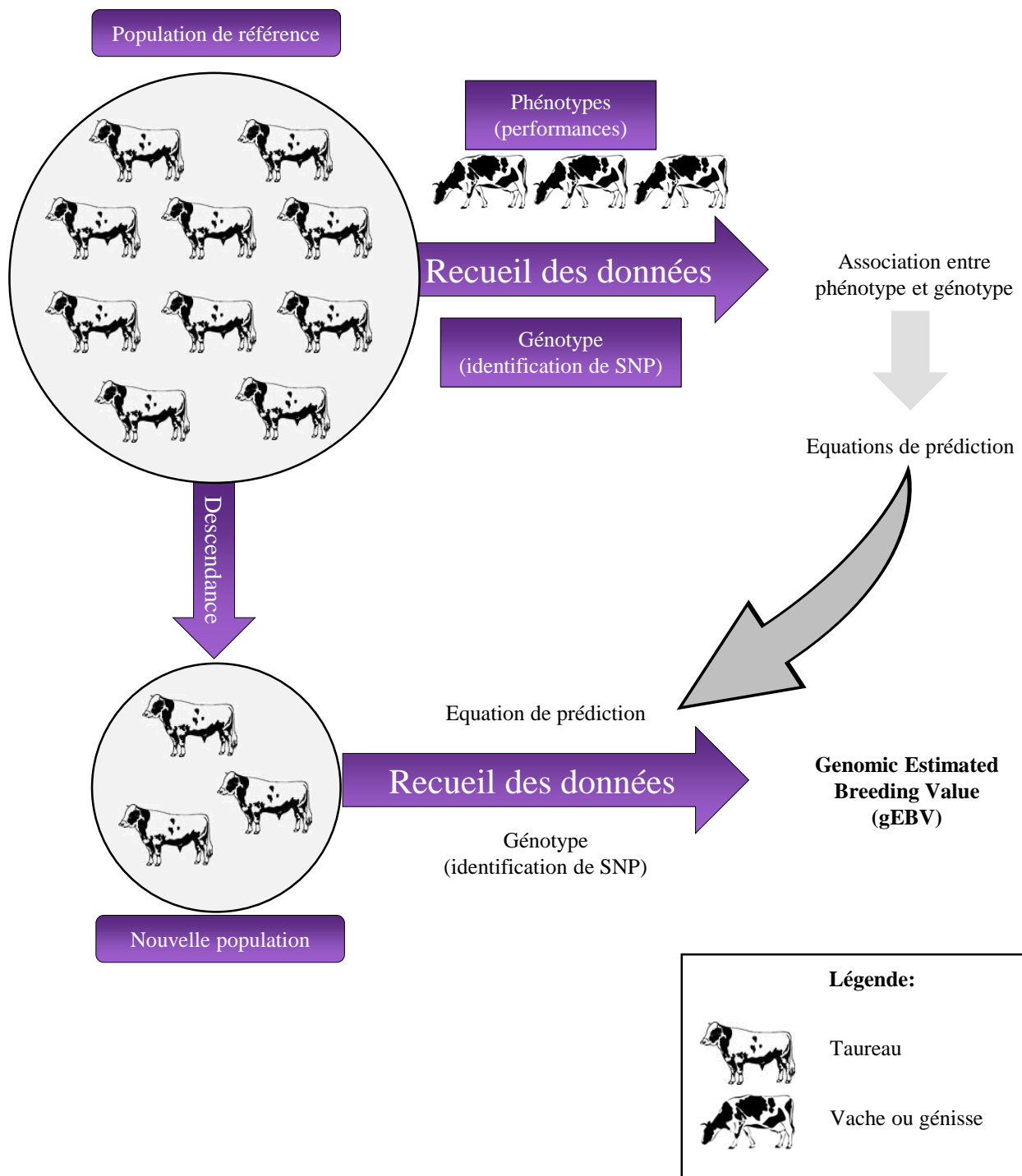


Figure 14. Principe de la sélection génomique

Sur une population de référence (composée de centaines de milliers d'individus) et pour chaque individu, sont collectées les données phénotypiques (performances en production par exemple), ainsi que les données de génotypes (grâce à un génotypage systématique). L'étape suivante est l'association entre le génotype et le phénotype via les marqueurs SNP. Les effets de chaque marqueur sont estimés et combinés pour produire les équations de prédiction. On applique les équations de prédiction à une nouvelle cohorte d'individus ne faisant pas partie de la population de référence, mais dont on dispose de l'information génétique. L'information qui est produite est la *genomic estimated breeding values* ou gEBV (adapté de Oldenbroek et Van der Waaij, 2015)

laitière et leur compilation. Cette période peut durer jusqu'à 3 ans. Pendant ce laps de temps, les candidats sont gardés en station d'élevage, où ils sont employés à produire une réserve de dose de semence en attendant les résultats du testage (de 2000 à 5000 doses selon les ES). L'ensemble des données de production des filles sert au calcul d'un index de valeur génétique du candidat. En moyenne, 10% des taureaux sont gardés à l'issue de l'étape de testage sur descendance.

De par son principe, le testage sur descendance permet également d'avoir un retour sur l'efficacité reproductive des candidats. Les taureaux infertiles ou sub-fertiles sont très rapidement identifiés, et peuvent être retirés du processus de sélection s'ils ne permettent pas d'obtenir un nombre suffisant de vêlages.

2.2.2. La révolution génomique

Depuis 2009, la filière bovine a connu un changement de paradigme pour l'évaluation de la valeur génétique des reproducteurs. Grâce à l'implication des chercheurs et ingénieurs de l'INRA, d'Allice et de l'IDELE, la France a été un pays moteur pour le développement de ce que l'on appelle la sélection génomique. La sélection génomique repose sur l'analyse génétique à haut-débit des candidats reproducteurs sur une palette de marqueurs SNP (pour *Single Nucleotide Polymorphism* ou polymorphisme de séquence, c'est-à-dire la présence d'une base nucléotidique différente en un *locus* donné entre 2 individus ou groupes d'individus)(Meuwissen, Hayes, et Goddard 2001). Les caractères phénotypiques d'intérêt pour les élevages sont associés à un ou plusieurs SNP : des caractères fonctionnels (santé, tempérament, fertilité, etc.), de production (quantité et composition du lait, etc.) ou de morphologie (taille et conformation de la mamelle, etc.). Si un individu possède à l'état homozygote un des SNP associé à un caractère donné, il le transmet à tous ses descendants (à la moitié de ses descendants si le SNP est présent à l'état hétérozygote). Les descendants ont ainsi plus de chances d'exprimer le caractère recherché.

L'association SNP-phénotype est établie sur la base d'une population dite de référence (Figure 14). La population de référence est composée de plusieurs milliers d'individus pour lesquels on possède les données génomiques, ainsi que les données de performance des descendants établies par testage. La caractérisation génomique de cette population permet de corréler des SNP à des caractères phénotypiques de leurs descendants. L'effet (positif ou négatif) de chaque SNP corrélié à ces caractères peut alors être estimé. La combinaison des SNP du génome d'un individu est révélatrice de sa valeur génétique globale.

Les données de génotypage des nouveaux candidats reproducteurs peuvent être confrontées aux indices établis grâce à la population de référence. On peut alors estimer indirectement la valeur génétique du candidat, sans passer par la phase de testage et donc sans retour sur les performances de ses descendants. L'évaluation génomique des candidats peut être effectuée seulement quelques semaines après leur naissance. L'ADN des candidats est extrait à partir d'une biopsie du cartilage auriculaire ou d'un prélèvement de poils à la base de la queue, puis envoyé à une plateforme de génotypage. En France, les analyses de génotypage sont conduites par Labogena. Les résultats de génotypage sont ensuite traités et enregistrés auprès du Centre de Traitement de l'Information Génétique (CTIG). Enfin, la société Valogène gère la diffusion des informations entre les différents organismes (ES et éleveurs).

En 2014, près de 30 millions de marqueurs génétiques d'intérêt avaient déjà été identifiés chez le bovin (Daetwyler et al. 2014). Seuls ceux présentant les effets les plus importants sur les caractères ont été conservés pour les analyses en routine (Hayes et al 2014) (Wiggans et al. 2016) (Brøndum et al. 2015). Ces dernières sont effectuées en utilisant une puce de génotypage permettant d'interroger 54 000 SNP (BovineSNP50 v2 BeadChip, Illumina). Les reproducteurs sélectionnés par la génomique ont ensuite une descendance. Les performances mesurées sur les filles viennent alors enrichir l'évaluation génétique des taureaux, ce qui permet entre autres d'avoir un retour sur l'efficacité de la sélection génomique et d'agrandir la population de référence. Au fil des générations, et par l'accroissement de la population de référence, de nouveaux caractères (notamment ceux peu héréditaires) sont ajoutés au calcul de la valeur génétique des individus.

La sélection génomique a été rendue disponible dès 2009 pour les races Holstein, Montbéliarde et Normande. Elle a depuis été étendue à d'autres races dont des races allaitantes (Limousine et Blonde d'aquitaine en 2011, Charolais en 2015, Abondance, Tarentaise, Brune, Vosgienne). La généralisation de la sélection génomique aux races allaitantes a nécessité plus de temps car les populations de référence sont moins grandes et les phénotypes moins précis et plus nombreux que pour les races laitières. En France pour l'année 2012, 60% des IA pour la race Holstein étaient réalisées avec la semence de taureaux génomiques (Hocdé et Joly 2013).

2.2.3. L'évolution des pratiques

L'utilisation de la génomique apporte un grand nombre d'avantages pour la sélection. Alors que le testage sur descendance n'est plus nécessaire, les reproducteurs peuvent être utilisés en production de semence dès leur puberté, sans avoir à attendre les retours de performances de leurs descendants. Ils sont donc utilisés beaucoup plus jeunes qu'avant. Ce gain de temps peut impliquer une réduction de l'intervalle entre les générations de plus de 30%, et par conséquent intensifier le progrès génétique (Scheffers et Weigel 2012).

Comme on connaît la valeur génétique des taureaux peu de temps après leur naissance, seuls les futurs reproducteurs de haute valeur sont conservés. L'économie financière est considérable. Avec le testage sur descendance, il était nécessaire pour les ES d'acheter les taureaux et de les entretenir jusqu'au retour sur les performances de leurs descendants, ce qui prenait 3 ans et avait un coût d'environ 50 000 euros par taureau (Le Mezec 2009). Cette économie financière, couplée à une palette plus large de caractères d'intérêt évaluables, rend possible et intéressante la sélection de plus de reproducteurs. La consanguinité des élevages est alors réduite. Par exemple, multiplier le nombre de reproducteurs par 3 réduit de 23% le taux de consanguinité (Colleau, Fritz et Guillaume 2009). La consanguinité est d'autant plus réduite que la connaissance des SNP associés à des anomalies génétiques permet d'utiliser quand même les taureaux qui en sont porteurs si les accouplements sont faits de façon à conserver l'hétérozygotie pour le SNP (80% des anomalies génétiques bovines sont en effet récessives).

Cependant, la sélection génomique présente tout de même certains inconvénients importants par rapport au testage sur descendance. Lors du testage, les reproducteurs étaient employés en IA. En plus du retour sur leur valeur génétique, le testage fournissait alors également des informations sur l'efficacité reproductive des reproducteurs. Ceux pour lesquels on n'arrivait pas à obtenir de descendant étaient nécessairement éliminés de la sélection. Comme en sélection génomique on ne passe plus par cette phase de testage, à leur puberté les taureaux génomiques sont directement envoyés à la production, et les semences diffusées sans retour sur l'efficacité reproductive des individus.

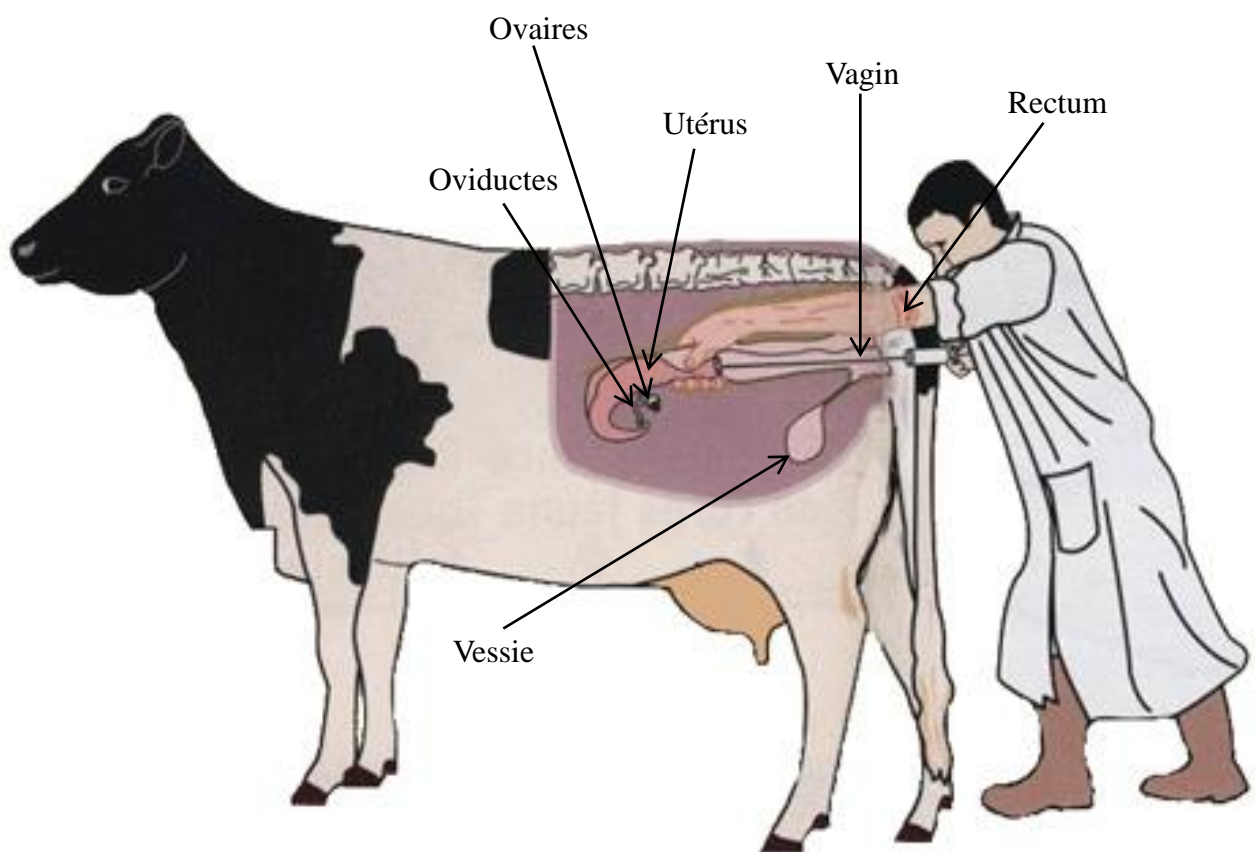


Figure 15. La mise en place de l'IA
(image provenant du site internet de l'ES Eliacoop)

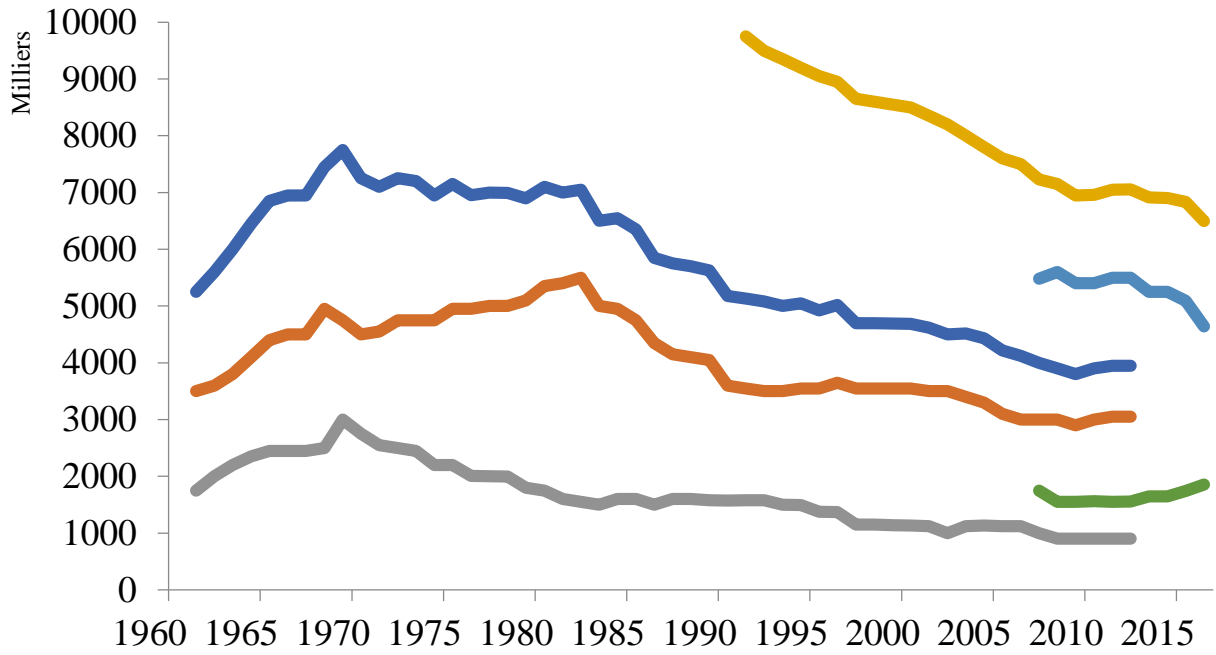
2.3. Les étapes de production de semence

2.3.1. La collecte et le conditionnement des éjaculats

Les jeunes candidats sélectionnés sont élevés dans les centres de collecte agréés des ES, après une phase initiale de quarantaine permettant de s'assurer de l'absence de maladies contagieuses. Les taureaux du centre sont soumis plusieurs fois par an à des tests sanguins pour contrôler leur statut sanitaire. La collecte de la semence est effectuée par le taurelier dans une salle de monte. Le taureau reproducteur est placé près d'un bœuf appelé « boute-en-train ». Les boute-en-train sont utilisés en remplacement des femelles, qui sont exclues des stations d'élevage pour des raisons sanitaires et de sécurité. Après 3 fausses montes sur le boute-en-train, le pénis du reproducteur est inséré dans un vagin artificiel gorgé d'eau à 37°C et muni d'un tube dans lequel est récoltée la semence après l'éjaculation. Un éjaculat bovin représente un volume moyen de 3 à 8 ml, et contient entre 4 et 8 milliards de spermatozoïdes (Hagen et Gayraud 2005). On compte en moyenne 2 collectes hebdomadaires en fonction de la race et du potentiel de l'individu (Gerard et al. 2008). L'éjaculat est immédiatement transmis au laboratoire d'analyse pour effectuer des contrôles qualité. Si l'éjaculat satisfait des conditions minimales de qualité, il peut être conditionné. Le conditionnement comprend une dilution de l'éjaculat dans des diluants de conservation, et son fractionnement en paillettes de 0.25 ml contenant chacune 20 millions de spermatozoïdes en moyenne (Vishwanath 2003). Chaque paillette est marquée d'un code-barre à 10 chiffres pour en garantir la traçabilité (avec le nom du taureau, le centre de production de semence et la date de conditionnement).

Le plus souvent dans la filière bovine, l'IA n'est pas réalisée en semence fraîche et les paillettes de semence doivent être conservées. La conservation se fait principalement par cryoconservation dans des cuves d'azote liquide à -196°C. La solution dans laquelle sont dilués les éjaculats avant leur conditionnement contient des éléments permettant de protéger les spermatozoïdes durant la descente et la remontée en température, en évitant la formation de cristaux de glace. Le lait ou le jaune d'œufs ont beaucoup été utilisés comme diluants, pour leurs propriétés isotoniques et nutritives. Des antibiotiques sont ajoutés pour protéger des éventuelles contaminations, et du glycérol pour faciliter la congélation. Un nouveau contrôle qualité des paillettes est effectué après l'étape de congélation. Pour chaque éjaculat, 2 paillettes sont prélevées au hasard et sont décongelées. Il s'agit de s'assurer qu'il reste un nombre suffisant de spermatozoïdes viables, car près de la moitié des spermatozoïdes sont morts après décongélation.

A



- Toutes IAT
- IAT taureaux - lait
- Toutes IAP
- IAP taureaux - lait
- IAT taureaux - boucherie
- IAP taureaux - boucherie

B

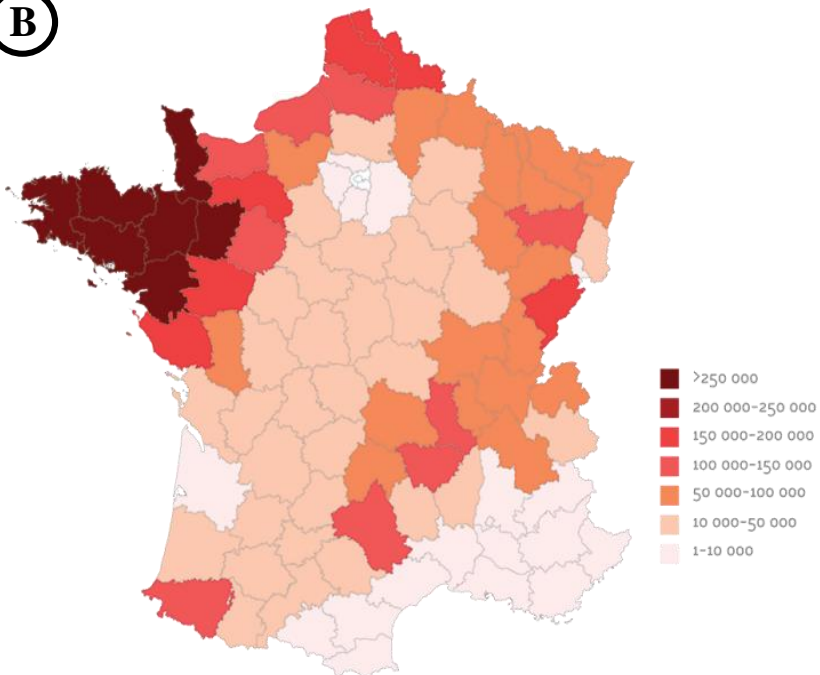


Figure 16. L'Insémination Artificielle (IA) en France

A. Evolution du nombre d'Insémination Artificielles Totales (IAT) et Premières (IAP), en France et sur la période 1960-2015.

B. IAT par département français en 2015-2016

(Source Alice)

2.3.2. La mise en place de l'IA

L'IA est réalisée après une étape de détection des chaleurs chez les femelles à inséminer, basée sur des observations comportementales faites par l'éleveur. Les élevages laitiers sont plus concernés car les éleveurs sont proches des vaches pour la traite. La procédure d'IA a été standardisée pour son utilisation à toutes les races ou élevages. La dose est sélectionnée grâce à son code-barre, est décongelée et montée sur un pistolet d'insémination. Il s'agit d'un dispositif composé d'une tige en métal et d'un piston à son extrémité. L'inséminateur introduit un bras dans le rectum de la vache pour maintenir le col de l'utérus. Avec sa main libre il insère le pistolet d'insémination dans le vagin de la femelle et le guide jusqu'au corps de l'utérus. Le contenu de la paillette est alors déposé quelques centimètres après la sortie du col utérin (Figure 15, page précédente). L'opération n'est pas douloureuse pour la vache et ne nécessite qu'une simple contention sans anesthésie.

2.3.3. Aspects économiques

Le cheptel bovin français est composé de 19,28 millions d'animaux dont 1 277 000 mâles de plus d'un an (Agreste, Ministère de l'Agriculture Français). L'effectif des cheptels est stable, et reparti dans près de 200 000 élevages (INTERBEV 2016). Les principales races en France en termes d'effectifs sont les Holstein (65%) les Montbéliard (17%) et les Normands (11%).

En 2016 en France, 7 100 607 IA ont été réalisées. Ce chiffre est cependant en quasi constante diminution depuis plusieurs décennies, suivant le déclin du cheptel français (Figure 16).

L'ensemble de la filière IA emploie en France plus de 2800 salariés. Son chiffre d'affaire représente 378 millions d'euros (source Coop de France).

Pour l'éleveur, le coût que représente l'IA comporte l'achat de la paillette (dont le prix varie en fonction de la valeur génétique du reproducteur), la mise en place de l'IA par un inséminateur, puis le diagnostic de gestation qui est réalisé par un vétérinaire. Chaque IA représente alors un coût qui peut varier de 50 à 100 euros par femelle inséminée.

Chapitre II : Les enjeux de la fertilité mâle pour les filières bovines

La fertilité peut être définie comme la capacité d'un individu à produire une descendance viable. Une bonne fertilité des reproducteurs est capitale pour un élevage durable. La naissance de veaux (et donc, une bonne fertilité des parents) est indispensable à la production laitière, assure le renouvellement des cheptels et la génération de nouveaux reproducteurs. Une diminution globale de la fertilité dans les élevages a été observée au cours des dernières décennies. Cette baisse de fertilité a des répercussions économiques sur l'ensemble des acteurs de la filière. En effet, la naissance d'animaux est directement liée à la production, que ce soit en industrie laitière (il faut une naissance pour enclencher une lactation) ou de la viande (naissance de veaux de boucherie). En conséquence, la fertilité est souvent considérée comme étant plus importante économiquement que les caractères de production (Thundathil, Dance, et Kastelic 2016). Tout comme les femelles, les mâles sont en partie responsables de la baisse de fertilité. Cependant, identifier les causes conduisant à une fertilité réduite est complexe, car elle résulte de la conjonction de multiples facteurs (environnementaux, biotechnologiques ou encore génétiques) (partie 1).

Dans la filière bovine, une première qualification du potentiel reproducteur est basée sur l'analyse standardisée de caractéristiques morphologiques et fonctionnelles de l'individu et sa semence. Les reproducteurs validant cette étape sont alors utilisés en IA. En fonction de leur résultat en IA, il est possible de caractériser leur fertilité réelle, qu'on nomme fertilité terrain. Chez les reproducteurs, elle peut varier d'une excellente fertilité à une totale stérilité, avec un grand nombre de profils intermédiaires. En particulier, les taureaux qualifiés de sub-fertiles ne sont pas totalement infertiles mais ont des difficultés à obtenir des descendants (partie 2).

Alors que les taureaux ayant un problème de spermogramme sont facilement identifiés et éliminés grâce aux contrôles qualité de routine, ce n'est pas le cas des taureaux sub-fertiles. Pour certains d'entre eux, le problème de spermogramme peut être pallié mais ces méthodes ne sont pas systématiquement efficaces. De plus, comme la fertilité a notamment une composante génétique, elle est aujourd'hui prise en considération dans le cadre de la sélection génomique des reproducteurs. De nombreuses études ont cependant montré que la fertilité est un caractère peu héritable. De ce fait, les contrôles qualité et le génotypage ne permettent pas totalement d'écarter les taureaux sub-fertiles. Dans ce contexte, il est capital de détecter précocement et efficacement ces taureaux en utilisant de nouveaux procédés (partie 3).

A

Taux de conception en IAP en France de 1999 à 2008

Année	Holstein			Normand			Montbéliard		
	Génisse	Vache	Total	Génisse	Vache	Total	Génisse	Vache	Total
1999	58,0%	42,9%	48,1%	59,5%	51,1%	54,0%	56,6%	52,3%	53,9%
2000	59,1%	42,3%	47,9%	61,0%	50,8%	54,4%	59,8%	53,7%	55,8%
2001	58,7%	41,1%	47,0%	61,3%	50,5%	54,1%	58,2%	53,4%	55,0%
2002	58,1%	39,6%	45,7%	61,2%	49,5%	53,5%	58,3%	52,8%	54,6%
2003	57,6%	38,3%	44,8%	60,6%	49,4%	53,2%	57,5%	51,9%	53,8%
2004	57,5%	37,6%	44,4%	60,8%	49,2%	53,2%	56,1%	51,9%	53,3%
2005	57,6%	37,8%	44,4%	60,1%	48,9%	52,8%	59,3%	51,9%	54,4%
2006	58,3%	37,6%	44,5%	60,3%	48,6%	52,6%	59,4%	52,3%	54,7%
2007	58,1%	37,7%	44,6%	60,3%	48,0%	52,2%	60,0%	52,9%	55,2%
2008	57,2%	37,7%	43,9%	59,6%	46,4%	50,6%	58,2%	51,6%	53,6%

Nombre d'IAP enregistrées par le SNIG en France de 1999 à 2008

Année	Holstein			Normand			Montbéliard		
	Génisse	Vache	Total	Génisse	Vache	Total	Génisse	Vache	Total
1999	532 499	1 020 987	1 553 486	85 008	157 431	242 439	90 938	159 685	250 623
2000	523 774	1 030 093	1 553 867	85 717	160 841	246 558	85 048	161 194	246 242
2001	543 328	1 073 121	1 616 449	83 328	164 844	248 172	89 125	174 669	263 794
2002	548 698	1 106 563	1 655 261	85 375	166 963	252 338	93 355	187 674	281 029
2003	545 596	1 070 452	1 616 048	83 895	162 362	246 257	96 557	187 903	284 460
2004	547 313	1 057 396	1 604 709	85 221	159 441	244 662	98 097	190 636	288 733
2005	528 589	1 045 334	1 573 923	82 186	158 495	240 681	99 712	193 649	293 361
2006	502 581	1 017 448	1 520 029	78 735	153 430	232 165	97 584	196 596	294 180
2007	506 756	1 002 631	1 509 387	79 086	152 150	231 236	101 069	202 698	303 767
2008	481 121	1 033 651	1 514 772	74 405	161 073	235 478	99 439	226 734	326 173

B

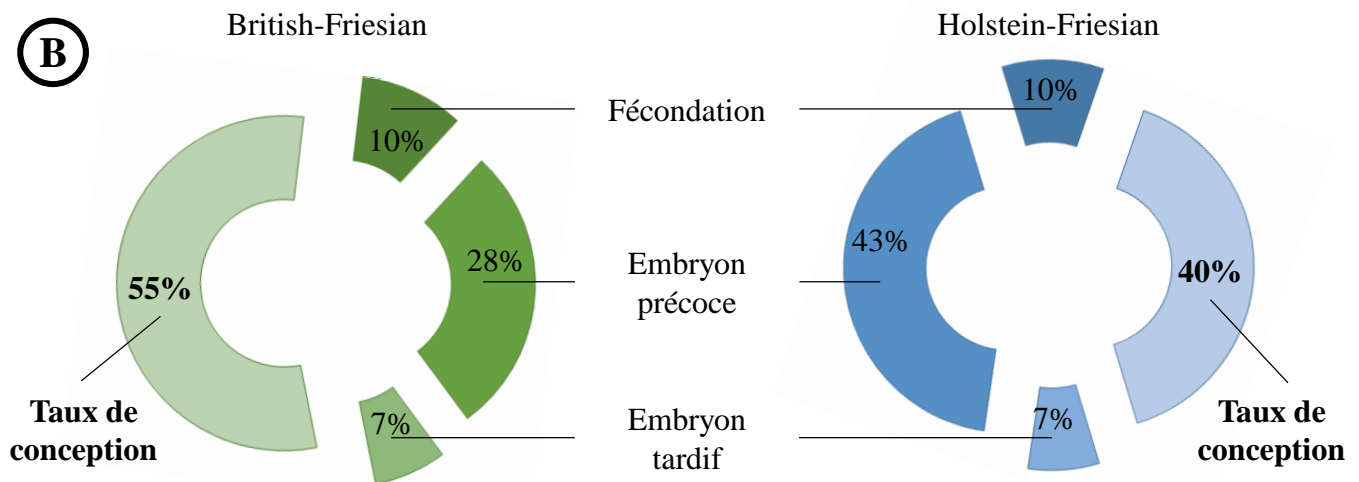


Figure 17. Le taux de conception bovin est en diminution, et est principalement lié à des pertes embryonnaires précoces

A. Evolution du taux de conception en inséminations artificielles premières (IAP, tableau 1) pour les vaches et génisses de races Holstein, Normande et Montbéliarde en France sur la période 1999-2008 (Le Mezec, 2010). Le tableau 2 correspond au nombre d'IAP correspondant d'après les données enregistrées dans le système national d'information génétique (SNIG) On observe globalement une diminution du taux de conception pour les races Holstein et Normand, en particulier pour les vaches. La diminution la plus implorante est observée chez les Holstein (sur vache comme sur génisse, diminution de 4,2% en 10 ans). La baisse est également importante chez les Normandes (sur vache comme sur génisse, diminution de 3,4% en 10 ans). Elle est en stagnation chez les Montbéliardes. Le taux de conception passe de 42,9 à 37,7% chez les Holstein et de 51,1 à 46,4% chez les Normandes. Au contraire, pour les Montbéliardes, le taux de conception sur génisse a augmenté, passant de 56,6 à 58,2%. Les tendances sont les mêmes en inséminations artificielles totales (IAT, non montré). **B.** Résultats attendus en terme de taux de conception après 100 IA sur vaches British-Friesian et Holstein-Friesian. Les pertes liées au processus de fécondation sont d'environ 10%. En fonction de la race, les pertes en période embryonnaire précoce (avant le 24^{ème} jour suite à la fécondation) varient de 28 à 43%. Ce sont les pertes les plus importantes au cours de la gestation. Les pertes en période embryonnaire tardive (entre le 24 et le 42^{ème} jour post-fécondation) sont faibles, de l'ordre de 7%. En conséquence, le taux de conception varie de 55 à 40% chez les 2 races concernées. (adapté de Diskin *et al*, 2006)

1. Causes et conséquences de la baisse de fertilité des élevages

1.1. La baisse de fertilité est délétère pour les filières bovines

1.1.1. Observations

Depuis plus de 50 ans, de nombreuses études ont souligné le fait qu'il existe une diminution généralisée de la fertilité dans les élevages. En 1998, Butler et son équipe ont compilé les données concernant le taux de conception (c'est-à-dire, le taux de naissances suite à l'IA) en Insémination Artificielle Première (IAP) de vaches Holstein dans l'état de New York aux Etats-Unis (Butler 1998). Le taux de conception est passé de 65% en 1951 à 40% en 1996. Dans la continuité de cette étude, Norman et ses collègues ont publié en 2009 des résultats provenant de 9 millions d'IAP réalisées aux Etats-Unis sur des vaches Holstein entre 1996 et 2007 (Norman et al. 2009). Là encore, le taux de conception est passé de 34% à 32%. Des données allant dans ce sens proviennent aussi d'Australie (MacMillan, Lean, et Westwood 1996), d'Irlande (Roche, Mackey, et Diskin 2000), du Royaume-Unis (Royal, Mann, et Flint 2000) ou d'Espagne (López-Gatius 2003), ce qui montre que la tendance est globale. En France, sur près de 21 millions d'IAP réalisées entre 1999 et 2008, le taux de conception est passé de 48,1 à 43,9% pour la race Holstein et de 54,0 à 50,6% pour la race Normande (Le Mezec 2010). Elle est restée stable pour la race Montbéliarde, passant de 53,9 à 53,6% sur la même période (Figure 17).

Lorsque l'IA n'aboutit pas à un vêlage, soit la fécondation a échoué, soit c'est le développement embryonnaire ou fœtal. Des recherches ont été réalisées pour tenter d'identifier à quels stades se produisent les échecs de gestation (Santos et al. 2004) (Diskin et Morris 2008). Les études montrent que la fécondation est efficace pour l'espèce bovine dans près de 90% des cas, que ce soit chez des vaches ou des génisses (Sartori et al. 2002). En revanche, en moyenne 35% des embryons sont perdus avant le 24^{ème} jour post-fécondation, ce qui correspond à l'étape de développement embryonnaire précoce avant l'implantation (Ryan et al. 1993) (Dunne, Diskin, et Sreenan 2000). Les échecs en période embryonnaire tardive (entre l'implantation et la vie fœtale débutant au 42^{ème} jour de gestation) ou en période fœtale sont de moins en moins importants à mesure de l'avancement du développement. Les échecs en période embryonnaire tardive vont de 3 à 14% suivant les études (Humblot 2001) (Silke et al. 2002) (Zobel et al. 2011). Enfin, les pertes en période fœtale sont inférieures à 10% (Diskin, Murphy, et Sreenan 2006).

1.1.2. Conséquences organisationnelles et économiques

Les échecs de gestation après IA ont des conséquences multiples sur les élevages, que ce soit en bovin laitier ou allaitant (Engelken 2008a) (Engelken 2008b).

Les éleveurs sont les premiers impactés par l'insuccès d'une tentative d'IA. La mise en place de l'IA a tout d'abord un coût pour l'éleveur : l'achat des paillettes, l'intervention de l'inséminateur, d'un vétérinaire, etc. Une IA non réussie implique ensuite un décalage dans les vêlages. Etant donné que les cycles de lactation sont initiés par une gestation, dans un élevage laitier le décalage des vêlages est lui-même associé à un décalage dans les lactations et donc à une perte de production de lait. Dans les élevages de bovins allaitants, une IA ratée implique une diminution de la production de veaux, qu'ils soient destinés au renouvellement du cheptel ou pour la production de veaux de boucherie. Quand l'IA a échoué et que le retour en chaleur des vaches a été détecté rapidement, on peut dans certains cas les inséminer à nouveau. Cependant, dans le cas où les pertes se produisent tardivement (par exemple, en période fœtale), les répercussions économiques sont plus sérieuses car il est souvent trop tard pour inséminer à nouveau et ainsi conserver la synchronicité des mises-bas et des lactations du cheptel. On observe aussi une augmentation du taux de réforme des génisses. Les différentes raisons qui amènent à réformer les génisses sont exposées dans l'étude de Burnside (Burnside et al. 1971). Une des raisons majeures du taux de réforme des génisses (dans au moins 25% des cas) concerne la fertilité de la génisse, qui peut sembler avoir des difficultés à concevoir en IA et qui sont alors retirées de la chaîne de reproduction (Vleck et Norman 1972).

En plus de l'impact sur les éleveurs, les échecs en IA se répercutent aussi sur les ES et les EMP. En effet, les ES élèvent les taureaux et produisent les paillettes, qui sont perdues si le taureau se révèle totalement inefficace. La diffusion des taureaux inefficaces représente aussi une perte économique pour les EMP, qui sont parfois engagées à effectuer de nouveau une IA si la première n'a pas abouti. Pour une EMP, le coût par paillette avec ou sans la mise en place est comprise entre 10 à 20 euros en fonction de l'entreprise. Le coût de la mise en place est lui fonction de la distance de déplacement à réaliser, qui est d'en moyenne 21,6km en France (Évaluation du service universel de la distribution et de la mise en place de la semence de ruminants, Conseil général de l'alimentation de l'agriculture et des espaces ruraux, 2014). Enfin, la commercialisation et la mise en place de paillettes non-fécondantes compromettent l'image des ES et des EMP auprès des éleveurs, qui seront plus réticents à employer leurs services.

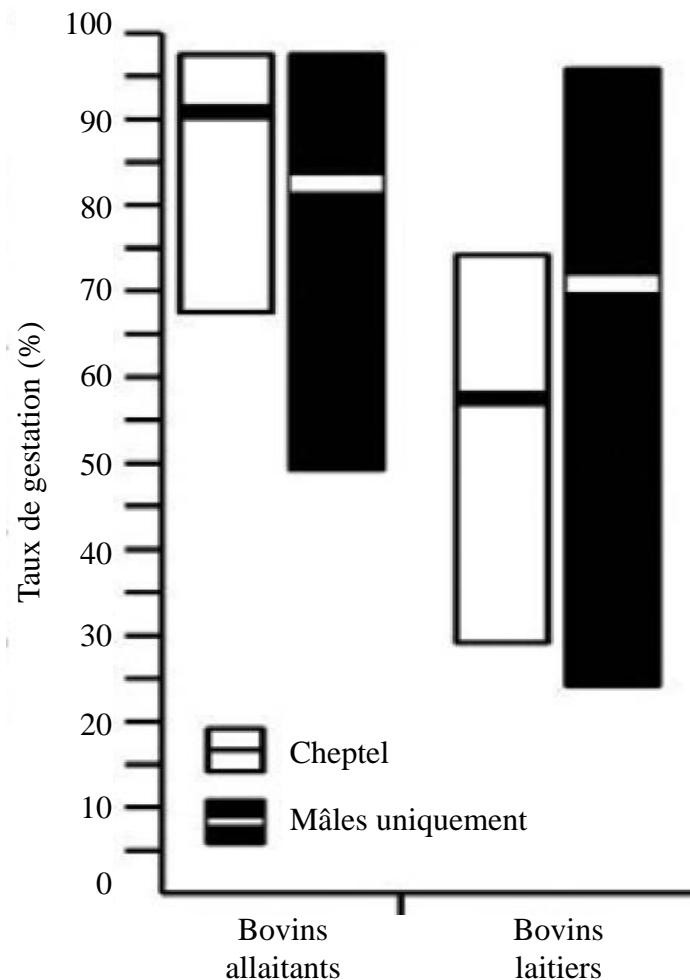


Figure 18. Estimation de la fertilité des cheptels et des mâles uniquement

Les rectangles blancs représentent les performances reproductives des cheptels pour les bovins allaitants (à gauche) et laitiers (à droite). Les rectangles noirs représentent les performances reproductives mais uniquement pour les mâles des mêmes cohortes. Les données ont été obtenues de nombreuses études répertoriées dans la publication de Flowers. La différence relative entre les performances des cheptels et des mâles uniquement donne une idée de la contribution des mâles dans les échecs en reproduction. Flowers postule que si la dispersion des résultats et les moyennes (barres noires et blanches) étaient identiques entre les cheptels et les mâles, alors les femelles et les mâles contribueraient de manière équivalente aux échecs de fertilité. En revanche, si les données diffèrent alors on peut en déduire qu'un des 2 sexes contribue de manière plus significative aux échecs en reproduction. Ce qu'on peut observer, c'est que pour les bovins allaitants le maximum de fertilité des mâles enregistrée est équivalent à celui des cheptels, alors que le minimum est largement inférieur pour les mâles que pour le cheptel. Les résultats sont opposés pour les bovins laitiers. Flowers en conclut que dans le cas de l'industrie bovine allaitante, les mâles contribuent de manière significative aux échecs de fertilité enregistrés pour les cheptels entiers, alors que pour les bovins laitiers ce sont les femelles qui contribuent de manière plus significative aux échecs de fertilité des cheptels. (adapté de Flowers *et al.*, 2013)

1.1.3. Estimation de la contribution des mâles à la baisse de la fertilité

La tendance à une baisse générale de la fertilité dans les élevages bovins repose sur une combinaison d'effets tels que la conduite d'élevage, l'environnement, mais aussi la pression de sélection sur les paramètres de production des vaches.

En effet, une sélection intensive pour améliorer des critères de production (par exemple, la production laitière) a pu se faire au détriment des autres fonctions, dont l'efficacité reproductive (Lucy 2001). De nombreuses études ont montré qu'il existe une corrélation génétique négative (autour de -0,35) entre la production laitière et le taux de conception (VanRaden et al. 2004) (Berry, Wall, et Pryce 2014). Cette corrélation génétique négative serait le résultat d'une pression métabolique mise sur la production au détriment de l'efficacité reproductive. Cependant, la corrélation génétique qui existe entre le taux de conception et la production laitière n'explique qu'en partie la baisse de la fertilité, et de plus uniquement une baisse de la fertilité femelle. De plus, la baisse de l'efficacité reproductive des génisses ne peut pas être associée à une pression de sélection sur les paramètres de production car celles-ci n'ont jamais connu de lactation. Il a aussi été montré que seulement la moitié de la diminution du taux de conception est expliquée par la sélection génétique sur les paramètres de production (Barbat et al. 2010). La diminution de la fertilité globale semble donc, au moins en partie, dépendante de la pression de sélection sur les caractères de production des femelles : d'autres facteurs seraient en jeu. Cette diminution de la fertilité globale pourrait être aussi en partie imputable à une baisse de la fertilité chez les mâles.

Une étude réalisée par Flowers en 2013 compile à la fois des données de fertilité globale des cheptels provenant d'industriels de la filière et d'agences gouvernementales, avec des données de fertilité mâle provenant d'études plus anciennes (Flowers 2013). Les résultats sont représentés Figure 18. Flowers a comparé la différence relative entre la fertilité globale des cheptels (mâles et femelle) avec la fertilité globale des mâles uniquement. Il en déduit que s'il n'y a pas de différence entre la fertilité du cheptel et la fertilité mâle, alors femelles et mâles contribuent équitablement à la fertilité générale du cheptel. En filière laitière comme en filière allaitante, le minimum de fertilité observé chez les mâles est inférieur au minimum de fertilité des cheptels. Flowers interprète cela comme étant une contribution significative des taureaux aux problématiques de fertilité globale. Cependant, la contribution mâle est beaucoup plus marquée en filière allaitante qu'en filière laitière, ce qui va dans le sens des données décrites plus haut, où il a été montré que la mauvaise fertilité dans la filière laitière est fortement induite par les femelles du fait de l'importante pression de production.

Le potentiel reproducteur d'un taureau résulte du bon déroulement de toutes les étapes nécessaires à l'acquisition de spermatozoïdes aptes à produire une descendance, de la vie fœtale à la vie adulte : différenciation des cellules germinales et des spermatides, leur relargage dans la lumière de l'épididyme, un bon transit et maturation dans les différentes sections de l'épididyme, et une bonne production des fluides spermatiques par les glandes accessoires, et enfin une bonne aptitude de la semence à subir des traitements technologiques. De plus, la longévité de la survie spermatique dans le tractus femelle est très variable d'un taureau à l'autre. Les taureaux les plus fertiles ont en général des spermatozoïdes qui survivent plus longtemps dans le tractus génital femelle (MacMillan et Watson 1975). De ce fait, les causes de la baisse de la fertilité mâle sont de natures très diverses, et peuvent avoir pour origine une perturbation à n'importe quelle étape du processus de production de la semence.

Alors que le développement embryonnaire tardif et fœtal est sous la dépendance de l'environnement maternel, au moins une partie des pertes au cours de la fécondation et du développement embryonnaire précoce peut être imputée au mâle. Il a notamment été montré que des spermatozoïdes de taureaux sub-fertiles sont compétents à la fécondation, mais que la formation du *pronucleus* paternel peut être impactée, conduisant à un premier clivage zygotique retardé (Ward et al. 2001). En effet, si le génome paternel présente des dommages, il peut jusqu'à un certain point être réparé par des facteurs ovocytaires avant l'EGA (Brandriff et Pedersen 1981). Cependant, si les dommages sont trop importants et persistent, le développement est arrêté après les premiers clivages et l'embryon entre en apoptose (Fatehi 2006). Des résultats similaires ont été décrits quand les spermatozoïdes utilisés sont morphologiquement anormaux (Walters et al. 2006). Enfin, il faut aussi prendre en compte les possibilités d'erreurs humaines dans la mesure de la fertilité mâle comme de la fertilité femelle. Par exemple, le succès de l'IA est fortement lié à une bonne détection des chaleurs de la femelle, nécessaire pour choisir le moment adéquat pour inséminer (Roelofs et al. 2010). Il apparaît qu'elle peut être mal évaluée dans 4 à 8% des inséminations en France (Freret et al. 2006). Il existe aussi un risque de mal évaluer la gestation de la vache, qui peut potentiellement être inséminée à nouveau alors qu'elle était déjà gravide, ou alors la considérer faussement comme gravide car les signes de retour en chaleur sont absents ou n'ont pas été détectés. Il est donc nécessaire de comprendre les causes des échecs en IA pour pouvoir dans la mesure du possible les éviter.

1.2. Les causes de la baisse de fertilité mâle

1.2.1. Effets environnementaux sur la baisse de la fertilité mâle

L'ensemble de la chaîne de production des spermatozoïdes est très sensible aux effets de l'environnement. Un grand nombre d'études concernent la mise en évidence de chacun de ces effets environnementaux sur le bon déroulement de la spermatogenèse et leurs conséquences sur la fertilité, chez l'Homme et la souris mais aussi chez le bovin.

Il a par exemple été montré dans une étude chez l'Homme que la pollution atmosphérique peut conduire à une fragmentation de l'ADN spermatique, sans conséquence sur la concentration ou la morphologie des spermatozoïdes (Rubes et al. 2005). Le même type d'observation a été fait chez des individus exposés à des pesticides (Sánchez-Peña et al. 2004), des perturbateurs endocriniens (Tavares et al. 2016) ou un niveau élevé de radiations chez la souris (Sailer et al. 1995). Chez l'Homme, des facteurs comme l'alcoolisme (Muthusami et Chinnaswamy 2005), le tabagisme (Prentki Santos et al. 2011), mais aussi la prise de traitements médicamenteux (Hayashi, Miyata, et Yamada 2008) peuvent induire une altération de la spermatogenèse et une baisse de la fertilité. De plus, il a été observé que la concentration spermatique et la motilité peuvent être altérées en cas de stress psycho-social (Ying Li et al. 2011). Il faut aussi garder à l'esprit que contrairement aux études se focalisant sur une unique source de perturbation environnementale, dans la réalité les individus sont plus ou moins exposés à une combinatoire de perturbateurs, dont les effets peuvent se cumuler.

En plus de ces effets environnementaux identifiés chez l'Homme et la souris qui peuvent également impacter les bovins, d'autres sources potentielles de perturbations peuvent être évoquées de manière plus spécifique à la filière bovine. Dans les pays d'Amérique du Nord, il est autorisé d'utiliser des hormones pour stimuler la croissance des jeunes taureaux. Bien que l'efficacité de tels traitements sur la croissance des individus soit démontrée (Greathouse et al. 1983), il a aussi été montré que l'utilisation de ces hormones de croissance est associée à une diminution du volume testiculaire et de la spermatogenèse à l'âge adulte (Ballachey et al. 1986).

La nutrition du jeune taureau peut profondément affecter son futur potentiel reproductif. Une augmentation du taux de protéines et de calories dans l'alimentation de jeunes taureaux de races allaitantes permet d'augmenter de 20 à 25% leur poids testiculaire une fois parvenus à l'âge adulte (Brito, Barth, Rawlings, Wilde, Crews, Boisclair, et al. 2007) (Brito, Barth, Rawlings, Wilde, Crews, Mir, et al. 2007a) (Brito, Barth, Rawlings, Wilde, Crews, Mir, et al.

2007b). Plus récemment chez des jeunes taureaux Holstein, il a été montré de la même façon une augmentation de la circonférence scrotale et du poids testiculaire à l'âge adulte, mais aussi une survenue du pic de LH plus précoce de 8 semaines et d'intensité plus importante au cours de la puberté. Les taureaux dont la ration est enrichie en protéines et calories atteignent leur puberté 45 jours plus tôt que les individus nourris selon un régime normal (Dance et al. 2015). L'âge à la puberté est une donnée d'intérêt capital pour la filière de la semence bovine. En effet, un taureau atteignant plus vite sa maturité sexuelle pourra produire de la semence plus tôt. Elle sera ainsi diffusée plus rapidement, ce qui aura pour conséquence un progrès génétique plus rapide par la diminution des intervalles entre les générations (Bezerra et al. 2013).

Une augmentation de la température scrotale perturbe le bon déroulement de la spermatogenèse. C'est d'ailleurs la raison pour laquelle, chez de nombreuses espèces mammifères, les testicules sont situés dans le scrotum à l'extérieur du corps, pour assurer une température inférieure de 3 à 4°C à celle du corps. Chez l'Homme, certaines professions (métallurgistes, conducteurs routiers, etc.) ou activités sportives pratiquées de façon intensive (cyclisme, sports motorisés, équitation, etc.) peuvent être associées à une élévation de la température scrotale (Józków et Rossato 2017). Chez le bovin, une augmentation de la température scrotale est associée à des altérations de la morphologie de la tête des spermatozoïdes et à plus de 50% de diminution du taux de développement embryonnaire mesuré au stade blastocyste après fécondation *in vitro* (FIV) (Lucio et al. 2016).

1.2.2. Effets des traitements technologiques de la semence

Les biotechnologies de la reproduction et toutes leurs étapes préparatoires ont aussi un impact négatif sur la fertilité mâle, depuis la collecte de l'éjaculat, le remplacement du plasma séminal par des éléments de dilution, le conditionnement en paillettes et leur cryoconservation, le tri des spermatozoïdes pour le sexage dans certains cas, et enfin l'IA en elle-même. Il faut aussi prendre en considération la variabilité induite par les manipulateurs qui effectuent toutes ces étapes.

Le plasma séminal contient des centaines de protéines, incorporées progressivement après le transit épидидymaire. Lors de la collecte des éjaculats, le plasma séminal est éliminé et remplacé par des éléments diluants. En effet et de manière paradoxale, il a été montré que l'incubation à long terme des spermatozoïdes bovins dans leur plasma séminal a un impact négatif sur leur viabilité (Way, Griel, et Killian 2000). Cependant, l'absence des protéines du plasma séminal pourrait affecter le pouvoir fécondant des spermatozoïdes dans le contexte de

l'IA (Moura 2006a) (Moura 2006b). Il a notamment été montré chez la souris que le plasma séminal contient des composants importants pour le développement embryonnaire et placentaire (Bromfield et al. 2014). De plus, la dilution et la mise en paillettes impliquent nécessairement une baisse de la concentration et du nombre de spermatozoïdes utilisés pour l'insémination, en comparaison à un éjaculat entier mis en jeu lors de montes naturelles.

Il est connu que la membrane spermatique est très sensible à la cryoconservation. La diminution de la viabilité des spermatozoïdes après cryoconservation est compensée par la concentration en spermatozoïdes de chaque paillette : il faut en moyenne 5 fois plus de spermatozoïdes pour atteindre la même fertilité qu'avec la semence fraîche, soit entre 1×10^6 à 11×10^6 spermatozoïdes pour les taureaux Holstein (Den Daas et al. 1998). Il est aussi décrit que la cryoconservation peut provoquer une capacitation prématurée des spermatozoïdes (Bailey, Morrier, et Cormier 2003). L'effet de la cryoconservation sur les paramètres spermatiques dépend cependant du type d'élément diluant utilisé (Hinsch et al. 1997).

L'utilisation du sexage de la semence, développé en 1983 par Garner et son équipe, semble diminuer la fertilité (Norman, Hutchison, et Miller 2010). Le sexage consiste en un tri des spermatozoïdes en fonction du chromosome sexuel X ou Y qu'ils contiennent. Les spermatozoïdes sont séparés en cytométrie en flux par une mesure de la fluorescence après leur coloration par un agent intercalant de l'ADN (généralement le Hoechst 33.342). Cependant, la coloration semble retarder la réplication de l'ADN au cours des divisions cellulaires après la fécondation ainsi que la production des acides ribonucléiques messagers (ARNm) chez l'embryon (Garner 2009). De plus, l'usage de la cytométrie en flux induit des dommages mécaniques sur les spermatozoïdes (Suh, Schenk, et Seidel 2005). L'utilisation de la semence sexée plutôt que de la semence conventionnelle est associée à une diminution d'environ 20% du taux de conception, chez les vaches comme chez les génisses, bien que les récentes améliorations de cette méthode semblent diminuer cet écart (S. T. Butler et al. 2014).

1.2.3. Effets génétiques

De nombreux effets génétiques ont été identifiés comme ayant une grande influence sur la fertilité. Depuis les premières évaluations de génotypage, il a été remarqué que certains haplotypes sont fréquents à l'état hétérozygote dans la population, mais ne sont jamais présents à l'état homozygote. Cela signifie que ces haplotypes ne sont pas viables à l'état homozygote. Ce phénomène est appelé la perte d'homozygotie. L'observation de *conceptus* arrêtés à des stades fœtaux ou le phénotype des cas d'animaux arrivés à terme a permis d'identifier et de classer certains syndromes pour diverses races bovines.

Race	Haplotype	Chromosome	Gènes associés	Fréquence de l'haplotype dans la population (%)	Syndrome associé	
Holstein	BY	21	FANCI	3,6	Brachyspina	
Holstein	Holstein Haplotype 1 (HH1)	5	APAF1	2,6		
Holstein	Holstein Haplotype 2 (HH2)	1		1,7		
Holstein	Holstein Haplotype 3 (HH3)	8	SMC2	2,5		
Holstein	Holstein Haplotype 4 (HH4)	1	GART	3,6		
Holstein	Holstein Haplotype 5 (HH5)	3	TFB1M	3,9		
Holstein	Holstein Haplotype 6 (HH6)	3	SLC35A3	4,6	Complex Vertebral Malformation (CVM)	
Holstein	Holstein Haplotype 7 (HH7)	6		6,9		
Holstein	Holstein Haplotype 8 (HH8)	7		2,1		
Holstein	Holstein Haplotype 9 (HH9)	10		2,9		
Holstein	Holstein Haplotype 10 (HH10)	11		2,2		
Holstein	Holstein Haplotype 11 (HH11)	12		3,2		
Holstein	Holstein Haplotype 12 (HH12)	15		5,5		
Holstein	Holstein Haplotype 13 (HH13)	18		3,7		
Holstein	Holstein Haplotype 14 (HH14)	19		2,4		
Holstein	Holstein Haplotype 15 (HH15)	20		2,7		
Holstein	Holstein Haplotype 16 (HH16)	26		2,0		
Holstein	Holstein Haplotype 17 (HH17)	26		1,8		
Montbéliard	Montbéliarde Haplotype 1 (MH1)	19	PFAS	9,0		
Montbéliard	Montbéliarde Haplotype 2 (MH2)	29	SLC37A2	7,0		
Montbéliard	Montbéliarde Haplotype 3 (MH3)	2		5,1		
Montbéliard	Montbéliarde Haplotype 4 (MH4)	4		3,5		
Montbéliard	Montbéliarde Haplotype 5 (MH5)	6		7,1		
Montbéliard	Montbéliarde Haplotype 6 (MH6)	7		2,6		
Montbéliard	Montbéliarde Haplotype 7 (MH7)	9		7,1		
Montbéliard	Montbéliarde Haplotype 8 (MH8)	13		3,5		
Montbéliard	Montbéliarde Haplotype 9 (MH9)	20		2,6		
Montbéliard	Montbéliarde Haplotype 10 (MH10)	24		2,5		
Montbéliard	Montbéliarde Haplotype 11 (MH11)	24		7,2		
Normand	Normande Haplotype 1 (NH1)	24		1,8		
Normand	Normande Haplotype 2 (NH2)	1		3,8		
Normand	Normande Haplotype 3 (NH3)	4		5,9		
Normand	Normande Haplotype 4 (NH4)	6		5,2		
Normand	Normande Haplotype 5 (NH5)	7		1,9		
Normand	Normande Haplotype 6 (NH6)	15		1,9		

Tableau 1. Liste des haplotypes létaux au stade embryonnaire chez le bovin

(données tirées de Fritz et al., 2013 et de la base de donnée de l'*Online Mendelian Inheritance in Animals* (OMIA), Université de Sydney, Australie)

Il a été découvert que certains de ces haplotypes correspondent à des mutations génétiques ou à des défauts chromosomiques létaux au stade embryonnaire (VanRaden et Miller 2006). L'utilisation des puces à SNP bovines ont alors permis de mettre en lien la perte d'homozygotie et les mutations causales dans les gènes associés (VanRaden et al. 2011). En 2013, Fritz et son équipe ont identifié 17, 11 et 6 haplotypes présentant une perte d'homozygotie significative chez les races Holstein, Montbéliard et Normand, respectivement (Fritz et al. 2013) (Tableau 1). Par exemple, pour les *Montbeliard Haplotype 1* (MH1), une mutation causale candidate a été identifiée sur le gène *phosphoribosylformylglycinamide synthase* (*PFAS*) codant pour un transporteur d'androgène. Sur 128 743 animaux génotypés et étudiés, aucun d'entre eux ne s'est révélé porteur de la mutation à l'état homozygote (Michot et al. 2017). Pour MH2, une mutation causale impliquerait le gène *Solute Carrier Family 37 Member 2* (*SLC37A2*) codant pour un transporteur membranaire de glucose-6-phosphate. Il a notamment été montré que la déficience en protéines impliquées dans le métabolisme du glucose-6-phosphate est également létale au stade embryonnaire chez la souris (Longo et al. 2002). De nombreuses autres mutations causales ont été identifiées à ce jour. Par exemple, la mutation causale du gène *Solute Carrier Family 35 Member 3* (*SLC35A3*) identifiée par Thomsen et ses collègues en 2006 serait responsable du *Complex Vertebral Malformation* (CVM), impliquant des pertes fœtales entre le 56^{ème} et le 168^{ème} jour post-fécondation (Thomsen 2005). La mutation causale du *Brachyspina syndrome*, identifiée par Charlier en 2012, est retrouvée dans 7 à 8% des individus de race Holstein (Charlier et al. 2012). La mutation entraînant ce syndrome serait responsable d'environ 0,14% des pertes embryonnaires de cette race.

Les mutations génétiques délétères pour la fertilité se sont notamment répandues dans la population du fait d'une hausse importante de la consanguinité des élevages au cours du XX^{ème} siècle (Cassell, Adamec, et Pearson 2003). Encore récemment, le taux de consanguinité augmentait de 0,10 à 0,25% par an dans la plupart des élevages (Kearney et al. 2004) (Mc Parland et al. 2007) (Hinrichs et Thaller 2011). Depuis, une meilleure planification des reproductions, en prenant davantage en considération le degré de parenté entre les individus, a permis d'endiguer en partie ce phénomène.

1.2.4. Autres effets

Divers effets peuvent avoir une influence négative sur la fertilité mâle mais ne sont pas détaillés dans ce document. Il peut s'agir entre autres, des maladies de type auto-immunes, d'infections ou d'obstructions du tractus génital.

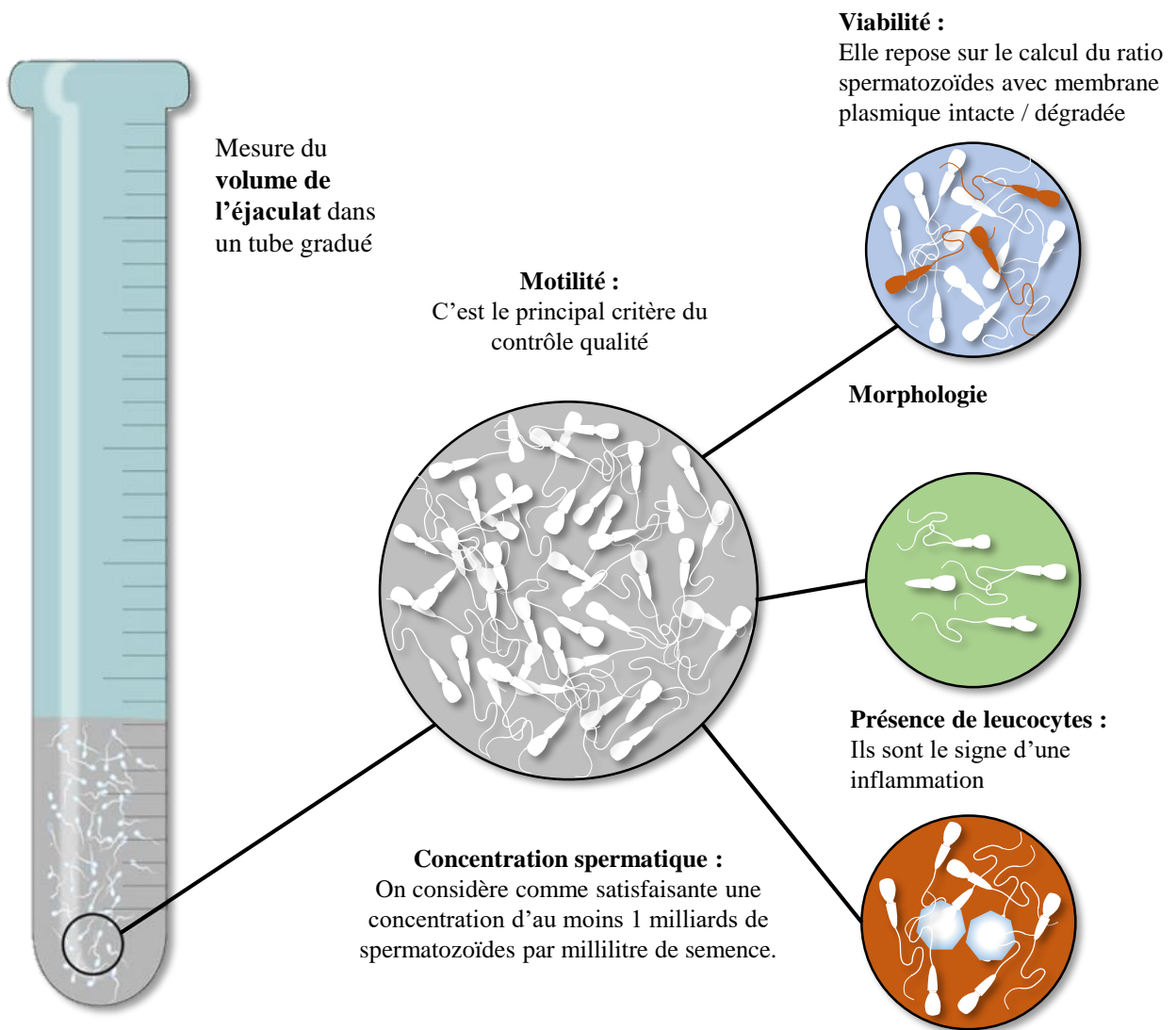


Figure 19. Le spermogramme

Ils sont effectués en routine avant le conditionnement des paillettes de semence.

2. Méthodes d'évaluation de la fertilité mâle

2.1. Analyses phénotypiques

2.1.1. Le spermogramme

Dans les ES, des contrôles qualité des éjaculats sont réalisés en routine avant leur conditionnement en paillette puis leur utilisation en IA. En effet, les ES souhaitent toujours maximiser le potentiel fécondant de leurs paillettes avant de les commercialiser. A l'issue des contrôles qualité de la semence, le taux de jeunes mâles éliminés pour inaptitude à produire de la semence de bonne qualité est de l'ordre de 15% (Gerard et al. 2008). Ces contrôles doivent être effectués au plus vite après la collecte de l'éjaculat. Le premier mis en œuvre et le plus simple est une mesure directe du volume de l'éjaculat dans un tube gradué. Les autres tests effectués en routine consistent principalement en la mesure de la concentration spermatique, de la motilité des spermatozoïdes et de leur viabilité (DeJarnette et al. 2004), regroupés sous le terme de spermogramme (Figure 19). Les paramètres mesurés par le spermogramme sont interconnectés ; il faut alors les mettre en perspective les uns par rapport aux autres pour évaluer d'où provient un éventuel spermogramme anormal.

La mesure de la motilité est le principal critère de contrôle qualité. Elle est faite par l'observation de la motilité massale (c'est-à-dire de la motilité affectée par la concentration en spermatozoïdes), à l'aide d'un microscope optique (avec grossissement X100). A l'issue de cette observation, une note allant de 1 à 5 est attribuée. La note minimale pour satisfaire le contrôle est une note de 3, qui correspond à 60% de spermatozoïdes motiles dans l'éjaculat. La concentration en spermatozoïdes est déterminée par spectrophotométrie. On considère comme satisfaisante une concentration d'au moins 1 milliards par millilitre de semence. La viabilité spermatique est observée au microscope optique (X100) à partir d'une gouttelette de semence pure déposée entre lame et lamelle à 37°C. La mesure de la viabilité repose sur le calcul du ratio spermatozoïdes avec une membrane plasmique intacte par rapport à ceux présentant une membrane plasmique dégradée. La viabilité des spermatozoïdes est liée aussi indirectement à leur motilité. Au cours de cette analyse, on peut aussi observer la présence ou non de leucocytes dans la semence (signe d'une inflammation) ou de spermatozoïdes immatures (signe d'une spermatogenèse incomplète).

De nombreux laboratoires sont aujourd'hui équipés d'un système automatisé dit *Computer Assisted Semen Analyzer* (CASA). Le CASA permet d'effectuer tous les types de mesures

précédemment cités, en s'affranchissant des biais liés à la subjectivité et l'erreur humaine (Garrett 2003). Par rapport aux analyses manuelles, de nouveaux paramètres peuvent être mesurés, comme la trajectoire des spermatozoïdes ou encore le type et la vitesse de leurs déplacements (Amann et Waberski 2014). Cette méthode fait cependant apparaître d'autres biais comme les conditions de préparation des échantillons, la programmation de la machine, etc. (Amann et Katz 2004).

Ces mesures sont faites au niveau de l'éjaculat, mais elles ne sont plus valables après décongélation de la paillette car le processus de cryoconservation a une influence négative sur les spermatozoïdes. C'est pourquoi pour chaque éjaculat conditionné, d'autres contrôles sont faits en décongelant 2 paillettes de semence prises au hasard. Pour que l'IA reste efficace après la cryoconservation, il faut observer au moins 6 millions de spermatozoïdes vivants et motiles par paillette après décongélation (ce qui représente une perte d'environ 50 % des spermatozoïdes).

2.1.2. Bull Breeding Soundness Evaluation

La mesure du spermogramme peut être complétée par des analyses phénotypiques de l'individu. L'ensemble des caractéristiques évaluées est combiné pour produire le *Bull Breeding Soundness Evaluation* (BBSE) (Alexander 2008). On pratique notamment un examen physique des organes reproducteurs (dont une mesure de la circonférence scrotale, CS), un examen physique général (taille de l'individu, musculature, etc.) et parfois un examen de santé et de libido. Il existe d'autres méthodes d'évaluation de l'intégrité testiculaire que la mesure de la CS, notamment l'échographie (Coulter et Bailey 1988) ou la thermographie (Lunstra et Coulter 1997) Dans le BBSE, les taureaux sont classés entre satisfaisants (*satisfactory*) et insatisfaisant (*unsatisfactory*). Pour être classé satisfaisant, le taureau doit remplir les conditions suivantes : une CS d'au moins 30 cm à 15 mois, 70% de spermatozoïdes sans anomalie morphologique (les anomalies sont détaillées section suivante), et 30% de motilité totale. La fertilité des taureaux varie de plus ou moins 20% entre les taureaux classés satisfaisants, et cette variation est due à des différences sub-microscopiques présentes dans les spermatozoïdes (Kastelic et Thundathil 2008).

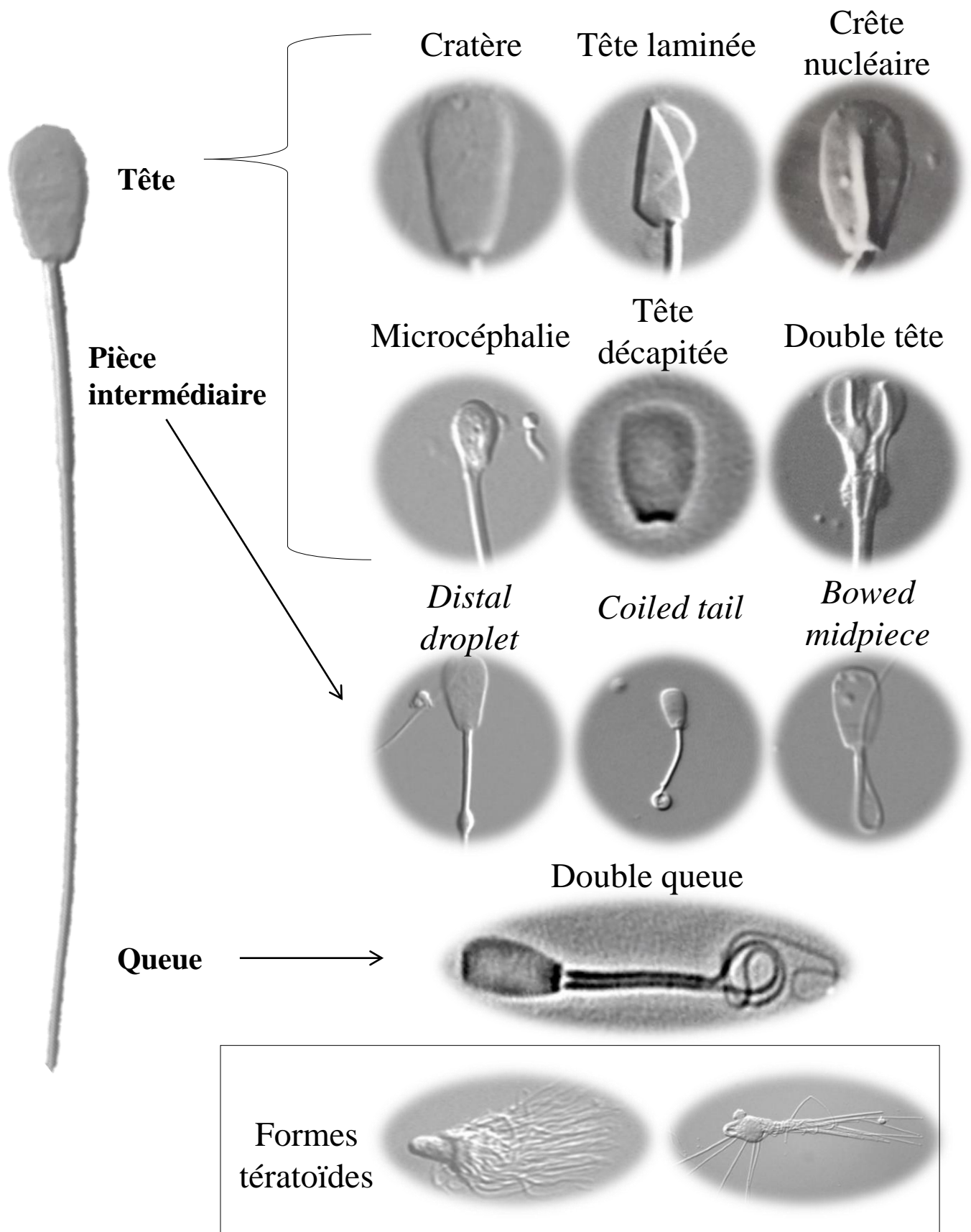


Figure 20. Exemple d'anomalies morphologiques des spermatozoïdes retrouvées chez le bovin

Les photographies ont été prises après observation par microscopie en contraste de phase, suivant différents grossissements.
 (Photos fournis par le LNCR)

2.1.3. Analyses morphologiques des spermatozoïdes

Les premières analyses concernant la bonne morphologie des spermatozoïdes bovins ont été effectuées il y a près d'un siècle. Tout comme les analyses macroscopiques de la qualité de la semence, la présence ou l'absence d'anomalies morphologiques des spermatozoïdes peuvent aussi être mesurées pendant les contrôles qualité. Pour que les structures spermatiques puissent être observées au microscope, la semence est diluée avec des solutions qui colorent les structures spermatiques (communément la coloration de Feulgen, d'éosine-aniline bleue, d'éosine-nigrosine). Il faut alors examiner au moins 100 spermatozoïdes sous un fort grossissement (entre x1000 et x1200), puis identifier et compter les anomalies.

Dans la filière bovine, l'analyse morphologique est secondaire par rapport aux analyses décrites précédemment car elle est très consommatrice en temps et dépendante de la subjectivité et de l'expérience de l'observateur. Les anomalies peuvent être plus ou moins marquées, et il est alors difficile de définir à partir de quel moment une structure spermatique doit être qualifiée d'anormale. Cependant, tout comme le CASA pour les analyses fonctionnelles, l'analyse morphologique s'est automatisée avec le développement d'un système *Computer-Assisted Sperm Morphology Analyzer* ou CASMA (Stephens, Hickman, et Hoskins 1988). Le CASMA, plutôt que l'observation manuelle, apporte une importante réduction des biais de mesure (Gravance et al. 1999).

Lorsqu'on analyse les spermatozoïdes, le premier point consiste à localiser la structure impactée (la tête, la pièce intermédiaire ou la queue). Il est alors possible d'identifier les éléments internes à la structure qui semblent impliqués. L'élément de structure impacté renseigne sur le stade de développement des spermatozoïdes auquel s'est produite l'anomalie. A partir de cela, on peut déterminer la cause de l'anomalie, et potentiellement prendre des mesures pour les pallier. Des photographies des anomalies spermatiques bovines les plus courantes sont données Figure 20.

Presque tous les éjaculats comportent des spermatozoïdes avec des profils anormaux. C'est à partir du moment où ces spermatozoïdes anormaux représentent une fraction importante de l'éjaculat qu'un problème de fertilité peut être diagnostiqué. Les taureaux sub-fertiles ou stériles présentent alors souvent une combinaison de différentes anomalies dont les morphologies sont les plus marquées. Bien que fortement dépendant du type d'anomalie, peu de taureaux sont efficaces à la reproduction s'ils ont un taux de spermatozoïdes anormaux de plus de 25-30%, dont 15-20% d'anomalies de la tête (Barth et Oko 1989).

2.1.4. Méthodes exploratoires

Dans la filière bovine les contrôles qualité de la semence doivent être effectués de manière rapide et économe, tout en étant le plus informatifs possible. Les évaluations en routine permettent d'identifier les taureaux présentant un profil spermatique distinctement anormal, qui peuvent alors être retirés de la chaîne de production de semence. Les taureaux qui valident les critères de contrôle qualité pourraient être testés avec d'autres méthodes pour évaluer des éléments plus subtils ayant un impact sur la fertilité. De nombreuses méthodes permettant d'évaluer la fonctionnalité des spermatozoïdes ont été développées mais ne sont pas utilisées en routine. On peut les classer en 2 types : celles qui permettent de mesurer l'intégrité spermatique, et celles dites fonctionnelles qui évaluent la compétence des spermatozoïdes à effectuer certaines tâches.

Des défauts dans l'intégrité de l'ADN spermatique et de la structure de la chromatine au cours de la spermiogenèse peuvent conduire à une diminution de la protection de l'ADN, le rendant plus accessible aux cassures qu'on qualifie de simple brin (*Single Strand Break*, SSB) ou de double brin (*Double Strand Break*, DSB). SSB et DSB sont associées à une réduction du succès de la fécondation (Simon et Lewis 2011), et interfèrent avec le développement embryonnaire et l'implantation (Lewis et Aitken 2005). Bien que la taille et la forme de la tête spermatique soit liée à l'organisation de la chromatine, un défaut de compaction n'est pas toujours détectable par les analyses morphologiques, de motilité, ou après FIV. Il existe principalement 3 méthodes qui permettent d'étudier l'état de fragmentation de l'ADN spermatique : la cytométrie en flux utilisant soit un *Sperm Chromatin Structure Assay* ou SCSA (Evenson, Darzynkiewicz, et Melamed 1980), soit un *Terminal deoxynucleotidyl transferase mediated dUTP Nick End Labelling* ou TUNEL (Gorczyca et al. 1993), l'électrophorèse (*Single-cell gel electrophoresis*, COMET (Aravindan et al. 1997)) et l'observation au microscope (*Sperm Bos Halomax*, SBH (García-Macías et al. 2007)). Ces méthodes reposent sur le marquage préalable des DSB et des SSB, notamment par l'acridine orange. Il apparaît que le SCSA permet une meilleure discrimination en fonction de la fertilité, et que c'est la méthode la moins coûteuse et la plus répétable (Evenson 2016).

Différentes techniques ont été développées pour mesurer l'aptitude à la fécondation des spermatozoïdes et sont décrites dans la revue de Vogiatzi (Vogiatzi et al. 2013). Certaines permettent d'examiner l'interaction entre les spermatozoïdes et la ZP, et si cette interaction permet d'enclencher la réaction acrosomique, avec les méthodes de la *Competitive Zona Binding* (D. Y. Liu et al. 1988), ou la méthode de l'*Hemizone Assay* (Burkman et al. 1988). La méthode de la *Sperm Penetration Assay* (Yanagimachi, Yanagimachi, et Rogers 1976)

permet quant à elle d'analyser les 4 étapes clés de la fécondation (capacitation, réaction acrosomique, fusion du spermatozoïde avec la membrane plasmique de l'ovocyte, et décondensation du noyau paternel). La fécondation est réalisée en utilisant un ovocyte de hamster dont la ZP a été préalablement retirée. Enfin, Krausz et ses collègues ont remarqué que la réponse des spermatozoïdes à la progestérone est très corrélée avec le taux de fécondation en FIV (Krausz et al. 1995)(Krausz et al. 1996).

L'intégrité membranaire est étroitement associée à la viabilité des spermatozoïdes. Elle peut être mesurée par l'*Hyposmotic Swelling Test* (Jeyendran et al. 1984). Les spermatozoïdes sont alors placés dans un milieu hypo-osmotique, ce qui induit une entrée d'eau dans le spermatozoïde. Quand la membrane plasmique est intacte, le flagelle se courbe d'une manière caractéristique. L'intégrité membranaire peut aussi être évaluée par coloration à l'éosine, le Syber-14 ou l'iodure de propidium : la pénétration de ces composés est reliée à l'état de la membrane. La translocation de la phosphatidylsérine membranaire, qui précède la perte d'intégrité membranaire après cryoconservation, peut être révélée par l'Annexine-V (Januskauskas, Johannisson, et Rodriguez-Martinez 2003).

Des espèces réactives de l'oxygène (ROS) sont produites par le métabolisme spermatique, comme le peroxyde d'hydrogène (H_2O_2), l'anion superoxyde (O_2^-) ou les radicaux d'hydrogènes (OH). Ces composés ont pour conséquence la peroxydation des lipides membranaires spermatiques, ce qui induit des dommages membranaires et réduit la motilité spermatique, impactant négativement la fertilité (Kasimanickam et al. 2007). L'état de peroxydation peut être évalué par le Test du *sperm MalonDiAldehyde* ou MDA (Gérard-Monnier et al. 1998), et la quantité de ROS en mesurant l'activité de la glutathion peroxydase ou de la superoxyde dismutase.

Le *Mucus penetration test* (Galli et al. 1991) mesure la distance parcourue par les spermatozoïdes à travers du mucus cervical ou d'un mucus artificiel. Cette traversée a été positivement corrélée avec la motilité et l'intégrité acrosomique : les spermatozoïdes de taureaux fertiles migrent plus loin dans le mucus (Taş, Bacinoglu, Cirit, Özdaş, et al. 2007) (Taş, Bacinoglu, Cirit, Özgümüş, et al. 2007).

2.2. Données de terrain après IA

Lorsqu'un reproducteur a passé avec succès les contrôles qualités, il est employé en production de semence, qui sera ensuite elle-même utilisée en IA. On peut alors évaluer la fertilité effective du reproducteur. Pour des raisons pratiques d'organisation de l'élevage (retour rapide sur l'état de gestation des vaches inséminées), des jalons ont été posés tout au long du processus entre IA et la naissance du descendant pour avoir une meilleure estimation du succès de l'IA.

2.2.1. Taux de fécondation

Le taux de fécondation mesure la capacité des spermatozoïdes à féconder un ovocyte. Il peut être évalué 48h après FIV à partir des caractéristiques morphologiques du zygote (extrusion du second globule polaire, l'observation des *pronuclei* ou des premières divisions cellulaires). Il peut aussi être évalué 3 à 4 jours après l'IA en récupérant les zygotes ou les ovocytes non-fécondés par rinçage (*flushing*) des oviductes de la femelle inséminée. Le taux de fécondation, avec le développement embryonnaire très précoce qui suit, est une bonne façon d'évaluer le potentiel fécondant d'un mâle. En effet, par la suite, l'environnement maternel joue un rôle de plus en plus important sur le développement embryonnaire (milieu maternel, sécrétions hormonales, endomètre permettant une bonne implantation, etc.) ; il est alors plus difficile d'évaluer si un défaut de fertilité provient du mâle ou de la femelle. Cette méthode n'est pas réalisée en routine car elle est difficile à mettre en œuvre sur des ovocytes provenant d'un grand effectif de femelles, alors que ce serait nécessaire pour limiter au maximum l'effet des femelles et se focaliser sur l'effet du mâle.

2.2.2. Taux de non-retour

Le taux de non-retour en chaleur (TNR) des femelles inséminées est la plus ancienne et la plus courante mesure d'évaluation du potentiel reproducteur des mâles sur le terrain. Le TNR consiste à diagnostiquer indirectement la gestation en observant si la femelle inséminée revient en chaleur après l'IA. Si c'est le cas, c'est que la gestation n'a pas été mise en place ou a été interrompue. Le TNR peut être évalué à différents points au cours de la gestation, principalement entre 60 et 90 jours après IA aux Etats-Unis et 56 jours en Europe (Amann et DeJarnette 2012). C'est une façon courante de mesurer la performance reproductive mais l'évaluation peut être biaisée par de nombreux effets. Ces effets peuvent être liés à la subjectivité du technicien, à l'échec de détection de l'œstrus, à l'absence d'ajustements avec d'autres effets impactant la gestation (production laitière par exemple, ou parité), etc.

2.2.3. Taux de conception

Le taux de conception géniteur (*Sire Conception Rate*, SCR) est basé sur un taux de naissance plutôt que sur un TNR. Le SCR a été implémenté aux US depuis 2008 par l'USDA *Animal Improvement Programs Laboratory* (Beltsville, Etats-Unis) (Kuhn, Hutchison, et Wiggans 2006) (Kuhn et Hutchison 2008). Pour que le taux de naissance soit indicatif il faut que les taureaux aient été utilisés au moins 300 fois en IA (dont 100 au cours des 12 derniers mois) sur un grand nombre de femelles différentes et dans au moins 10 cheptels différents. Le taux de conception est alors corrigé de certains effets « taureau » pouvant influencer sa fertilité (comme son âge, son degré de consanguinité, etc.) et des effets « femelle » (âge, nombre de lactations, date à l'IA, etc.). Il est alors exprimé de manière relative à une moyenne par rapport à une population de référence (Kuhn, Hutchison, et Norman 2008). Il s'agit de la mesure la plus efficace car elle prend en considération toutes les éventuelles pertes liées à la gestation.

2.2.4. Fertilité et sub-fertilité

Les taureaux employés à la production de semence présentent des performances à la reproduction pouvant être assez variables. Peu de taureaux sont complètement stériles. Ces taureaux sont généralement rapidement identifiés et retirés de la chaîne de production de semence. Cependant, on estime que de 20 à 40% des taureaux sont affectés par une fertilité réduite (Kastelic et al. 2000). Ce sont des taureaux qu'on qualifie de sub-fertiles. Ces taureaux ne présentent souvent pas d'anomalie majeure détectée lors des contrôles qualité de la semence et ne sont alors pas identifiables, puis employés en production. En revanche, les taux de conception de ces taureaux peuvent être réduits d'entre 20 à 25 % (Larson et Miller 2000).

Ce phénomène est d'autant plus marqué depuis l'utilisation à grande échelle de la sélection génomique. Lorsque la valeur génétique des reproducteurs était évaluée par testage sur descendance, on pouvait évaluer au même moment la valeur de l'individu et sa capacité à donner naissance à une descendance viable. Aujourd'hui et dans le contexte de la sélection génomique, c'est-à-dire sans passer par une période de testage sur descendance, on ne peut mesurer les paramètres de fertilité terrain seulement après la diffusion du reproducteur. De plus, le taureau aura déjà été utilisé à grande échelle par les éleveurs avant d'avoir effectué suffisamment d'IA pour pouvoir les calculer les taux. Il n'y a donc aucun retour sur la fertilité effective des reproducteurs avant leur diffusion. Leur fertilité est uniquement estimée sur la base de leur valeur génétique pour la fertilité, et sur les contrôles qualité de la semence.

Caractère	Circonférence scrotale	Nombre de spermatozoïdes	Concentration spermatique	Motilité
Volume de l'éjaculat	0,2	0,83 ± 0,13	- 0,16 ± 0,10	0,06 ± 0,13
Nombre de spermatozoïdes	/		0,60 ± 0,07	0,50 ± 0,13
Concentration spermatique	0,77	/		/
Pourcentage de spermatozoïdes viables	0,63	/	/	/
Motilité	0,76 ± 0,07	/	0,61 ± 0,10	
Ratio spermatozoïdes anormaux	0,31 ± 0,09	/	0,36	0,87 ± 0,08

Tableau 2. Corrélations génétiques entre paramètres spermatiques et phénotypiques liés à la fertilité chez le bovin

Ces données ont été mesurées sur 25 cohortes bovines
(données tirées de Berry *et al.*, 2014)

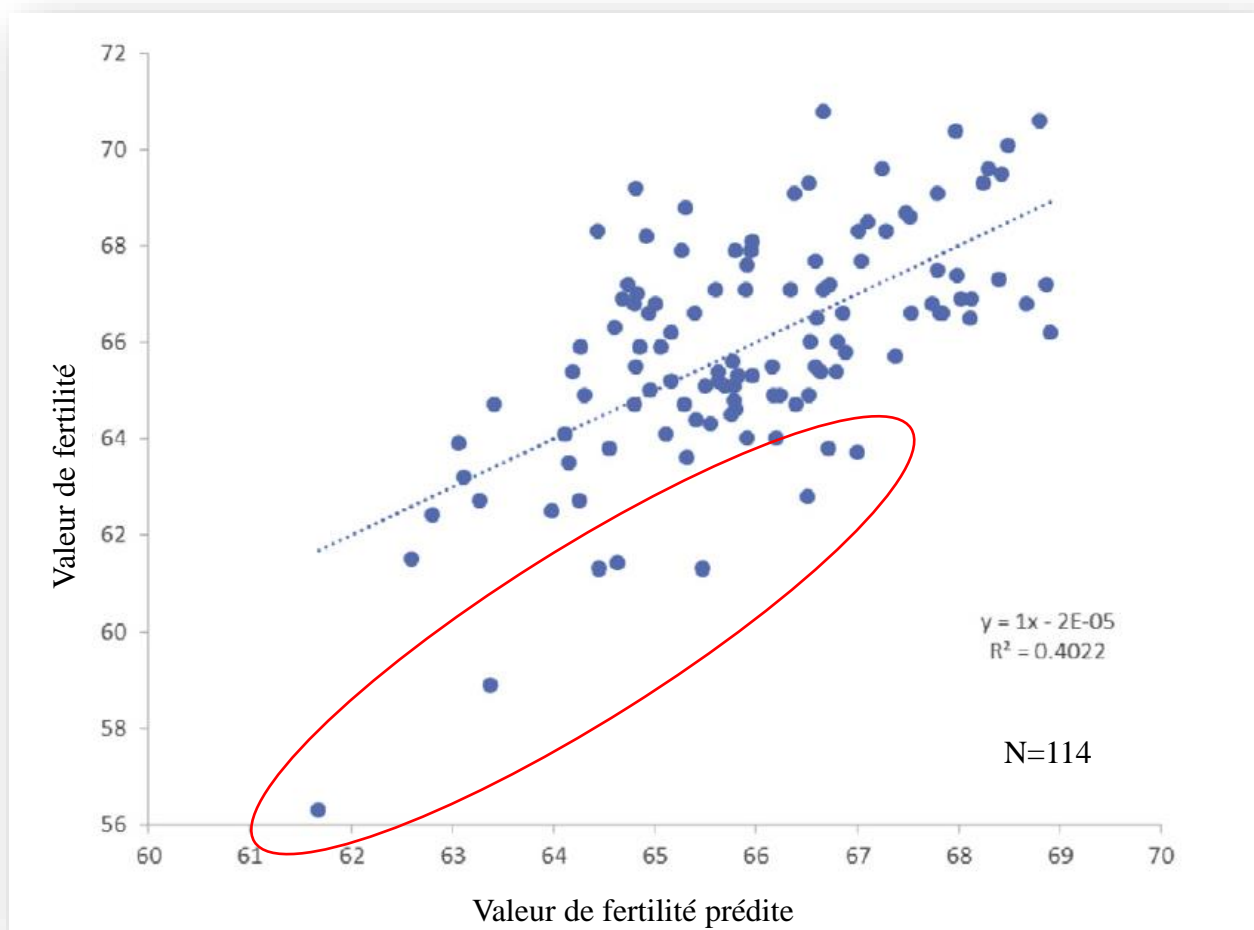


Figure 21. Relation entre la valeur de fertilité prédite et le taux de non-retour à 56 jours

Sellem a établi un modèle de prédiction de la fertilité basé sur des données de fonctionnalités de la semence. La valeur de fertilité prédite peut alors être comparée à la valeur de fertilité réelle observée sur le terrain. Le modèle permet d'expliquer 40% de la fertilité (R^2), donc encore 60% de la variabilité reste inexpliquée. En conséquence, on peut observer qu'un certain nombre d'individus se révèlent moins fertiles sur le terrain que ce qui a été estimé. Bien que donnant de bons résultats, les analyses des données de fonctionnalité ne sont alors pas suffisantes pour écarter totalement les individus sub-fertiles.
(données tirées de Sellem *et al.*, 2015)

3. Les limites des méthodes d'évaluation de la fertilité mâle

3.1. Les analyses phénotypiques de la semence ne permettent pas de prédire la sub-fertilité

En 2014, Berry et son équipe ont publié des données issues d'une méta-analyse sur 25 populations bovines (Berry, Wall, et Pryce 2014). Les résultats sont résumés dans le Tableau 2. On observe une forte corrélation pour les paramètres spermatiques mesurés lors du spermogramme. Par exemple, la corrélation génétique entre la motilité et la présence d'anomalies morphologiques des spermatozoïdes est de $0,87 \pm 0,08$. De la même façon, les caractères phénotypiques tels que la Circonférence Scrotale (CS), sont aussi fortement corrélés aux paramètres de spermogramme, comme avec la concentration spermatique (0,77), la motilité ($0,76 \pm 0,07$), ou encore le nombre de spermatozoïdes viables (0,63). Elle l'est moins avec le volume de l'éjaculat (0,20) ou le taux de spermatozoïdes anormaux ($0,31 \pm 0,09$). Une CS importante est alors associée avec un volume d'éjaculat plus grand, qui contient plus de spermatozoïdes motiles et viables.

Ces corrélations ont été mesurées sur 25 cohortes différentes, et on peut en revanche observer que la variabilité des corrélations entre les cohortes analysées est parfois très importante. De plus, d'autres études se sont focalisées sur l'étude des corrélations entre les paramètres spermatiques ou la CS avec cette fois les résultats en termes de fertilité. Des revues compilent les résultats de ces études (Parkinson 2004) (Burns et al. 2011). Leurs résultats sont souvent contradictoires et montrent que les paramètres spermatiques ne sont pas souvent corrélés à la fertilité. Par exemple, Farrell a montré en 1998 que des caractéristiques en termes de motilité et de vitesse mesurées par CASA étaient hautement corrélées avec la fertilité calculée sur la base d'un TNR à 59 jours (0,87) (Farrell et al. 1998). En 2015, Sellem a conduit une étude sur la corrélation entre paramètres spermatiques évalués par le CASA et une valeur de fertilité, dans le but de pouvoir développer une équation de prédiction de la fertilité mâle pour les centres de production de semence (Figure 21) (Sellem et al. 2015). Chaque paramètre spermatique pris indépendamment (motilité, viabilité, présence d'anomalies morphologiques, activité mitochondriale, etc.) ne permet pas de prédire de façon significative la fertilité (dans cet article, calculée à partir d'un TNR à 56 jours). Un modèle basé sur la combinaison de variables provenant de la mesure de plusieurs paramètres spermatiques donne de meilleurs résultats. Néanmoins, le modèle est calculé en utilisant des données pour des taureaux déjà employés avec un nombre important d'IA pour estimer avec précision la valeur de TNR 56, et

non avec des données de jeunes taureaux ayant peu d'IA à leur actif. Dans le contexte de la sélection génomique les taureaux sont utilisés de plus en plus jeunes, les données de la fertilité terrain sont peu nombreuses, donc cet indicateur n'est pas forcément adapté.

De très nombreuses équipes de recherche se sont penchées sur le lien qui peut exister entre les paramètres spermatiques et la compétence au développement de l'embryon. Par exemple, Andersson et collègues ont publié en 1992 une étude sur le développement embryonnaire en comparant des taureaux fertiles ou sub-fertiles (Andersson, Aalto, et Gustavsson 1992). Bien qu'ayant des spermogrammes et des paramètres testiculaires normaux, les taureaux sub-fertiles avaient des TNR faibles. Un des taureaux présentait notamment une mauvaise aptitude à la fécondation, et un autre plutôt une mauvaise aptitude au développement embryonnaire. Ce dernier présentait une translocation réciproque entre les chromosomes X et Y qui peut expliquer les pertes embryonnaires. De manière similaire, Ducos et collègues en 2000 ont mené une étude sur 3 taureaux de race Montbéliarde présentant une fertilité très faible (Ducos et al. 2000). Les individus présentaient cependant une qualité de semence normale. Un des taureaux montrait une translocation réciproque entre les chromosomes 12 et 17.

Ces résultats montrent bien, qu'au moins dans certains cas, il est impossible de prédire les performances à la reproduction des individus en ne tenant compte que de leur spermogramme ou de leurs paramètres phénotypiques. De plus, la mesure de ces paramètres ne peut être mise en œuvre que tardivement dans le développement de l'animal, qui doit avoir atteint la puberté. D'autres types d'analyse de la qualité de la semence, comme la mesure de l'intégrité membranaire ou de l'état de la chromatine, donnent de meilleures corrélations avec la fertilité et seraient plus adaptés pour prédire la fertilité d'un individu (voir section sur les autres méthodes d'évaluation de la fertilité mâle). Par contre, ces méthodes ne sont pas utilisées en routine dans la filière bovine. Il était alors nécessaire de prédire autrement et plus précocement la fertilité des individus, notamment dans le contexte de la sélection génomique où les individus sont sélectionnés peu de temps après la naissance. Si possible, l'estimation de la fertilité doit aussi prendre en compte les informations de génotype qui sont capitales à un bon développement embryonnaire.

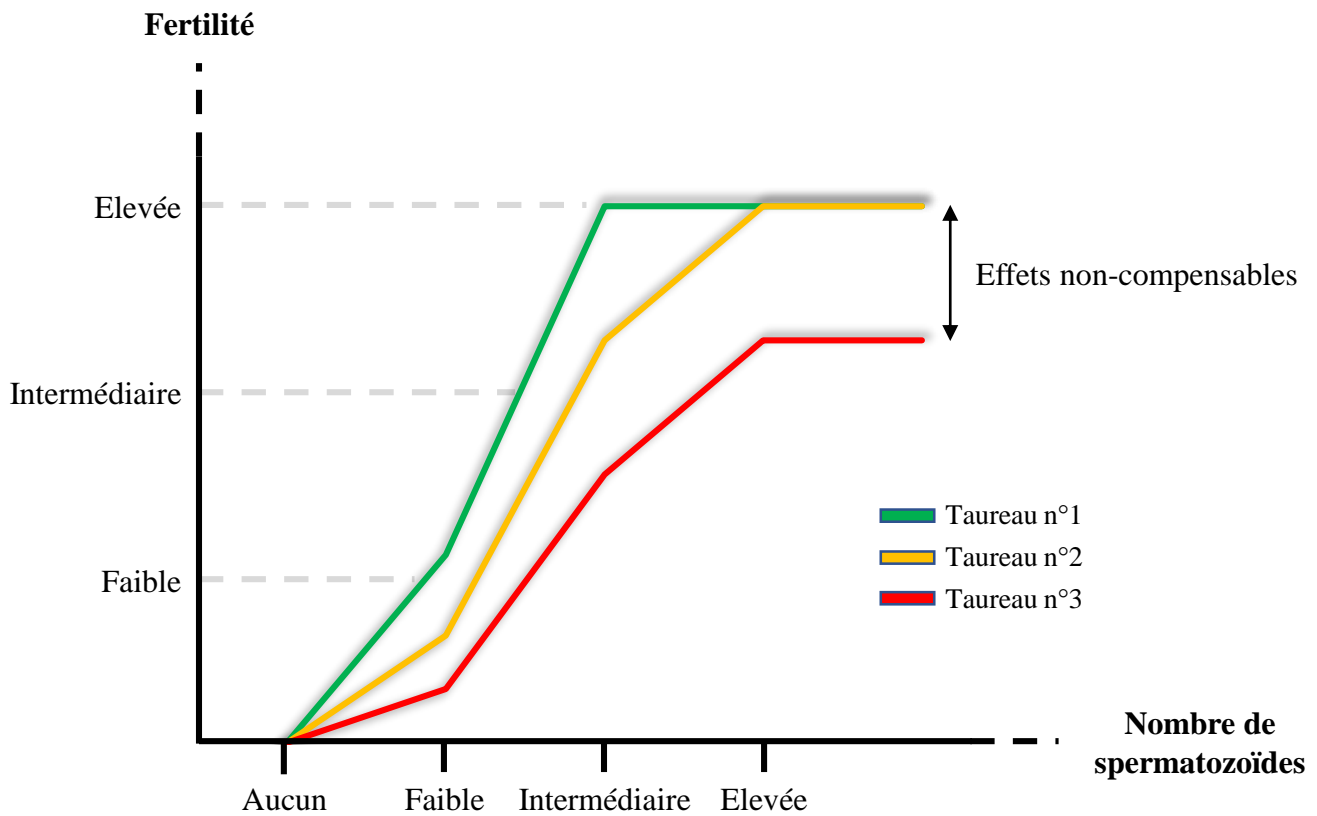


Figure 22. Relation entre le nombre de spermatozoïdes par paillette et la fertilité des taureaux

Les taureaux, numérotés de 1 à 3, contiennent différentes proportions de spermatozoïdes avec des effets compensables et non-compensables. Le taureau n°2 présente un nombre plus important de spermatozoïdes à effets compensables que le taureau n°1. Il est alors nécessaire d'augmenter le nombre de spermatozoïdes par paillette pour atteindre le plateau de fertilité. Le taureau n°3 ne pourra jamais atteindre le plateau, même en augmentant le nombre de spermatozoïdes par paillette. Son éjaculat présente en effet une proportion importante de spermatozoïdes à effets non-compensables.

(adapté de Salisbury et VanDemark, 1961)

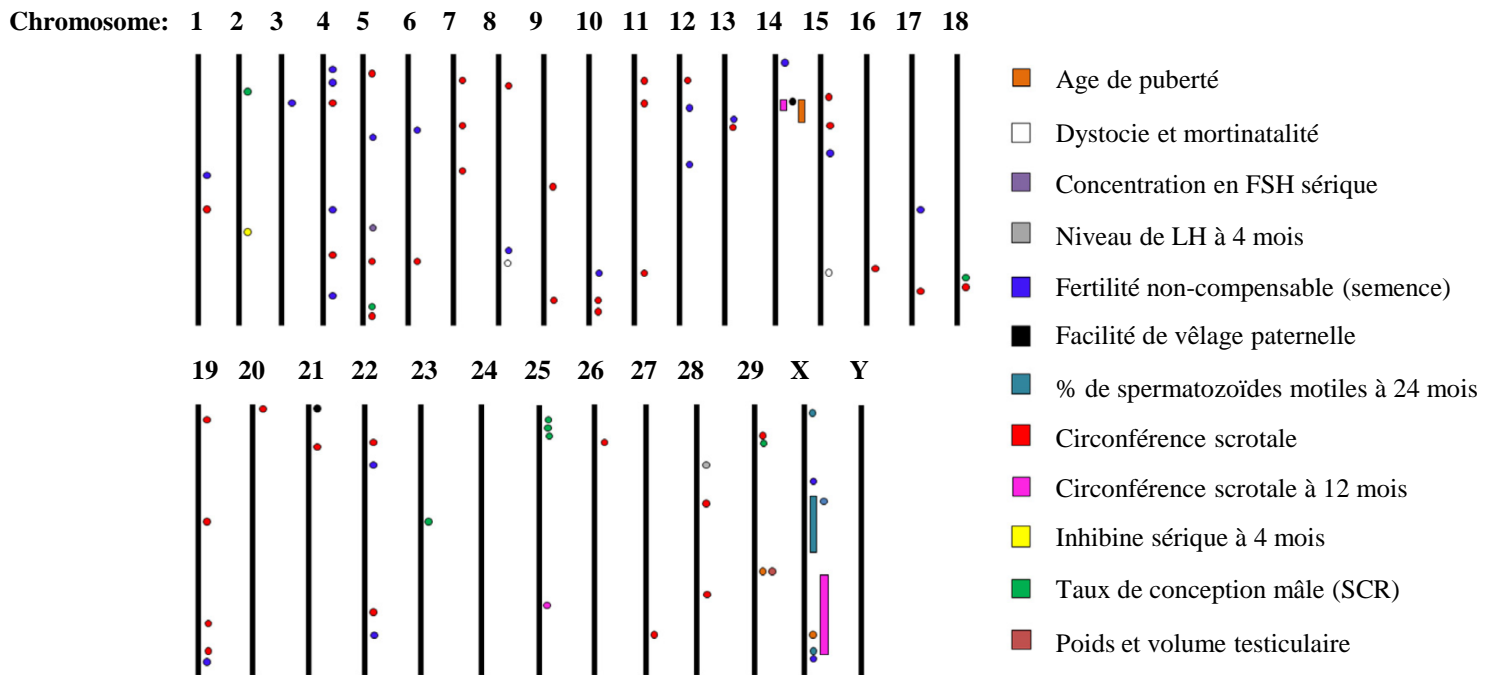


Figure 23. Représentation schématique des QTL associés à la fertilité mâle bovine, chromosome par chromosome

(Tiré de Fortes et al. 2013)

3.2. Effets compensables et non-compensables

Quand on parle de fertilité, il faut savoir différencier les effets compensables des effets non-compensables (Saacke 2008). Chez certains taureaux, malgré le fait qu'une part importante des spermatozoïdes possède des anomalies morphologiques mineures ou une mauvaise motilité, une fraction d'entre eux reste parfaitement fécondante. La qualité médiocre de la semence peut être alors palliée simplement en concentrant les spermatozoïdes dans la paille (Figure 22). Il a été montré que ce procédé augmentait le TNR jusqu'à atteindre un palier (Salisbury et al. 1961) (Den Daas et al. 1998). Les ES ont souvent recours à cette méthode pour pouvoir utiliser des taureaux de haute valeur génétique.

Paradoxalement, certaines études montrent une corrélation négative entre le nombre de spermatozoïdes par paille et le TNR (Amann et DeJarnette 2012). En fait, certains effets impactant la fertilité mâle ne sont pas compensables par une augmentation de la concentration des spermatozoïdes dans les pailles. Au contraire des effets compensables qui semblent avoir une implication avant la fécondation, les effets non-compensables semblent jouer après la fécondation. Les spermatozoïdes présentent alors des défauts d'intégrité de l'ADN, du ratio histone/protamine, ou des mutations génétiques délétères et n'ont pas la capacité de conduire un développement embryonnaire normal. Les ES vont avoir tendance à augmenter le nombre de spermatozoïdes par paille pour les taureaux sub-fertiles, ayant pour conséquence cette corrélation négative entre concentration en spermatozoïdes et TNR, car ces individus présentent des éjaculats à fort taux d'effets non-compensables.

3.3. L'évaluation génétique ne permet pas de prédire la sub-fertilité

En sélection génomique, les reproducteurs sont identifiés peu après leur naissance d'après l'analyse de leur génotype. Depuis plus d'une décennie, des tentatives d'identification de *Quantitative Trait Loci* (QTL), les *Genome-Wide Association Studies* (GWAS) ou les analyses en gène-candidat ont permis d'identifier des régions génomiques ou des gènes associés à la fertilité. En 2013, Fortes et collègues ont publié une revue listant les QTL et les régions génomiques associées à la fertilité bovine mâle (Fortes et al. 2013) (Figure 23). En 2014, Hering a publié des résultats provenant d'études en GWAS et a identifié 34 SNP significativement associés avec la motilité spermatique, et 3 significativement associés au volume de l'éjaculat et au nombre de spermatozoïdes dans l'éjaculat (Hering, Olenski, et Kaminski 2014) (Hering et al. 2014). La prise en compte de ces données pour la sélection

Type de paramètre	Sous-catégorie	Minimum	Maximum	Héritabilité moyenne
Phénotypique	SC	0,36	0,75	0,420 ± 0,0002
Spermogramme	Volume de l'éjaculat	0,04	0,65	0,197 ± 0,019
	Nombre de spermatozoïdes	0,03	0,54	0,222 ± 0,018
	Concentration spermatique	0,10	0,56	0,169 ± 0,028
	Pourcentage de spermatozoïdes viables	0,00	0,22	0,128 ± 0,044
	Motilité	0,01	0,51	0,054 ± 0,009
	Ratio spermatozoïdes anormaux	0,07	0,35	0,194 ± 0,029
Fertilité	TNR	0	0,002	0,001 ± 0,001

Tableau 3. Héritabilité des paramètres spermatiques et de fertilité bovine
(tiré de Berry et al. 2014)

génomique permettrait d'identifier précocement les individus fertiles ou non, et aussi d'améliorer la fertilité des élevages en prenant la fertilité comme critère de sélection. En effet, si l'héritabilité d'un caractère est bonne, alors le caractère sera susceptible d'être transmis à la génération suivante. L'héritabilité est définie comme la proportion de variation phénotypique (après de multiples ajustements des effets de l'environnement) attribuable à des différences génétiques entre les individus. Selon les caractères, la part de la variabilité due à la génétique peut varier énormément. Pour être précise sa mesure doit être faite sur une très grande population de référence (Hayes et al. 2009). Cela prend beaucoup de temps car il y a peu de reproducteurs disponibles chaque année pour alimenter la population de référence.

Par exemple, certains caractères phénotypiques liés à la performance reproductive des femelles sont héréditaires et corrélés à la fertilité, comme par exemple le BCS (*Body Condition Score*). Les paramètres de fertilité femelle sont aujourd'hui pris en considération et évalués dans le cadre de la sélection génomique (Berry et al. 2003). Moins d'études ont été réalisées concernant la fertilité mâle. L'héritabilité des caractères phénotypiques associés à la production de semence a été évaluée moyennement à hautement héréditaire (Cammack et al. 2009), notamment la CS qui possède une hérédabilité estimée de 0,42 (Barth 2007).

La sélection génomique peut prendre en compte les caractères associés à la qualité de la semence dans le calcul de performance d'un candidat (Tableau 3). Ces caractères sont cependant très peu héréditaires : par exemple de 0,05 pour la motilité, 0,23 pour les anomalies morphologiques (Berry, Wall, et Pryce 2014). De la même façon, l'héritabilité pour le TNR ou le SCR est de l'ordre de 0,001 et 0,08, respectivement (Kuhn et Hutchison 2008). Ce sont donc des hérédibilités extrêmement faibles. Cela montre que les marqueurs de la fertilité mâle n'expliquent qu'une fraction de la fertilité, et que ces marqueurs ne peuvent aujourd'hui être utilisés pour améliorer la fertilité générale des élevages par la sélection génomique.

Aujourd'hui, l'estimation de la fertilité est basée précocement sur les données génotypiques des individus, puis à la puberté sur le spermogramme. Or, nous avons vu que ces mesures ne permettent pas d'écarter totalement les individus sub-fertiles. Alors qu'ils pouvaient être identifiés pendant la phase de testage sur descendance, avec l'avènement de la sélection génomique les taureaux sub-fertiles sont maintenant diffusés sans retour sur leur fertilité terrain. Dans ce contexte, il est nécessaire de développer de nouveaux critères pour qualifier la semence. Au laboratoire, l'objectif est d'utiliser la méthylation de l'ADN comme biomarqueur épigénétique de la qualité de la semence, et plus particulièrement d'analyser la méthylation de l'ADN et de corréliser altérations de la méthylation et baisse de la fertilité.

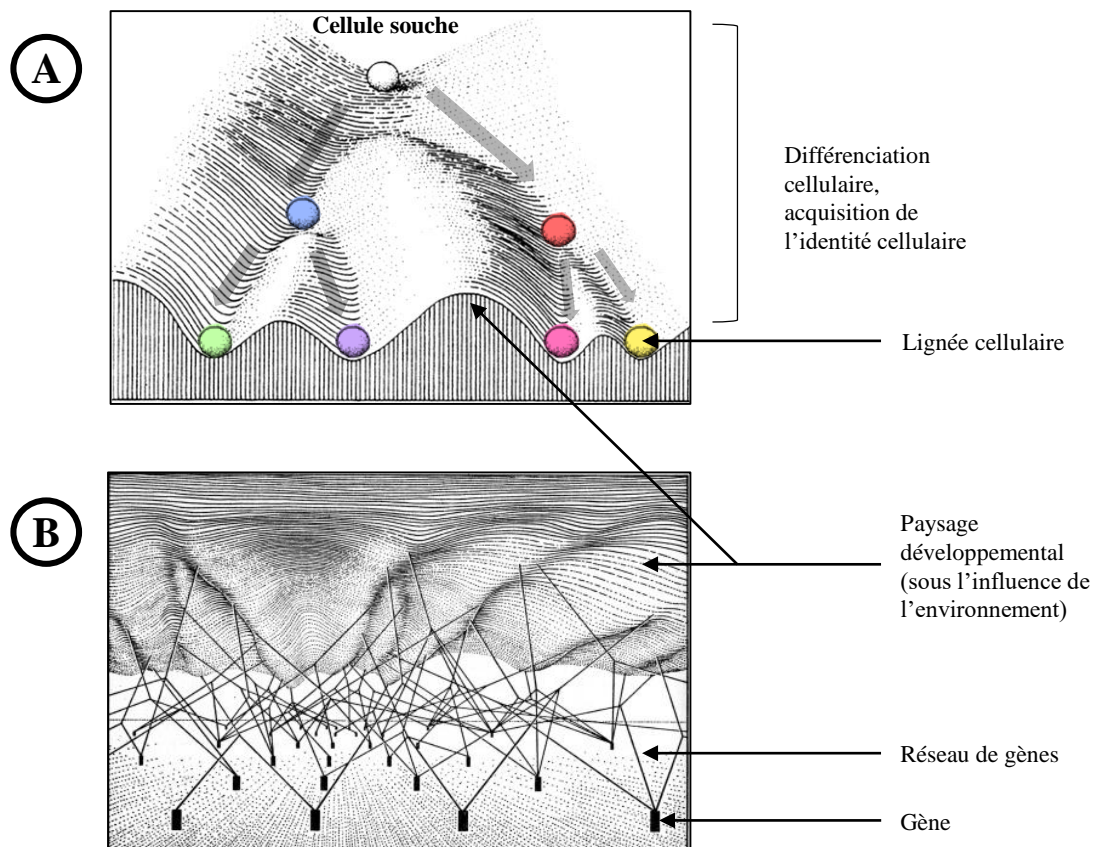


Figure 24. Le paysage développemental selon Conrad Waddington (1942)

A. La boule blanche au sommet du paysage schématise une cellule souche totipotente. Au gré des divisions cellulaires et du développement, les cellules acquièrent une identité et se différencient (boules bleues et rouges, qui elles-mêmes se différencieront en boules verte, violette, rose et jaune). Selon cette représentation, la différenciation est un processus linéaire dans le temps, sans possibilité de retour en arrière. Ceci a été démenti depuis par le clonage somatique (passage d'un noyau de cellule différenciée à un noyau de cellule totipotente) et l'existence des IPS, cellules pluripotentes induites. La fécondation elle-même sous entend une reprogrammation puisque les gamètes mâles et femelles sont des cellules très différenciées, leur génome subit une phase de reprogrammation permettant la formation d'un noyau zygotique aux propriétés de totipotence. **B.** Le paysage est sous-tendu par des réseaux de gènes. L'environnement extérieur peut aussi avoir un impact sur le paysage et le modifier, ce qui va avoir un impact sur le réseau de gènes sous-jacent et en modifier l'expression. De cette façon, on résume souvent le phénotype d'un individu à la résultante de son génotype et de l'environnement, l'épigénétique étant à l'interface entre les 2. (adapté de Waddington, 1957)

Chapitre III : L'épigénétique de la fertilité mâle

L'épigénétique est définie comme l'étude de l'ensemble des modifications structurales de la chromatine ayant une influence sur l'expression des gènes sans modification directe de la séquence nucléotidique de l'ADN (Bird 2007). L'état épigénétique général d'une cellule ou d'un type cellulaire donné, à un temps t , est ce qu'on appelle son épigénome. Le concept de l'épigénétique a été pour la première fois formulé par Conrad Waddington en 1942 (Figure 24). Il postula alors qu'entre le génotype d'un individu et son phénotype, il devait exister une interface qu'il proposa de nommer « épigénotype ». Alors que toutes les cellules de l'organisme sont porteuses du même patrimoine génétique, chaque type cellulaire n'exploite qu'une fraction de l'ensemble de l'information contenue dans son génome. C'est l'ensemble des mécanismes épigénétiques qui permettent de maintenir l'identité cellulaire à travers un profil d'expression de gènes précis. Les mécanismes épigénétiques reposent principalement sur un ensemble de modifications chimiques covalentes apposées sur la chromatine ; ce qui leur confère un caractère héritable au cours des divisions cellulaires.

Trois grands mécanismes épigénétiques sont généralement décrits : les modifications post-traductionnelles (MPT) des histones, l'intervention des ARN non-codants et la méthylation de l'ADN (Figure 25). Ces mécanismes agissent sur l'expression génique en influençant notamment le degré de compaction de la chromatine. L'apposition de marques épigénétiques dites répressives induit une compaction de la chromatine au niveau des régions régulatrices de la transcription, bloquant la fixation des facteurs de transcription (*Transcription Factors*, TF) et en conséquence empêchant la transcription des gènes associés. La chromatine compactée est alors appelée hétérochromatine et se retrouve souvent dans les régions pauvres en gènes et les séquences répétées. Au contraire, l'apposition de marques épigénétiques permissives conduit à un état de relâchement de la chromatine. La transcription des gènes associés est alors possible. Cette configuration chromatiniennne de type ouvert caractérise l'euchromatine, principalement localisée dans des régions riches en gènes du génome (Figure 26).

L'épigénome caractéristique des spermatozoïdes est acquis progressivement au cours de la spermatogenèse, et permet l'extrême compaction de leur génome. Dans les spermatozoïdes matures, la transcription des gènes est arrêtée. Dans ce contexte, des perturbations au cours de la différenciation des gonades fœtales ou des différentes étapes de la spermatogenèse peuvent conduire à un épigénome spermatique altéré, et ayant potentiellement des répercussions sur la

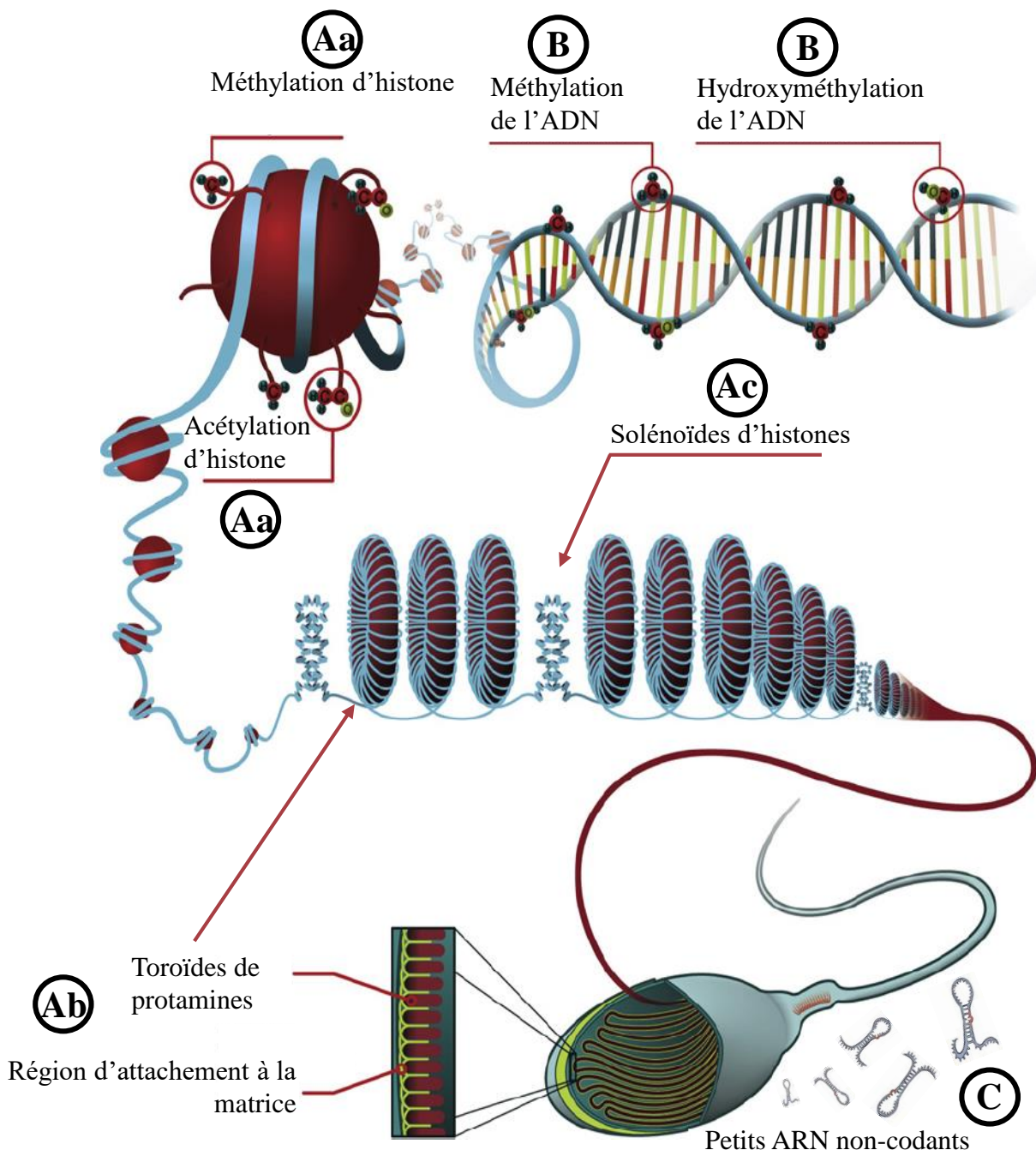


Figure 25. Principales marques épigénétiques impliquées dans l'architecture de la chromatine de spermatozoïde.

Au sein des cellules, le noyau contient l'information génétique portée par l'ADN. **Aa**. L'ADN s'enroule autour d'octamères d'histones, pour former la chromatine. Sur les résidus libres des histones (queues des histones) sont apposées des modifications post-traductionnelles, dont les plus étudiées sont les acétylations et les méthylation. **Ab**. Dans les spermatozoïdes, les histones sont remplacées au cours du temps par des protamines, qui s'organisent pour former des macrostructures toroïdales. Les toroïdes sont fixés à la membrane nucléaire par les régions d'attachement à la matrice. **Ac**. Les histones restantes (de 5 à 15% chez le bovin), sont retrouvées sur certains loci spécifiques et organisent la chromatine sous forme de solénoïdes. **B**. Des groupements méthyles et hydroxyméthylse sont liés de manière covalente à l'ADN au niveau des sites CpG. **C**. Enfin, il a été montré que la tête spermatique ne contient pas seulement l'ADN mais aussi des ARN non-codants, qui seront transmis à l'ovocyte à la fécondation. Ces petits ARN non-codants sont également présents dans les cellules somatiques pour réguler l'expression des gènes au niveau post-traductionnel. L'ensemble de ces marques épigénétiques, regroupées sous le terme d'épigénome, participe à la régulation transcriptionnelle. Dans le cas des gamètes, ces marques constituent aussi une part d'information transmise à la descendance au moment de la fécondation et leur nature et leur distribution génomique jouent un rôle important dans le bon développement embryonnaire. (adapté de Carrell et al, 2012)

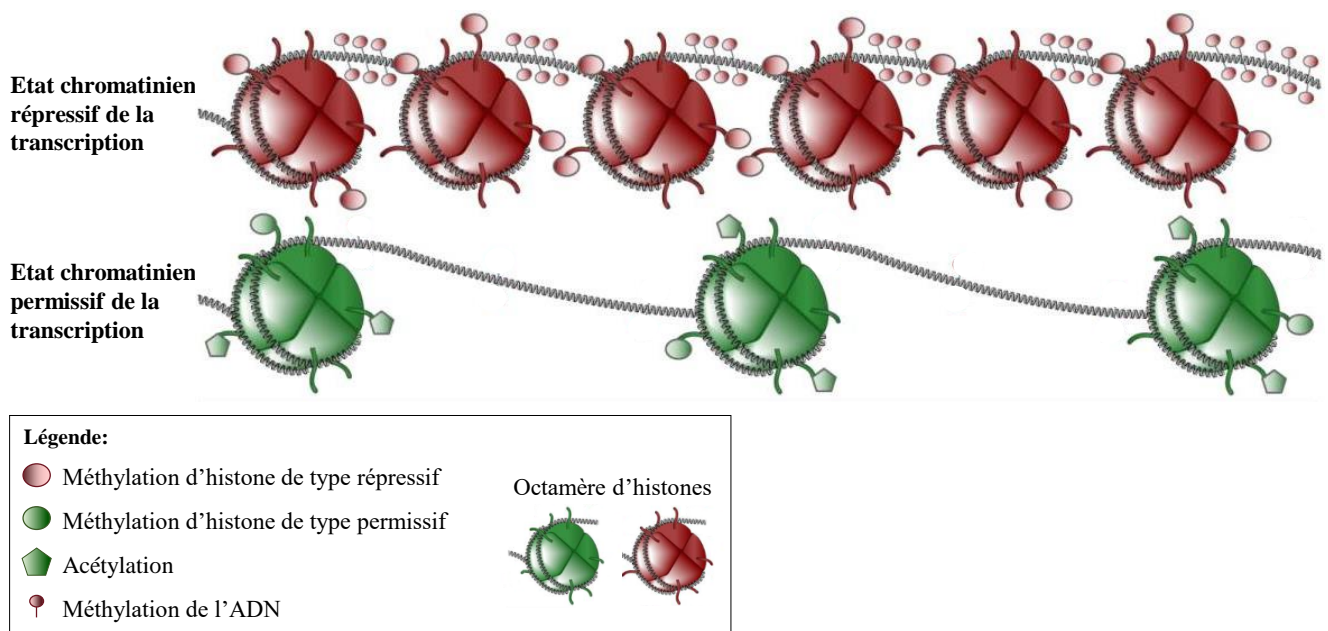


Figure 26. Etats chromatinien et régulation de la transcription

Dans les cellules somatiques, l'ADN est enroulé autour des octamères d'histones. Les marques épigénétique sont apposées directement sur l'ADN (méthylation de l'ADN) ou sur les queues des histones (modification post-traductionnelles des histones). Une chromatine compactée ne permet pas la fixation des facteurs de transcription, empêchant la transcription des gènes de la région (hétérochromatine, octamères en rouge). Au contraire, une chromatine relâchée libère l'espace nécessaire à la fixation de ces facteurs, et permet la transcription des gènes associés (euchromatine, octamères en vert).

Pour ne pas surcharger le schéma, seule la méthylation de l'ADN entre les nucléosomes est représentée, mais il existe aussi de la méthylation sur l'ADN enroulé autour des nucléosomes (adapté de Wierda *et al*, 2012)

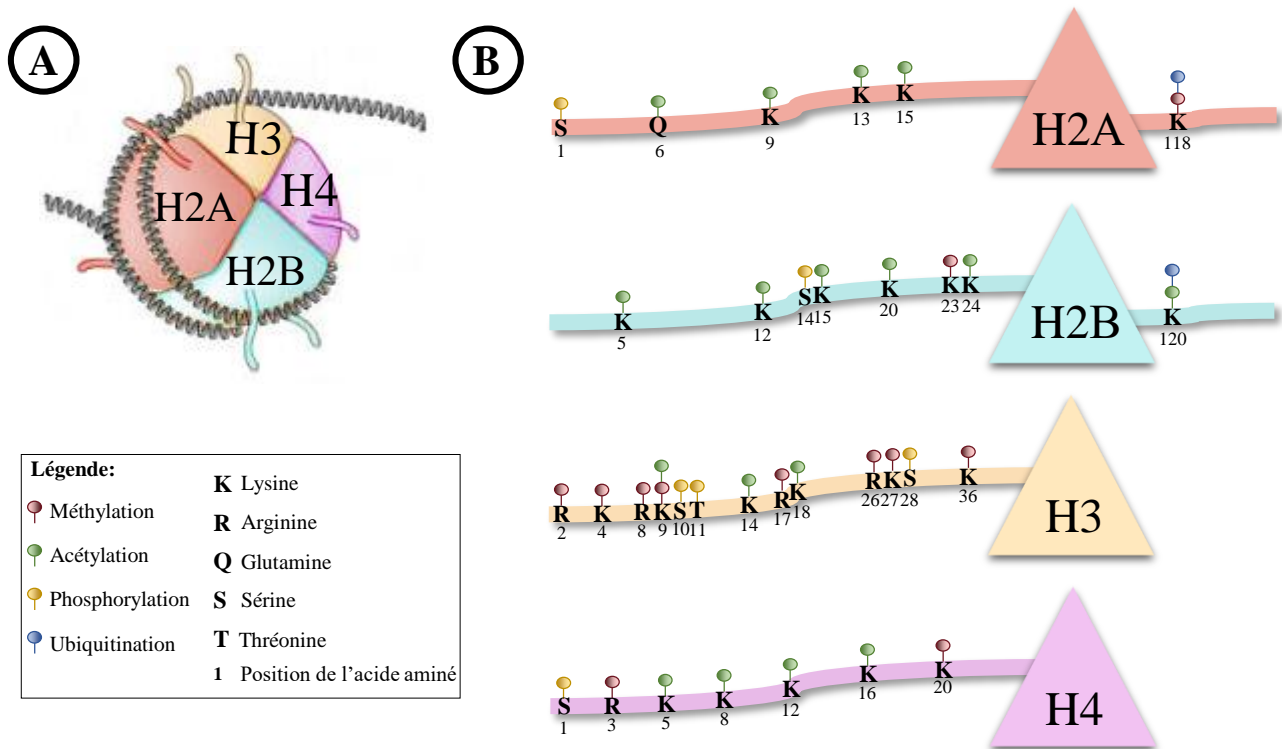


Figure 27. Les modifications post-traductionnelles des histones

A. Le nucléosome constitue l'unité de base d'organisation de la chromatine, composé par un octamère d'histones canoniques (avec 2 copies des protéines H2A, H2B, H3 et H4), autour duquel s'enroule 147 paires de base d'ADN. Chaque histone possède une ou des chaînes protéiques appelées queues d'histones, et qui font saillie vers l'extérieur du nucléosome. **B.** Représentation des 4 types d'histones canoniques, avec leurs queues protéiques sur lesquelles peuvent s'apposer des MPT. Sont alors notés le type de modification, le type et la position de l'acide aminé sur lesquels peuvent s'apposer chaque marque. Un même acide aminé est susceptible d'être modifié par plusieurs types de marques différentes, ou être marqué plusieurs fois de la même marque (non-représenté). Par exemple, la lysine n°4 de l'histone H3 (H3K4) peut recevoir une, deux, ou trois marques de méthylation (noté respectivement H3K4me, H3K4me2 et H3K4me3).

fertilité. Ces perturbations peuvent même être transmises dans certaines conditions aux descendants.

Après une description des principaux mécanismes épigénétiques, l'accent sera mis sur la méthylation de l'ADN, qui est la marque épigénétique faisant l'objet du travail de thèse. Sont ensuite successivement abordés la répartition génomique des sites méthylables, la régulation de la méthylation, les interactions entre la méthylation et les autres marques épigénétiques, et enfin les nombreux rôles biologiques de la méthylation de l'ADN. La reprogrammation épigénétique du génome, élément clef du processus de la reproduction animale, est détaillée dans un troisième temps. Enfin, nous aborderont l'état des connaissances scientifiques sur le lien qui peut exister entre les altérations de l'épigénome et l'infertilité ou la sub-fertilité mâle, qui a été mis en évidence chez de nombreuses espèces comme l'Homme, la souris ou encore le bovin.

1. Les différentes marques épigénétiques

1.1. Les modifications post-traductionnelles des histones

Dans les cellules, l'ADN contenu dans le noyau s'associe avec des protéines pour former la chromatine, dont l'unité structurelle est le nucléosome. Chaque nucléosome est formé par un octamère de protéines histones, soit 4 dimères d'histones H2A, H2B, H3 et H4, autour duquel s'enroule une séquence d'ADN de 147 paires de bases. Les histones H1 ne participent pas à la formation du nucléosome mais en maintiennent la structure en s'associant avec l'ADN. Certains acides aminés particuliers des chaînes protéiques N- et C- terminales des histones (ou queues d'histones) sont modifiés chimiquement par des enzymes de remodelage de la chromatine. A ce jour, des dizaines de types de MPT d'histones ont été décrites : la méthylation, l'acétylation, la phosphorylation, la carbonylation... les plus étudiées étant l'acétylation et la méthylation. L'état des connaissances à leur sujet est résumé dans les revues de Sadakierska-Chudy et Filip (Sadakierska-Chudy, Kostrzewa, et Filip 2014) (Sadakierska-Chudy et Filip 2015).

Dans un souci de clarification, les histones sont notées selon une nomenclature propre. Après le nom de l'histone en question, est noté l'abréviation et la position de l'acide aminé sur lequel est apposée la MPT, et enfin le type et le nombre de MPT. Par exemple, H3K4me3 signifie qu'il y a 3 groupements méthyle sur la lysine n°4 de l'histone H3.

Diverses MPT d'histones peuvent être apposées en même temps sur les queues d'histones, ce qui donne naissance à une combinatoire de modifications très importante (Figure 27). Le nombre de MPT et le résidu sur lequel elles sont apposées, sont responsables de l'activation ou de l'inactivation de la chromatine (i.e., son degré de compaction). De ce fait, l'effet global de cette combinatoire de MPT sur la chromatine est très complexe à évaluer. Certaines MPT sont le plus souvent retrouvées dans l'hétérochromatine et sont considérées comme des marques épigénétiques répressives, comme par exemple H3K9me3 ou H3K27me3. Inversement, d'autres sont retrouvées principalement dans l'euchromatine et donc ont été décrites comme des MPT permissives, comme par exemple H3K4me3 et H3K36me3, de même que l'acétylation des histones.

Les MPT d'histones, comme les autres types de marques épigénétiques, sont modulées à l'échelle génomique et au cours des processus biologiques par l'intermédiaire de plusieurs familles d'enzymes, que l'on regroupe sous le nom de machinerie épigénétique. Les familles d'enzymes agissant sur l'épigénome peuvent être réparties en 3 catégories suivant leur fonction. On distingue les enzymes permettant l'apposition sur le génome des marques (enzyme écrivain ou *writer*), les enzymes permettant le retrait de ces marques (enzyme effaceur ou *eraser*), et enfin les enzymes qui interprètent les marques épigénétiques afin de recruter d'autres complexes protéiques (enzyme lecteur ou *reader*). Concernant les MPT, les acétylations et les méthylation d'histones sont apposées par les enzymes de la famille des histones lysines acétyltransférases (KAT) ou des méthyltransférases (HMT et SUV39H1), respectivement. Ces mêmes marques sont retirées par d'autres familles d'enzymes, les histones désacétylases (HDAC et Sirtuines) et les histones déméthylases (HDM).

1.2. Les ARN non-codants

Une large part du génome autrefois qualifiée d'ADN poubelle a en fait une fonction biologique, voire est même transcrite en ARN. Ces ARN, regroupés sous le terme ARN non-codants, ne sont pas traduits en protéine mais ont un rôle structural dans l'architecture chromatinienne ou dans la régulation de l'expression génique. On distingue 2 sous-catégories d'ARN non-codants en fonction de leur taille. Les longs ARN non-codants (lncRNA, pour *long non-coding RNA*) sont d'une taille supérieure à 200 pb. Au contraire, les petits ARN non-codants sont d'une taille inférieure à 200 pb. Cette famille de petits ARN comprend des miRNA (microRNA) ou encore des piARN (*PIWI-interacting RNA*) et des si-RNA (*small interfering RNA*).

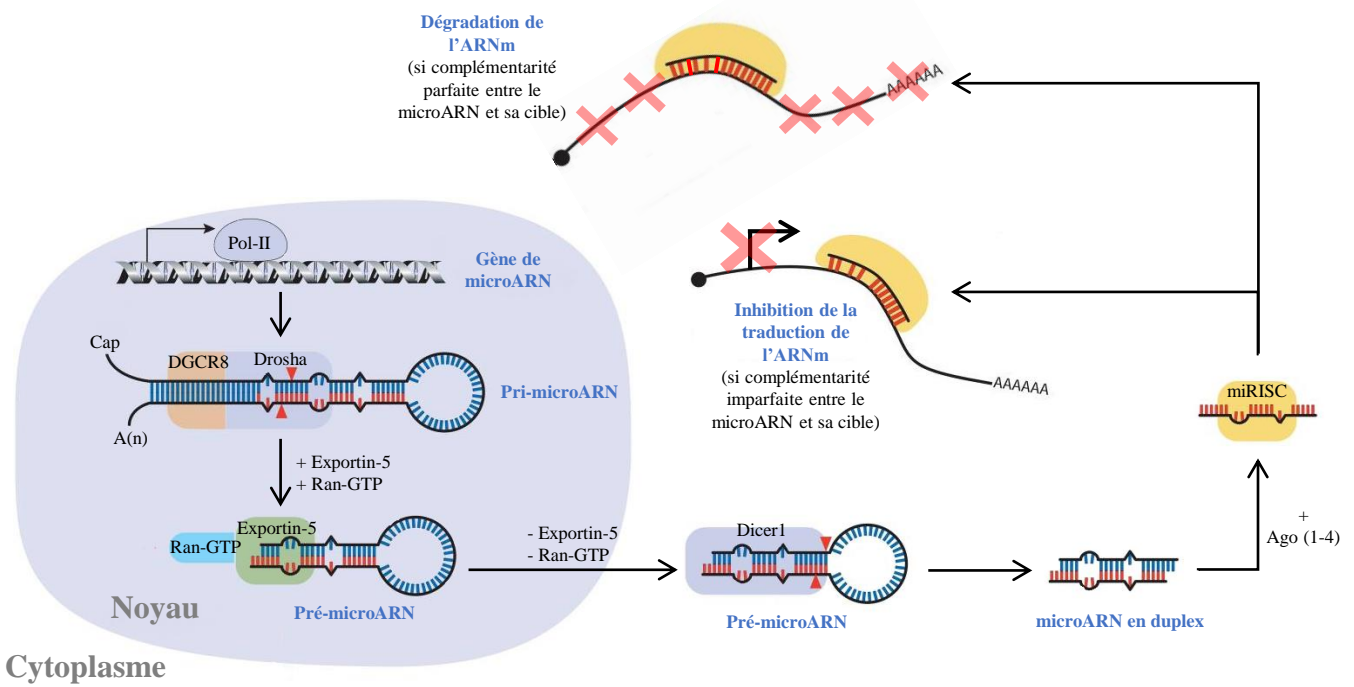


Figure 28. Modes d'action des microARN

Les précurseurs des microARN sont transcrits à partir de gènes de microARN par l'ARN polymérase II. Ceux-ci sont sous la forme d'un ARN double brin avec une boucle terminale (pri-microARN). Après avoir été transformés par le complexe DGC8/DROSHA en pré-microARN, ils sont exportés dans le cytoplasme. Ils sont alors maturés par Dicer pour former le miRNA mature qui pourra s'associer au complexe RISC, avec les protéines Argonautes 1 à 4 (Ago). Lorsque la complémentarité de séquence entre le miRNA et sa cible est parfaite, l'ARNm est dégradé. Si la complémentarité de séquence est imparfaite, il y a simplement inhibition de la traduction de l'ARNm. Dans les 2 cas, l'action du miRNA sera de diminuer l'expression de sa cible. (adapté de Luense, Carletti et Christenson 2009)

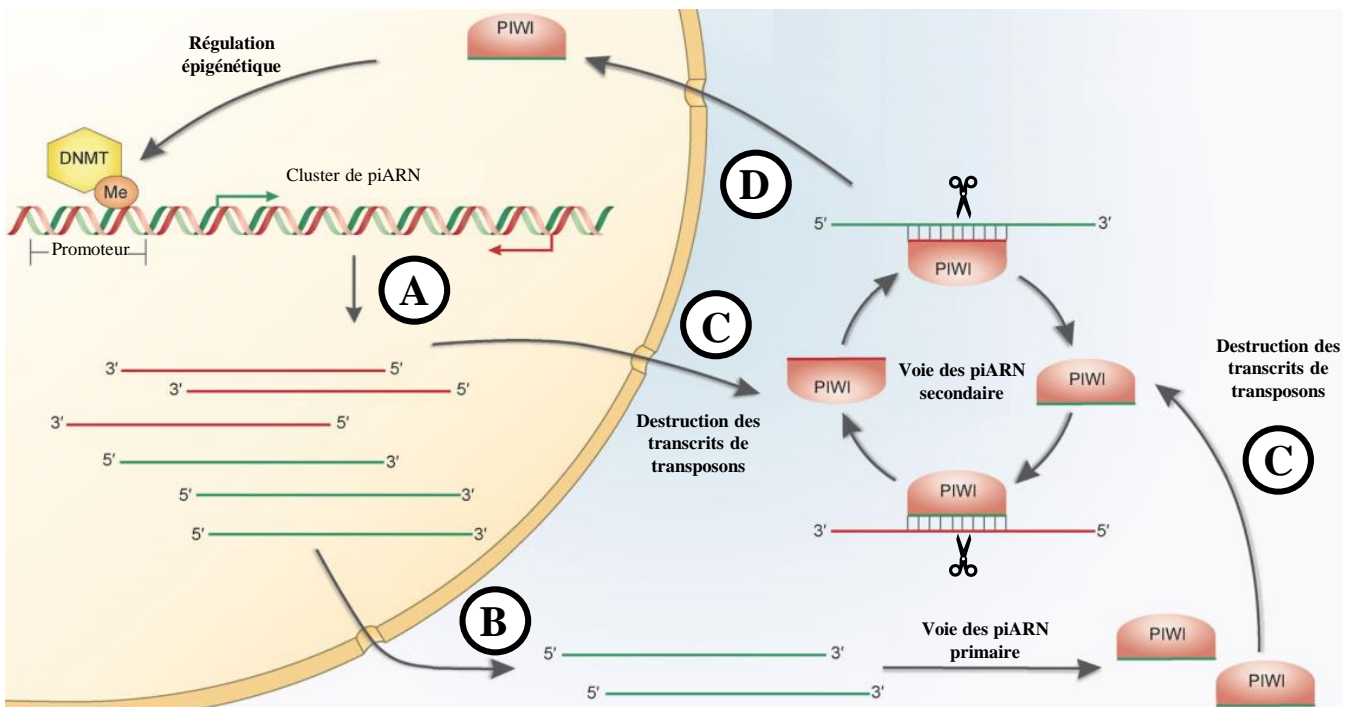


Figure 29. Mécanismes de régulation des transposons par les piARN

A. Production des piARN précurseurs sens et antisens. **B.** Les piARN précurseurs antisens sont transportés dans le cytoplasme, subissent une maturation pour générer des piARN sens, puis s'associent avec les protéines PIWI. **C.** Le complexe ainsi formé clive les précurseurs de piARN sens et les transcrits de transposons, ce qui constitue l'inactivation post-transcriptionnelle des transposons. Les transcrits de piARN sens ainsi formés sont pris en charge par les protéines PIWI. Le nouveau complexe va alors cliver les précurseurs de piARN antisens, ce qui amplifie la production des piARN. **D.** Les complexes peuvent entrer dans le noyau et, en association avec les DNMT, apposer des marques de méthylation de l'ADN sur leurs gènes cible pour en réguler la transcription (voie d'inactivation transcriptionnelle des piARN). (adapté de Ku et Lin, 2014)

Les miRNA peuvent réguler la traduction des gènes en réprimant spécifiquement la traduction d'ARNm. Les précurseurs des microARN sont transcrits par l'ARN polymérase II, puis sont maturés par DROSHA et DGCR8 dans le noyau, puis par DICER dans le cytoplasme. Ils sont ensuite incorporés dans le *RNA-Induced Silencing Complex* (RISC) pour assurer leur fonction (Figure 28). Les ARNm cibles sont alors soit inactivés post-transcriptionnellement, soit dégradés, en fonction du degré d'appariement des miRNA avec les ARNm. Il a été montré que l'inactivation de DICER ou de DROSHA conduit à un arrêt de la spermatogenèse au stade spermatogonie (Hayashi et al. 2008).

Les siRNA sont impliqués dans le contrôle de la mobilité des transposons ; s'ils présentent une complémentarité parfaite avec leur cible ils induisent une inhibition de la transcription par modification de la chromatine ou induisent un clivage de l'ARNm cible.

Les piRNA (Figure 29) sont quant à eux impliqués dans la répression des éléments transposables (Ku et Lin 2014). D'abord identifiés et étudiés chez la Drosophile, les piRNA ont principalement 2 modes d'action. Le premier mode d'action permet de prévenir l'expression des éléments transposables en dirigeant leur inactivation transcriptionnelle. En effet, les piRNA guident la protéine nucléaire MIWI2 vers les transposons ; se forme alors un complexe protéique permettant l'apposition de MPT répressives d'histones associée à une méthylation *de novo* de l'ADN médiée par DNMT3a/b/l. Le second mode d'action des piRNA est l'inactivation post-transcriptionnelle des transposons. Ce mécanisme se fait par le biais du cycle d'amplification en ping-pong des piRNA, assurant la génération d'un nombre important de piRNA pour contrer les éléments transposables très actifs du génome.

On sait aujourd'hui qu'une partie des ARN non-codants est issue de la transcription d'éléments répétés du génome. Par exemple, les séquences satellites (ADNsat) sont principalement situées autour des centromères et en bordure des télomères. Les ADNsat péri-centromériques sont transcrits au cours de l'embryogenèse précoce et participent à la stabilité de la formation des chromocentres (Probst et al. 2010). De la même façon, la transcription de l'élément transposable LINE1 (*Long Interspersed Nuclear Element 1*) est impliquée dans le développement embryonnaire précoce. En effet, l'interruption de son activité entraîne un arrêt embryonnaire au stade 2 à 4 cellules chez la souris (Beraldi et al. 2006). Il a été montré que des LTR (*Long Terminal Repeat*), une classe de rétrotransposons, sont également exprimés, cette fois pendant l'EGA chez la souris et le bovin (Bui et al. 2009).

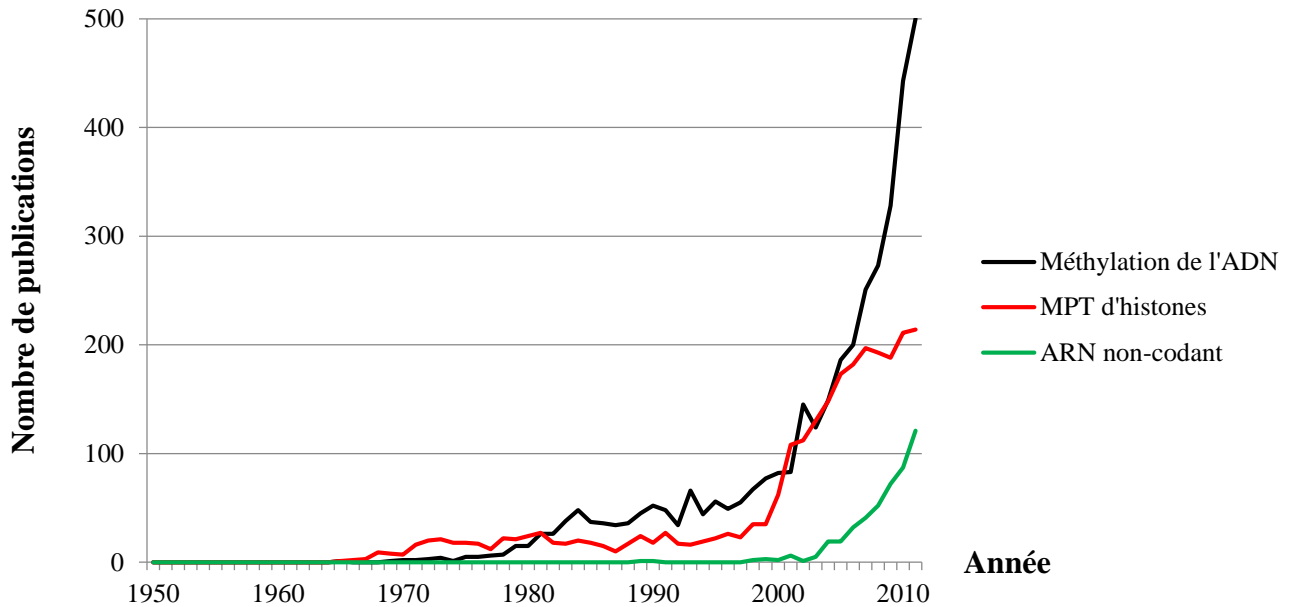


Figure 30. La méthylation de l'ADN est la marque épigénétique la plus étudiée

Le graphique représente le nombre de publications contenant les termes associés à méthylation de l'ADN, modifications post-traductionnelles des histones ou ARN non-codants, en fonction de l'année de publication. Le nombre de publications en ordonnée est normalisé en divisant le résultat par le nombre total de publications de l'année, pour corriger l'effet de l'augmentation générale du nombre de publications publiées chaque année. On observe que la méthylation de l'ADN est le terme le plus cité dans les publications en 2010, devant les MPT d'histones et les ARN non-codants.

(les données proviennent de la base de donnée Medline, traitées avec l'application MLTrends (Palidwor et Andrade-Navaroo 2010) (termes utilisés pour la recherche: DNA methylation ; histone AND (acetylation methylation phosphorylation sumoylation glycosylation carbonylation ubiquitination ubiquitylation citrullination deimination biotinylation ribosylation crotonylation isomerization formylation propionylation butyrylation); non-coding RNA, ncRNA)

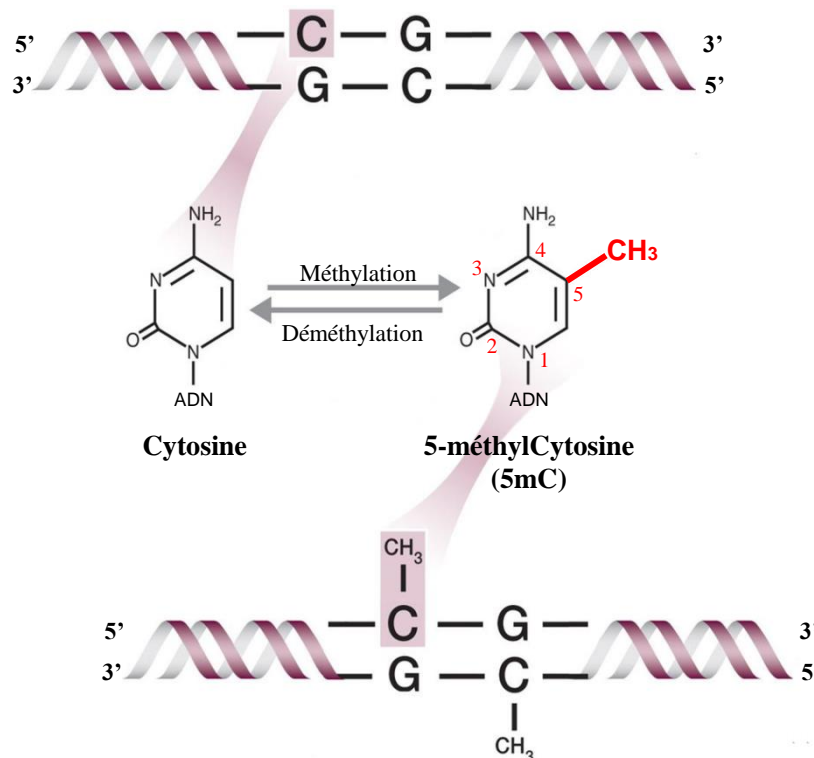


Figure 31. La méthylation de l'ADN

Les sites CpG sont des cytosines suivies d'une guanine dans le génome. Les cytosines des sites CpG sont susceptibles d'être chimiquement modifiées pour former les 5-méthylcytosines (5mC). Cette modification consiste en l'apposition d'un groupement méthyle (CH₃) sur le carbone en position 5 de la cytosine. Par nature les sites CpG sont des sites palindromiques, car dans le génome une guanine fait toujours face à une cytosine sur le brin antisens, et inversement. Chez les mammifères, les cytosines du brin sens et du brin antisens sont toutes 2 méthylées ou non-méthylées sur les sites CpG. L'apposition et l'élimination des marques de méthylation, mettent en jeu des machineries enzymatiques spécifiques.

(adapté de Olivia May, <https://www.caymanchem.com/article/2153>)

1.3. La méthylation de l'ADN

Dès 1948, Rollin Hotchkiss fut le premier à découvrir que des cytosines pouvaient être modifiées chimiquement. L'implication de ces modifications dans des processus biologiques n'a été cependant démontrée qu'à partir des années 1970 (Holliday et Pugh 1975). Jusqu'à présent, la méthylation de l'ADN est la marque épigénétique la plus étudiée (Figure 30). Son implication dans de nombreux processus biologiques fait que la cytosine méthylée est souvent décrite comme étant la cinquième base nucléotidique du génome. Il existe également de la méthylation des ARN, qui ne sera pas abordée ici.

La méthylation de l'ADN consiste en la liaison covalente d'un groupement méthyle sur le carbone 5' des cytosines (Figure 31). Les cytosines méthylées sont alors notées 5mC, pour 5-méthylCytosine. Dans le règne animal, et dans la grande majorité des cas, les 5mC sont trouvées dans un contexte de dinucléotides CpG. On note CpG les cytosines directement suivies par une guanine dans la séquence génomique (le p de CpG représentant la liaison phosphate entre les 2 nucléotides). Tout comme le fait que le terme épigénome fait référence à l'état épigénétique général d'un génome, le terme méthylome de l'ADN définit l'état de méthylation général d'un génome dans une cellule à un temps t.

Chez les plantes, les champignons, et dans certains cas plus rares dans le règne animal, les 5mC peuvent être trouvées dans un contexte autre que CpG (Ramsahoye et al. 2000). Cela a notamment été révélé grâce aux nouvelles techniques d'analyse de la méthylation par séquençage à haut-débit (qui sont explicitées en détail dans le chapitre Méthodologie). Dans le règne animal, les CpH (un C suivi d'un A, T ou C) ont surtout été étudiés dans les cellules souches embryonnaires et les cellules pluripotentes induites (Lister et al. 2009), les cellules nerveuses (Lister et al. 2013), ainsi que dans les cellules germinales mâles. Cependant, il semblerait qu'une portion substantielle de méthylation en contexte CpH soit présente dans tous les types cellulaires (de 0 à 1%), comme cela a été montré chez l'Homme (Schultz et al. 2015), et que cette méthylation est enrichie sur certains motifs, comme le motif TNCAC (où N représente n'importe quelle base). Parmi les sites CpH, la méthylation des sites CpA est la plus fréquente après celle des sites CpG chez l'Homme et la souris (Ziller et al. 2011).

Des groupements méthyles peuvent aussi être apposés sur des bases nucléotidiques autres que cytosines. Par exemple, la N6-méthyladénine (N6mA), retrouvée de manière prépondérante chez les procaryotes et chez certains eucaryotes, a été identifiée dans les cellules souches embryonnaires de souris. La N6mA est surtout présente en région intergénique et est associée avec le variant d'histone non-canonique H2A.X. Notamment, N6mA serait impliquée dans la

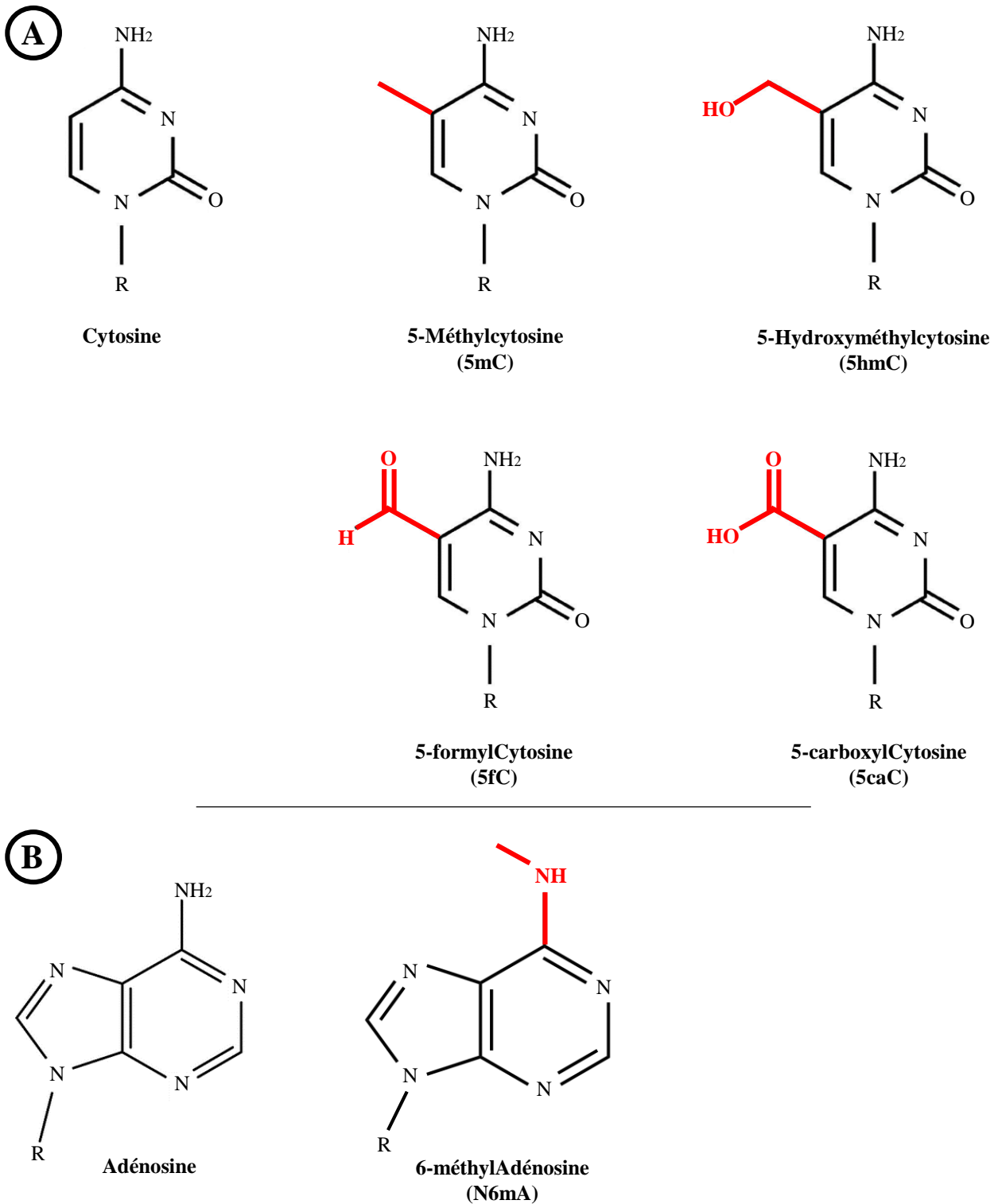


Figure 32. Les diverses formes principales de méthylation de l'ADN

A, Les formes de méthylation de l'ADN principalement décrites chez les mammifères sont les modifications sur le carbone 5 des bases nucléotidiques de type cytosine B. Il existe aussi des phénomènes de méthylation de l'ADN sur les bases nucléotidiques de type adénosine, et dont l'une des principale est la 6-méthylAdénosine (N6mA).

régulation transcriptionnelle, et en particulier pour les gènes du chromosome X. Cette méthylation jouerait un rôle dans l'inactivation des gènes et des transposons LINE1, notamment ceux qui sont les plus récents dans l'évolution (Wu et al. 2016). Il est possible que d'autres types de méthylation existent chez les eucaryotes, comme la N1-méthyladénine (N1mA), la N3-méthyleadénine (N3mA) ou encore la N3-méthylcytosine (N3mC) (Sedgwick 2004). Hormis la méthylation, il existe d'autres types de marques qui peuvent être retrouvées sur les cytosines, par exemple la 5-hydroxyméthylcytosine (5hmC), la 5-carboxylcytosine (5caC) et la 5-formylcytosine (5fC) (Figure 32). Ces marques de méthylation interviennent notamment lors du processus de retrait des marques de méthylation (qu'on qualifie de processus de déméthylation), mais certaines d'entre elles sembleraient avoir des fonctions biologiques à part entière.

1.3.1. Répartition génomique des sites CpG

De manière générale, les génomes des mammifères sont appauvris en sites CpG. On observe approximativement 4 fois moins de sites CpG que le nombre attendu par rapport au nombre de bases G et C du génome chez l'Homme (Lander et al. 2001). Cette déplétion serait le résultat du fort potentiel mutagénique des 5mC (Duncan et Miller 1980). La désamination des C en U est reconnue et réparée par la machinerie de réparation de l'ADN ; ce n'est pas le cas des désaminations spontanées de 5mC en T. Ainsi, au cours de l'évolution des mammifères de larges portions du génome se sont retrouvées très appauvries en sites CpG. Certaines régions génomiques en ont cependant conservé une forte densité. Cette conservation de régions denses en CpG pourrait donc être le fruit d'une forte pression de sélection.

On nomme les régions denses en sites CpG, les îlots CpG (*CpG Islands*, CGI) (Bird et al. 1985). Pour être considérée comme un CGI, une région doit avoir une longueur d'au moins 200 pb, contenir un taux de GC de plus de 50%, et un ratio en sites CpG observés / sites CpG attendus supérieur à 0,6 (Gardiner-Garden et Frommer 1987). Il s'agit de la définition la plus courante. Il existe cependant des algorithmes de prédiction basés sur d'autres paramètres faisant considérablement varier le nombre de CGI estimés (Tableau 4). En fonction de la définition et de l'algorithme utilisé, il existe chez l'Homme entre 350 000 et 25 000 CGI (Glass et al. 2007), et environ 23 000 chez la souris (Illingworth et al. 2010). Plus récemment, il est apparu que les régions qui encadrent les CGI à respectivement ± 2 kb et ± 4 kb ont aussi une importance dans la régulation génomique : on les appelle les « rivages CpG » (*CpG shores*), et les « plateaux CpG » (*CpG shelves*) (Figure 33, page suivante). Les *CpG shores* et

Base de donnée / Prédiction	Longueur (pb)	G + C (%)	Ratio CpG observés/attendus	Repeat Masked	Nombre de CGI	Références
ENSEMBL	≥ 400	≥ 50	≥ 0,6	Non	/	Birney <i>et al.</i> , 2004
NCBI peu stringeant	≥ 200	≥ 50	≥ 0,6	Non	307 193	/
NCBI stringeant	≥ 500	≥ 50	≥ 0,6	Non	24 163	/
USCS	> 200	≥ 50	> 0,6	Oui	28 226	Karokhik <i>et al.</i> , 2008
EMBOSS	Définie par l'utilisateur	Défini par l'utilisateur	Défini par l'utilisateur	NA	Variable	Rice <i>et al.</i> , 2000
CpGProD	> 500	> 50	> 0,6	Oui	76 793	Ponger <i>et al.</i> , 2002
CpGCluster	NA	NA	NA	Non	197 727	Hackenberg <i>et al.</i> , 2006

Tableau 4. Principaux algorithmes de prédiction des îlots CpG utilisés pour analyser le génome Humain

Le nombre d'îlots CpG (CGI) prédits par les algorithmes varie considérablement en fonction des paramètres utilisés. Le terme *Repeat Masked* correspond aux « éléments répétés » dont on connaît la localisation génomique qui sont masqués et ne sont pas pris en compte pour estimer le nombre de CGI.

pb, paires de bases; NA, Non-Applicable

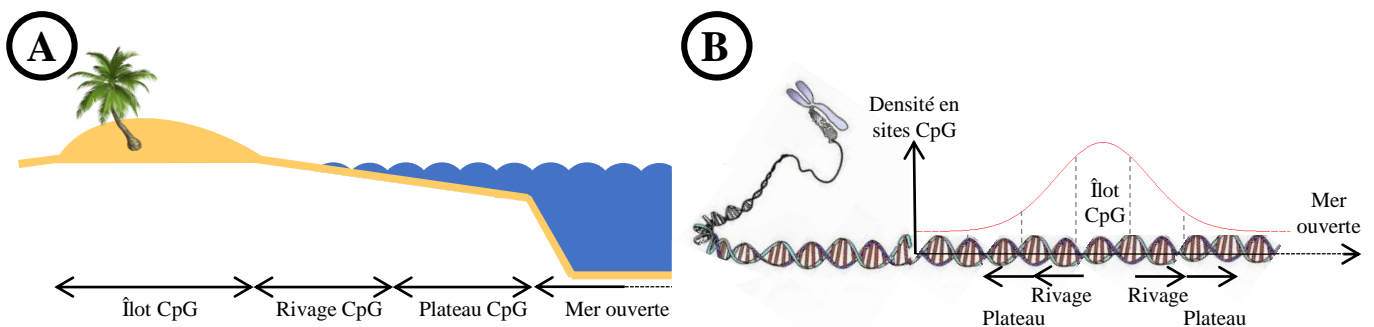


Figure 33. Concepts d'îlots CpG, de rivages CpG (CpG shore), de plateaux CpG (CpG shelf) et de mer ouverte (Open Sea)

A. Représentation schématique des appellations « d'îlot », « rivage », « plateau » et « mer ouverte »

B. Application de ce concept à la génomique. En fonction de la densité en sites CpG, on décrit traditionnellement 4 types de régions génomiques. Les îlots CpG (*CpG island*) sont les régions génomiques les plus denses en sites CpG (régions d'au moins 200pb, avec un taux de G+C > 50%, et un ratio site CpG observés / sites CpG attendus > 0,6). Autour des îlots, la densité en sites CpG reste plus importante que dans le reste du génome mais décroît avec la distance. Les rivages CpG (*CpG shore*) sont situés à ± 2kb de part et d'autre des îlots, et sont moins denses en sites CpG que les îlots, mais sont plus denses que les plateaux CpG (*CpG shelves*) qui sont eux situés à ± 2kb des rivages CpG (donc ± 4 kb des îlots). Le reste du génome, pauvre en sites CpG, est qualifié de mer ouverte (*Open sea*). Les régions les plus riches en sites CpG comme les îlots et les rivages sont retrouvés dans les éléments régulateurs de la transcription, comme les promoteurs ou les *enhancers* des gènes. On les retrouve aussi particulièrement enrichis dans les séquences répétées.

shelves sont moins denses en CpG que les CGI, mais plus que le reste du génome. Le reste du génome, quant à lui très peu dense en CpG, est qualifié de « mer ouverte » (*Open Sea*).

Leur répartition sur le génome est hautement conservée entre l'Homme et la souris, ce qui dénote leur importance biologique (Figure 34, page suivante). La moitié des CGI sont présents dans des séquences répétées, et parmi elles principalement les éléments transposables (Grandi et al. 2015). Ils sont aussi fréquemment retrouvés à ± 2 kb autour du site d'initiation de la transcription des gènes (*Transcription Start Site*, TSS), contenus dans le promoteur des gènes. Ils sont aussi particulièrement enrichis dans les exons. Chez l'Homme, 72% des gènes comportent un CGI, les 28% restant ayant une densité en CpG semblable au reste du génome (Saxonov, Berg, et Brutlag 2006). Les gènes contenant un CGI sont souvent des gènes dits de ménage (*house-keeping genes*), dont l'expression est ubiquitaire. Les gènes dont l'expression est tissu-spécifique sont plutôt associés à des promoteurs sans CGI (Glass et al. 2007).

1.3.2. Régulation de l'expression génique par la méthylation

La régulation de l'expression des gènes est en partie contrôlée par l'état de méthylation. On parle alors d'hypométhylation pour un site CpG ou une région riche en CpG donnée lorsque cette région est pauvre en méthylation pour une condition donnée par rapport à une autre condition. On parle au contraire d'hyperméthylation lorsque le site est riche en méthylation. L'effet exercé par l'hypo- ou hyperméthylation d'une région sur l'expression génique dépend de la densité en site CpG, mais aussi du contexte génomique des CpG.

Comme vu dans la partie précédente, les gènes dont le promoteur présente un ou des CGI sont principalement des gènes de ménage, dont la particularité est de présenter des profils d'expression très similaires entre les types cellulaires (Saxonov, Berg, et Brutlag 2006). Il est décrit que l'hypométhylation du promoteur et du premier exon d'un gène en favorise l'activation transcriptionnelle (Brenet et al. 2011). Au contraire, quand les CGI présents dans le promoteur d'un gène sont méthylés, la transcription du gène est réprimée. Ainsi, dans les cellules, la production de plus de 90% des ARNm est sous le contrôle de la méthylation dans les promoteurs contenant un CGI (Shen et al. 2012).

Nous avons aussi vu que près de 30% des gènes ne présentent pas de CGI dans leur promoteur, et que ce sont ces gènes qui présentent des profils d'expression tissu-spécifique. Dans ce contexte, il existe une corrélation négative forte entre la méthylation du promoteur et la transcription du gène (Gal-Yam et al. 2008). Il faut cependant noter qu'environ 30% des gènes dont le promoteur comporte un CGI présentent aussi un profil d'expression tissu-

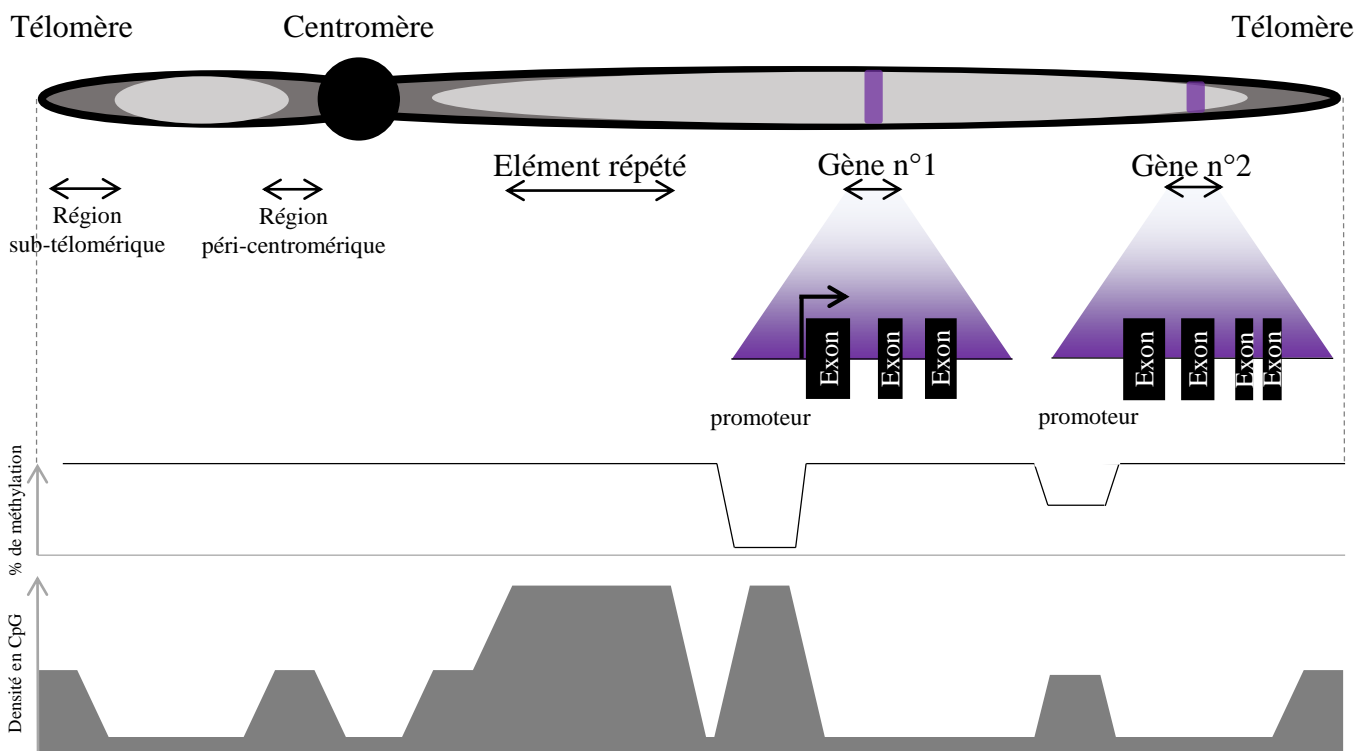


Figure 34. Répartition génomique de la méthylation de l'ADN

Représentation d'un chromosome, avec un centromère de type acrocentrique (tous les chromosomes sont acrocentriques chez le bovin). Autour du centromère sont situées les régions péri-centromériques, composées en grande partie de séquences satellites. Chacune des 2 extrémités du chromosome contient les régions sub-téломériques puis les téломères. Ici est représenté un gène n°1 dont le promoteur contient un CGI. Si ce promoteur est hypométhylé, alors le gène est transcrit. Un gène n°2 possède un promoteur en amont mais ne contenant pas de CGI. Ces promoteurs sont généralement hyperméthylés. En général on retrouve aussi un haut niveau de méthylation dans le corps des gènes. Les régions sub-téломériques, péri-centromériques, et les divers types de séquences répétées sont souvent riches en CpG et sont hyperméthylés dans les cellules somatiques.

spécifique, comme c'est le cas par exemple des gènes de la famille Hox qui sont impliqués dans la régulation du développement (Mohn et Schübeler 2009).

Seuls 20% des sites intragéniques (les introns et les exons autres que le premier) possèdent un ou plusieurs CGI. La méthylation est dans ce cas également corrélée à l'expression (Varley et al. 2013). La manière dont la méthylation intragénique permet l'expression des gènes est encore sujet à discussion. Il a été montré qu'au moins certains de ces CGI résident en fait dans des promoteurs alternatifs au sein des gènes, et que la méthylation servirait à la suppression de ces promoteurs (Maunakea et al. 2010) (Maunakea et al. 2013). Les changements importants de méthylation au niveau de la transition entre introns et exons suggèreraient plutôt que la méthylation est impliquée dans des mécanismes de régulation des épissage alternatifs (Laurent et al. 2010).

Dans le corps des gènes (introns et exons autres que le premier) et en absence de CGI, il existe une corrélation entre hyperméthylation et expression des gènes (Schultz et al. 2015). Ce n'est cependant pas le cas dans les cellules qui ne se divisent peu ou pas (Aran et al. 2011). De manière générale les exons sont plus méthylés que les introns. On observe cependant des profils de méthylation particuliers au niveau de la jonction entre introns et exons. En effet, la transition entre un exon suivi par un intron est marquée par un pic suivi d'une chute du niveau de méthylation (Laurent et al. 2010). En parallèle, à la jonction entre un intron et un exon, on observe cette fois une chute puis une augmentation du niveau de méthylation.

Dans le contexte des CpG *shores*, c'est-à-dire les régions bordant les CGI et moins denses en CpG, la méthylation est alors associée à l'inhibition transcriptionnelle du gène associé (Irizarry et al. 2009). Ces régions ont attiré l'attention de nombreuses équipes de recherche, car il semble que les CpG *shores* sont fréquemment différenciellement méthylés dans diverses pathologies comme le cancer (Qu et al. 2015), ou encore entre les différents types cellulaires (Doi et al. 2009). Au contraire des CpG associés à des CGI, les CpG distribués dans l'*open sea* sont globalement hyperméthylés.

La régulation des gènes par la méthylation ne se fait pas exclusivement à proximité des gènes. Les amplificateurs (*enhancers*) sont des régions régulatrices de l'expression situés à distance du ou des gènes qu'ils régulent (jusqu'à 1000 kb en amont ou en aval du gène). Il en existe potentiellement des centaines de milliers dans le génome (Shen et al. 2012). Ces régions ne sont généralement pas associées à des CGI et présentent un taux de méthylation intermédiaire quand ils sont actifs (en moyenne 30%). Ils peuvent alors fixer des facteurs de transcription et stimuler l'expression de leurs gènes (Stadler et al. 2011). Les facteurs de transcription (TF)

sont des protéines de liaison à l'ADN ayant la capacité d'interpréter l'information de méthylation contenue dans les promoteurs ou les *enhancers* des gènes. Chez l'Homme, il en existerait plus de 1400 (Vaquerizas et al. 2009). Ces facteurs reconnaissent des motifs génomiques pour s'y fixer et ainsi participer à la régulation de l'expression des gènes. Certains d'entre eux sont sensibles à l'état de méthylation de leurs séquences cibles, que ce soit pour l'ADN méthylé ou non-méthylé. Yin et collègues ont montré que 60% des 519 TF qu'ils ont analysés sont sensibles à l'état de méthylation de leur cible (Yin et al. 2017). Par exemple, le facteur *CCCTC-Binding factor* (CTCF) se fixe sur la région consensus CCGCGNGGNGGCAG lorsque celle-ci est hypométhylée (N étant n'importe quelle base nucléotidique). Il s'agit d'un facteur de type isolateur (*insulator*), c'est-à-dire qui a la capacité de modifier la structure de la chromatine en interagissant avec d'autres régions génomiques et d'autres protéines, d'isoler un gène en bloquant l'interaction entre le promoteur et un *enhancer* et en conséquence d'agir sur la transcription des gènes associés (Figure 35).

1.3.3. La machinerie de la méthylation de l'ADN

Apposition de la méthylation de l'ADN : les enzymes écrivain (*writer*)

Les différents membres de la famille des *ADN MéthylTransférases* (DNMT) ont pour rôle le transfert des groupements méthyles à l'ADN. Les différentes DNMT partagent une structure protéique similaire mais ont des fonctions et des profils d'expression spatio-temporels différents (Figure 36).

L'enzyme DNMT1, qualifiée de DNMT de maintenance, assure la transmission de la 5mC pendant la phase de réplication de l'ADN au cours des divisions cellulaires (Figure 37.A). En effet, au cours de la réplication, le brin néosynthétisé est dépourvu de méthylation. DNMT1 cible alors préférentiellement les sites CpG hémi-méthylés, pour méthyler symétriquement le brin néosynthétisé (Leonhardt et al. 1992). Des études ont montré que DNMT1 aurait aussi une fonction de réparation de l'ADN (Mortusewicz et al. 2005), et qu'elle pourrait aussi être impliquée dans la méthylation *de novo* (Egger et al. 2006).

DNMT3a et DNMT3b sont 2 enzymes qui présentent des structures protéiques et des fonctions similaires. On les qualifie d'enzymes de méthylation *de novo*, c'est-à-dire ciblant des CpG non-méthylés (Figure 37.B). Alors que DNMT3a est exprimée de manière ubiquitaire, DNMT3b est exprimée préférentiellement dans la thyroïde, la moelle osseuse et les testicules. DNMT3b joue aussi un rôle fondamental au cours du développement embryonnaire précoce, démontré par la létalité au stade embryonnaire en cas de perte de fonction (Okano et al. 1999). DNMT3l est exprimé préférentiellement au cours du

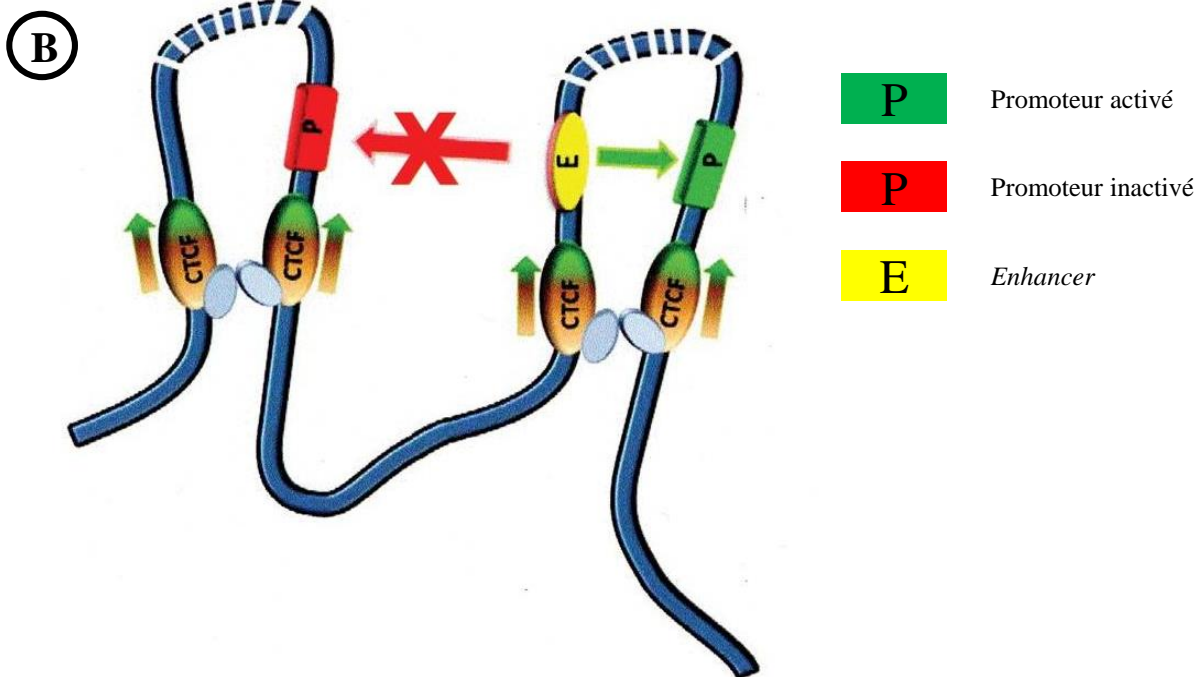
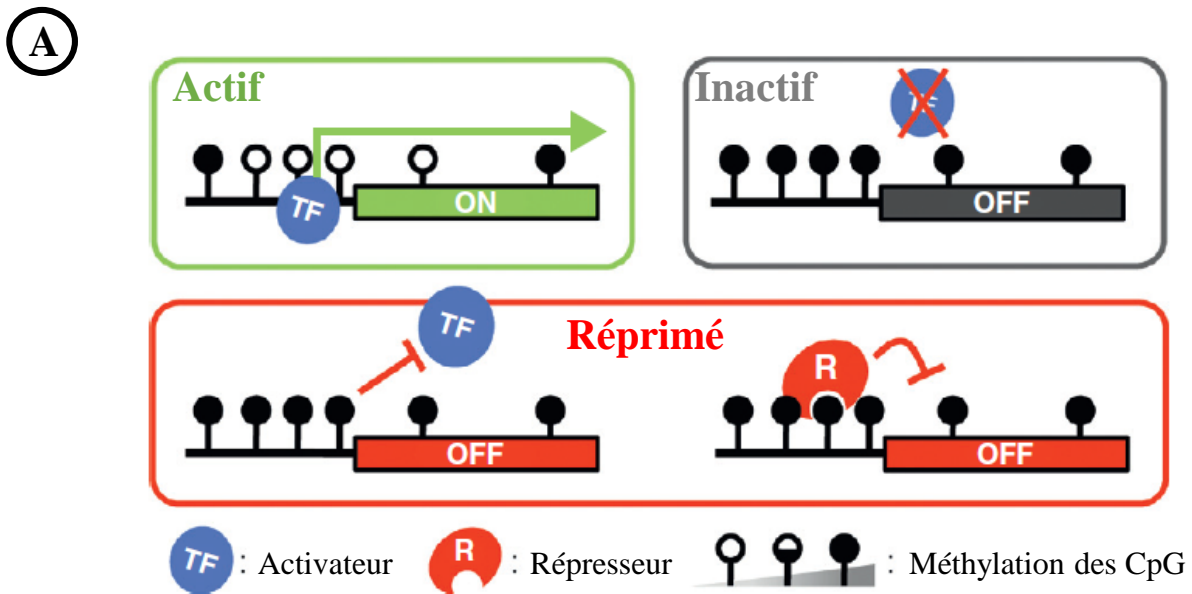


Figure 35. Degré de méthylation du promoteur et transcription génique

A. Les promoteurs et le premier exon des gènes actifs sont généralement hypométhylés. L'absence d'activateur conduit à une augmentation de la méthylation aux promoteurs de gènes inactifs. Il y a répression d'un gène lorsque la méthylation de l'ADN interfère directement ou indirectement avec des activateurs. Ce peut être par le biais du recrutement de répresseurs méthyl-dépendants. (adapté de Baubec et Schübeler, 2014).

B. Boucles d'ADN bordées par des sites de fixation du facteur CTCF. Le facteur CTCF, protéine en doigts de zinc, se fixe sur son site consensus lorsque celui-ci n'est pas méthylé. La fixation de CTCF permet la formation d'une boucle d'ADN, au sein de laquelle, il peut y avoir création d'un environnement favorable aux interactions entre promoteurs et *enhancers* favorisant la transcription soit au contraire exclusion des *enhancers* de la boucle et silence transcriptionnel. (adapté de Ghirlando et Felsenfeld, 2016)

PBD: proliferating cell nuclear antigen (PCNA) Binding Domain

PHD: Polybromo Homology Domain

NLS: Nuclear Localization Signal

TRF: Targeting Replication Foci

CXXC: Cystein rich, zinc finger DNA-binding motif

PWWP: tetrapeptide domain containing proline-tryptophan-tryptophan-proline motif

GK: glycine-lysine

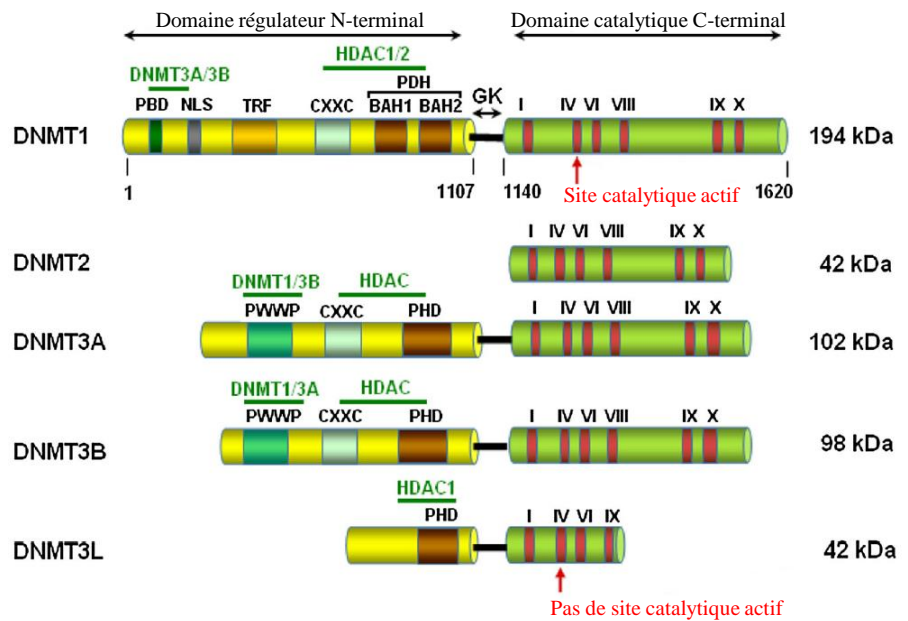


Figure 36. Représentation schématique de la structure des protéines de la famille des DNMT

Toutes les DNMT possèdent un site N-terminal de fixation à l'ADN excepté DNMT2. DNMT3L ne possède pas le site C-terminal catalytique canonique composé de 6 motifs d'acides aminés conservés. En vert, sites d'interaction avec d'autres enzymes de la machinerie épigénétique. Il existe aussi une DNMT3C chez la souris qui n'est pas représentée sur cette figure.

(adapté de Sadakierska-Chudy *et al.*, 2014)

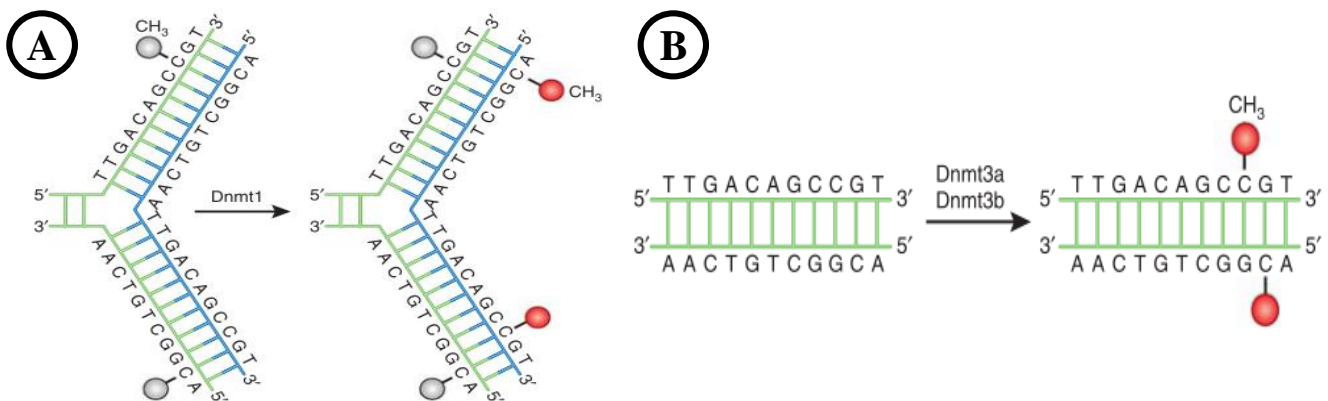


Figure 37. Modes d'action de la DNMT de maintenance (DNMT1), et des DNMT *de novo* (DNMT3a et DNMT3b)

A. Mode d'action de DNMT1. Lors de la réplication, le brin néo-synthétisé (en bleu) est dépourvu de groupement méthyle sur les cytosines en contexte CpG. DNMT1 cible préférentiellement ces sites et méthyle le brin néo-synthétisé (groupement méthyle rouge sur le schéma) ; cette caractéristique assure donc l'héritabilité de la méthylation de l'ADN à travers la mitose. **B.** Mode d'action de DNMT3a et DNMT3b. Ces enzymes de méthylation *de novo* ciblent de manière préférentielle les sites CpG non-méthylés, pour méthyle les 2 brins d'ADN au contraire de DNMT1.

(Moore *et al.*, 2012)

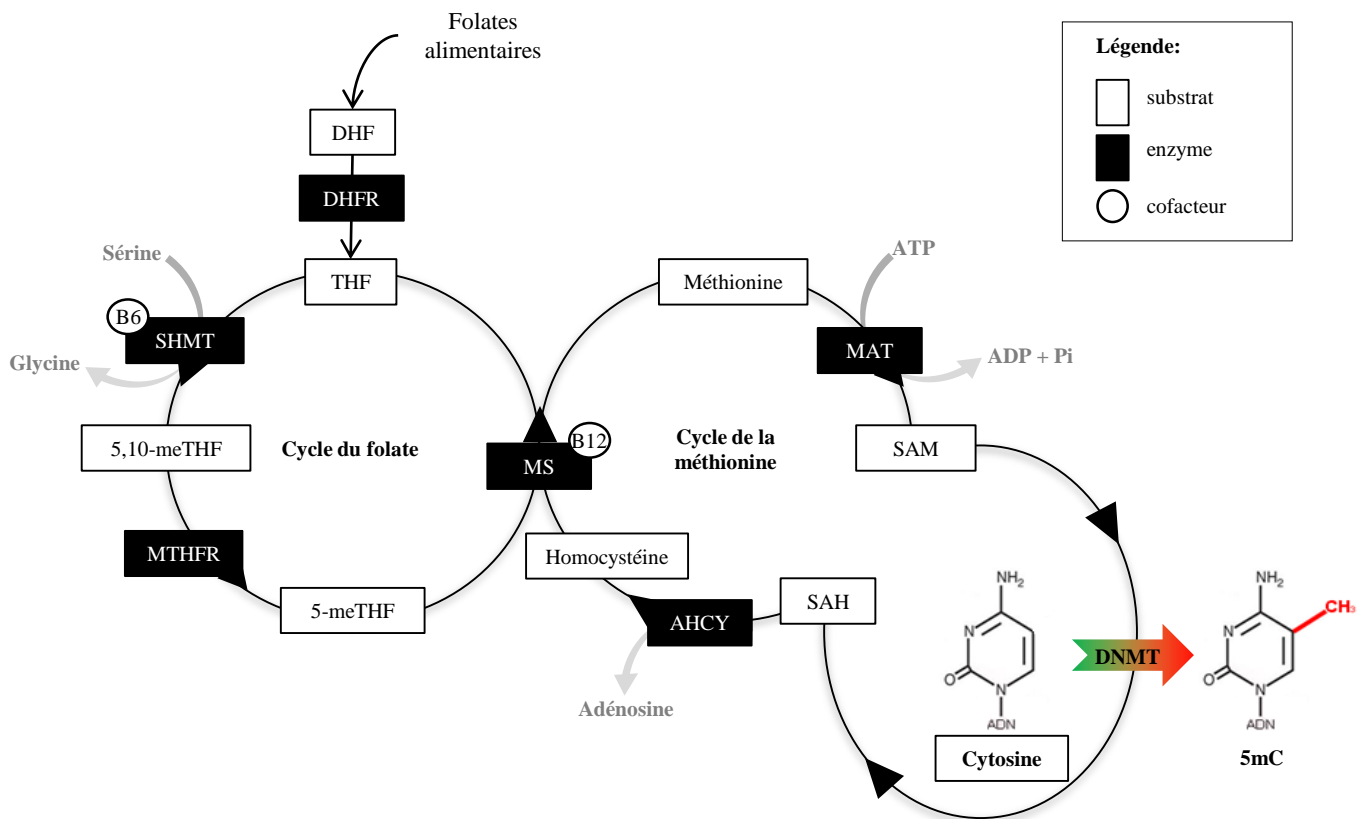


Figure 38. Représentation schématique du cycle du folate et de la méthionine

Le 5-méthyltétrahydrofolate est la forme circulante du folate. Il est synthétisé sous l'action de la méthylène-tetrahydrofolate réductase par réduction du 5,10-méthylène-tetrahydrofolate. Le groupement méthyle est ensuite transféré par la méthionine synthase à l'homocystéine pour former la méthionine et du tétrahydrofolate. En retour, la méthionine fournit le groupement méthyle nécessaire à la formation de la S-adenosylméthionine, qui est le donneur universel de groupement méthyle, pour l'ADN mais aussi pour les protéines et les lipides

Abréviations: AHCY, adénylhomocystéinase; B6, vitamine B6; B12, vitamine B12; BHMT, bétaine-homocystéinéméthyltransférase; DHF, dihydrofolate; DHFR, dihydrofolateréductase; MAT, méthionine adényltransférase; MTHFR, méthylène-tetrahydrofolate réductase; MS, méthionine synthase; 5-meTHF, 5-méthyl-tétrahydrofolate; 5,10-meTHF, 5,10-méthylène-tétrahydrofolate; TPMT, thiopurine S-méthyltransférase; SAH, S-adenosylhomocystéine; SAM, S-adenosylméthionine; SHMT, sérine hydroxyméthyltransférase; THF, tétrahydrofolate.

DNMT1 active

DNMT1 inactive

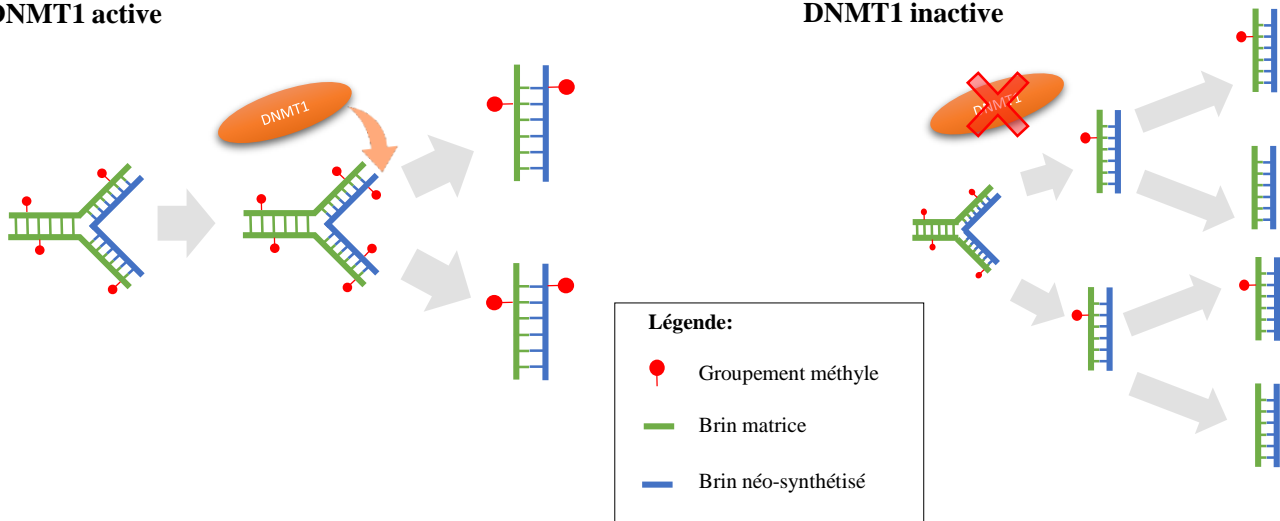


Figure 39. Représentation de la voie passive de la déméthylation de l'ADN

Est représenté un brin d'ADN en cours de réplication. Si l'enzyme DNMT1 est active (partie de gauche), cela va conduire à son recrutement sur le site hémi-méthylé et à la méthylation du brin néo-synthétisé. Si DNMT1 est inactive (partie de droite), la méthylation globale de l'ADN est progressivement diluée au cours des réplifications successives.

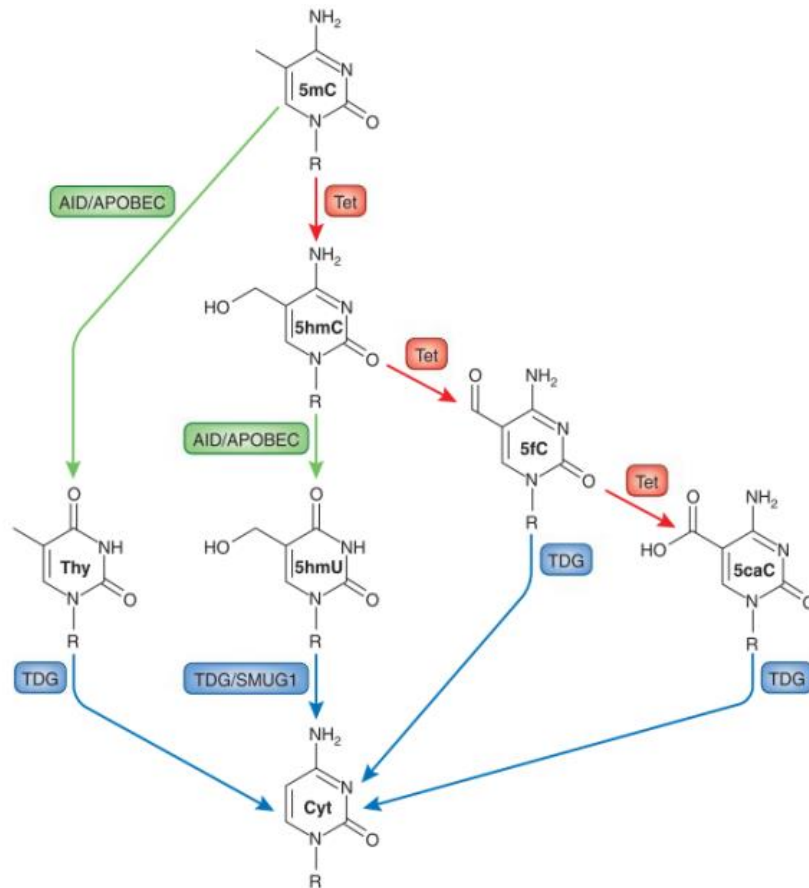


Figure 40. Voie active de la déméthylation de l'ADN

Les 5mC peuvent être désaminées par le complexe AID/APOBEC pour former une thymine. Le mésappariement que cette transformation engendre, est corrigé par un mécanisme d'excision de base avec la TDG. De façon différente, les 5mC sont aussi susceptibles d'être oxydées en 5hmC par les enzymes Tet. La 5hmC peut être encore oxydée par les Tet pour générer successivement la 5fC puis le 5caC, qui peuvent toutes 2 être corrigées par la TDG. La 5hmC est aussi transformable par le complexe AID/APOBEC pour former la 5hmU, qui sera aussi réparée par la TDG ou SMUG1. A la fin de toutes ces voies différentes, la 5hmC a été convertie en cytosine.

5mC, 5-méthylcytosine; 5hmC, 5-hydroxyméthylcytosine; 5hmU, 5-hydroxyméthyluracile, 5caC, 5-carboxylcytosine, Cyt, cytosine; AID/APOBEC, *Activation-induced cytidine deaminase / Apolipoprotein B mRNA Editing Enzyme Catalytic*; TDG, *Thymine DNA Glycosylase*, Tet, *Ten-Eleven Translocation*; SMUG1, *Single-Strand Selective Monofunctional Uracil-DNA glycosylase 1*.

(adapté de Moore *et al.*, 2012)

développement embryonnaire précoce, et dans les cellules germinales fœtales et adultes (La Salle et al. 2007). DNMT3L est de structure similaire aux autres DNMT, mais ne possède pas de domaine catalytique. Elle a un rôle de cofacteur, en s'associant à l'ADN pour guider l'activité des DNMT3a/b. Enfin, un dernier membre de la famille des DNMT est spécifiquement actif dans la lignée germinale mâle chez la souris et impliquée dans le contrôle de rétrotransposons, DNMT3C (Barau et al. 2016).

Le substrat des DNMT est la S-adénylméthionine (SAM), produit majeur de la suite de réactions enzymatiques faisant partie du cycle du folate, couplé au cycle de la méthionine (Figure 38). Sa synthèse nécessite l'apport de folate (ou vitamine B9) par l'alimentation. SAM est le donneur universel de groupement méthyle, puisqu'il est requis pour méthyler l'ADN, les protéines et les lipides.

Le retrait des marques de méthylation de l'ADN ; les enzymes effaceur (*eraser*)

Aujourd'hui, 2 voies de déméthylation de l'ADN sont principalement décrites : la voie passive et la voie active. La voie passive consiste en une simple dilution de la méthylation au cours de la division cellulaire, lorsque DNMT1 est absente ou inactive. DNMT1 ne peut alors pas méthyler le brin non-méthylé pendant sa néosynthèse (Figure 39).

La voie active de déméthylation de l'ADN se fait par le biais de plusieurs voies enzymatiques impliquant des réactions séquentielles pour déméthyliser les 5mC (Figure 40). Les 2 voies enzymatiques impliquent que les 5mC soient en premier lieu chimiquement modifiées soit au niveau du groupement amine, soit sur le groupement méthyle. Dans le premier cas, la 5mC est convertie en thymine par l'*Activation-Induced cytidine Deaminase / APOlipoprotein B mRNA-editing Enzyme Complex* (AID/APOBEC). Le mésappariement G/T engendré est reconnu et corrigé par le mécanisme d'excision de base (*base excision repair*, BER). Dans le second cas, les enzymes de la famille des *Ten-Eleven Translocation* (TET) catalysent l'oxydation du groupement méthyle en 5 hydroxyméthylcytosine (5hmC). La 5hmC est elle-même convertie en cytosine suivant 2 mécanismes. La première l'oxydation supplémentaire des 5hmC en 5-formylcytosine (5fC) puis en 5-carboxylcytosine (5caC). Dans la seconde la 5hmC est désaminée par AID/APOBEC pour former la 5-hydroxyméthyluracile (5hmU). A la fin de chacun de ces processus, le mécanisme BER utilise la *Thymine DNA Glycosylase* (TDG) pour cliver la base modifiée qu'il remplace par une cytosine. Un autre membre de la famille des DNA glycosylase, la *Single-strand-selective Monofunctional Uracil-DNA Glycosylase 1* (SMUG1), participe aussi au mécanisme de BER et est impliqué dans le processus de déméthylation (Moore, Le, et Fan 2012).











Famille MBD	Cibles identifiées <i>in vitro</i>
MeCP2 	NMGG AATT; NMGG TTAA NGMCTTAA NGMCAATT
MBD1 	TCMCGCA AGGMCCT
MBD2 	NMGG NGMC
MBD4 	MG; MG GM GT
Famille SRA	Cibles identifiées <i>in vitro</i>
UHRF1 	NMGN; NMGN; NHGN NGMN NGCN NGHN
UHRF2 	NMGN; NMGN NGMN NGCN
Famille Zinc Finger	Cibles identifiées <i>in vitro</i>
Kaiso 	TCCTGCNA; CMGMGT; CMGMGT AGGACGNT GGMGMA GCGCA
ZBTB4 	CTGCCAT; MGCCAT; MGCCAT GACGGTA GGGTA GCGTA
ZBTB38 	MGCCAT GGMGTA
ZFP57 	TGCMGC ACGCMG

Figure 41. Les trois familles de protéines de liaison aux sites CpG méthylés chez les mammifères

Sur la partie gauche sont représentées chacune des protéines de lecture de la méthylation de l'ADN connue chez les mammifères, avec leurs domaines fonctionnels. Les domaines en doigts de zinc (ZF) soulignés en noir sont ceux qui participent à la reconnaissance de la méthylation. Sur la partie droite du tableau sont représentées les cibles connues pour chacune des protéines CXXC, CXXC-type zinc finger; BTB/POZ, broad complex, Tramtrack and bric-à-brac/pox virus and zinc finger Domain; KRAB, Krüppel-associated box; MBD, Methyl-Binding Domain; PHD, Plant Homeo Domain; RING, Really Interesting New Gene domain; SRA, SET and RING-finger Associated domain; TTD, Tandem Tudor Domain; Ubl, Ubiquitin-like domain; ZF, Cys2His2 zinc finger domain. Les bases notées M représentent les 5mC, les H les 5hmC et les N n'importe quel nucléotide (adapté de Buck-Koehntop et Defossez, 2013)

Contrairement aux 5fC et 5caC, la 5hmC n'est pas uniquement considérée comme un produit intermédiaire du processus de déméthylation de l'ADN mais comme une base nucléotidique ayant ses propres fonctions biologiques. En effet, la 5hmC semble intervenir dans la régulation génique, notamment par le recrutement de la protéine de liaison UHRF1 (*Ubiquitin-like PHD and RING Finger*) (Frauer et al. 2011). Les taux de 5hmC chez les mammifères sont particulièrement élevés dans le système nerveux et les cellules souches embryonnaires, de l'ordre de 0,5% (Kriaucionis et Heintz 2009). Les taux de 5fC et de 5caC sont plus faibles encore, de 10 à 100 fois moins que la 5hmC dans les embryons préimplantatoires murins (Pfaffeneder et al. 2011).

L'interprétation des marques de méthylation de l'ADN : les enzymes lecteur (*reader*)

Trois familles de protéines peuvent se lier à l'ADN, et ainsi médier les conséquences moléculaires de la méthylation : les protéines à domaine de liaison à l'ADN méthylé (*Methyl-CpG-Binding Domain*, MBD), les protéines de la famille *SET and RING-finger Associated domain* (SRA), et les protéines à domaines en doigt de zinc (*Zinc Finger*, ZnF) (Figure 41).

Les MBD ont été les premières protéines d'interaction avec la méthylation à avoir été identifiées. Cette famille d'enzymes comprend 7 membres dont la *Methyl-CpG-binding Protein 2* (MeCP2) et les MBD1 à 6. Toutes (sauf MBD5 et 6) contiennent un site de liaison de haute affinité aux sites CpG méthylés, et interagissent avec des protéines. MeCP2, MBD1 et MBD2 peuvent recruter des complexes protéiques répresseurs de la transcription grâce au *Transcriptional Repression Domain* (TRD). Il a aussi été montré récemment que MeCP2 peut se lier aux CpH dans le cerveau adulte chez la souris (Chen et al. 2015). MBD4 serait elle impliquée dans la réparation de l'ADN et la réponse au stress oxydatif (Laget et al. 2014).

Les protéines de la famille SRA sont au nombre de 2, UHRF1 et UHRF2. Ces protéines se fixent sur l'ADN hémi-méthylé et permettent le recrutement de DNMT1 aux sites de réplication de l'ADN (Bostick et al. 2007) et participent donc au maintien de la méthylation à travers la mitose donc à l'héritabilité de cette marque épigénétique.

La dernière famille est celle des protéines à domaine en doigt de zinc, dont les membres importants sont Kaiso, ZBTB4, ZBTB38 et Zpf57. Ils présentent une expression tissu-spécifique et sont impliqués dans la répression transcriptionnelle. Par exemple, Kaiso peut se lier sur une séquence spécifique non-méthylée TCCTGCNA ou sur 2 sites CpG consécutifs méthylés (Daniel et al. 2002). Récemment, il a été montré que Zpf57 conjointement à UHRF1 et DNMT1, joue un rôle dans la maintenance de l'empreinte lors de la reprogrammation épigénétique au cours du développement embryonnaire précoce (Quenneville et al. 2011).

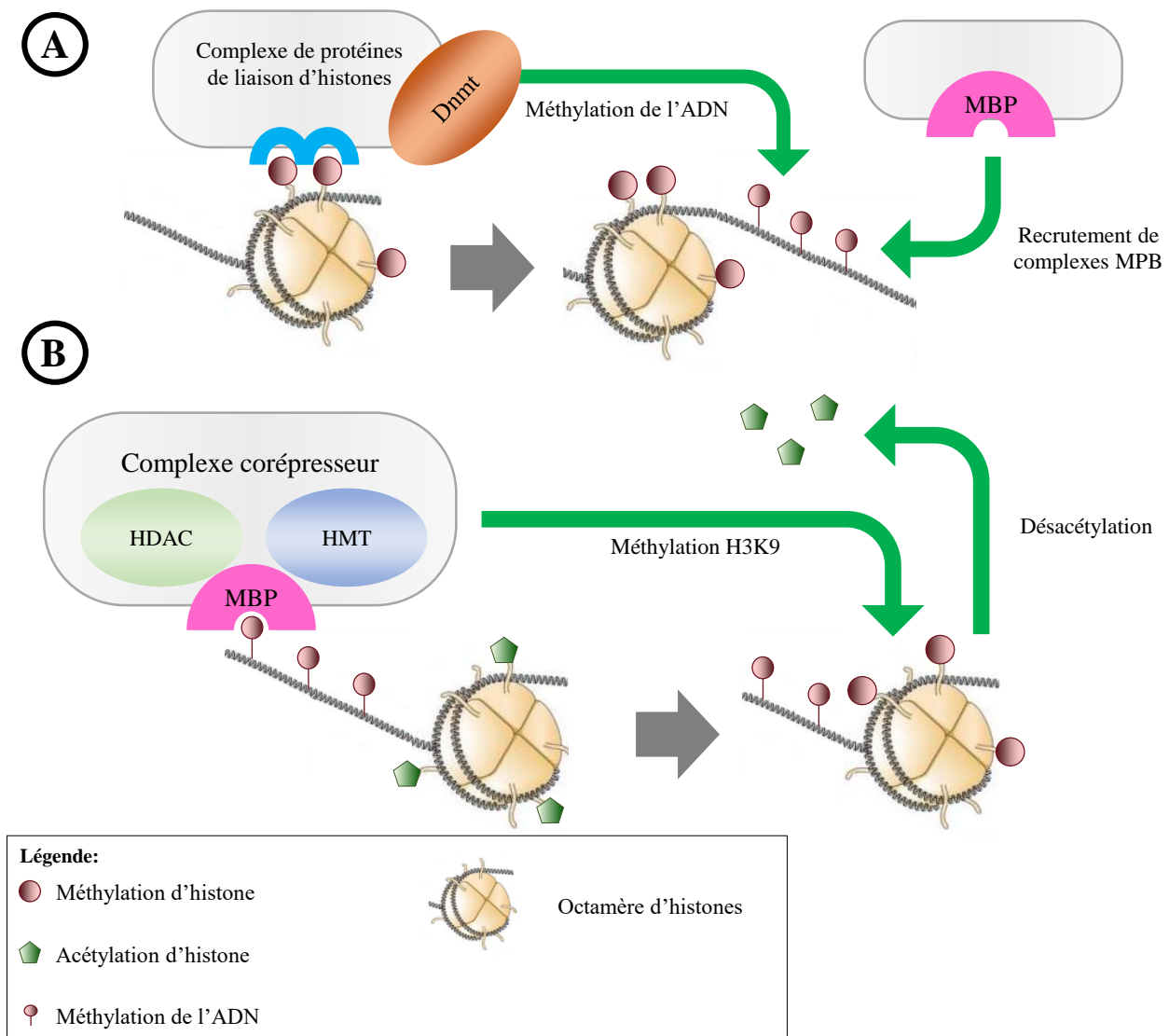


Figure 42. Interactions entre la méthylation de l'ADN et les modifications post-traductionnelles de histones

A. Modèle d'apposition de la méthylation de l'ADN dirigée par les MPT d'histones. Des complexes protéiques composés entre autre de protéines de liaison aux MPT d'histones ont la capacité de recruter la machinerie de la méthylation de l'ADN. L'ADN méthylé peut ensuite être reconnu par les protéines de liaison à l'ADN méthylé (MBP). **B.** Modèle d'apposition et de retrait de MPT d'histones dirigé par la méthylation. L'ADN méthylé est reconnu par les MBP. Ces protéines peuvent recruter un complexe protéique composé d'histone déacétylases (HDAC) et méthyletransférases (HMT). En conséquence, l'apposition de MPT d'histones de type répressive (désacétylation, H3K9me3) et la méthylation de l'ADN participent à la formation de l'hétérochromatine. (adapté de Sasai et Defossez, 2009)

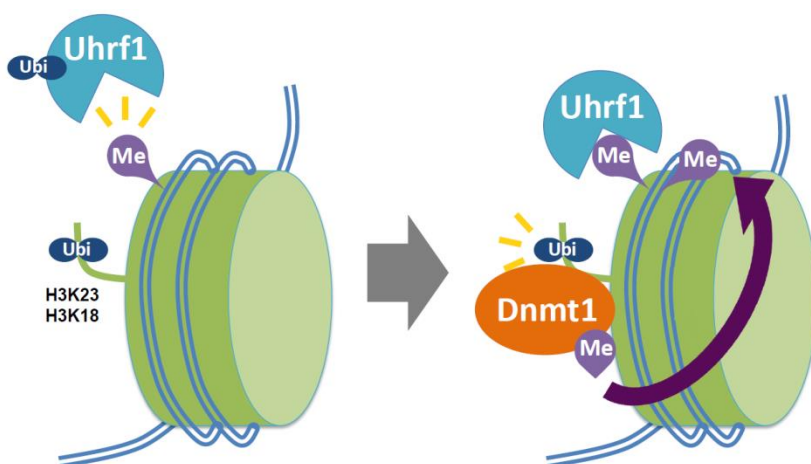


Figure 43. Maintenance de la méthylation de l'ADN par DNMT1 et Uhrf1

Après la réplication de l'ADN, Uhrf1 se lie à l'ADN hémiméthylé et ubiquitine H3K23 et/ou H3K18. Ces marques d'histones permettent le recrutement de DNMT1 au site de réplication, qui peut en retour méthyler l'ADN hémiméthylé, et ainsi concourir à la maintenance de la méthylation de l'ADN pendant la réplication.

(ubi, ubiquitine; Me, groupement méthyle)
(adapté de Nishiyama, Yamaguchi et Nakanishi, 2015)

1.3.4. Dialogue avec les autres marques épigénétiques

L'ensemble des interactions complexes entre les marques épigénétiques est schématisé par l'expression « dialogue épigénétique ». Les 3 types de marques épigénétiques décrites dans la partie précédente œuvrent de concert dans le but de contrôler la transcription des gènes et de maintenir la structure chromatinienne.

En effet, le niveau de méthylation de l'ADN sur certaines régions génomiques est corrélé ou anti-corrélé avec la présence de certains types de marques d'histones. De nombreuses études ont montré que les DNMT coopèrent directement avec les enzymes de la machinerie des histones, et il semblerait que ce soit la machinerie des MPT des histones qui guide la machinerie de la méthylation de l'ADN. Par exemple, la méthylation de H3K4 (H3K4me2 et H3K4me3) empêche le recrutement de la machinerie de méthylation de l'ADN comme DNMT3a et DNMT3l, et donc la méthylation de l'ADN de la région concernée (Ooi et al. 2007) (Noh et al. 2015) (Vlachogiannis et al. 2015). Il a été montré que ce mécanisme protège les CGI et les TSS des gènes exprimés (Mikkelsen et al. 2007). Une autre MPT d'histone de type permissive, H3K36me3 recrute DNMT3a et DNMT3b au niveau des exons de gènes *via* son domaine PWWP (Dhayalan et al. 2010) (Baubec et al. 2015).

Inversement, la méthylation de l'ADN peut aussi permettre le recrutement de la machinerie des MPT d'histones. Il existe une liaison directe entre DNMT1 et DNMT3a avec l'histone méthyltransférase SUV39H1 et les histones désacétylases (Fuks et al. 2000) (Fuks 2003). L'ADN méthylé peut provoquer le recrutement de l'histone méthyltransférase SUV39H1 qui va induire la triméthylation de la position H3K9, et des HDAC pour la désacétylation des histones. Ensemble, ces marques provoquent la compaction de la chromatine de la région concernée, et donc un accès restreint à l'ADN pour les facteurs de transcription (Figure 42).

Les protéines lectrices de la méthylation, qui reconnaissent les sites CpG méthylés, ont la capacité de recruter des répresseurs protéiques. Par exemple, la protéine UHRF1, peut se fixer sur les sites CpG hémiméthylés après la réplication de l'ADN (Figure 43). Elle est ensuite capable d'ubiquitinyler les positions H3K23 et H3K18, ce qui permet le recrutement de DNMT1 sur ces mêmes sites pour maintenir la méthylation (Nishiyama et al. 2013).

MeCP2 est impliqué dans le recrutement d'histones désacétylases (complexe Sin3A/HDAC), d'histones méthyltransférases pour méthyler notamment H3K9, et du complexe de remodelage de la chromatine ATRX (Nan et al. 2007). MBD2 s'associe avec le complexe NuRD et MBD1 avec les H3K9 méthyltransférases SETDB1 et SUV39H1 (Fujita et al. 2003)

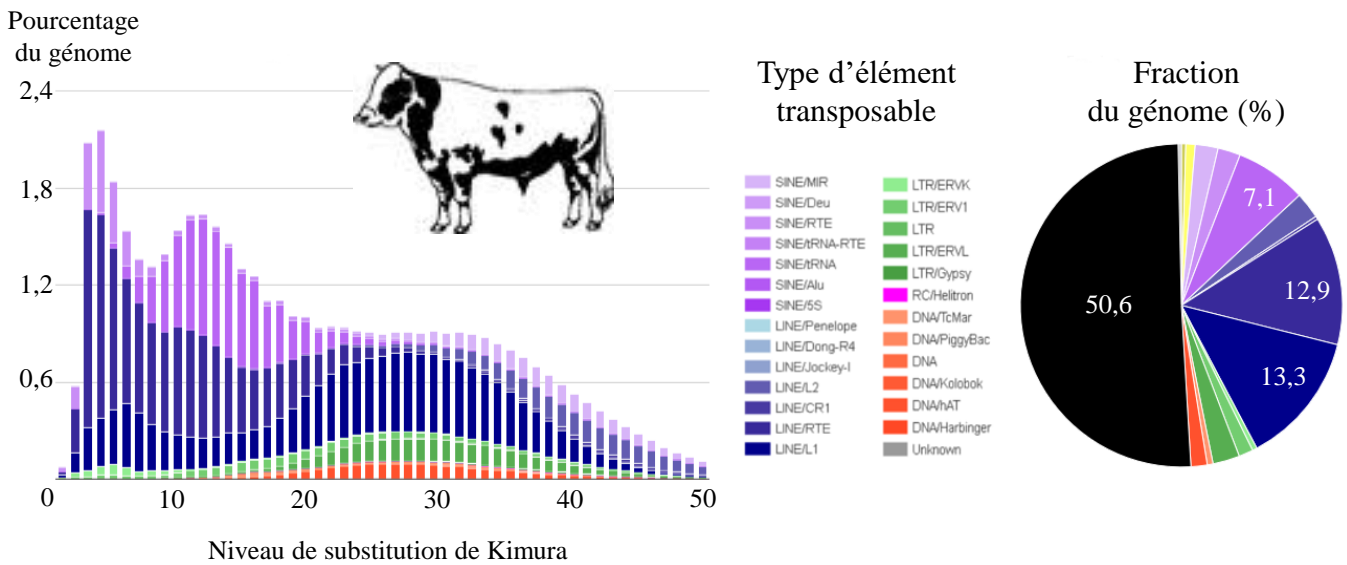


Figure 44. Les éléments de type transposables dans le génome bovin

Le graphique représente sur l'axe des ordonnées la couverture génomique pour chaque type d'élément transposable. L'axe des abscisses correspond à la distance de Kimura de ces éléments par rapport à leur séquence consensus (allant de 0 à 50). Les éléments situés sur la partie gauche du graphe n'ont pas beaucoup divergé par rapport à la séquence consensus et correspondent à des éléments relativement récents, alors que les séquences situées sur la droite du graphe correspondent à des éléments ayant divergé ou dégénéré à partir de leur séquence consensus depuis longtemps. Chez le bovin, on observe un pic situé sur la gauche du graphe, donc une part importante des éléments transposables est composée d'éléments récents et potentiellement actifs. L'ensemble des éléments transposables du génome représente près de 50% du génome bovin.

(adapté de l'outil RepeatMasker, Institute for Systems Biology)

(Sarraf et Stancheva 2004). De la même façon, MBD4 s'associe avec le complexe Sin3A/HDAC pour réprimer la transcription (Kondo et al. 2005).

La méthylation de l'ADN peut aussi réguler l'expression des microARN (Saito et Jones 2006). Il a par exemple été montré que lorsque les DNMT sont inhibées, des microRNA qui sont normalement inactivés par l'hyperméthylation d'un CGI régulateur sont de nouveau actifs. De manière opposée, les ARN non-codants peuvent réguler l'expression des DNMT et des enzymes de la machinerie des histones, et donc avoir une influence sur l'état épigénétique général. Par exemple, l'inactivation de DICER dans les cellules souches embryonnaires de souris induit une diminution de la quantité de microARN, dont le microARN-290 régulant l'expression de DNMT3a et b (Sinkkonen et al. 2008).

1.4. Rôles biologiques de la méthylation de l'ADN

1.4.1. Rôle structural et stabilité du génome

Chez les vertébrés, en fonction des espèces, plus de la moitié du génome est constituée de séquences répétées. Ce sont des motifs nucléotidiques dupliqués des centaines voire des milliers de fois dans le génome. On en distingue plusieurs familles en fonction de leur rôle biologique, de leur localisation ou de leur mode de réplication (Figure 44).

Les séquences satellites (ADNsat) ont un rôle dans l'organisation spatiale du génome, la stabilité de la chromatine, et sont aussi transcrites sous forme d'ARN non-codants au cours de l'embryogenèse (Probst et al. 2010). On distingue 2 types de séquences satellites, les majeures et les mineures, qui représentent environ 3,5 % du génome chez la souris, 4 à 5% chez l'Homme et 23% chez le bovin (Chaves et al. 2003). Elles sont préférentiellement dans les régions péri-centromériques de type hétérochromatique et les régions sub-téломériques, où elles sont très méthylées dans les cellules somatiques. L'hypométhylation de ces séquences est associée chez l'Homme à la survenue du syndrome de *Immunodeficiency with Centromeric instability and Facial anomalies* (ICF). Cette hypométhylation est liée dans 60% des cas du syndrome d'ICF à des mutations dans le gène de DNMT3b. L'hypométhylation des séquences satellites péri-centromériques (notamment des chromosomes 1, 9 et 16) est associée une décondensation de ces séquences et une modification du positionnement des chromosomes dans le noyau. Alors que l'hétérochromatine est normalement associée à la membrane nucléaire dans les cellules somatiques, la distance qui sépare l'hétérochromatine des chromosomes touchés dans le cas du syndrome est augmentée (Jefferson et al. 2010). L'hypométhylation est aussi associée à une transcription aberrante de ces séquences satellites

péri-centromériques qui ne sont normalement pas exprimées à l'âge adulte (Eymery et al. 2009), ainsi que celle de nombreux autres gènes (Jin et al. 2008). L'augmentation de l'expression de ces derniers n'est pas due à une altération de la méthylation du promoteur. Elle résulterait d'un contrôle à distance induit par la perturbation de l'organisation nucléaire. Enfin, l'hypométhylation affecte également les régions sub-télomériques, induisant une augmentation de l'expression des *Telomeric Repeat Containing RNA* (TERRA), dont la résultante est le raccourcissement des télomères (Yehezkel et al. 2008). L'ensemble de ces données soulignent donc le fait que la méthylation de l'ADN, en particulier dans les régions péri-centromériques et sub-télomériques riches en séquences satellites, participe à la structure de la chromatine et l'organisation nucléaire du génome.

Contrairement aux ADNsat, les éléments répétés de type transposables (*Transposable Elements*, TE) ont le potentiel de se répliquer et de se transloquer dans le génome. Si cette insertion se fait dans un gène ou dans un des éléments régulateurs de ce gène, ce dernier voit son expression perturbée. Les TE sont les éléments répétés les plus nombreux du génome : ils représentent à eux seuls 46,5% de la taille totale du génome chez le bovin (Adelson, Raison, et Edgar 2009). Un haut niveau de méthylation de l'ADN permet l'inactivation transcriptionnelle des TE, et fait aussi intervenir la machinerie des piARN (Tóth et al. 2016). Il a été montré que les piARN ont la capacité de recruter et de diriger la machinerie de méthylation vers les TE et leurs régions régulatrices pour les inactiver transcriptionnellement (Aravin et al. 2008). L'inactivation de protéines impliquées dans le fonctionnement des piARN, notamment MILI ou MIWI2, empêche la méthylation *de novo* des éléments transposables dans les spermatocytes embryonnaires chez la souris (Kuramochi-Miyagawa et al. 2008). D'autres protéines semblent impliquées dans le processus d'inactivation transcriptionnelle des TE, comme *Maelstom* (MAEL) (Sienski, Dönertas, et Brennecke 2012). Il faut aussi noter que l'inactivation transcriptionnelle des TE n'est pas seulement le fait de la méthylation de l'ADN mais repose aussi sur le recrutement d'autres marques épigénétiques de type répressif par la voie des piARN (notamment H3K9me3). Une perturbation du mécanisme de méthylation de l'ADN par les piARN conduit à un arrêt de la méiose et une stérilité chez le mâle (Bourc'his et Bestor 2004), ce qui dénote son importance biologique dans le contexte de la fertilité mâle.

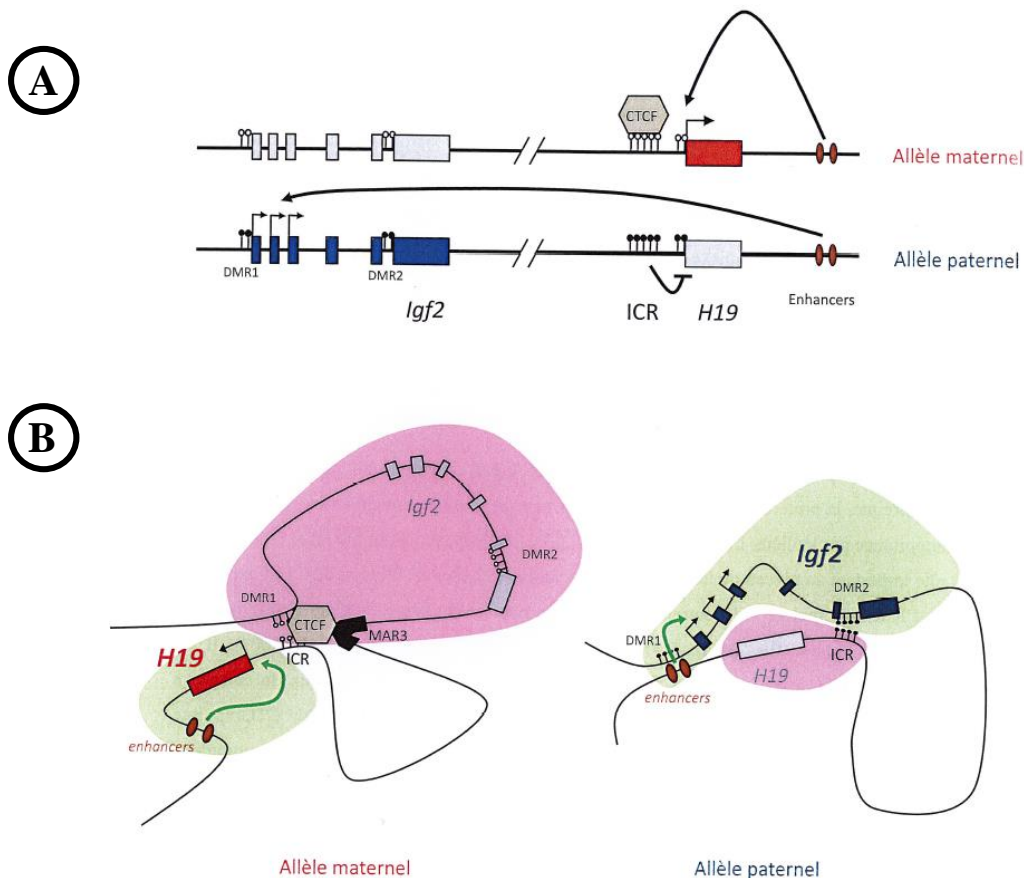


Figure 45. Le phénomène de l’empreinte parentale

A. Représentation de la régulation des gènes IGF2-H19 chez l’Homme. Sur l’allèle paternel, la région de contrôle de l’empreinte (ICR) est méthylée, ce qui empêche la fixation du facteur CTCF. En conséquence, le promoteur du gène IGF2 peut interagir avec les *enhancers* et sa transcription est activée. Sur l’allèle maternel, l’ICR est déméthylée et permet la fixation de CTCF, ce qui permet cette fois la transcription de H19 par son interaction avec les *enhancers* (tiré de la thèse de Paul Monnier: Locus H19 et contrôle épigénétique du développement embryonnaire, 2013). **B.** Structure tridimensionnelle du locus. Sur l’allèle maternel, l’ICR interagit avec la DMR1 et une région d’attachement à la matrice (MAR3), ce qui isole IGF2. Sur l’allèle paternel, l’ICR interagit avec la DMR2. IGF2 se trouve cette fois dans une zone transcriptionnellement active.

Gène (Alias)	Allèle exprimé
Achaete-Scute Family BHLH Transcription Factor 2 (ASCL2/MASH2)	Maternel
Cyclin-dependent kinase inhibitor 1C (CDKN1C)	Maternel
GNAS complex locus (GNAS/NESP55)	Maternel
H19	Maternel
Insulin-like growth factor 2 (IGF2)	Paternel
Insulin-like growth factor 2 receptor (IGF2R/CIMPR)	Maternel
KCNQ1 opposite strand/antisense transcript 1 (KCNQ1OT1/LIT1)	Paternel
MAGE-like 2 (MAGEL2)	Paternel
Maternally expressed 1 (MEG1/GRB10)	Maternel
Maternally expressed 3 (MEG3/GTL2)	Maternel
Maternally expressed gene 8 (MEG8)	Maternel
MER1 repeat containing imprinted transcript 1 (MIMT1/ITUP1/USP29)	Paternel
Mesoderm specific transcript homolog (MEST/PEG1)	Paternel
Neuronatin (NNAT)	Paternel
Nucleosome assembly protein 1-like 5 (NAPIL5)	Paternel
Paternally expressed 3 (PEG3)	Paternel
Paternally expressed 10 (PEG10)	Paternel
Paternally expressed 10 (PEG11/RTL1)	Paternel
Pleckstrin homology-like domain, family A, member 2 (PHLDA2)	Maternel
Pleiomorphic adenoma gene-like 1 (PLAGL1/ZAC1)	Paternel
Small nuclear ribonucleoprotein polypeptide N (SNRPN)	Paternel
Tumor suppressing subtransferable candidate 4 (TSSC4)	Maternel
X (inactive)-specific transcript (XIST)	Paternel

Tableau 5. Listes des gènes soumis à l’empreinte connus chez le bovin
(sources: Magee *et al.*, 2014, ainsi que le site www.geneimprint.com).

1.4.2. L'empreinte parentale

En 1984, McGrath et Solter ont mené des expériences consistant à énucléer des ovocytes de souris, puis à micro-injecter soit 2 *pronuclei* mâles, soit 2 *pronuclei* femelles (McGrath et Solter 1984). Les embryons ainsi produits sont appelés respectivement androgénotes ou gynogénotes. Ils ont observé un arrêt du développement embryonnaire ne se manifestant pas de la même manière pour les androgénotes et gynogénotes. Les embryons gynogénotes présentent un retard de croissance et un arrêt du développement dû à une hypoplasie (réduction du nombre de cellules) du placenta et des annexes embryonnaires. Les embryons androgénotes présentent un développement embryonnaire anormal, conduisant à un arrêt développemental précoce et la présence d'un placenta hyperplasique (augmentation du nombre de cellules). Bien que les embryons gynogénotes ou androgénotes possèdent un génome diploïde complet, ces travaux soulignent la complémentarité des génomes parentaux. Suite à ces travaux fondateurs, il est apparu que certains gènes, dits soumis à l'empreinte parentale, sont exprimés de manière mono-allélique en fonction de l'origine parentale des allèles. Certains gènes sont exprimés uniquement à partir de l'allèle paternel, d'autres à partir de l'allèle maternel (Figure 45). Cette expression différentielle est médiée par une méthylation différentielle de régions de contrôle de l'empreinte (*Imprinting Control Regions*, ICR). L'empreinte parentale est apposée au niveau des cellules germinales en fonction du sexe. Ainsi, les gènes d'expression maternelle présentent généralement une ICR hyperméthylée dans les spermatozoïdes ; les gènes d'expression paternelle présentent une ICR méthylée dans les ovocytes.

A ce jour, plus de 120 gènes soumis à l'empreinte ont été identifiés chez les souris, 90 chez l'Homme et 24 chez le bovin. Beaucoup d'entre eux sont impliqués dans la croissance fœtale et le développement (Gabory et al. 2009). La liste des gènes soumis à l'empreinte chez le bovin est donnée Tableau 5.

1.4.3. L'inactivation du chromosome X

Comme les individus femelles (XX) ont un X surnuméraire par rapport aux mâles (XY), il existe un mécanisme de compensation de dose connu sous le nom d'inactivation de l'X. En effet, sans l'inactivation d'un des 2 chromosomes X, il y aurait chez la femelle un déséquilibre génétique pour les gènes portés par ce chromosome, alors 2 fois plus nombreux que chez le mâle. Chez la souris, le phénomène de l'inactivation de l'X intervient rapidement après la fécondation, avec des séquences d'inactivation dès le stade 2 - 4 cellules. La cinétique et la stratégie d'inactivation est cependant assez variable entre les espèces (Okamoto et al.

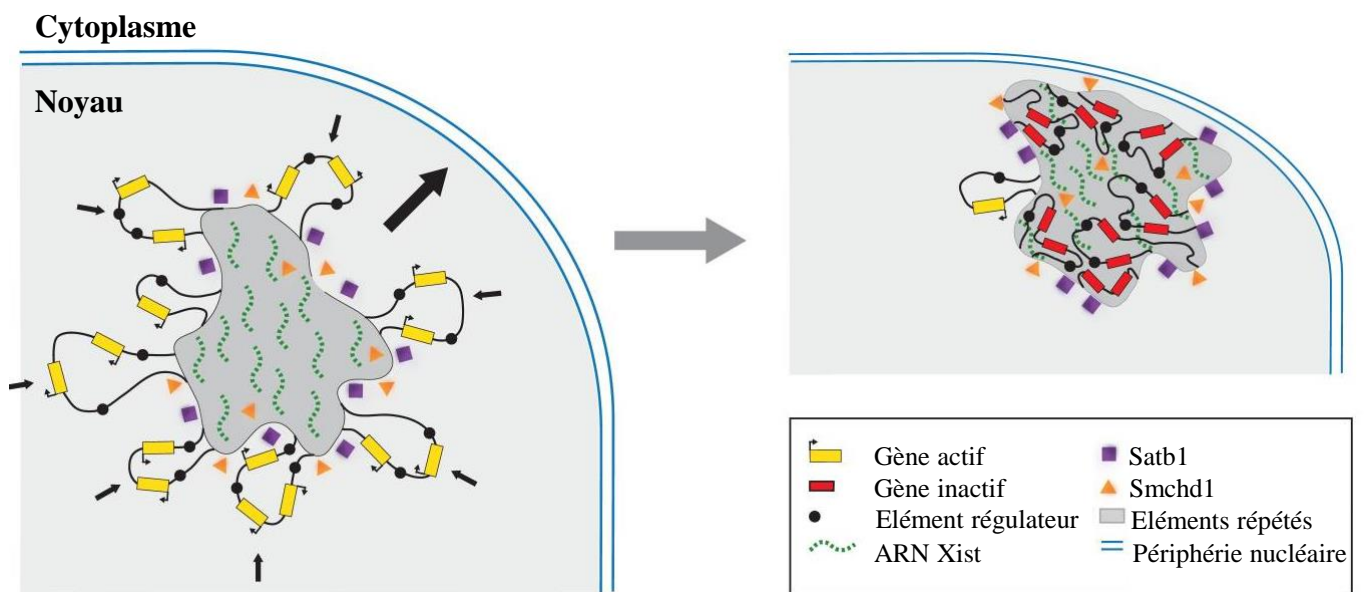


Figure 46. Représentation de l'inactivation des gènes et de la réorganisation spatiale du chromosome X

Lors du processus de l'inactivation de l'X, l'ARN Xist couvre tout d'abord les régions chromosomiques contenant des éléments répétés. Au cours de ce processus, les gènes et les autres éléments du génome, qui étaient précédemment en position externe, sont internalisés dans les territoires couverts par Xist. Le chromosome est également déplacé vers la périphérie du noyau. Les facteurs de structure de la chromatine comme SATB1 et Smchd1 sont potentiellement impliqués dans ce déplacement. (adapté de Brockdorff et Turner, 2015)

2011). Elle commence avec l'expression du long ARN non-codant Xist par le chromosome X à inactiver (Figure 46). La liaison de Xist au chromosome est suivie par le recrutement de facteurs de remodelage de la chromatine, dont les enzymes de la machinerie de la méthylation. D'autres facteurs de remodelage impliquent l'apposition de marques d'histones répressives (comme H2Ak119ub ou H3K27me3), et le retrait de marques permissives (comme les acétylations d'histones). L'organisation du chromosome et sa localisation nucléaire est profondément modifiée. La conséquence de ces remaniements est l'inactivation progressive d'une grande partie des gènes du chromosome X concerné. Cependant, certains gènes semblent échapper à cette inactivation, comme cela a été démontré chez l'Homme (Carrel et Willard 2005) ou chez la souris (Yang et al. 2010). L'échappement de certains gènes à l'inactivation ferait intervenir la protéine de liaison à l'ADN, CTCF (Giorgetti et al. 2016). Chez la souris, l'inactivation de l'X concerne dans un premier temps l'X paternel jusqu'au stade blastocyste, après une transition sans inactivation, elle se remet en place de manière aléatoire dans les cellules embryonnaires. Ainsi, le chromosome inactivé peut être d'origine paternelle ou maternelle ; au fil des divisions cellulaires, le chromosome inactivé reste le même.

1.4.4. Différenciation cellulaire et expression tissu-spécifique des gènes

Nous avons vu que la méthylation de l'ADN est un élément dynamique du génome, pouvant être apposée ou effacée par de multiples facteurs. Cette dynamique joue un rôle clef lors du processus de différenciation cellulaire. Au cours du développement, les cellules totipotentes, qui ont comme caractéristique la possibilité de donner naissance à n'importe quel lignage cellulaire, vont progressivement se différencier : tout d'abord en cellules pluripotentes (au potentiel de différenciation restreint) puis en cellules différenciées appartenant aux différents lignages embryonnaires. Au final, chaque type cellulaire doit pouvoir assurer les fonctions qui lui sont propres, et donc présenter un profil d'expression adapté.

En 2010, Laurent et son équipe ont comparé la méthylation des génomes de cellules souches embryonnaires humaines (hESC, de type pluripotentes), de lignées fibroblastiques issues de la différenciation des hESC, et de fibroblastes néonataux. Près de 80% des DMR identifiées entre les différents types cellulaires présentent une augmentation de la méthylation au cours de la différenciation. Les DMR hypométhylées dans les hESC sont associées à des gènes de la pluripotence, tels que ceux de la famille HOX. Pour cette famille de gènes, 286 DMR sont hypométhylées dans les hESC, contre 71 dans les fibroblastes néonataux. Finalement, 4 fois plus de DMR sont hypométhylées dans les cellules pluripotentes par rapport au premier stade

de différenciation. Ces DMR sont associées aux régions promotrices, TSS et exons de gènes de la pluripotente et du développement, ce qui suggère que la différenciation est associée à une réduction du profil d'expression des gènes par la méthylation.

Dans la plupart des tissus somatiques humains, en moyenne seuls 4% des CpG restent déméthylés quel que soit le type cellulaire (Varley et al. 2013). Ceux-ci sont associés au TSS des gènes de ménage dont l'expression est ubiquitaire. Le taux de méthylation des autres CpG est donc susceptible de varier d'un type cellulaire à l'autre. Il a également été montré que 21,8% des CpG sur les autosomes présentaient une méthylation variant de plus de 30% en fonction du tissu (Ziller et al. 2013). Les 2 tiers de ces CpG ont été retrouvés à distance des gènes, dans des régions identifiées comme sites de fixation de TF ou des *enhancers*. De plus ces CpG sont retrouvés de manière plus importante dans des *CpG shores* plutôt que dans des CGI.

En conclusion, chaque type cellulaire présente un profil de méthylation spécifique, qui a été mis en place au cours de la différenciation cellulaire. La méthylation participe alors à la régulation de l'expression des gènes tissus-spécifiques. Seule une fraction des CpG intervient dans ce processus, et sont situés à distance des gènes qu'ils régulent, rassemblés dans des régions de densité intermédiaire en CpG plutôt que dans des CGI.

1.4.5. Adaptation à des conditions environnementales

- L'exposition à des toxiques

L'environnement peut avoir un effet direct sur l'épigénome des individus. Par exemple, l'exposition chronique à de l'arsenic a été associée à la survenue de cancers de la peau, des poumons, des reins, du foie et de la vessie (*Arsenic in drinking water*. Washington, National Academy of Sciences, 2001). Il a été montré que les individus exposés présentent des altérations du méthylome de l'ADN lymphocytaire à l'âge adulte (Pilsner et al. 2007). Cependant, un point important à considérer est que l'environnement peut aussi avoir un effet multigénérationnel sur les individus. En effet, les embryons et fœtus peuvent être exposés indirectement aux facteurs environnementaux au travers de l'environnement maternel pendant la gestation. Toujours concernant l'arsenic, l'exposition *in utero* des fœtus a également été associée à une hausse du niveau de méthylation de l'ADN dans le sang de cordon, notamment des éléments répétés LINE1 et du promoteur des gènes *CDKN2A* ou *TP53* (Kile et al. 2012) (Intarasunanont et al. 2012) (Pilsner et al. 2012). Les observations faites au sujet de l'environnement *in utero* en association avec la survenue de pathologies après la naissance, a fondé le concept de la *Developmental Origins of Health and Disease* (DOHaD), qui postule

Espèce	Type de régime	Effets épigénétiques étudiés	Résultat	Références
Ovin	Déficiences en vitamine B12, folate et méthionine	Méthylation de l'ADN	Hypométhylation globale	Sinclair <i>et al.</i> , 2007
Poulet	Supplémentation en bétaine	Méthylation de l'ADN	Hyperméthylation globale	Hu <i>et al.</i> , 2015
Porcin	Supplémentation en bétaine	Méthylation de l'ADN et MPT d'histones	Hypométhylation et augmentation de H3K27me3 pour le gène PEPCK1. Hyperméthylation et augmentation de H3K4me3 pour les gènes PEPCK2 et FBP1	Cai <i>et al.</i> , 2014
Poulet	Supplémentation en zinc	Méthylation de l'ADN	Hypométhylation et augmentation de l'expression du gène A20	Li <i>et al.</i> , 2015
Porcin	Taux variables de zinc	Méthylation de l'ADN	Hyperméthylation dans les régions régulatrices du gène ZIP4	Karweina <i>et al.</i> , 2015
Porcin	Régime restrictif	MPT d'histones	Augmentation du taux global de H3K9ac, réduction de H3K9me3	Nebendahl <i>et al.</i> , 2016
Porcin	Supplémentation en acide folique	Méthylation de l'ADN	Hyperméthylation globale	Jing-bo <i>et al.</i> , 2013
Ovin	Insuffisance maternelle en protéine	Méthylation de l'ADN	Hypométhylation dans le cerveau	Begum <i>et al.</i> , 2012
Porcin	Supplémentation en micronutriments à effet méthylant	Méthylation de l'ADN	Différentiel de méthylation du gène IYD entre les groupes. Phénotypes obèses	Braunschweig <i>et al.</i> , 2012

Tableau 6. Liste des études faisant le lien entre nutrition et méthylation de l'ADN chez les espèces d'intérêt agronomique

que la survenue d'évènements pendant la période conceptionnelle peut avoir des conséquences sur la santé après la naissance et pour le reste de la vie de l'individu.

De nombreux chercheurs se sont penchés sur les impacts des composés environnementaux sur la méthylation de l'ADN. Il a été montré que les particules générées par le trafic routier peuvent induire une hypométhylation des transposons LINE1 dans les cellules sanguines chez l'Homme (Baccarelli et al. 2009). L'exposition à la fumée de cigarette chez la femme enceinte est associée à une augmentation de la méthylation de l'ADN pour la DMR du gène soumis à l'empreinte *IGF2* chez les descendants (Murphy et al. 2012). Cette augmentation est plus importante chez les descendants masculins que féminins, tendant à montrer que les effets de l'environnement sur le méthylome sont dans certains cas, spécifiques au sexe. Une étude parue en 2015 a montré que chez des souris exposées quotidiennement à la nicotine, on observe une expression différentielle de 15 gènes dans les testicules. Un d'entre eux, le gène de la Profiline 1 (*PFNI*) présentait un promoteur hypométhylé. Les spermatozoïdes des souris exposées avaient une motilité augmentée (Dai et al. 2015). Un grand nombre de composés peuvent avoir de la même façon une influence sur les niveaux de méthylation des ICR des gènes soumis à l'empreinte ou sur les éléments répétés, comme l'éthanol (Stouder, Somme, et Paoloni-Giacobino 2011), le benzène (Bollati et al. 2007), ou le bisphénol-A (Zhang et al. 2012).

- La nutrition

Un autre paramètre environnemental ayant une grande influence sur le méthylome est la nutrition. Un exemple connu chez l'Homme est le cas de la famine Hollandaise de 1944-1945. Les fœtus des femmes enceintes qui ont subi la famine, ont également été exposés à une restriction calorique très importante. Devenus adultes, ils ont présenté une hypométhylation de la DMR d'*IGF2* dans les cellules sanguines (Heijmans et al. 2008). Chez la souris, le pelage de type Agouti est sous contrôle de gènes dont le promoteur est influencé par le régime maternel (Waterland et Jirtle 2004). Chez les abeilles, alors que toutes les larves sont de génotype identique, une nutrition différentielle (avec ou sans gelée royale à tous les stades larvaires) va conférer un phénotype de Reine ou d'ouvrières, respectivement (Thompson et al. 2008). La plupart des dispositifs expérimentaux concernant l'étude de l'impact nutritionnel sur le méthylome sont basés sur la supplémentation ou la restriction en donneur de groupements méthyles ou en cofacteurs du cycle du folate. De nombreuses études de ce type ont été effectuées chez les espèces d'intérêt agronomique (répertoriées dans le Tableau 6). Mais aucune n'a analysé le méthylome spermatique de la descendance dans ce contexte de nutrition différentielle.

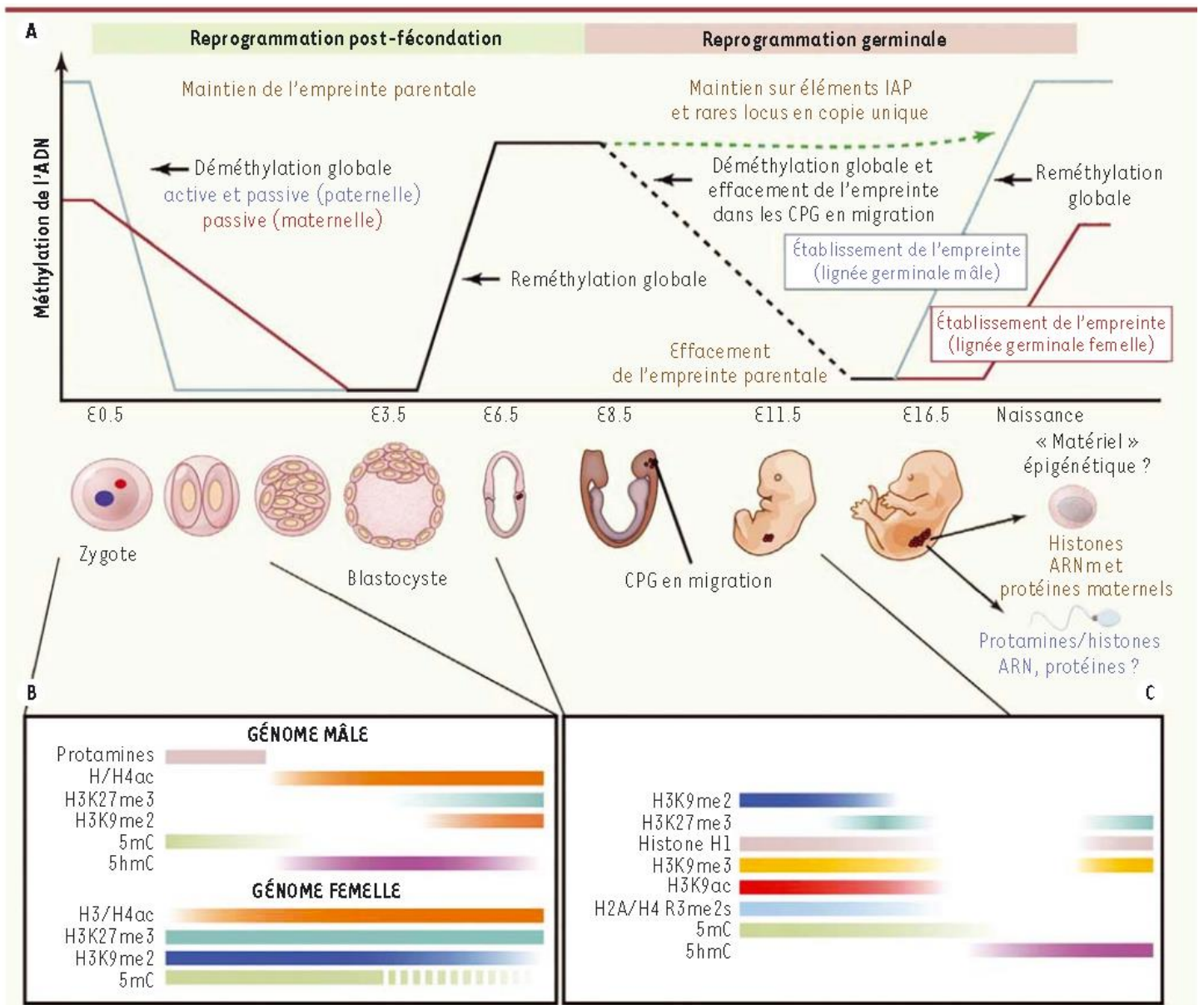


Figure 47. Les différentes phases de la reprogrammation épigénétique au cours du développement chez la souris

Le graphique **A** représente l'évolution de la méthylation de l'ADN au cours de la phase de reprogrammation post-fécondation puis dans la lignée germinale. L'encadré **B** représente quant à lui l'évolution des marques de méthylation de l'ADN, des modifications post-traductionnelles des histones et du remplacement des histones par les protamines dans les génomes paternel et maternel entre le stade zygotique et le stade 2 cellules. L'encadré **C** représente les modifications épigénétiques dans la lignée germinale au cours du développement embryonnaire.

(Tiré de Junien *et al.*, 2016)

- L'Aide Médicale à la Procréation (AMP)

En 2015 en France, 24 839 enfants sont nés grâce à l'aide médicale à la procréation (d'après l'Agence de la biomédecine), principalement par fécondation *in vitro* ou l'injection intracytoplasmique de spermatozoïde (*Intracytoplasmic Sperm Injection*, ICSI). Cependant, la manipulation des gamètes et la culture *in vitro* des embryons dans le cadre des AMP induit une augmentation de l'incidence de certains syndromes comme celui de Beckwith-Wiedemann ou d'Angelman. Pour les enfants nés après AMP, ces syndromes sont associés à des altérations de la méthylation sur les gènes soumis à l'empreinte, notamment *H19*, *IGF2* et *KCNQ1OT1* pour le syndrome Beckwith-Wiedemann, et *SNRPN* pour le syndrome d'Angelman. Alors que certaines études épidémiologiques sur de larges cohortes humaines n'ont pas montré d'augmentation significative de la fréquence d'apparition de troubles de l'empreinte consécutifs à l'AMP (Lidegaard, Pinborg, et Andersen 2005) (Bowdin et al. 2007), d'autres ont montré l'inverse (Sutcliffe et al. 2006) (Sanchez-Albisua et al. 2007). De la même façon, certaines études ont montré qu'il pouvait exister des altérations du méthylome chez des enfants sains et conçus par AMP, tandis que d'autres ne voient pas de différences (Puumala et al. 2012).

2. La reprogrammation épigénétique, un processus biologique fondamental pour la fertilité mâle

A certaines étapes clefs de la vie d'un individu, on assiste à de profondes restructurations de l'épigénome. Ces mécanismes sont connus sous le nom de reprogrammation épigénétique du génome. La reprogrammation consiste essentiellement en un effacement des marques épigénétiques préexistantes (méthylation de l'ADN et MPT des histones), suivi par l'apposition de nouvelles marques épigénétiques. Deux phases de reprogrammation majeures de l'épigénome sont décrites. La première se produit au cours de la différenciation des cellules germinales primordiales (PGC) et la gamétogenèse ; elle permet la transition d'un épigénome caractéristique de l'épiblaste à celui des PGC, puis à celui des gamètes. L'épigénome gamétique qui est ainsi mis en place joue un rôle capital dans les fonctions gamétiques, le développement embryonnaire après la fécondation et la régulation de l'expression des gènes à la génération suivante. La seconde phase de reprogrammation a lieu après la fécondation au cours du développement embryonnaire précoce, et permet de reprogrammer les épigénomes spermatiques et ovocytaires ultraspécialisés en un épigénome embryonnaire conférant un état totipotent aux cellules conduisant à la différenciation de tous

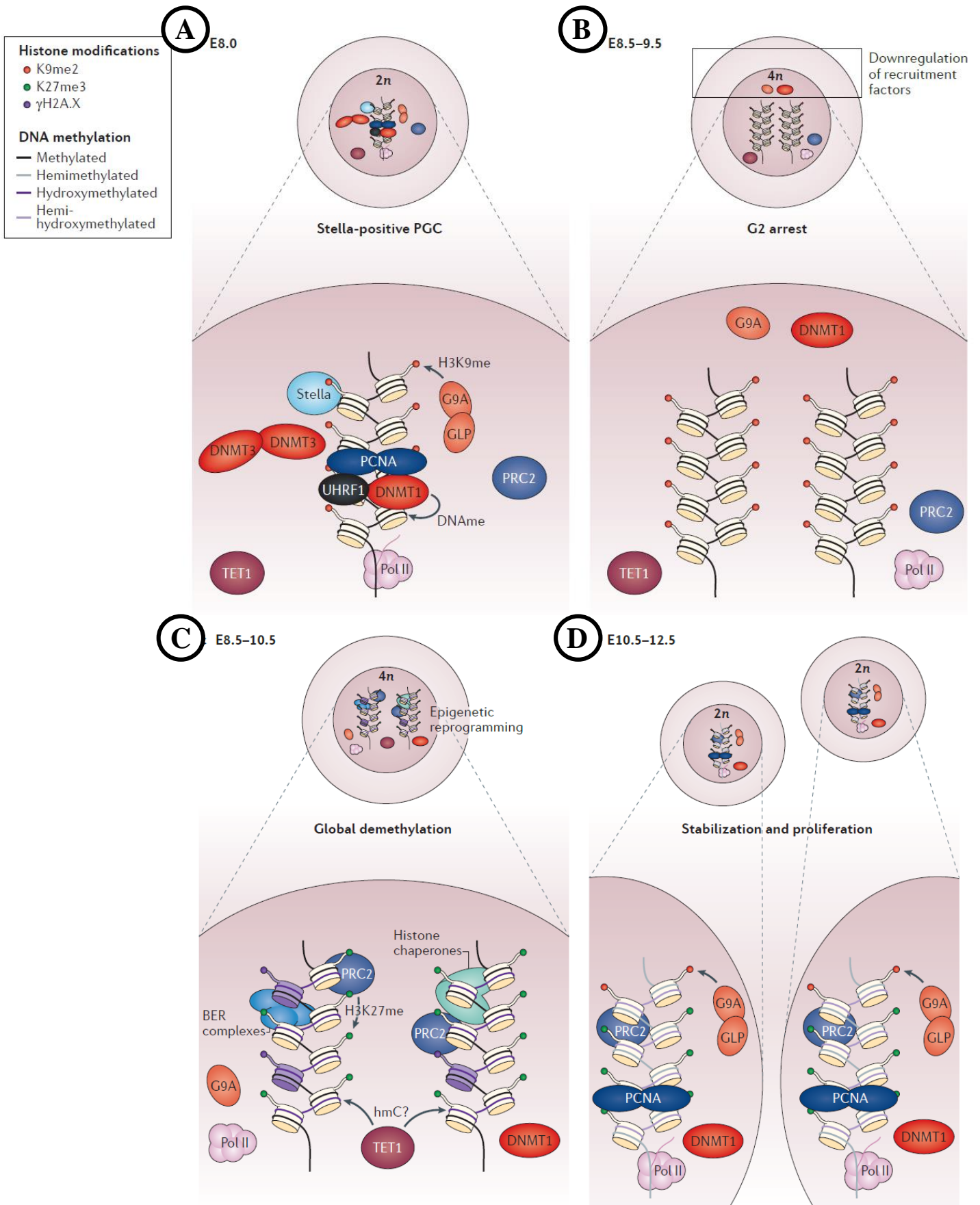


Figure 48. La reprogrammation épigénétique dans les PGC

A. A E8,0 chez la souris, les PGC présentent encore un profil épigénétique de type somatique, avec une maintenance de la méthylation de l'ADN par DNMT1. **B.** Après la réplication de l'ADN, on assiste à l'exclusion de DNMT1 et la répression transcriptionnelle des DNMT3. **C.** Mise en route du phénomène de déméthylation globale. Tet1 est exprimée. On assiste également au retrait des marques sur H3K9 et à l'apposition des marques sur H3K27. **D.** La déméthylation globale se poursuit de manière passive au cours des divisions cellulaires du fait de l'absence d'UHRF1. La transcription est ré-établie. H3K27 persiste au cours des divisions et H3K9 est ponctuellement apposée.

(tiré de Smith *et al.*, 2013)

les lignages cellulaires au cours du développement. Seuls des épigénomes gamétiques correctement apposés pourront être efficacement reprogrammés lors de cette seconde phase.

Les mécanismes épigénétiques mis en jeu au cours des phases de reprogrammation épigénétique pendant la gamétogenèse mâle sont décrits dans ce chapitre, en portant une attention particulière à la méthylation de l'ADN. Les données présentées sont principalement issues d'études réalisées chez l'Homme et la souris. A l'initiation de ce travail de thèse, peu de données étaient disponibles chez le bovin. Un schéma général des remaniements épigénétiques au cours de la reprogrammation chez la souris est donné Figure 47 (page précédente).

2.1. La reprogrammation dans la lignée germinale mâle

La reprogrammation épigénétique du génome des PGC est indispensable pour la suppression du profil épigénétique de type épiblastique acquis durant l'embryogenèse précoce, et l'apposition d'un profil épigénétique de type germinale permettant l'activation du programme transcriptionnel propre à la gamétogenèse en fonction du sexe de l'individu.

2.1.1. Déméthylation puis méthylation *de novo* dans les gonades fœtales

Comme vu au chapitre I, les PGC sont issues de la différenciation de cellules épiblastiques. L'effacement de l'épigénome épiblastique se déroule en 2 phases principales : une première pendant la migration des PGC, et une seconde après leur entrée dans les crêtes génitales. L'ensemble du processus intervient entre E7,5 - 13,5 chez la souris, entre la 2^{ème} et la 4,5^{ème} semaine de gestation chez l'Homme, et entre 25 et 31 jpc chez le bovin (Figure 48).

La première phase de déméthylation est rendue possible par la répression de la machinerie de méthylation, incluant DNMT3a, 3b et 3l, ainsi que par l'exclusion nucléaire d'UHRF1 (Ohno et al. 2013). La répression d'UHRF1 empêche le recrutement de DNMT1 aux sites de réplication de l'ADN pendant les mitoses, limitant son action de méthylation de maintenance. La première phase de déméthylation des PGC implique donc un blocage des mécanismes de méthylation *de novo*, en parallèle d'une déméthylation passive pendant la prolifération cellulaire. Des analyses de la méthylation globale dans les PGC chez la souris ont montré que le taux de méthylation passe successivement de 71 à 30% entre E6,5 et E9,5 (Seisenberger et al. 2012). A ce stade, les promoteurs de certains gènes comme *Nanog* sont concernés par cette déméthylation et sont exprimés.

Cette première phase de déméthylation globale n'affecte pas les gènes soumis à l'empreinte, en particulier les gènes soumis à l'empreinte maternelle. Un certain nombre de gènes impliqués dans la méiose et la différenciation des cellules germinales échappe également à cette phase de déméthylation précoce. Il a par exemple été montré que *DEAD-Box Helicase 4 (Ddx4)*, *Deleted in AZoospermia like (Dazl)* ou *SYnaptosomal Complex Protein 3 (Sycp3)* sont hyperméthylés à E10,5 chez la souris (Maatouk 2006). Le blocage de la déméthylation sur ces gènes pourrait impliquer le facteur Zpf57, car les gènes concernés présentent des sites de fixation pour ce facteur.

La seconde phase de déméthylation implique cette fois des mécanismes actifs. Entre E9,5 et E10,5 chez la souris, les gènes *Tet1* et *Tet2* sont fortement exprimés dans les PGC (Hackett et al. 2013) (Vincent et al. 2013). On observe alors une augmentation du taux d'5hmC (Tang et al. 2015). En parallèle de la voie des Tet, AID, une enzyme de la voie de déméthylation active du complexe BER, et la glycosylase TDG, sont également exprimées dans les PGC et participent à la déméthylation active (Popp et al. 2010) (Cortellino et al. 2011). En conséquence, la déméthylation globale des PGC se poursuit et touche la plupart des gènes, en intragénique et en région promotrice (Popp et al. 2010) (Guibert, Forne, et Weber 2012), et la plupart des CGI (Kobayashi et al. 2013). Cet effacement actif de la méthylation concerne également les gènes soumis à empreinte (comme *H19* ou *Kcnqlot1*) (Popp et al. 2010), ainsi que la plupart des éléments transposables. Les PGC conservent cependant une méthylation substantielle, et notamment pour certaines familles de transposons actifs comme l'*Intra-cisternal A Particle (IAP)* chez la souris (Hajkova et al. 2002), ainsi que des séquences satellites mineures et des rétrotransposons de type LINE1 (Kobayashi et al. 2013). La conservation d'une méthylation substantielle sur ces éléments aurait pour fonction la prévention de leur activation transcriptionnelle et donc assurerait une protection contre un risque accru de mutations par la dérégulation des gènes adjacents et la transposition.

En parallèle de l'effacement des marques de méthylation, on assiste à la suppression de la plupart des MPT d'histones de type répressif comme H3K9me2, H3K9me3, H3K64me3 ou H3K27me3, et de type permissif comme H3K9ac (Seki et al. 2007) (Daujat et al. 2009) (Mansour et al. 2012).

À E13,5 chez la souris, la méthylation atteint son niveau le plus bas, de l'ordre de 15% dans les PGC mâles et 7% chez les PGC femelles (Popp et al. 2010) (Seisenberger et al. 2012), ce qui en fait les cellules les plus hypométhylées chez les mammifères. Les gènes précédemment hyperméthylés à E10,5 comme *Ddx4*, *Dazl* ou *Sycp3* sont alors hypométhylés et exprimés (Maatouk 2006).

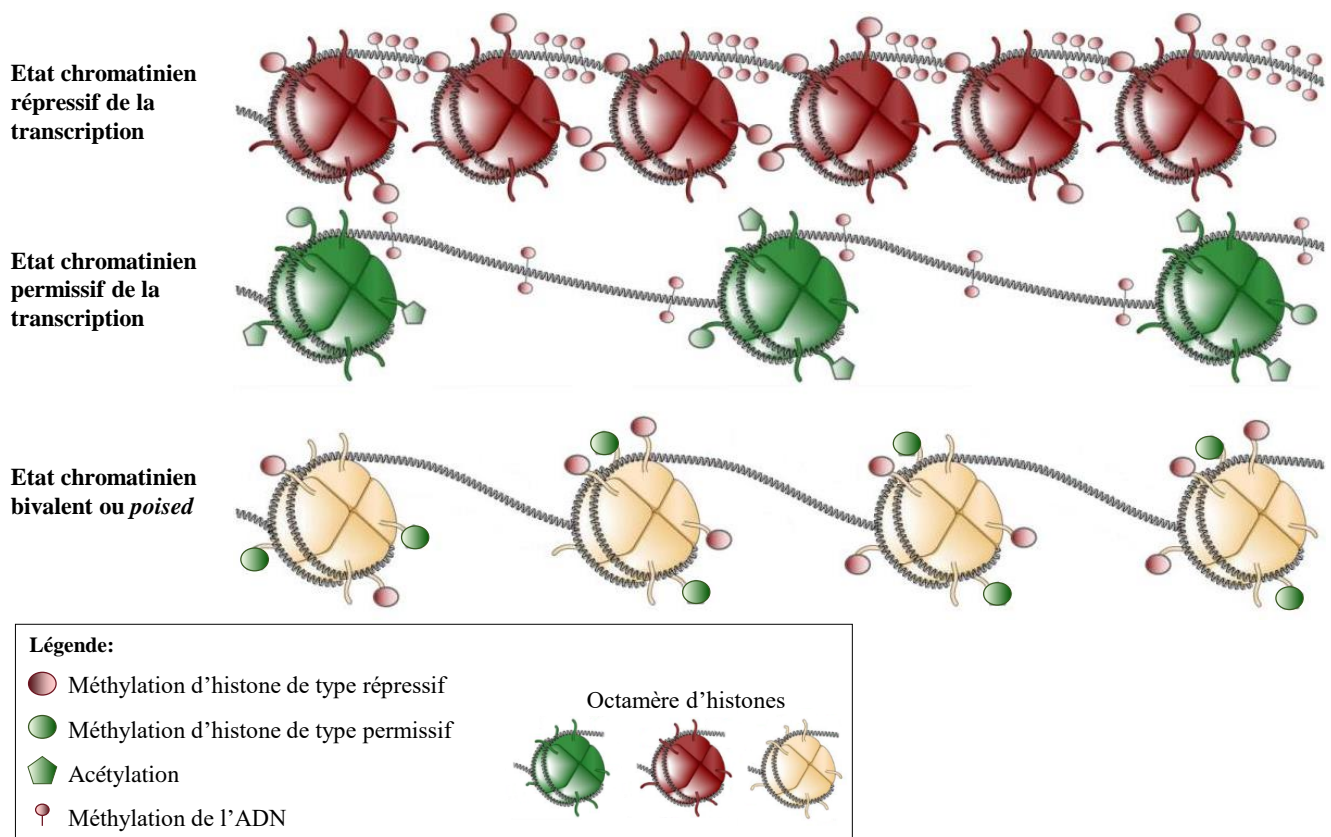


Figure 49. Etat chromatinien de type bivalent

Dans la lignée germinale chez les mammifères, les promoteurs de gènes associés à la régulation du développement présentent à la fois des MPT d'histones de types permissif et répressif. Par ce biais, les gènes de la pluripotence sont inactifs au cours de la gamétogénèse mais sont prêts à entrer rapidement en activation au cours du développement embryonnaire précoce. (adapté de Wierda *et al*, 2012)

Dès ce stade, on peut observer certaines régions chromatinienne dites bivalentes et qui sont caractérisées par la présence simultanée de MPT d'histones répressives (H3K27me3) et permissives (H3K4me3) (Sachs et al. 2013) (Figure 49). Les régions bivalentes présentent également une hypométhylation de l'ADN. Ces régions, qui n'ont pas subi la vague de reprogrammation, sont particulièrement retrouvées dans les promoteurs de facteurs de transcription impliqués dans la différenciation cellulaire et le développement embryonnaire, comme les facteurs de la famille *Hox*, *Fox*, *Sox*, *Tbx*, *GATA*, etc. On dit alors que les gènes en question sont « *poised* », c'est-à-dire « en position ». Cela signifie que ces gènes ne sont pas encore exprimés à ce stade, mais sont destinés après la fécondation à être activés très rapidement pour assurer leur fonction.

Au cours de la suite du développement, le génome subit une vague de méthylation *de novo*. De E14.5 à E16.5 chez la souris, les PGC reviennent à un taux de méthylation global de 50% (Seisenberger et al. 2012). A partir de E16,5 chez la souris et la semaine 19 de développement chez l'Homme, les PGC sont arrêtées en phase G1 de mitose, au stade pro-spermatogonies. L'apposition *de novo* des marques de méthylation nécessite l'expression de DNMT3a et DNMT3l, ainsi que DNMT3C chez la souris (Sakai et al. 2004) (Kaneda et al. 2004) (Kato et al. 2007). Cette étape de reméthylation est également marquée par l'acquisition progressive du profil parental des gènes soumis à l'empreinte. On observe par exemple la méthylation des ICR des gènes *H19* et *Dlk1-Meg3* (Henckel et al. 2012). La méthylation des gènes soumis à l'empreinte ne sera définitivement établie qu'après la naissance. La vague de méthylation en période foétale concerne aussi les séquences répétées. Aux stades plus précoces et durant la déméthylation massive du génome des PGC, nous avons vu que les IAP, les rétrotransposons LINE1 et certaines séquences satellites de type mineur (en position centromérique) conservaient un haut degré de méthylation. A E17,5, leur niveau de méthylation a encore augmenté (Lees-Murdock, De Felici, et Walsh 2003). Les séquences répétées de type SINEB1 ainsi que les séquences satellites majeures, qui elles n'ont pas échappé à la déméthylation, sont également reméthylées à ce stade. Les éléments répétés SINEB1 sont notamment méthylés par DNMT3a, les séquences satellites par DNMT3b, et les séquences IAP et LINE1 par DNMT3a et DNMT3b. La méthylation de l'ensemble des éléments répétés nécessite aussi la collaboration de DNMT3l avec les autres DNMT3, comme l'ont montré les expériences de KO chez la souris (Kato et al. 2007).

2.1.2. Remaniements épigénétiques après la naissance et dans la gonade adulte

Entre le troisième et le dixième jour après la naissance chez la souris, les spermatogonies se réactivent, et recommencent à proliférer pour coloniser les tubes séminifères. Une partie d'entre elles formera la réserve de spermatogonies (*spermatogonial stem cells*, SSC), tandis que les autres entreront en spermatogenèse ou en apoptose.

La production finale des gamètes mâles est mise en place à la puberté. On assiste à d'importantes restructurations de l'épigénome. Le spermatozoïde est en effet une cellule unique, que ce soit morphologiquement ou dans sa fonction biologique. L'acquisition de caractéristiques fonctionnelles particulières, concourant à la protection du patrimoine génétique paternel, à la motilité et à l'aptitude à pénétrer l'ovocyte confère au spermatozoïde un statut de cellule ultraspécialisée, nécessitant un programme épigénétique très différent de celui de l'ovocyte ou des cellules somatiques.

Au niveau de la méthylation de l'ADN, nous avons vu que celle-ci a été apposée au niveau des gènes soumis à l'empreinte au cours du développement fœtal au stade spermatogonie. Alors que les spermatogonies adultes sont largement hyperméthylées par rapport aux PGC ou aux cellules germinales fœtales, on observe encore la présence de plus de 330 promoteurs de gènes hypométhylés. Ces gènes ont la particularité d'intervenir dans le développement de la lignée germinale (notamment pour le processus de méiose ou pour la synthèse des piARN), ou dans le cadre de l'auto-renouvellement des cellules. Cela signifie qu'entre la période fœtale et post-natale, les gènes importants pour la méiose perdent leur méthylation (pour préparer leur activation pendant la spermatogenèse), alors que les gènes impliqués dans la prolifération sont méthylés et inactivés.

Des mécanismes de modification de la méthylation de l'ADN semblent encore actifs même en période adulte. En effet, chez l'Homme, les enzymes DNMT3a, DNMT3b et surtout DNMT1 sont exprimées tout au long de la spermatogenèse (Chen et Chan 2014). On n'observe cependant pas d'expression de DNMT3l. Les enzymes sont détectées dans le noyau jusqu'au stade de spermatides allongées, voire même dans différents compartiments des spermatozoïdes matures (Marques et al. 2011). L'expression de ces enzymes entraîne des changements subtils de méthylation sur certains gènes dont des gènes soumis à l'empreinte. Ce phénomène avait déjà été observé avant le stade pachytène des spermatocytes chez la souris, et concernait notamment des gènes comme *Tnp1*, *Prm1* et *Prm2*. De façon globale, le taux de méthylation est modifié de moins de 10% pour l'ensemble du génome au cours de la spermatogenèse chez l'adulte (Oakes et al. 2007).

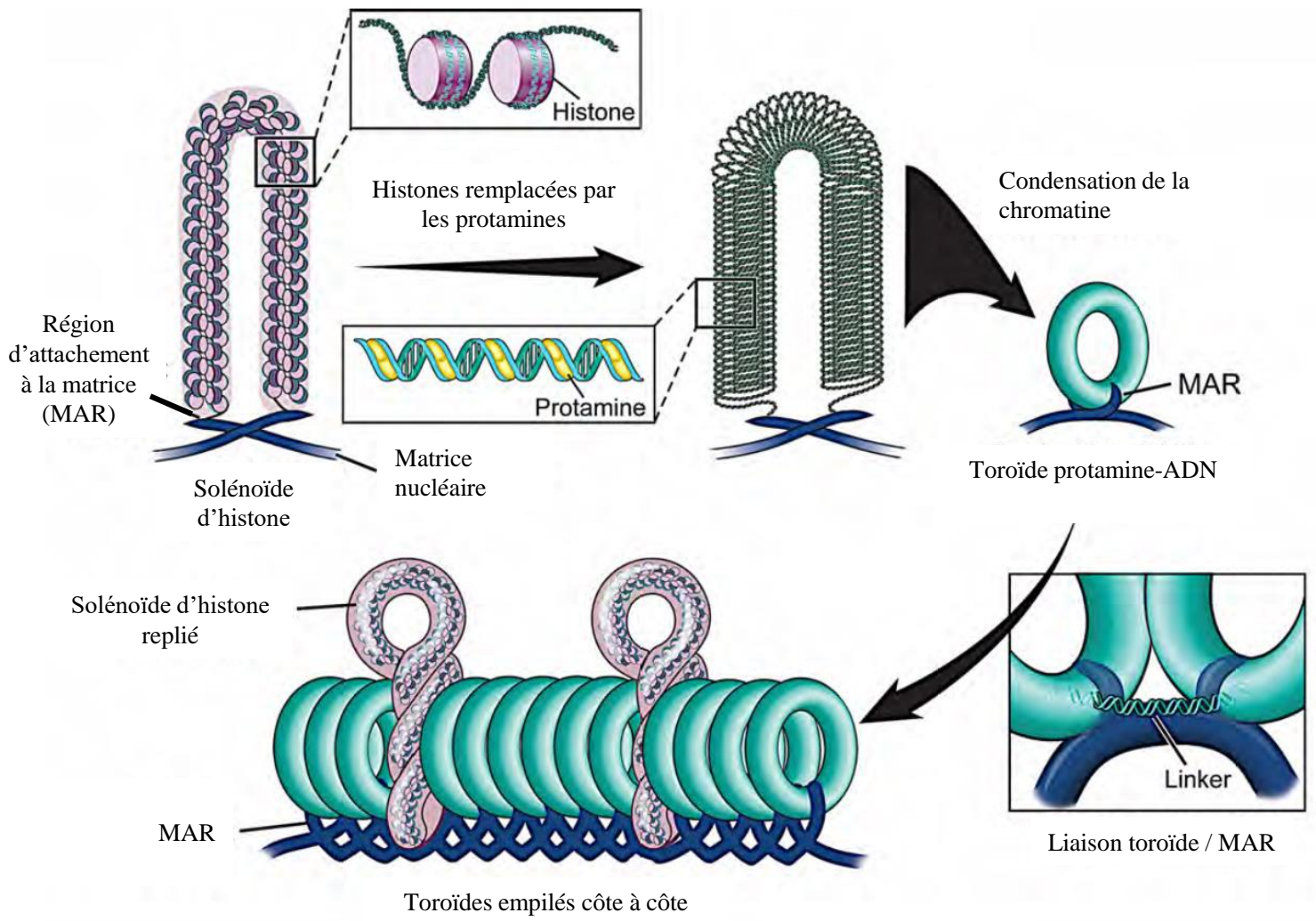


Figure 50. Remplacement des histones par les protamines
 (Traduit de la présentation de Sandro Esteves, *Sperm DNA Fragmentation in Male Infertility*)

Les promoteurs de la plupart des gènes de la pluripotence restent largement hypométhylés, dans les spermatozoïdes murins. D'autres gènes clefs de la pluripotence sont cependant hyperméthylés comme *Nanog*, *Lefty1*, *Brd1*, *Slc5a4a* dont la déméthylation se produira après la fécondation (Farthing et al. 2008).

Globalement, le génome spermatique est hypométhylé par rapport aux cellules somatiques (Hammoud et al. 2011) (Molaro et al. 2011) (Krausz et al. 2012), mais il est hyperméthylé par rapport à celui d'un ovocyte (Erkek et al. 2013). Nous ne décrivons pas davantage les séquences hypométhylées dans le spermatozoïde mature ici, puisque cela sera abordé dans la discussion incluse dans notre article intitulé « *Bull spermatozoa exhibit genome-wide DNA hypomethylation with both species peculiarities and conserved tissue-specific features* ».

Au stade spermatide, entre 85 à 99% des histones suivant les espèces, sont remplacées d'abord par des protéines de transition (TNP1 et TPN2), puis par des protamines (Brykczynska et al. 2010) (Figure 50). Le remplacement des histones par les protamines est aussi facilité par l'hyperacétylation des histones (notamment H3K4ac), permettant un état de relâchement de la chromatine (Hazzouri et al. 2000). Cette hyperacétylation est rendue possible par la dégradation d'HDAC1 dans les spermatides allongées (Choi et al. 2008).

Les protamines sont des protéines indispensables à la compaction finale du noyau spermatique et à l'obtention d'un spermatozoïde fonctionnel. Deux types de protamines ont été décrits, mais jusqu'à présent seule la protamine I a été isolée chez le bovin (Maier et al. 1990). Les protamines sont responsables de l'extrême compaction de la chromatine ; jusqu'à 20 fois plus que celle des cellules somatiques. En effet, les protamines contiennent un domaine peptidique très conservé et riche en acides aminés basiques comme l'arginine. Cette basicité, responsable d'une charge globale hautement positive des protamines, renforce leur affinité pour l'ADN qui est globalement porteur de charges négatives. De plus, les cystéines portées par les protamines sont responsables de la formation de ponts disulfure intra- et inter-protamines. La chromatine s'assemble alors sous la forme d'une macrostructure de 50kb spécifique aux spermatozoïdes appelée « toroïdes ». Les structures toroïdales sont attachées à la matrice nucléaire par les régions d'attachement à la matrice (*matrix attachment regions*, MAR).

Une partie des histones résiduelles est composée d'histones alternatives spécifiques ou non au spermatozoïde, (Talbert et Henikoff 2010). Il semblerait que les histones soient préférentiellement maintenues dans des régions contenant les gènes clefs du développement embryonnaire, notamment codants pour des TF embryonnaires, des gènes soumis à

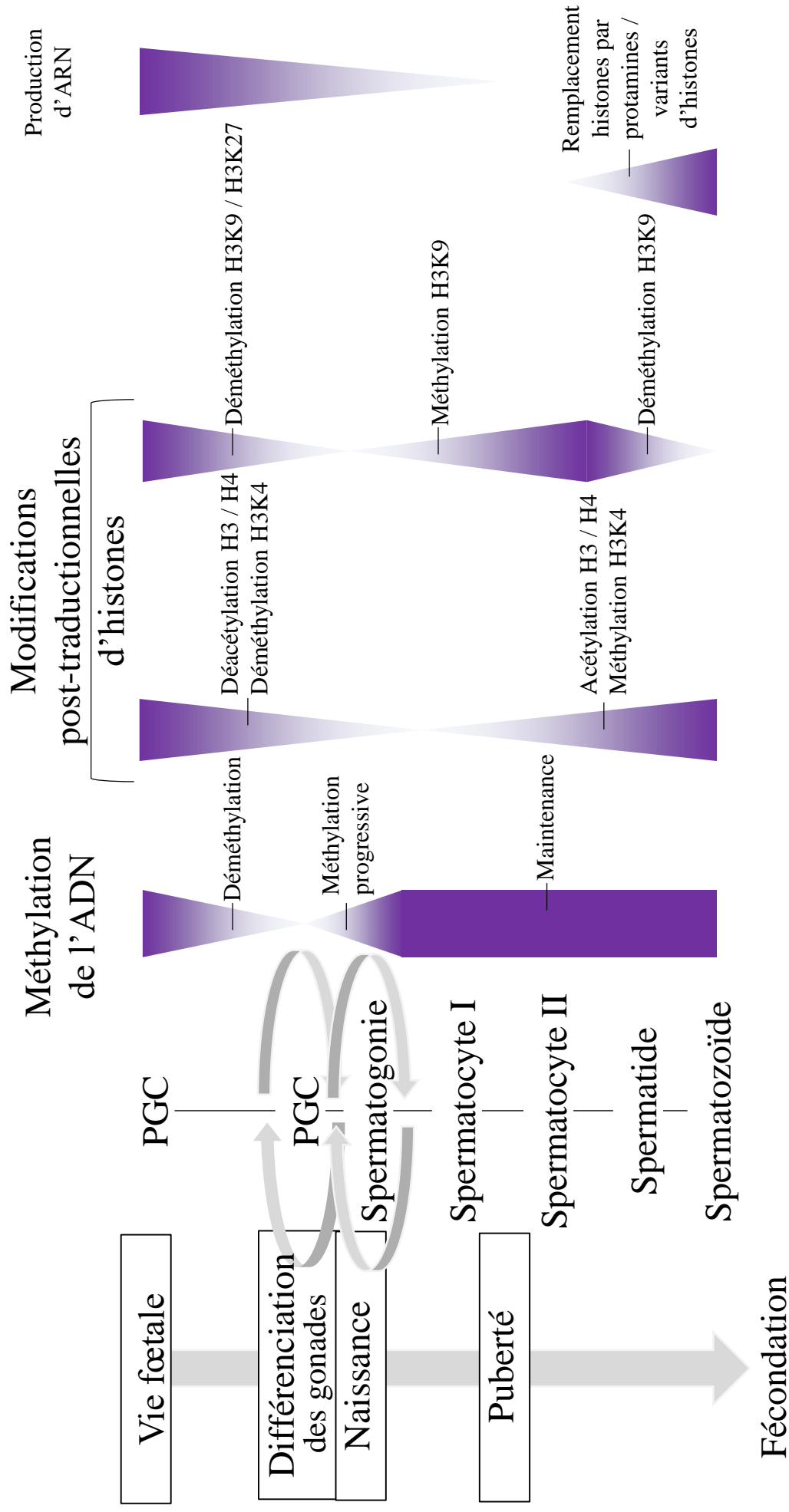


Figure 51. Modifications des marques épigénétiques au cours de la spermatogénèse

Au cours de la spermatogénèse, on assiste à de profond remaniements des marques épigénétiques (méthylation de l'ADN, modifications post-traductionnelles des histones). La spermatogénèse est également marquée par le remplacement de la plupart des histones par des protéines de transition, puis par des protamines. La production d'ARN est arrêtée au stade spermatozoïde. (adapté de: Boissonnas *et al.*, 2013)

l’empreinte et des promoteurs de gènes à microRNA (Hammoud et al. 2009). Mais leur localisation est encore sujet à débat : en 2014, Samans a publié une étude montrant que chez l’Homme et le bovin, les histones restantes sont retrouvées dans des régions intergéniques et introniques, et associées à des éléments répétés centromériques ou de type rétrotransposons. Chez la souris et l’Homme, les régions associées à ces histones semblent majoritairement hypométhylées (Erkek et al. 2013) (Samans et al. 2014). D’autres équipes ont montré la présence d’histones dans les séquences satellites et centromériques mais aussi dans les gènes des ARN ribosomiaux dans les spermatozoïdes bovins (Sillaste et al. 2017). Au cours de la spermatogenèse on assiste aussi au profond remaniement des MPT d’histones (Singh et al. 2013) (Lesch et al. 2013). On observe toujours la présence de domaines bivalents chromatinien, que ce soit dans les cellules germinales ainsi que dans les spermatides (Erkek et al. 2013) (Lesch et al. 2013). Les bivalents sont cependant plus nombreux que dans les cellules germinales fœtales, notamment sur des gènes de FT *Zinc Finger*, de protocadhérines, de la voie de signalisation Wnt, de la famille des PRDM, etc. (Hammoud et al. 2014).

Les remaniements décrits ci-dessus conduisent à un arrêt de la transcription au stade spermatide (Figure 51). Dans le noyau du spermatozoïde mature il reste cependant encore un grand nombre d’ARN non-codants comme des miRNA ou des piRNA, ainsi que des fragments d’ARNm, ARNr et ARNt. Ces ARN sont transmis à l’ovocyte lors de la fécondation et sont nécessaires au développement embryonnaire (Hayashi 2003). On observe par exemple la présence d’ARNm qui permettront la production de récepteurs hormonaux (Dadoune 2009) L’expression de ces ARN est mise en place avant la phase pachytène de la méiose (Hammoud et al. 2014) (Li et al. 2013).

2.2. La reprogrammation de l'épigénome embryonnaire

Suite à la fécondation, une profonde réorganisation de l'épigénome a lieu dans le zygote. Cette réorganisation est nécessaire pour établir un état de totipotence nécessaire aux premières étapes du développement embryonnaire et permettant la différenciation des lignages cellulaires constituant le zygote.

La reprogrammation consiste tout d’abord en une déméthylation globale des génomes paternel et maternel, de manière asynchrone (Figure 52, page suivante). Le génome paternel est déméthylé de manière active par la machinerie des TET, et notamment TET3. Cela induit une rapide diminution du niveau de 5mC, ainsi qu’une accumulation de 5hmC, et dans une moindre mesure de 5fC et 5caC (Wossidlo et al. 2011). Le destin de la 5hmC est encore sujet

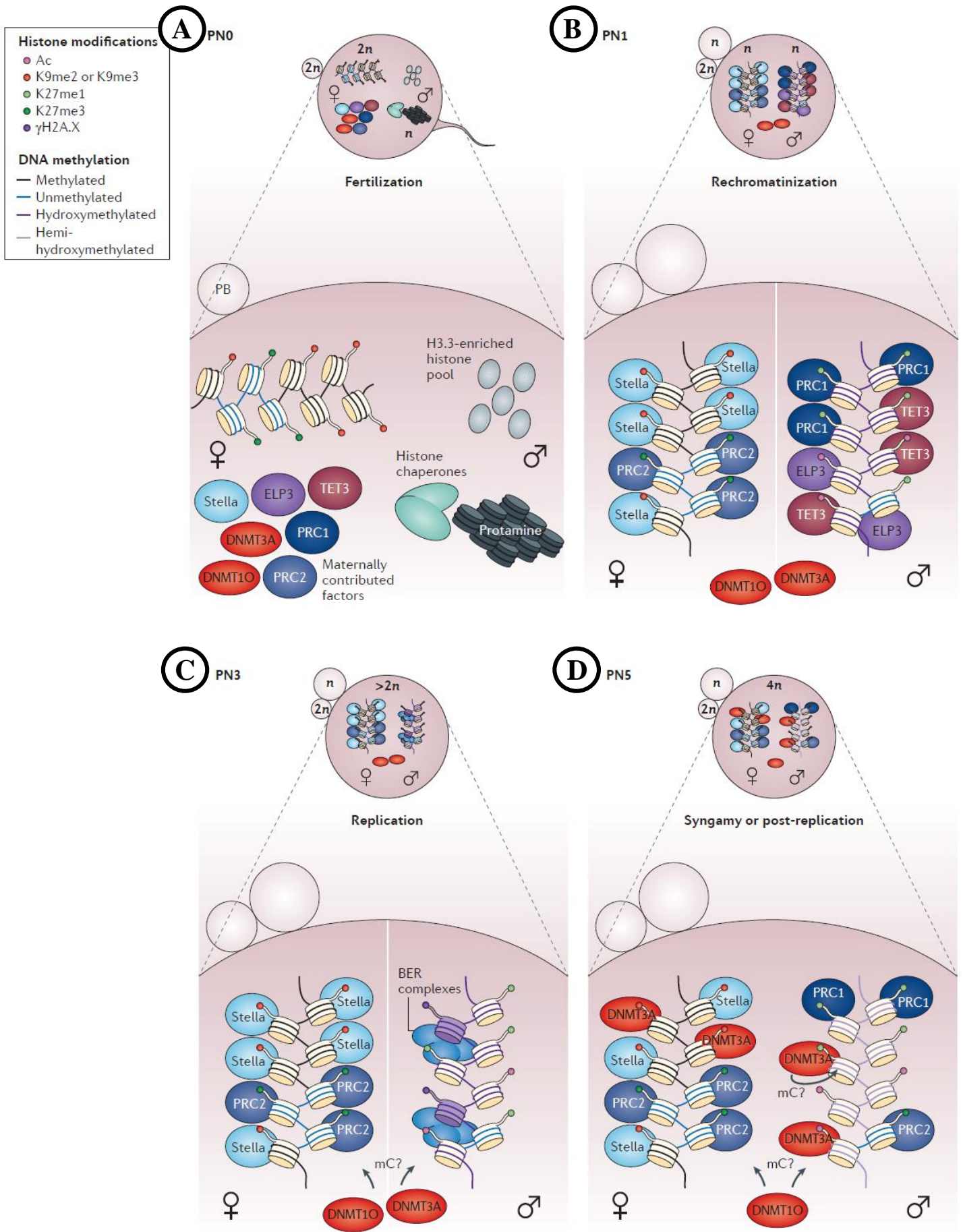


Figure 52. La reprogrammation épigénétique dans le zygote

A. Après la fécondation, les protamines paternelles sont rapidement désassemblées. **B.** Pendant l'incorporation d'histones d'origine maternelles dans le *pronucleus* mâle se déroule également leur acétylation, l'apposition de marques de méthylation H3K27 et l'hydroxyméthylation de l'ADN par Tet3. Le génome maternel est protégé de la déméthylation par Stella et PRC2. **C.** On observe le recrutement de complexe BER au génome paternel, qui est globalement hydroxyméthylé. **D.** Après la syngamie, la méthylation *de novo* des génomes est conduite par DNMT3A.

(tiré de Smith *et al.*, 2013)

à débat, et pourrait impliquer le mécanisme d'excision-réparation de base (Hajkova et al. 2010). Dans les cellules souches embryonnaires de souris, la 5hmC est principalement retrouvée au niveau des TSS et dans les exons des gènes (Ficz et al. 2011). La déméthylation des gènes soumis à l'empreinte et de certains éléments transposables comme les IAP est réprimée, notamment par des facteurs comme *Dppa3* ou *Zfp57* (Nakamura et al. 2007) (Li et al. 2008). D'autres éléments répétés, comme LINE1 et certaines classes de LTR, sont cependant activement déméthylés. La déméthylation du génome maternel se fait de manière passive au cours des premières divisions cellulaires, probablement par l'exclusion nucléaire de DNMT1 ovocytaire (Cardoso, Sporbert, et Leonhardt 1999). Il semble cependant conserver une activité limitée pour le maintien des profils de méthylation sur les gènes soumis à l'empreinte et de certaines séquences répétées (Hirasawa et al. 2008).

La cinétique de déméthylation varie selon les espèces : alors qu'elle est déjà à un stade très avancé au stade 1 cellule chez la souris, elle n'est réalisée qu'au stade 2 cellules chez l'Homme (Guo et al. 2014), et ne commence distinctement qu'au stade 4 cellules chez le lapin (Reis e Silva et al. 2012). Il semble alors que la cinétique de déméthylation coïncide avec l'activation du génome embryonnaire (qui a lieu chez le bovin au stade 8 cellules).

Très rapidement après la fécondation, les protamines du génome paternel sont remplacées par les histones ovocytaires (Burton et Torres-Padilla 2010). Les MPT d'histones sont également reprogrammées dans les génomes parentaux. Au stade une cellule chez la souris, le génome maternel est riche en H3K4me3, H3K9me2/3, H3K27me2/3, H4K20me3, H3K64me3. La présence de H3K9me3 permet le recrutement de la protéine Stella, qui protège le génome maternel de la déméthylation active (Nakamura et al. 2012). Ces MPT sont cependant absentes du génome paternel, qui au contraire présente plutôt des marques telles que H3K9me1 et H4K20me1, ainsi qu'une hyperacétylation de l'histone H4.

Au stade morula le génome embryonnaire est globalement hypométhylé. C'est au cours de la différenciation des premiers lignages cellulaires que la méthylation est réapposée. L'ICM apparait globalement hyperméthylée, alors que le trophoctoderme reste à un niveau de méthylation intermédiaire (Santos et Dean 2004). A ce stade les niveaux de 5hmC redeviennent faibles (Okamoto et al. 2016). Les profils de méthylation établis au cours des lignages deviennent progressivement de type somatique. Cependant, tout comme lors de la reprogrammation des gamètes, la reprogrammation des marques de méthylation post fécondation dans l'embryon n'est pas totale, en plus des gènes soumis à l'empreinte parentale, certaines régions ou *loci* sont protégés de cette reprogrammation épigénétique ce qui laisse la

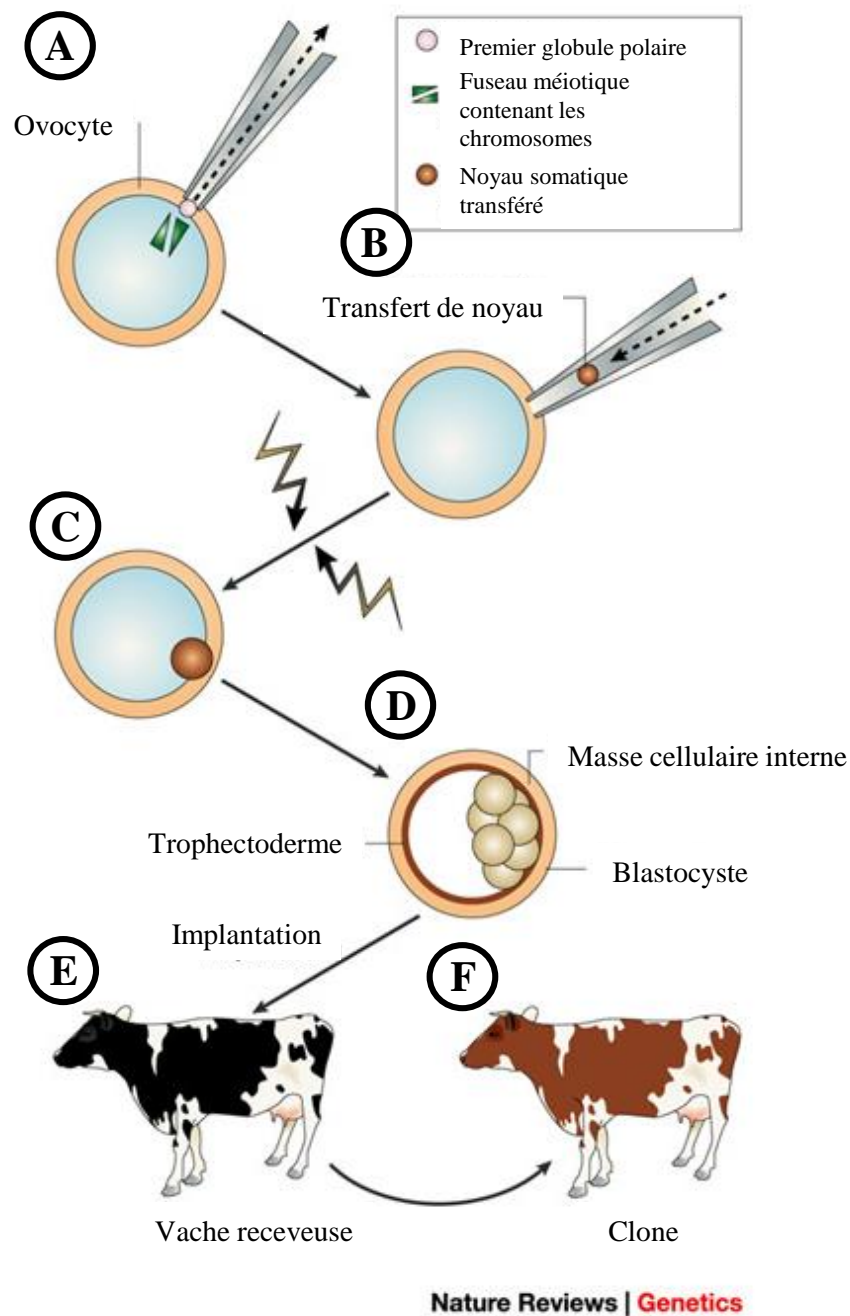


Figure 53. Principe du clonage par transfert nucléaire (*Somatic Cell Nuclear Transfer*)

A. Un ovocyte accepteur est énucléé. B. Le noyau d'une cellule somatique, prélevée sur l'organisme donneur, est microinjecté dans l'ovocyte énucléé. C. Un choc électrique est appliqué dans le but de faire fusionner la membrane de l'ovocyte accepteur avec celle de la cellule somatique donneuse de noyau. D. Les embryons ainsi reconstitués sont mis en culture jusqu'au stade blastocyste. E. L'embryon est transféré à une mère porteuse. F. L'embryon issu du transfert nucléaire peut ainsi donner naissance à un individu génétiquement identique à l'individu donneur de la cellule somatique.

(Tiré de Rhind *et al.*, 2003)

place à une transmission d'informations épigénétiques multigénérationnelle (voir chapitre suivant).

Un modèle important de perturbation de la reprogrammation épigénétique au cours du développement embryonnaire précoce est porté par les travaux sur le clonage (Figure 53). Le clonage reproductif (*Somatic Cell Nuclear Transfer*, SCNT) est un mode de reproduction assistée qui consiste à fusionner une cellule somatique et un ovocyte préalablement énucléé et à transférer ainsi le noyau somatique par micromanipulation. Après l'enclenchement des divisions cellulaires du zygote par traitement chimique ou par électrostimulation, l'embryon en développement est la copie génétique de l'individu dont est issue la cellule somatique. Ce processus, quoique très peu efficace, est rendu possible par la faculté de l'ovocyte à reprogrammer le noyau de la cellule somatique, pour obtenir un état de totipotence, ce qui permet d'initialiser le développement. Les facteurs ovocytaires permettant cette reprogrammation, sont toujours sujets à débat, mais pourraient concerner la machinerie épigénétique. En effet, de nombreux travaux montrent que des drogues épigénétiques administrées dans le milieu de culture (notamment la Trichostatine A) permettaient d'améliorer l'efficacité de la reprogrammation (Oh et al. 2012). L'état épigénétique du génome somatique pourrait également influencer l'efficacité de la reprogrammation. Chez la souris, il semblerait que la marque H3K9me3 soit un frein à la reprogrammation, et que sa suppression dans les cellules donneuses par surexpression d'une histone déméthylase pourrait favoriser la reprogrammation (Matoba et al. 2014). L'importance de la composante paternelle dans la reprogrammation épigénétique est quant à elle soulignée par des travaux montrant que la « protaminisation » du génome des cellules donneuses par surexpression du gène PRM1 améliorait la reprogrammation (Iuso et al. 2015). En utilisant le clonage somatique, les pertes embryonnaires et fœtales sont cependant très importantes. Chez le bovin, seuls 30 à 50% des embryons clonés atteignent le stade blastocyste, et seuls 11,5 % arriveront à terme (Keefer 2015). Il semblerait qu'une reprogrammation épigénétique incomplète soit à l'origine de ces pertes embryonnaires et des anomalies de gestation rencontrées à des stades plus tardifs. Il a notamment été montré que les embryons pré- et post-implantatoires issus de SCNT étaient hyperméthylés par rapport à des blastocystes issus d'IA (Couldrey et Lee 2010) (Couldrey, Wells, et Lee 2011) (Couldrey et Wells 2013). Différentes études ont montré que les perturbations de la méthylation pouvaient être retrouvées à un niveau global y compris chez des clones adultes sains (De Montera et al. 2010) comme locus spécifique, et notamment sur des gènes soumis à l'empreinte (Smith et al. 2015) ou du métabolisme hépatique (Kiefer et al. 2016).

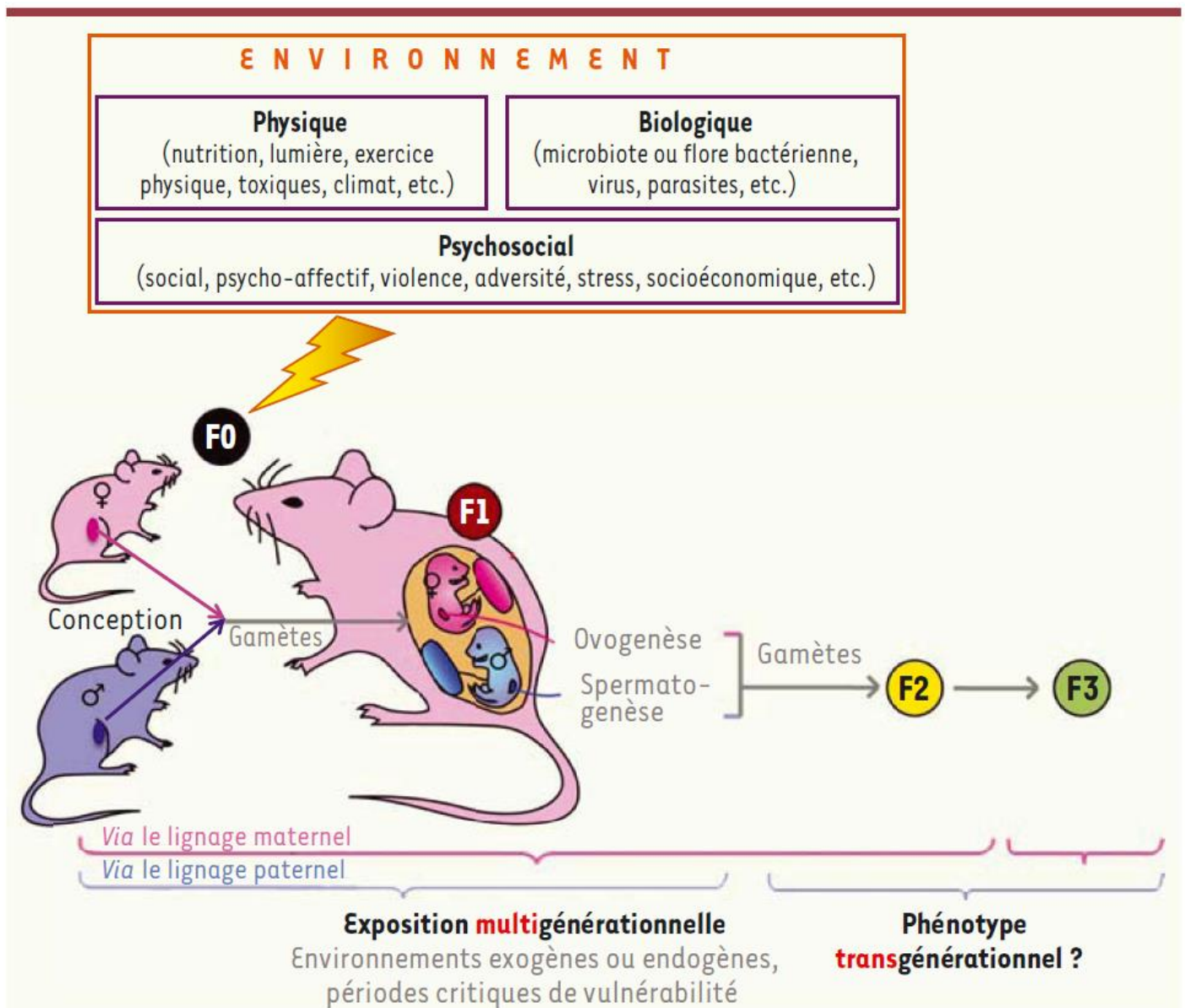


Figure 54. Effets multi et transgénérationnels

L'environnement pris au sens large (que ce soit l'exposition à des toxiques, la nutrition, ou encore des facteurs psychosociaux comme le stress) peut modifier l'épigénome de la génération F0, dont l'épigénome gamétique. Ces modifications peuvent être transmises à la génération F1. Si l'individu F0 est une femelle gestante, les modifications épigénétiques peuvent également impacter directement les cellules germinales des embryons en développement. Les modifications de l'épigénome sont alors directement transmises à la génération F2, qui sera issue des cellules germinales de la génération F1. On parle alors d'effets multigénérationnels lorsque des modifications épigénétiques sont transmises à l'embryon ou aux cellules germinales qu'il porte, et d'effets transgénérationnels lorsque les modifications sont transmises aux générations suivantes qui n'ont jamais été exposées. Ainsi, l'exposition de la mère entraîne des effets multigénérationnels pour les générations F1 et F2, alors que l'exposition paternelle entraîne uniquement des effets multigénérationnels pour la génération F1. On ne peut parler d'effets transgénérationnels qu'à partir de la génération F3 pour la voie femelle, et F2 pour la voie mâle.

(tiré de Junien *et al.*, 2016)

Ces travaux soulignent le fait qu'un bon déroulement des étapes de reprogrammation, et une bonne apposition des marques de méthylation et notamment sur les gènes soumis à l'empreinte, sont fondamentaux pour le bon déroulement du processus de reproduction. Ces travaux suggèrent aussi fortement que si un embryon se développe à partir d'un spermatozoïde dont l'épigénome est altéré (représenté dans le cas du SCNT par une cellule avec un épigénome de type somatique), les pertes embryonnaires seront importantes.

2.3. Effets multigénérationnels et transgénérationnels

On a longtemps considéré que les phases de reprogrammation épigénétique empêchaient toute forme de transmission des marques épigénétiques acquises durant la vie de l'individu à la génération suivante. Cependant, il est possible que certaines d'entre elles puissent être transmises directement aux générations exposées *in utero* ainsi qu'à leurs gamètes (un phénomène appelé transmission multigénérationnelle), voir même aux générations ultérieures qui n'ont pas été exposées (transmission transgénérationnelle). La transmission multi- ou transgénérationnelle des expositions à l'environnement diffère en fonction du sexe des parents et de leurs descendants. (Figure 54).

Les transmissions concernent toutes les marques épigénétiques. Par exemple, les spermatozoïdes sont porteurs d'ARN qui sont transmis à l'embryon. Les histones conservées dans les spermatozoïdes matures (de 1 à 15% des histones suivant l'espèce considérée), ainsi que les MPT qu'elles portent, sont également transmises par ce biais. De la même façon, on sait aujourd'hui qu'au cours de la reprogrammation épigénétique dans les cellules germinales certains CpG échappent à cette reprogrammation.

En 2010, l'équipe de recherche de Skinner a étudié l'exposition de rattes gestantes à des perturbateurs endocriniens (vinclozoline et méthoxychlore), et ses effets sur la fertilité et le méthylome spermatique des descendants mâles exposés *in utero* (Anway et al. 2005). Ils ont observé des perturbations de la méthylation jusqu'à la 4^{ème} génération, c'est-à-dire chez des individus qui n'ont pas été exposés aux perturbateurs, ce qui mettrait en lumière la transmission de profils de méthylation de type transgénérationnel. Il faut cependant souligner le fait que ces travaux sont encore aujourd'hui controversés, notamment dans la conduite de l'expérimentation.

D'autres travaux de même type ont été réalisés et leurs auteurs ont conclu ne pas observer de différences significatives dans la méthylation de l'ADN des cellules germinales des mâles exposés *in utero* aux mêmes composés chimiques (F1), ni de leur descendants (F2) (Iqbal et al.

2015). De la même façon, des critiques concernant cette dernière étude ont été émises du point de vue du nombre d'échantillons utilisés (trop faible) et de l'interprétation des données (Guerrero-Bosagna 2016).

Enfin, d'autres travaux importants ont consisté à exposer des souris mâles à de l'acétophénone (composé ayant une odeur semblable à l'amande) 5 fois par jours pendant 3 jours. A chaque exposition, les souris recevaient également un choc électrique afin d'associer l'odorant à un stress. Après 10 jours sans exposition à l'acétophénone ni choc électrique, les mâles ont été accouplés avec des femelles non-exposées. Il a été montré que les descendants (F1), ainsi que les descendants des descendants (F2), présentaient des modifications neuroanatomiques au niveau du bulbe olfactif (Dias et Ressler 2014). Les individus F1, tous comme les individus F0, présentaient également une hypométhylation du gène *Olfr151*. Ce gène code pour le récepteur olfactif impliqué dans la reconnaissance de l'acétophénone. En conséquence, des altérations du méthylome spermatique paternel et consécutives à des expositions environnementales, peuvent dans certains cas être transmises au travers des générations et avoir une influence fonctionnelle sur les descendants.

Les régions qui échappent à l'effacement du méthylome au cours des phases de reprogrammation sont qualifiées de « fugitives » (*escapes*). Récemment, il s'est avéré que beaucoup d'entre elles concernent des gènes qui participent au développement neural (Tang et al. 2015). Les *escapes* semblent aussi situés dans des régions proches des séquences répétées comme les IAP chez la souris (Seisenberger et al. 2012) (Guibert, Forne, et Weber 2012).

3. Méthylation de l'ADN et infertilité mâle

Depuis une décennie, un grand nombre d'études a montré que des perturbations épigénétiques pouvaient être retrouvées chez des individus à fertilité altérée. Nous présenterons ici les études qui traitent de la méthylation de l'ADN, mais il existe beaucoup d'autres études portant sur un ou plusieurs autres types de marques épigénétiques. Sont présentées tout d'abord les manipulations pharmacologiques et génétiques qui ont été effectuées chez la souris, puis dans un second temps les études sur cohortes humaines, pour terminer avec l'état de l'art chez le bovin. Les méthodes d'analyses de la méthylation aperçues ici dans les différentes études décrites sont détaillées dans le chapitre Méthodologie.

3.1. Manipulations pharmacologiques et génétiques chez la souris

Chez les animaux modèles comme la souris ou le rat, de nombreuses équipes de recherche ont étudié l'impact d'altérations du méthylome sur la spermatogenèse et le développement embryonnaire.

L'inactivation conditionnelle de *Dnmt3a* ou *Dnmt3l* dans les cellules germinales embryonnaires de souris mâle conduit à une totale stérilité, marquée par une absence de production de spermatozoïdes à l'âge adulte (Bourc'his 2001) (Bourc'his et Bestor 2004). En effet, les cellules germinales ne se développent pas et s'arrêtent au stade de spermatogonies (Kaneda et al. 2004). La stérilité serait due soit à des échecs de la méiose soit à une augmentation importante de l'apoptose. Dans les spermatogonies à l'âge adulte, on observe avec l'inactivation de *Dnmt3a* une hypométhylation des gènes *H19* et *Dlk1*, et dans une moindre mesure de *Rasgrfl* (Kaneda et al. 2004) (Kato et al. 2007). Alors qu'avec l'inactivation de *Dnmt3b*, seul *Rasgrfl* est légèrement affecté (Kato et al. 2007). Cela montre que les différentes DNMT3, a et b, n'ont pas les mêmes cibles, et que le gène *Rasgrfl* nécessite l'action des 2 types de DNMT3 pour être méthylé. L'inactivation de *Dnmt3l* engendre l'hypométhylation de *H19*, *Rasgrfl*, ainsi que de *Dlk1* (Kaneda et al. 2004), mais aussi celle d'éléments répétés comme des LTR et des IAP (Bourc'his et Bestor 2004).

D'autres études se sont focalisées sur l'impact des piRNA sur la méthylation de l'ADN. En 2007, Aravin a montré qu'une déficience en Mili ou Miwi2, qui sont impliquées dans la voie des piRNA induit une surexpression des éléments répétés L1 et IAP d'un facteur 5 ou 10. Le phénotype des individus est similaire à celui qu'on observe avec l'inactivation de *Dnmt3l* (Aravin, Hannon, et Brennecke 2007).

La Décitabine (5-aza-2'-deoxycytidine, 5-azadC) est une molécule utilisée comme anticancéreux. Il s'agit d'un dérivé de 5mC, qui après son incorporation dans l'ADN au cours de la réplication lie de façon covalente les DNMT. Les DNMT, et notamment les DNMT3a et 3b, sont alors dégradées, ce qui conduit à une diminution de la méthylation avec le temps (Song et al. 2016). L'utilisation de ce composé est généralement associée avec une réduction de la fertilité et des paramètres spermatiques, comme la concentration spermatique et la motilité (Oakes et al. 2007). Son injection chez des souris mâles adultes semble montrer que ce composé a des répercussions phénotypiques directes, comme une réduction du poids testiculaire et de la taille des tubes séminifères. On observe aussi la diminution du nombre de cellules haploïdes par rapport aux cellules diploïdes, probablement causée par une apoptose massive des cellules au stade spermatocyte (Song et al. 2016). Les spermatozoïdes produits

sont aussi moins nombreux et présentent plus d'anomalies morphologiques (Kläver et al. 2015). Il semblerait également que la 5-azadC bloque l'activité de méthylation *de novo* dans les spermatogonies (Oakes et al. 2007). Il a par exemple été observé une hypométhylation de gènes comme *H19*, *Tcf3* et *Abt1* dans les spermatozoïdes après le traitement. Une analyse pangénomique a révélé qu'entre des individus traités par la 5-azadC et des individus témoins, 6821 CpG sont différentiellement méthylés (Kläver et al. 2015). Il semble cependant que la transmission transgénérationnelle de la réponse au traitement à la 5-azadC soit excessivement limitée. En effet, à la génération F3, seulement 50 CpG restent différentiellement méthylés sur l'ensemble du génome (analyse RRBS) entre descendants de contrôles et descendants d'individus exposés. A la génération F3, seule une hypométhylation du gène *Oct4* est observée dans les spermatozoïdes murins.

3.2. Troubles de la fertilité et de la spermatogenèse chez l'Homme en lien avec la méthylation de l'ADN

Les premières études faisant le lien entre méthylation de l'ADN spermatique et infertilité chez l'Homme ont été publiées il y a une dizaine d'années. La plupart des études consiste en la comparaison du taux de méthylation entre un groupe d'individus contrôles n'ayant aucun problème de fertilité avéré, et un groupe d'individus ayant des difficultés à concevoir et entrant dans des protocoles de procréation médicalement assistée, et/ou présentant des paramètres spermatiques altérés. Les premières approches ont analysé l'état de méthylation d'un gène dont on connaît l'implication dans la spermatogenèse ou le développement embryonnaire (le plus souvent un gène soumis à l'empreinte). Chez les patients atteints d'oligospermie (OS, une réduction de la concentration spermatique), les chercheurs ont observé une hyperméthylation de certains gènes d'expression paternelle, (attendus hypométhylés dans les spermatozoïdes comme *LIT1*, *MEST/PEG1*, *PEG3*, *PLAG1* ou encore *SNRPN*) (Kobayashi et al. 2007) (Marques et al. 2008) (Poplinski et al. 2009) (Sato et al. 2011). Cette hyperméthylation a aussi été retrouvée chez des individus présentant des ratios Protamine1/Protamine2 anormaux (Hammoud et al. 2010). En parallèle a été observée une hypométhylation de certains gènes soumis à l'empreinte maternelle (attendus hyperméthylés sur le génome paternel), notamment le gène *GTL2* ou une ICR du locus *IGF2/H19* (Marques et al. 2008). Un différentiel de méthylation, cette fois entre des groupes d'individus tératozoospermiques (les spermatozoïdes ont une morphologie anormale) ou oligoasthénotératozoospermiques (OAT, c'est-à-dire avec une réduction de la concentration

spermatique, une mauvaise motilité et une morphologie anormale), a aussi été retrouvé sur le site de liaison au CTCF de l'ICR du locus *IGF2/H19* (Boissonnas et al. 2010).

Ces approches gène-candidat ont été complétées par des analyses des séquences répétées du génome. Chez les mêmes individus des groupes infertiles, qui présentaient une méthylation différentielle sur des gènes soumis à l'empreinte, aucun différentiel de méthylation n'a été retrouvé sur les séquences LINE1 ou Alu (Kobayashi et al. 2007) (Kobayashi et al. 2009) (Marques et al. 2008) (Boissonnas et al. 2010). Les auteurs en ont conclu que les différentiels de méthylation identifiés sont spécifiques aux gènes soumis à l'empreinte. Cependant, d'autres études ont pu identifier des différentiels de méthylation sur des séquences spécifiques de type *Alu* (El Hajj et al. 2011), comme par exemple une hypométhylation de la séquence *Alu Yb8* (Urduingio et al. 2015), ou encore d'une séquence de type satellite (Houshdaran et al. 2007). Ces exemples montrent qu'au moins dans certains cas, les différentiels de méthylation retrouvés chez des individus à spermogramme altéré ne sont pas uniquement localisés dans des gènes soumis à l'empreinte mais pourraient avoir un caractère plus global.

D'autres chercheurs se sont alors attelés à identifier des différentiels de méthylation sur des gènes impliqués dans la spermatogenèse ou dans des processus développement. Houshdaran a en 2007 identifié *NTF3*, *MT1A*, *PAX8*, *HRAS* ou *SFN* comme étant différentiellement méthylés entre individus avec paramètres spermatiques anormaux et individus fertiles. De la même façon, *DAZI* (*Deleted in Azoospermia-like*) a été retrouvé hyperméthylé chez des individus OAT (Navarro-Costa et al. 2010), mais ces résultats n'ont pas été confirmés dans une autre étude (Wu et al. 2010). Le promoteur du gène *cyclic AMP-responsive element modulator* (*CREM*) a été identifié comme hyperméthylé chez des individus OS ou à ratio Protamine1/Protamine2 anormal, résultat qui a été confirmé sur une cohorte d'individus plus importante (Nanassy et Carrell 2011a) (Nanassy et Carrell 2011b).

Suite à ces approches régions candidates, le développement des micropuces a permis d'interroger un plus grand nombre de CpG à la fois, situés sur de nombreux gènes et répartis sur l'ensemble du génome. Ce type d'analyses est sans *a priori*, ce qui permet d'identifier des gènes dont on n'aurait pas soupçonné l'implication. En 2011, Pacheco a utilisé une puce *Illumina Infinium HumanMethylation27* qui permet d'interroger plus de 27 000 CpG (Pacheco et al. 2011). Les 1000 CpG les plus variables entre les groupes de fertilité, ont permis de séparer les individus en fonction de la motilité de leurs spermatozoïdes, et 80% de ces CpG sont hypométhylés dans le groupe des individus sub-fertiles. Des altérations de la méthylation pour les gènes de la machinerie épigénétique *HDAC1*, *SIRT3* et *DNMT3a* ont été identifiées chez ces mêmes individus infertiles. Selon la même méthode, Aston en 2012 a

identifié des DMC associés à des gènes impliqués dans des processus biologiques liés au développement, aux pathologies du système reproducteur et à une mauvaise qualité embryonnaire après FIV. La même approche avec une micropuce 27K a permis d'observer l'hyperméthylation de gènes de la machinerie des piRNA comme *PIWIL2* (*Piwi-like RNA-mediated gene silencing 2*) et *TDRD1* (*Tudor domain containing 1*) dans des testicules des patients atteints de problèmes de spermatogenèse (Heyn et al. 2012), ou encore des gènes *INSL6*, *MAEL*, *SLC25A31* et *SPO11* chez des patients OS (Schütte et al. 2013).

A partir de 2015, les études utilisent une puce qui interroge cette fois 450 000 CpG. Aston identifie alors 8500 CpG présentant un différentiel de méthylation significatif entre individus normozoospermiques et infertiles. Utilisant cette même puce, l'étude d'Urduingio en 2015 est particulièrement intéressante. Elle se focalise en effet sur des individus présentant une sub-fertilité idiopathique. Cette étude a permis d'identifier 2752 CpG différenciellement méthylés par rapport au groupe contrôle, dont une fraction est retrouvée sur 48 gènes soumis à l'empreinte. On peut donc en conclure que des individus présentant un spermogramme normal, et qui se révèlent sub-fertiles, peuvent toutefois présenter des altérations de la méthylation au niveau de gènes cruciaux pour le développement embryonnaire.

D'autres études se sont intéressées à la transmission des perturbations de la méthylation spermatique aux embryons. En 2005, Benchaib n'a pas pu établir de corrélation entre méthylation globale et taux de fécondation ou qualité embryonnaire après FIV (Benchaib 2005). En revanche, en 2009, Kobayashi a montré que dans la moitié des cas étudiés, des altérations de profils de méthylation spermatiques chez des individus OS se retrouvaient dans les embryons produits à partir des mêmes éjaculats. Cela implique d'une part que les perturbations de la méthylation, dans les spermatozoïdes chez les patients OS, ne sont pas effacées après FIV, et d'autre part qu'elles ne sont probablement pas reprogrammées correctement dans l'embryon ; elles peuvent donc potentiellement perdurer au cours du développement embryonnaire voire au-delà. Les résultats de cette étude vont cependant à l'encontre d'autres études, montrant notamment qu'une hypométhylation du gène *H19* retrouvée dans des embryons pré-implantatoires pourrait ne pas être présente à l'origine dans les spermatozoïdes (Chen et al. 2010), ou encore qu'il puisse ne pas exister de différences après FIV ou ICSI en terme de taux de fécondation, de cinétique de clivage embryonnaire, de taux d'implantation des blastocystes ou de taux de conception (Boissonnas et al. 2010).

3.3. Altération épigénétique et fertilité bovine

L'implication de quelques gènes dans la spermatogenèse des bovidés a été démontrée par l'utilisation de modèles originaux.

Les hybrides bovin-yacks sont stériles dès la première génération du fait d'un arrêt total de la spermatogenèse. Liu et son équipe ont montré que le promoteur du gène *DAZL* est hyperméthylé chez l'hybride mâle en comparaison avec les mâles bovins ou yack (Liu et al. 2011). Cette hyperméthylation est associée à une légère diminution de l'expression de *DAZL* dans les testicules des hybrides en comparaison avec ceux des mâles bovins ou yack.

Un autre gène de la famille *DAZ*, le gène *Boule*, a été étudié en utilisant le même modèle (Yao et al. 2015). Ce gène est normalement exprimé dans les spermatoocytes pendant la première division de méiose. Une hyperméthylation dans le CGI du promoteur de ce gène a pu être associée à une diminution de son expression dans les testicules des hybrides.

En 2013 Gu et son équipe ont analysé la méthylation du promoteur du gène *PIWIL1* et son état transcriptionnel dans les testicules de bovins, de yacks et d'hybrides, ainsi que l'état de méthylation des séquences répétées LINE-1. Ils ont pu noter une diminution très significative (d'un facteur 5) de l'expression de *PIWIL1* chez les hybrides. Cette diminution de l'expression de *PIWIL1* était associée à une hyperméthylation très importante et significative du promoteur du gène chez les hybrides (avec une méthylation sur les CpG analysés de l'ordre de 90% chez les hybrides contre 20% chez les mâles de race pure). Ils ont détecté une expression d'ARNm de LINE-1 plus forte chez les hybrides. Il n'y avait cependant pas de différentiel significatif de méthylation sur LINE-1 entre les groupes, mais ils ont noté une légère augmentation chez les hybrides (90% de méthylation contre environ 83% chez les mâles de race pure) (Gu et al. 2013).

En 2014 Jena a analysé par bisulfite-PCR l'état de méthylation du locus soumis à l'empreinte *IGF2 / H19* dans les spermatozoïdes de taureaux issus d'un croisement *Bos taurus X Bos indicus*. Les taureaux sont répartis en 2 groupes en fonction de leur taux de conception après plus de 100 IA réalisées avec leur semence. Globalement, aucune différence significative de méthylation n'a été observée entre les 2 groupes. En revanche, un des 7 sites de liaison de CTCF sur le gène *H19* présentait un différentiel de méthylation significatif entre les groupes de fertilité (Jena et al. 2014).

En 2014, Verma a réalisé une étude concernant 6 buffles répartis en 2 groupes entre individus fertiles et sub-fertiles. Le classement des taureaux est fonction du taux de conception calculé

après 50 IA, du taux de clivage en FIV et du taux d'obtention de blastocystes. L'analyse de la méthylation dans les spermatozoïdes a été réalisée par l'utilisation d'une micropuce bovine ciblant les promoteurs ou les régions codantes de 2967 gènes. Au total, 151 gènes sont trouvés différemment méthylés, dont 73 hyperméthylés pour le groupe fertile et 78 gènes pour le groupe sub-fertile. Ces gènes se sont révélés être impliqués dans la régulation transcriptionnelle et la prolifération cellulaire, et 13 d'entre eux sont directement impliqués dans la spermatogenèse, la capacitation ou le développement embryonnaire (Verma et al. 2014).

Tout récemment, Kropp et collègues ont publié une étude effectuée sur 2 groupes de taureaux avec des taux de conception très bons ou très mauvais ; l'analyse a été conduite en utilisant des 3 pools pour chaque groupe de taureaux, avec 2 taureaux par pool. L'analyse de la méthylation dans les spermatozoïdes est faite par *Methyl Binding Domain capture assay* suivi d'un séquençage haut-débit (MethylCap-seq). Cette analyse a permis d'identifier 76 régions différemment méthylées, dont 25 sont situées dans des gènes. La plupart des régions (60/76) et des gènes (20/25) différemment méthylés sont hypométhylés chez les taureaux sub-fertiles. Une analyse du développement embryonnaire par FIV pour les mêmes taureaux n'a pas permis de mettre en évidence de différences en termes de taux de fécondation ou de formation de blastocystes entre les 2 groupes de fertilité (Kropp et al. 2017).

Objectifs de la thèse

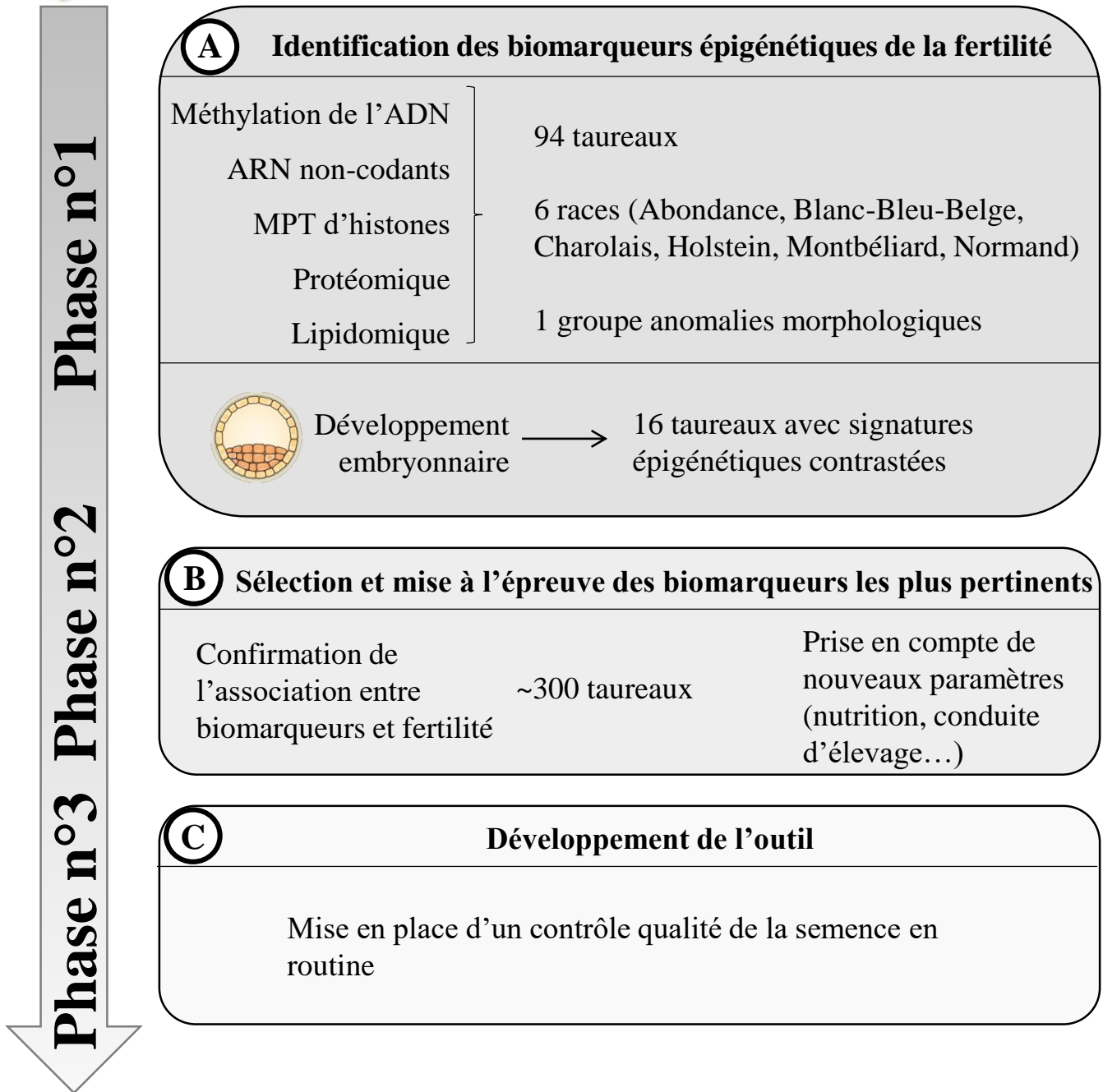


Figure 55. Les phases du projet Sequamol

A. La première phase consiste à étudier de façon exhaustive l'épigénome de la semence bovine (en plus de données de protéomique et de lipidomique) afin d'établir un catalogue de biomarqueurs moléculaires. Cette analyse est menée sur la semence de 94 taureaux, de 6 races différentes (Abondance, Blanc-Bleu-Belge, Charolais, Holstein, Montbéliard, Normand). Pour les taureaux Holstein et Montbéliard, les individus sont classés en fonction de leur performance réelle sur le terrain (TNR, Taux de non retour au niveau de l'éjaculat utilisé pour les analyses). Les taureaux de race Blanc-Bleu-Belge ont également des TNR variables et peuvent ainsi être répartis en plusieurs catégories de fertilité. Les effectifs de taureaux des autres races ne permettent pas d'évaluer la relation entre épigénome et fertilité, mais alimentent les comparaisons entre races. Enfin, un dernier groupe de taureaux Holstein présente des anomalies majeures de morphologie spermatique. Un second temps de la phase 1 est consacré à produire des embryons en FIV à partir de la semence de 16 taureaux Holstein présentant les profils épigénétiques les plus contrastés, afin d'étudier la cinétique développementale du stade une cellule au stade blastocyste, d'évaluer 4 marques épigénétiques par immunofluorescence aux stades 1 et 4 cellules ainsi qu'au stade blastocyste, et enfin de réaliser une étude transcriptomique et de la méthylation par pyroséquençage au stade blastocyste. **B.** La seconde phase du projet consiste à confirmer le catalogue de biomarqueurs identifiés lors de la phase 1 sur une cohorte plus importante d'individus (une centaine par race pour les races Holstein et Montbéliard). Les cohortes étant plus importantes, plus d'informations peuvent être prises en considération comme la nutrition des animaux, le climat, etc., dans le but de raffiner le catalogue de biomarqueurs. **C.** Lors de la troisième et dernière phase du projet, l'objectif est de développer un outil qui pourra être utilisé en routine par les éleveurs pour évaluer le potentiel de fertilité d'un éjaculat. Cet outil sera vraisemblablement une puce d'épigénotypage.

La présence de taureaux sub-fertiles sur le marché de l'IA influence l'efficacité et la durabilité des élevages, entraînant des pertes de production, une augmentation du taux de réforme des génisses et un gaspillage des ressources. Aujourd'hui, l'évaluation de la fertilité des taureaux est basée sur l'analyse combinée de marqueurs génétiques, morphologiques, cinétiques et métaboliques de la semence. Ces marqueurs ne sont cependant pas suffisants pour identifier les taureaux sub-fertiles avant leur entrée en production. Alors que l'ensemble des connaissances acquises en épigénétique commencent à être exploitées en santé humaine, elles ne le sont toujours pas en filière animale.

L'objectif général de mon travail de thèse est d'utiliser la méthylation de l'ADN pour identifier de nouveaux biomarqueurs fiables et précoces de la fertilité mâle. Cette thèse s'inscrit dans le cadre projet SeQuaMol (pour Qualité Moléculaire de la Semence), développé au sein d'un laboratoire commun entre l'unité de Biologie de la Reproduction et du Développement de l'INRA et d'Allice. Le but du projet SeQuaMol est de développer de nouveaux outils moléculaires de diagnostic pour mieux décrire et optimiser la fertilité chez le taureau. Ce projet se déroule en 3 phases, décrites Figure 55.

Phase n°1 : identification des biomarqueurs épigénétiques de la fertilité

Cette identification est réalisée sur une cohorte de 94 taureaux, de 6 races différentes (Abondance, Blanc-Bleu-Belge, Charolais, Holstein, Montbéliarde et Normande). Au sein des races Holstein et Montbéliarde, les individus sont catégorisés en fonction de l'adéquation entre un indicateur génétique pour la fertilité mâle et les performances réelles de l'individu. Les 4 autres races nous permettent d'évaluer l'impact de la race sur le méthylome. La cohorte contient également une catégorie de taureaux Holstein dont les spermatozoïdes présentent un taux d'anomalies morphologiques important. L'épigénome de la semence est étudié de manière exhaustive, avec une recherche de biomarqueurs de méthylation de l'ADN, de microARN, de MPT d'histones, du ratio histone/protamine...pour évaluer comment ces biomarqueurs évoluent en fonction de la fertilité et de la race des individus.

Lorsque le catalogue de biomarqueurs sera complet, les 16 taureaux Holstein présentant les signatures épigénétiques les plus contrastées seront sélectionnés pour effectuer des FIV. Cette étape a pour objectif d'observer l'influence des marques épigénétiques sur l'expression des gènes au cours du développement embryonnaire, ainsi que la cinétique développementale. Les marques épigénétiques (5mC, 5hmC, H3K27me3 et H3K9me3) seront analysées par marquage fluorescent avec des anticorps puis observation en microscopie confocale, aux

stades 1 et 8 cellules et au stade blastocyste. De plus, une analyse transcriptomique et épigénétique (approche gène-candidat déterminée selon les résultats d'expression par conversion bisulfite-pyoséquençage) sera conduite au stade blastocyste en distinguant trophoctoderme et cellules de la masse interne. La cinétique de développement de ces embryons sera analysée par *Time-Lapse video* (sur embryons vivants du stade 1 cellule au stade blastocyste).

Le but de la phase n°1 du projet est donc de constituer un catalogue, pertinent et fiable, de biomarqueurs épigénétiques de la fertilité mâle.

Phase n°2 : Sélection et mise à l'épreuve des biomarqueurs les plus pertinents

Le catalogue de biomarqueurs établi, la seconde phase du projet consiste en leur validation sur une cohorte de plusieurs centaines d'individus. Au moment de la rédaction de ce manuscrit, cette cohorte est en cours de constitution. Les individus proviendront de multiples centres d'élevages différents, et donc ont été soumis à des paramètres environnementaux différents (nutrition, climat, personnel, etc.). La prise en compte de ces paramètres permettra de raffiner le catalogue de biomarqueurs, et de mettre en lumière l'influence de chaque composante environnementale sur le méthylome spermatique.

Phase n°3 : Développement de l'outil

L'objectif final du projet SeQuaMol est le développement d'un outil utilisant les différents types de biomarqueurs épigénétiques identifiés et validés lors des étapes précédentes. Cet outil, sous forme de puce d'épigénotypage, pourra être utilisé en routine par les ES pour contrôler la qualité épigénétique des éjaculats avant leur distribution.

Après une présentation méthodologique justifiant notre approche technique, les résultats seront présentés en 2 parties (la première partie ayant abouti à la soumission d'un article). La présentation de ces travaux est suivie d'une discussion générale sur mon travail de thèse, et d'une présentation des perspectives de ce travail et du projet SeQuaMol.

Méthodologie

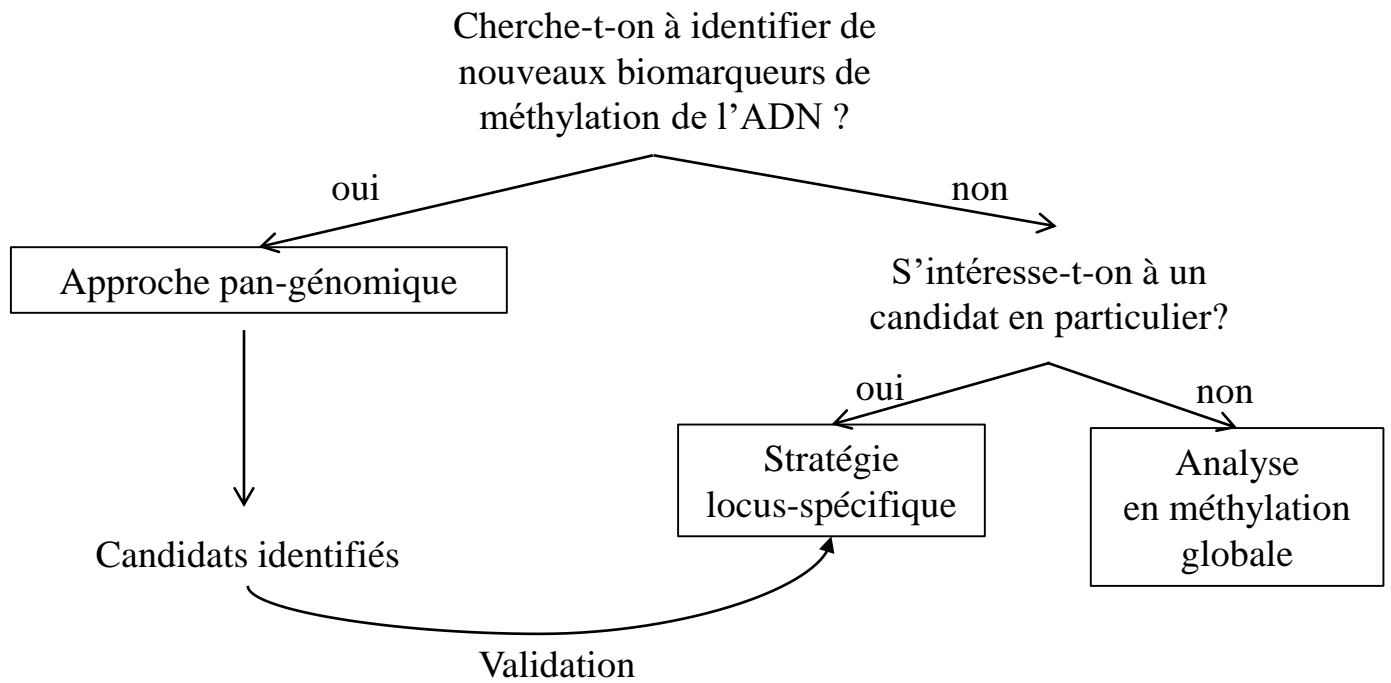


Figure 56. Arbre décisionnel pour l'analyse de la méthylation de l'ADN

Afin de choisir la méthode d'analyse la plus adaptée, il est possible d'établir un arbre décisionnel basé sur la question biologique posée. Si l'objectif est d'identifier de nouveaux biomarqueurs de méthylation, il faut choisir une méthode permettant de cibler le plus grand nombre de CpG possible. L'approche à employer dans ce cas est l'analyse pangénomique. Dans le cas contraire, il peut alors s'agir d'analyser l'état de méthylation de l'ADN sur une séquence déjà identifiée du génome, sur un gène d'intérêt par exemple. Ce sont les analyses *locus*-spécifique. Enfin, si l'on souhaite évaluer l'impact général d'un traitement ou d'une condition sur l'état de méthylation du génome, il faut alors s'orienter vers une stratégie d'analyse globale de l'ADN. Pour chacune de ces stratégies de nombreuses techniques d'analyse ont été développées, répondant plus spécifiquement aux conditions particulières de chaque type d'étude (coût, nombre d'échantillons à analyser, sensibilité, quantité et qualité d'ADNg disponible...).

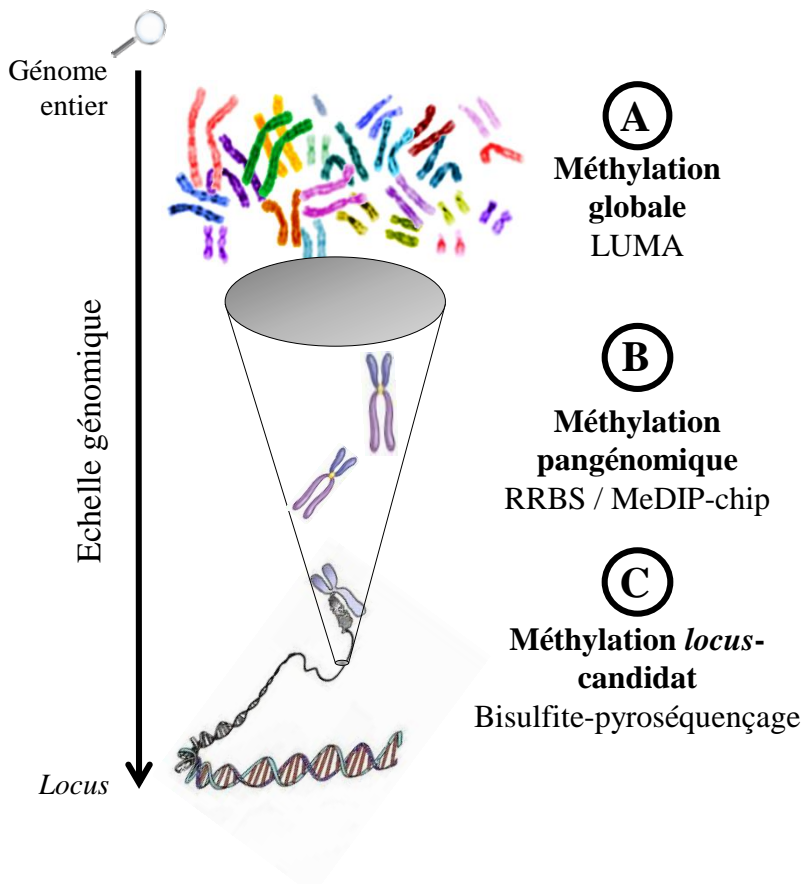


Figure 57. Schéma représentatif de l'analyse de la méthylation de l'ADN à différentes échelles génomiques

A. Analyse de la méthylation globale. Elle consiste en un calcul d'un pourcentage moyen de méthylation représentatif de l'état de méthylation de l'ensemble des sites MspI du génome. Cette échelle est couramment utilisée pour identifier des modifications majeures du méthylome pour un type cellulaire donné (tumorigénèse, vieillissement, exposition à un toxique, etc.). **B.** Analyse de la méthylation pangénomique. Ce type d'analyse consiste à déterminer les cytosines (DMC) ou régions différemment méthylées (DMR) entre deux groupes d'échantillons. Cette échelle est couramment utilisée pour identifier des biomarqueurs de méthylation associés à un état physiologique d'un individu ou un type cellulaire donné (paramètres phénotypiques, maladies, etc.). **C.** Analyse de la méthylation *locus*-spécifique. Cette pratique repose l'analyse précise de profils de méthylation associés à une région candidate donnée. Elle est particulièrement employée pour valider des données issues d'analyse pangénomiques, sur un gène d'intérêt fonctionnel ou dans le cas de gènes impliqués dans certaines pathologies (cancers, syndrome de Beckwith-Wiedemann...).

1. Méthylation de l'ADN : quelle stratégie pour quel objectif ?

La finalité de toute analyse de méthylation de l'ADN est de déterminer, pour un échantillon et à un temps donné, l'état de méthylation de tout ou partie du génome. Les recherches en méthylation ont été rendues possibles par la mise au point de nombreuses stratégies d'analyse développées depuis les années 1970. Le progrès technologique, et notamment suite à l'avènement du séquençage à haut-débit, a de plus permis d'obtenir continuellement une meilleure couverture en CpG du génome ainsi qu'une meilleure sensibilité des mesures. Il existe un nombre important de revues publiées qui cataloguent et catégorisent les méthodes d'analyses de la méthylation ((Olkhov-Mitsel et Bapat 2012) (Kurdyukov et Bullock 2016) (Yong, Hsu, et Chen 2016), pour ne citer que quelques revues récentes). D'autres études se focalisent sur un nombre plus restreint de méthodes pour en comparer l'efficacité (Lisanti et al. 2013) (Bock et al. 2016). Le traitement de l'ensemble des informations disponibles permet de s'orienter précisément vers la stratégie la plus adaptée aux besoins de notre étude. Cette partie aura pour objectif de décrire les stratégies d'analyse de la méthylation disponibles, et ce dans le but de justifier notre approche méthodologique. Un schéma représentatif du processus décisionnel à employer pour sélectionner la stratégie la plus adaptée est donné Figure 56.

Lorsqu'il est nécessaire d'analyser la méthylation de l'ADN dans un projet de recherche, plusieurs éléments clefs sont à considérer *a priori* pour sélectionner l'approche la plus adaptée. La question primordiale à se poser est quelle est précisément la question biologique à laquelle on souhaite répondre. Il peut s'agir d'établir *de novo* un catalogue de DMC entre échantillons traités et échantillons témoins. Ou bien de mettre en évidence une corrélation entre le niveau d'expression d'un gène d'intérêt et l'état de méthylation d'éléments régulateurs de son expression. Dans tous les cas, le premier aspect à étudier est donc de savoir si l'on cherche à identifier de nouveaux biomarqueurs, ou si des cibles sont connues par avance. Dans le premier cas il faut alors s'orienter vers des analyses pangénomiques de l'ADN, dont les différentes approches techniques permettent de parcourir tout ou partie du génome pour identifier des DMC ou des DMR d'intérêt. Si l'on ne cherche pas à identifier de nouveaux biomarqueurs de méthylation, on peut avoir pour intérêt de cibler un gène ou une région en particulier (par exemple, pour générer des données de validation) ou non. Dans ce cas, il faut s'orienter vers une stratégie *locus*-spécifique ou d'analyse en méthylation globale, respectivement. Finalement, la stratégie à employer dépend de l'échelle génomique à laquelle on veut se placer : globale, pangénomique ou *locus*-candidat (Figure 57).

Stratégies d'extraction de l'information de méthylation

Stratégies de révélation					
	PCR / qPCR	NGS	Micropuce	Pyroséquençage	Anticorps secondaire
Bisulfite de sodium	<ul style="list-style-type: none"> ● PCR - clonage séquençage 	<ul style="list-style-type: none"> ● WGBS ● RRBS 	<ul style="list-style-type: none"> ● Infinium 	<ul style="list-style-type: none"> ● PCR - pyroséquençage 	/
Enzyme de restriction	<ul style="list-style-type: none"> ● PCR méthylsensible ● COBRA 		<ul style="list-style-type: none"> ● CHARM 	<ul style="list-style-type: none"> ● LUMA 	/
Enrichissement MBD / Anticorps	<ul style="list-style-type: none"> ● MeDIP-qPCR 	<ul style="list-style-type: none"> ● MeDIP-seq 	<ul style="list-style-type: none"> ● MeDIP-chip 	/	<ul style="list-style-type: none"> ● ELISA ● Immunofluorescence

Tableau 7. Liste des principales techniques développées pour l'analyse de la méthylation de l'ADN

Une fois identifiée l'échelle d'analyse nécessaire pour répondre à la question biologique posée, il faut sélectionner une méthode de préparation de l'ADN pour en extraire l'information de méthylation (partie 1). Des aspects plus pratiques sont alors à évaluer, comme le nombre d'échantillons à analyser et les contraintes de temps, la quantité et la qualité d'ADNg à disposition, la disponibilité des équipements et des réactifs, les coûts engendrés, ou encore la disponibilité et la maîtrise des logiciels de bio-informatique pour l'analyse et l'interprétation des résultats. Il est aussi important de se rappeler que de manière générale chaque stratégie va invariablement introduire des biais. En avoir conscience permet d'opter pour des mesures pour réduire ces biais, ou au moins de les prendre en compte lors de l'interprétation des résultats (Robinson et al. 2010). Ils peuvent en effet avoir une grande influence sur les calculs de pourcentages de méthylation et de ce fait fausser les résultats.

Il faut enfin choisir un procédé technique de traduction des profils de méthylation en information qui pourra ensuite être traitée (partie 2). Le choix du procédé dépend du niveau de sensibilité de la mesure requis pour l'étude, et c'est particulièrement vrai dans le cas des analyses pangénomiques. Le choix de la méthode de traduction dépend également de l'état de connaissance du génome de l'espèce d'intérêt (une méconnaissance se traduisant par des défauts d'assemblage ou d'annotation), la quantité de matériel biologique (de 10ng à 5µg), ainsi que de la forme de l'échantillon biologique (ADNg extrait, cellule unique, échantillon en bloc paraffine...). Les principales technologies utilisées pour la traduction des profils de méthylation en information sont la réaction de polymérisation en chaîne (*Polymerase Chain Reaction*, PCR), l'utilisation d'anticorps secondaires, l'hybridation sur puce ou le séquençage haut-débit.

La combinatoire de ces aspects (échelle génomique, méthode d'enrichissement et méthode de traduction de l'information) a donné naissance à une multitude de procédés techniques d'analyse adaptés à tous les besoins. Ceux-ci sont répertoriés dans le Tableau 7. L'évaluation des méthodes disponibles nous a conduits pour ce travail de thèse à sélectionner le LUMA, le RRBS et le pyroséquençage pour les analyses en méthylation globale, pangénomique et *locus-candidat*, respectivement. Après avoir justifié ce choix méthodologique, chacune de ces méthodes est décrite dans la partie 2 de ce chapitre. Nous avons été amenés à comparer notre analyse pangénomique (RRBS) à une autre technique (MeDIP-chip) mise en œuvre au laboratoire. Les données sont présentées dans le 1^{er} article.

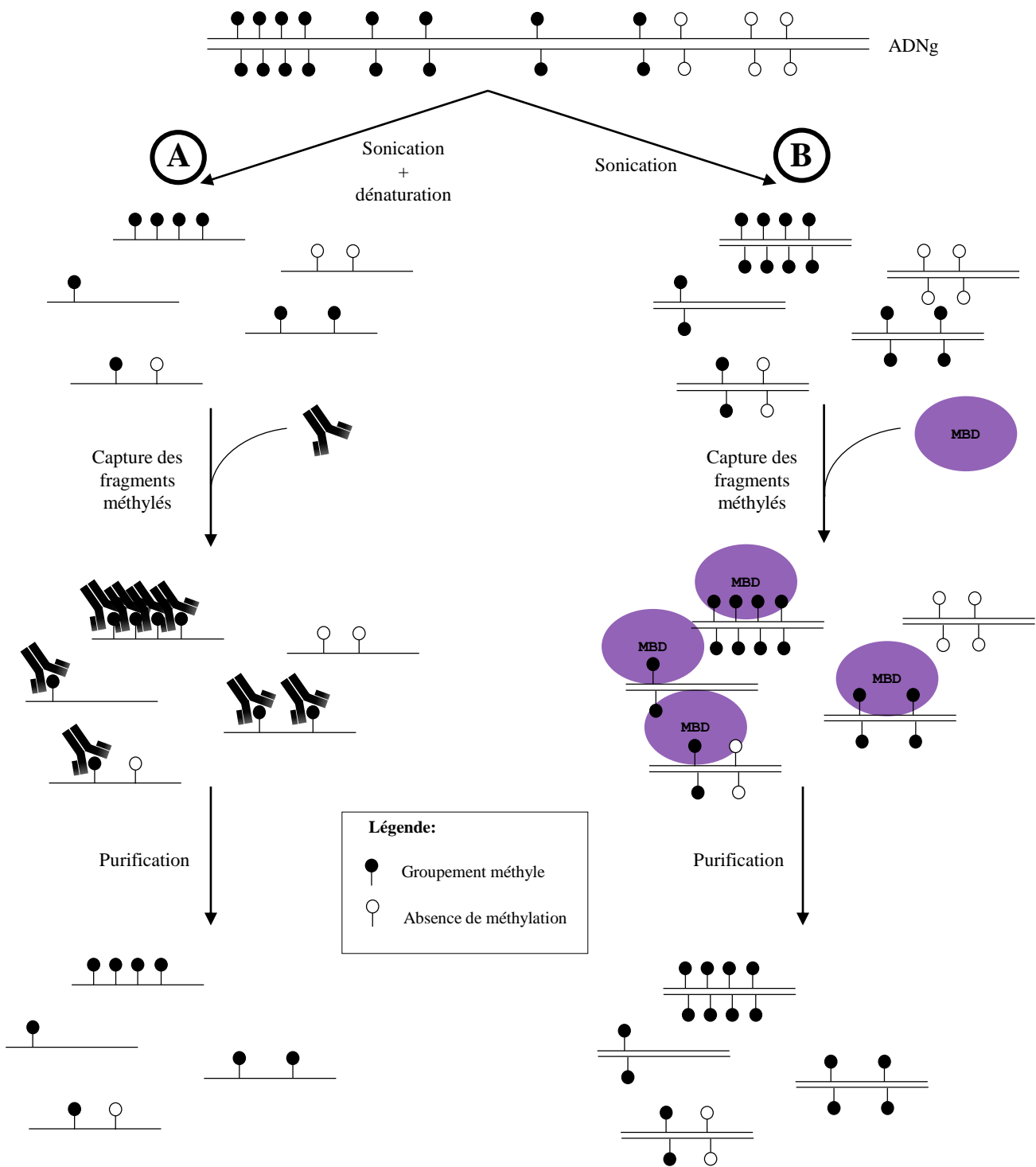


Figure 58. Schéma représentatif de la stratégie d'analyse de la méthylation de l'ADN par affinité

A. Méthode par immunoprécipitation de l'ADN avec des anticorps ciblant les cytosines méthylées (Anti 5meC). L'ADN est traité par sonication et dénaturé en vue d'obtenir des fragments simples brins. Les fragments comportant des sites CpG méthylés sont ensuite capturés par un anticorps spécifique des sites CpG méthylés. Après purification, seuls les fragments méthylés sont précipités, et peuvent être utilisés pour la suite du protocole. **B.** Méthode utilisant des protéines à MBD. Le protocole est similaire à celui employé avec les anticorps mis à part le fait que l'ADN ne nécessite pas d'être simple brin (on ne passe donc pas par une étape de dénaturation de l'ADN).

1.1. Les stratégies d'extraction de l'information de méthylation

1.1.1. L'utilisation d'enzymes de restriction

Cette pratique repose sur la digestion différentielle de l'ADN avec des endonucléases sensibles ou non à l'état de méthylation. En effet, certaines enzymes ont la propriété de couper l'ADN indépendamment de l'état de méthylation de leur site de restriction. Par exemple, MspI clive les sites 5'- C[^]CGG -3' du génome indépendamment de l'état de méthylation du CpG au centre du site (Cedar et al. 1979). Au contraire, des isoschizomères de ces enzymes (c'est-à-dire d'autres enzymes coupant sur le même site de restriction) ne peuvent cliver l'ADN que si le ou les CpG contenus dans le site de restriction sont méthylés ou non-méthylés (Tarasova, Nayakshina, et Degtyarev 2008). Ces enzymes sont alors qualifiées de méthyl-sensibles (*Methylation-Sensitive Restriction Enzymes*, MSRE). Par exemple, un isoschizomère de MspI est HpaII, qui ne clive 5'- C[^]CGG -3' que s'il n'est pas méthylé. Dans le cadre de l'analyse de la méthylation, ce potentiel peut être exploité de diverses façons. Les premières stratégies d'analyse reposaient sur une révélation par *Southern blot*. En comparant les profils obtenus en digérant l'ADN avec MspI ou HpaII, puis en faisant migrer les produits par électrophorèse, il est possible de décrire de façon qualitative l'intensité du clivage de l'ADN, et par extension l'état de méthylation général d'un échantillon biologique. Développées plus récemment, les technologies de séquençage de nouvelle génération (*Next Generation Sequencing*, NGS) ont considérablement augmenté le potentiel que représente l'utilisation de ces enzymes.

1.1.2. Les stratégies par affinité

Les stratégies par affinité sont des techniques permettant de capturer spécifiquement les séquences d'ADN méthylées (Figure 58). L'utilisation d'anticorps ciblant les CpG méthylés permet d'immunoprécipiter uniquement les fragments d'ADN méthylés générés après sonication et dénaturation de l'ADN. Des enrichissements relatifs entre conditions peuvent alors être mesurés par rapport au matériel de départ (non-immunoprécipité), qu'on appelle *input*. Une des méthodes d'analyse utilisant cette approche technique est le *Methylated DNA ImmunoPrecipitation* ou MeDIP (Weber et al. 2005). Cette méthode est utile quand on ne souhaite pas une résolution quantitative à la base près, car l'information de méthylation est obtenue pour un fragment d'ADN donné. En revanche, les méthodes par immunoprécipitation tendent à être biaisées par la densité en CpG : à méthylation équivalente, les fragments riches en CpG sont susceptibles d'être plus immunoprécipités que les fragments

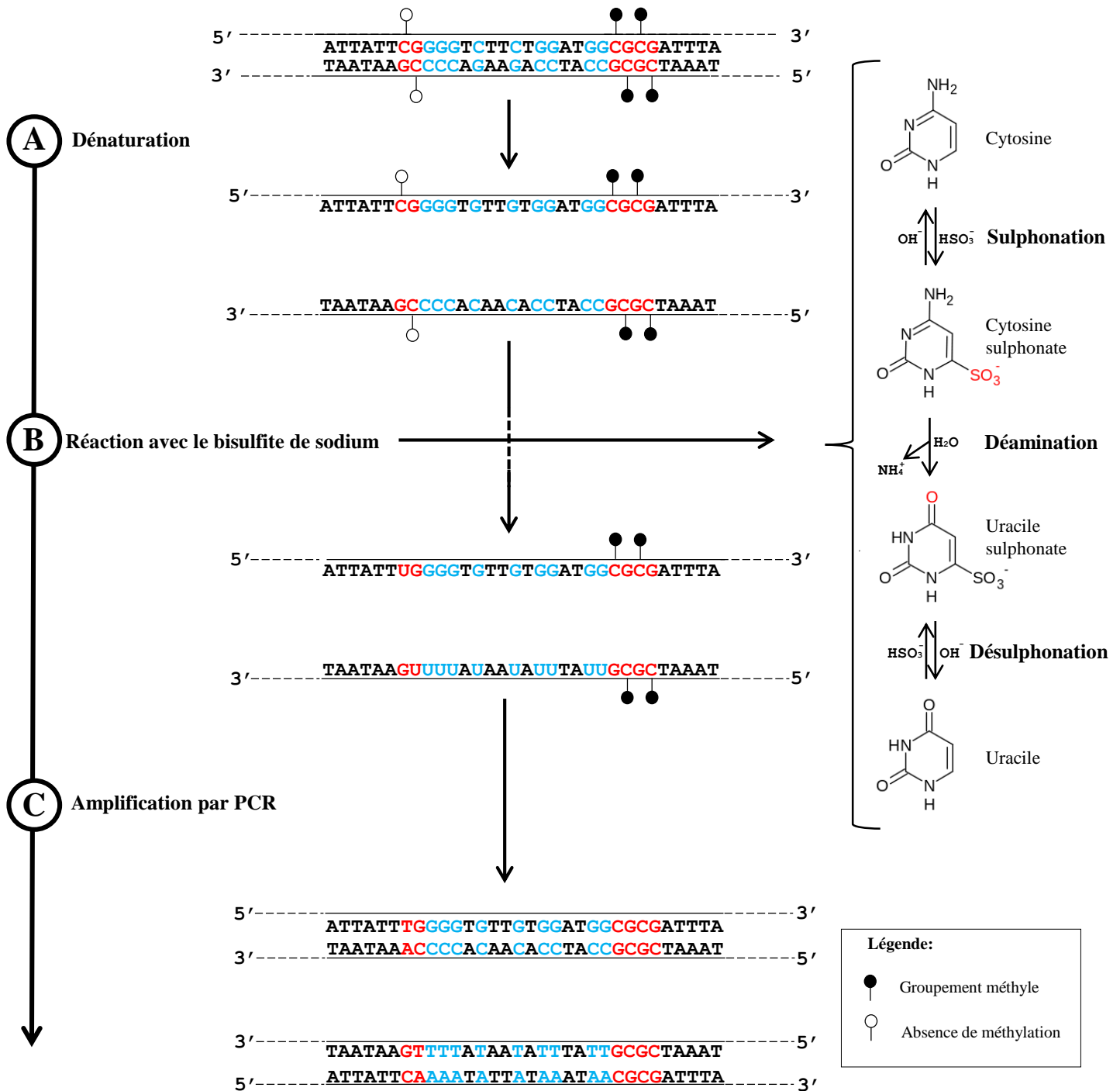


Figure 59. La conversion de l'ADN par bisulfite de sodium

La conversion de l'ADN par bisulfite de sodium se déroule en 3 étapes. **A.** La première étape consiste en la dénaturation de l'ADN par l'action conjuguée de la température et de la soude. **B.** L'ADN est ensuite traité en présence de bisulfite de sodium. Les cytosines non-méthylées (ici représentées en rouge lorsqu'elles sont suivies d'une guanine, et en bleu lorsqu'elles sont suivies par une base nucléotidique autre qu'une guanine) sont alors converties en uraciles. La réaction chimique se fait en trois étapes (représentées dans l'accolade à droite de la figure), avec i) une étape de sulphonation, puis ii) de déamination, et enfin iii) de désulphonation. Les cytosines méthylées sont protégées ; le groupe méthyle retarde la sulphonation. En limitant le temps réactionnel, les cytosines méthylées ne sont pas converties. **C.** La dernière étape du traitement est l'amplification par PCR de l'ADN. Les uraciles résultants de la conversion des cytosines non-méthylées sont alors remplacés par des thymines. L'information de méthylation a alors été transformée en polymorphisme de séquence C/T, qui peut être révélé par des méthodes de séquençage.

pauvres en CpG. Enfin, l'efficacité de cette stratégie repose également sur l'efficacité et la spécificité des anticorps.

Pour contourner les problématiques liées à l'utilisation d'anticorps, des stratégies utilisant des protéines à domaine dit à *Methyl-CpG-Binding Domain* (MBD) ont été développées (Serre, Lee, et Ting 2010). Les protéines à MBD comme MBD2 ou MBD3L1 se lient spécifiquement aux fragments d'ADN double brins méthylés générés après la sonication de l'ADNg. Comme avec l'utilisation d'anticorps, l'information de méthylation reste qualitative au niveau de la région. Les MBD présentent une meilleure affinité quand il y a de multiples sites CpG méthylés dans une région génomique de courte longueur. Cette méthode va alors préférentiellement précipiter des blocs de CpG méthylés plutôt que des CpG méthylés isolés. Un des inconvénients de cette méthode est qu'elle peut de ce fait tendre à sur-représenter les séquences répétées riches en CpG (Nair et al. 2011). Cependant, les protéines à MBD seraient plus sensibles et plus efficaces que l'utilisation d'anticorps (Li et al. 2010) (Bock et al. 2010). Plusieurs kits commerciaux sont disponibles pour les étapes d'élution, et qui présentent des résultats très différents en termes de rendement (De Meyer et al. 2013).

Un des avantages des stratégies par affinité par rapport à celles avec enzymes de restriction est qu'elles font la distinction entre les 5mC et les 5hmC. De plus, alors que l'utilisation d'enzymes de restriction génère une fragmentation séquence-spécifique du génome, la capture des fragments par affinité est indépendante de la séquence.

1.1.3. La conversion de l'ADN au bisulfite de sodium

Les 2 stratégies présentées précédemment permettent d'extraire l'information à partir d'ADN natif. Cependant, pour utiliser certaines méthodes l'état de méthylation de l'ADN doit être préalablement converti par l'utilisation du bisulfite de sodium. Il s'agit d'une approche technique développée en 1992 par Frommer et son équipe (Frommer et al. 1992) pour l'analyse quantitative de la méthylation. Elle est basée sur la réaction différentielle entre un réactif chimique et les cytosines en fonction de leur état de méthylation. Le traitement de l'ADN au bisulfite de sodium s'attaque en priorité aux cytosines non-méthylées et les transforme en uraciles par désamination. Les cytosines méthylées sont protégées ; leur désamination requiert un temps réactionnel plus important (Figure 59). Après une étape d'amplification par PCR les uraciles sont remplacés par des thymines. Cette méthode permet de convertir une information de méthylation en polymorphisme de séquence. Cette étape est indispensable si l'on souhaite ensuite séquencer les données, car les séquenceurs ne peuvent pas identifier la méthylation. L'étape de conversion est assez longue mais doit être limitée

dans le temps (4h), nécessite l'obscurité et l'absence de contact avec l'air pour limiter l'oxydation du bisulfite de sodium, et requiert de multiples changements de tube ce qui augmente les risques d'erreurs techniques et de contaminations. De plus la perte de matériel est importante durant ce processus : d'une part une certaine fraction de l'ADN (pouvant aller jusqu'à 96%) peut être perdue pendant les étapes de purification (Grunau, Clark, et Rosenthal 2001), d'autre part la conversion induit une fragmentation importante de l'ADN. Cependant, de nombreux kits de conversion sont commercialisés par des entreprises et facilitent cette étape (durée de conversion moindre et perte de matériels réduite) (Leontiou et al. 2015). Une autre problématique importante soulevée par cette approche est qu'après traitement au bisulfite de sodium il est impossible de distinguer la 5mC de la 5hmC (Booth et al. 2012). En effet des études ont montré que le bisulfite de sodium réagit avec les 5hmC pour aboutir à la cytosine 5-méthylènesulfonate, qui n'est ensuite pas désaminé (Nestor et al. 2010) (Jin, Kadam, et Pfeifer 2010). D'autres études suggèrent que la cytosine 5-méthylènesulfonate peut ralentir ou même bloquer la Taq polymérase dans les réactions PCR (Huang et al. 2010) La méthylation mesurée après conversion de l'ADN est donc la résultante des informations en 5mC et en 5hmC confondues. Enfin, une mauvaise conversion au bisulfite aura un impact non négligeable sur les calculs de méthylation, et peut conduire à une mauvaise interprétation des résultats si cette information n'est pas prise en considération.

1.2. Les stratégies de révélation

Les stratégies de révélation regroupent l'ensemble des technologies permettant, à partir des sites CpG identifiés par les stratégies d'extraction, d'en mesurer l'état de méthylation. Les premières stratégies, basées sur l'utilisation de la PCR puis diverses méthodes de séquençage de l'information, ne permettaient d'interroger que quelques CpG. Les méthodes plus modernes comme les micropuces ou le séquençage à haut-débit ont permis de considérablement augmenter la puissance des analyses. Une autre méthode basée sur une technologie de séquençage est le pyroséquençage. Comme il s'agit d'une des méthodes sélectionnées, le pyroséquençage est décrit en détail dans la section 2.1 de ce chapitre.

1.2.1. Les méthodes basées sur la PCR

Les méthodes d'analyse reposant sur la PCR sont des méthodes simples et efficaces à mettre en œuvre pour les approches gène-candidat. Toutes reposent sur la conversion préalable de l'ADN au bisulfite de sodium. Les amorces de PCR sont ensuite dessinées pour amplifier la région d'intérêt (un gène ou un îlot CpG particulier par exemple).

La méthode la plus simple pour évaluer l'état de méthylation d'un CpG est la *Bisulfite Sequencing PCR* (BSP) (Clark et al. 1994). L'ADN, préalablement converti au bisulfite de sodium, est soumis à un séquençage direct de Sanger. La méthylation est alors mesurée par la mesure relative des pics de C et de T des sites CpG sur l'électrophorégramme. La BSP peut être réalisée directement avec les produits de PCR ou en passant par leur clonage dans un plasmide bactérien. Cette méthode n'est pas sans présenter certains inconvénients.

- Dans le dessin des amorces. La baisse de complexité de la séquence d'ADN après conversion diminue la spécificité des amorces. De plus, l'amorce ne doit pas dans la mesure du possible chevaucher de CpG, ce qui peut être délicat pour les régions qui en sont denses
- Il peut y avoir amplification préférentielle envers l'allèle méthylé ou non-méthylé
- Des polymorphismes de séquence peuvent également biaiser l'amplification

Une autre méthode utilisant la PCR est la *Methylation Specific PCR* (MSP) (Herman et al. 1996). Elle repose cette fois sur l'amplification sélective des allèles méthylés ou non-méthylés, en utilisant 2 couples d'amorces. Les produits de PCR sont ensuite chargés sur un gel d'électrophorèse : la présence d'une bande d'amplification pour l'un ou l'autre des produits est révélatrice de l'état de méthylation ou non de la région d'intérêt. Comme la BSP, il s'agit d'une méthode qualitative.

L'utilisation de la *High Resolution Melting* (HRM) peut également être envisagée pour l'analyse de la méthylation (Wojdacz et Dobrovic 2007). Après PCR, les amplicons sont traités avec un intercalant qui se fixe sur l'ADN double brin. En chauffant progressivement les amplicons, on atteint le point de fusion de l'ADN, qui correspond à la température pour laquelle les 2 brins d'ADN se séparent, et où l'intercalant est alors libéré. Le polymorphisme C/T généré par la conversion de l'ADN au bisulfite de sodium affecte la courbe de fusion de l'ADN, ce qui est révélateur de l'état de méthylation. Une amélioration de cette méthode est la *Methylation-Sensitive High Resolution*, qui combine une PCR quantitative sur l'ADN converti au bisulfite de sodium indépendamment de l'état de méthylation, puis d'une HRM.

Les méthodes par PCR présentent quelques désavantages. Ciblant des régions assez courtes (d'environ 500 pb), elles ne permettent d'interroger qu'un nombre limité de CpG à la fois, en comparaison des méthodes plus récentes comme l'utilisation de micropuces ou avec les techniques de NGS. Il est parfois délicat de dessiner des amorces de PCR du fait de la baisse de complexité de la séquence génomique après la conversion au bisulfite de sodium, ce qui peut dans certains cas augmenter considérablement les amplifications non-spécifiques.

1.2.2. L'hybridation sur micropuce

L'hybridation sur micropuce peut révéler l'état de méthylation soit par clivage de fragments d'ADN issus de leur préparation avec enzymes de restriction, soit capturés par affinité (anticorps anti-5mC, MBD). La fraction enrichie en ADN méthylé et l'ADNg total (*input*) sont marqués avec des fluorochromes différents, puis hybridés sur une micropuce à ADN. Après une étape de rinçage pour éviter les signaux non-spécifiques, les micropuces sont scannées par un microscope confocal à fluorescence. La fluorescence est mesurée pour chacun des puits de la micropuce. Les sondes sur lesquelles s'hybrident les fragments d'ADN sont dessinées pour couvrir l'ensemble des régions génomiques que l'on souhaite étudier (le plus fréquemment les promoteurs des gènes ou les îlots CpG, parfois l'ensemble du génome).

Différentes technologies de micropuces sont commercialisées, et varient selon la taille des sondes qui peuvent être utilisées. En règle générale, plus les sondes utilisées sont longues plus la spécificité est importante, mais cela réduit le nombre de sondes différentes qui peuvent être interrogées à la fois. Certaines proposent des technologies permettant de co-hybrider l'*input* et les préparations sur la même puce, ce qui réduit le nombre de répliques techniques à effectuer du fait de la variation inter-micropuces.

Au laboratoire, cette technologie a été utilisée pour révéler l'état de méthylation de fragments capturés par la technologie de MeDIP (décrite avec les stratégies par affinité). La micropuce a été dessinée spécifiquement pour cibler les promoteurs (-2000 à +1360 pb par rapport au *gene start*) de 21 296 gènes bovins. Des résultats d'analyse du méthylome de la semence bovine utilisant cette technologie sont donnés dans l'article n°1 de ce manuscrit.

D'autres technologies utilisent la conversion préalable de l'ADN au bisulfite de sodium, comme la technologie *Bead array* de la société Illumina. Chaque sonde est dessinée en double, une pour cibler les C méthylées, l'autre les C non-méthylées (converties en T). L'incorporation de ddNTP et leur coloration avec un agent fluorescent permet de calculer le ratio de signal lumineux provenant des sites méthylés ou non-méthylés.

Jusqu'à récemment le coût de ces méthodes était avantageux par rapport aux stratégies NGS. C'est moins le cas aujourd'hui du fait de la baisse continue des coûts de séquençage. La méthode par hybridation présente également des inconvénients. L'information générée est qualitative et est limitée au dessin des sondes. Enfin, la quantité d'ADN requise est assez importante (entre 500ng à 1µg).

1.2.3. Le séquençage à haut-débit

Le développement des techniques de NGS a rendu possible l'analyse de la méthylation de manière quantitative au CpG près, et ce en balayant les CpG répartis sur tout ou partie du génome. Les fragments d'ADN produits à partir des méthodes d'extraction de la méthylation sont déposés sur une cellule de mesure (*flow cell*). Une extrémité du fragment peut s'hybrider à la *flow cell* grâce à la présence d'un adaptateur sur le fragment d'ADN (ajouté pendant la préparation des banques) et de son complémentaire sur la *flow cell*. L'étape suivante est l'amplification en *cluster* pour chaque fragment hybridé afin d'amplifier le signal lors de la mesure finale. Le séquençage en lui-même consiste alors à incorporer successivement des nucléotides couplés à des fluorochromes différents en fonction de leur type. Les nucléotides s'incorporent et l'intensité de fluorescence mesurée détermine la nature du nucléotide incorporé. Les méthodes NGS présentent de nombreux avantages sur les méthodes présentées précédemment. Elles permettent de mettre en place des analyses sans *a priori*, c'est-à-dire sans préjuger des régions génomiques méthylées comme ce doit être le cas avec les micropuces. Des désavantages de cette méthode sont d'abord son coût et la disponibilité des équipements de séquençage. De plus, les méthodes NGS nécessitent de gérer et d'analyser l'importante quantité de données générées.

Plus récemment, de nouvelles technologies de séquençage, dites de 3^{ème} génération, ont vu le jour. Elles permettent notamment de séquencer directement une unique molécule d'ADN, sans passer par des étapes d'amplification par PCR. Des exemples de telles technologies sont la technologie *Single Molecule Real-Time Sequencing* (SMRT) ou Oxford Nanopore. Ce type de technologie présente encore un taux d'erreur de séquençage assez important comparé aux séquençages de 2nde génération.

2. Choix des méthodes retenues pour notre étude

Pour mener à bien nos études il a fallu sélectionner les méthodes les plus adaptées à nos besoins. Notre objectif était de caractériser de façon la plus exhaustive possible le profil de méthylation de l'ADN spermatique bovin. Nous avons alors sélectionné une méthode d'analyse globale pour comparer différents types cellulaires et pour différentes espèces, 2 méthodes pangénomiques pour avoir une information plus précise des différentiels qui existent entre types cellulaires, et une approche gène-candidat pour valider les données pangénomiques. Dans un second temps, l'approche consistait à identifier des biomarqueurs de

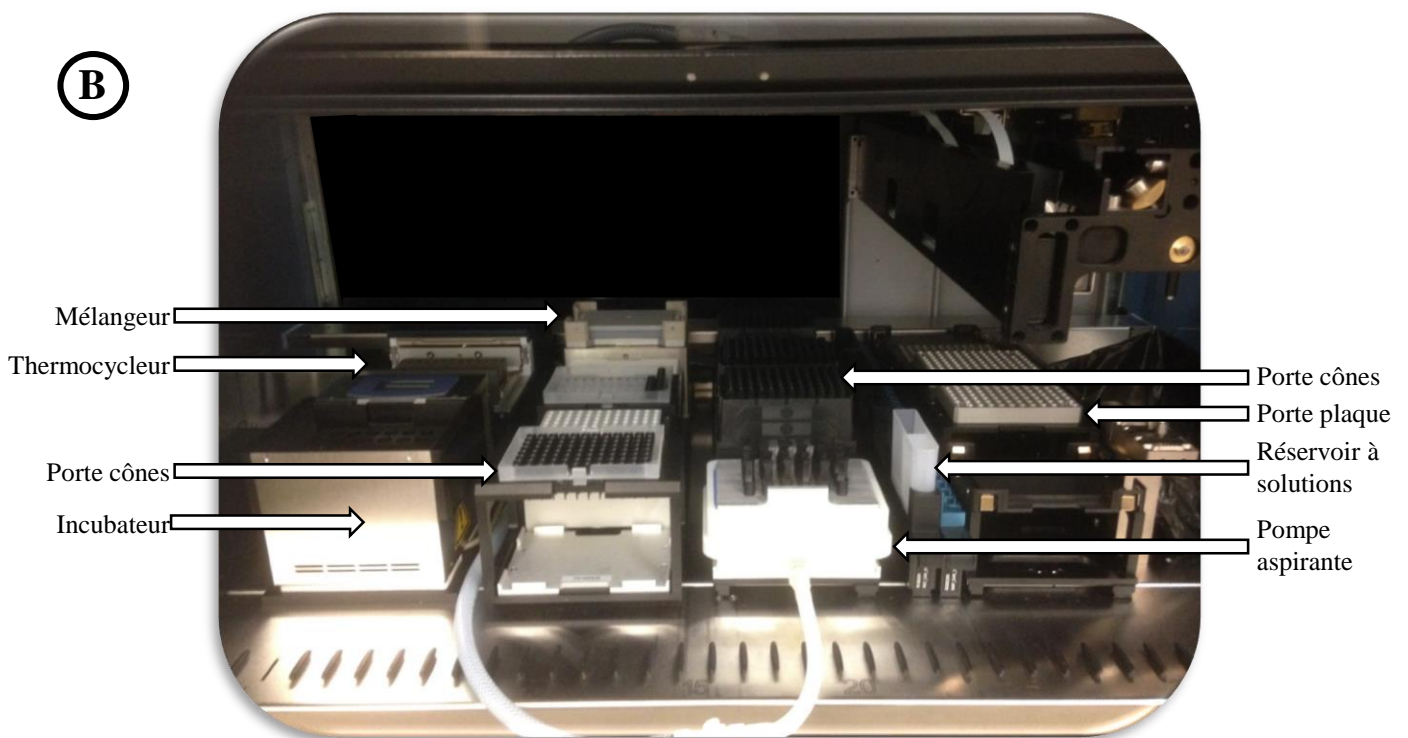
A**B**

Figure 60. Automatisation du protocole de RRBS sur un automate de pipetage

A. Automate de pipetage de modèle MICROLAB Star (Hamilton) tel que celui utilisé au laboratoire. **B.** Vue intérieure de l'automate. Les flèches pointent vers les différents éléments ajoutés sur le plan de travail de l'automate afin de réaliser l'ensemble du protocole de préparation des banques RRBS. Il permet ainsi de générer jusqu'à 24 banques de RRBS à la fois. L'ensemble du protocole s'effectue sur 4 jours. L'incubateur permet de maintenir les solutions à 4°C pendant les pipetages. La pompe aspirante est quant à elle utilisée pendant l'étape de purification sur colonne qui suit la double conversion au bisulfite de sodium. La PCR finale de production de banques ainsi que la purification des produits de PCR est effectuée en dehors de l'automate pour éviter les contaminations (Thermocycleur Eppendorf).

la fertilité bovine, puis de valider ces biomarqueurs par une approche gène-candidat. D'un point de vue technique, il a fallu prendre en compte le nombre important d'échantillons à analyser (dont notamment 58 individus pour le seul axe d'analyse 2). La méthode à sélectionner devait donc permettre de générer les banques rapidement, et de manière peu coûteuse. La quantité de matériel n'est pas un frein car l'ADN est extrait à partir d'une paille de semence pour chaque individu, et une grande quantité d'ADN peut être extraite à partir d'une seule de ces pailles.

Pour les méthodes globales et gène-candidat, nous avons sélectionné respectivement la méthode du *Luminometric Methylation Assay* (LUMA) et du bisulfite-pyroséquençage. Le LUMA est une technique couramment utilisée au laboratoire, nous disposons alors de l'expertise et du matériel nécessaire. Le bisulfite-pyroséquençage a été retenu pour des raisons identiques, ainsi que pour la fiabilité des résultats que cette méthode permet d'obtenir. Pour effectuer les analyses pangénomiques, nous avons mis en œuvre la méthode du *Reduced Representation Bisulfite Sequencing* (RRBS). Cette méthode est basée sur du séquençage à haut-débit, mais sur une préparation d'ADN où seules les régions riches en CpG sont conservées. Ne séquencer qu'une fraction du génome permet d'augmenter la profondeur de séquençage pour chaque CpG, tout en réduisant les coûts. Il s'agit d'une bonne alternative au *Whole Genome Bisulfite Sequencing* (WGBS), qui est considéré comme la meilleure méthode mais qui se révèle onéreuse. En effet, elle nécessite de séquencer l'ensemble du génome, donc l'effort de séquençage à fournir est important pour que la profondeur de séquençage par CpG soit suffisante. Pour la caractérisation du méthylome (article n°1), des données de MeDIP-chip générées par Hélène Kiefer sont également présentées. L'utilisation du RRBS sur les mêmes échantillons que ceux utilisés en MeDIP-chip a alors permis de pouvoir comparer l'efficacité et la complémentarité des 2 méthodes.

Le nombre important de banques de RRBS à produire pour l'identification des biomarqueurs de fertilité (Résultat partie n°2) nous a conduit à implémenter le protocole de RRBS sur un automate de pipetage (Figure 60). Cet automate a permis de produire l'ensemble des banques à haut-débit (généralement 12 banques à la fois, jusqu'à un maximum de 24) et de manière hautement reproductible (réduction du biais dû au manipulateur). Pour traiter les données issues du séquençage des banques de RRBS, il a également été nécessaire de mettre au point l'ensemble de la procédure de traitement bio-informatique des données.

Les méthodes du bisulfite-pyroséquençage, de LUMA et du RRBS sont détaillées dans cet ordre dans la section suivante. Pour le RRBS sera également détaillé l'ensemble de la procédure de traitement bio-informatique des données.

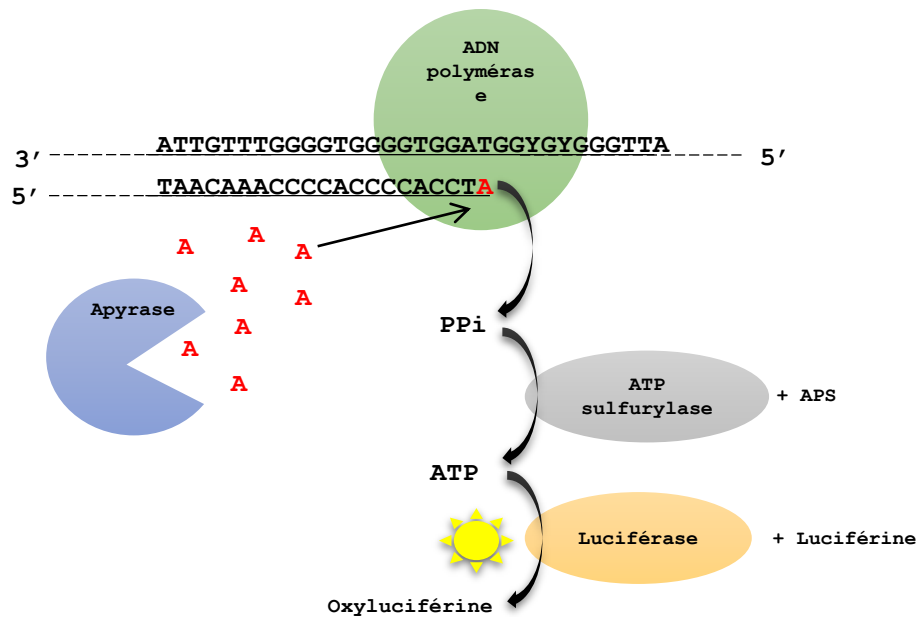
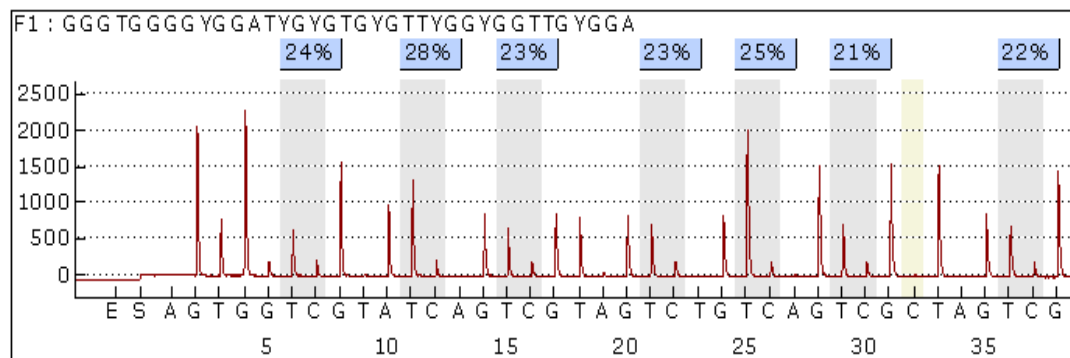
A**B****C**

Figure 61. Principe de la mesure de la méthylation de l'ADN par pyroséquençage

A. Photographie d'un appareil de mesure PyroMark Q24 (Qiagen) tel que celui utilisé au laboratoire. **B.** Principe du pyroséquençage. L'amorce de pyroséquençage est hybridée à une matrice (ADN simple brin ; souvent le brin antisens du produit de PCR) biotinylée et mis en présence d'un mélange d'enzymes (ADN polymérase, ATP sulfurylase, luciférase et apyrase) et de substrats (adénosine 5' phosphosulfate (APS) et luciférine). Les dNTP sont ajoutés de manière séquentielle au mélange réactionnel. Un pyrophosphate (PPi) est relargué à l'incorporation de chaque dNTP au brin d'ADN synthétisé par la polymérase. Le PPi est converti en ATP par l'ATP sulfurylase en présence d'APS. Cet ATP permet la conversion de la luciférine en oxyluciférine par la luciférase. Cette réaction enzymatique produit un rayonnement en lumière visible détecté par un dispositif couplé à un photomultiplicateur et une caméra. La lumière détectée est donc proportionnelle à la quantité d'ATP produite, elle-même proportionnelle à la quantité de PPi libéré et donc au nombre de dNTP incorporés au brin nouvellement formé. L'apyrase est une enzyme dégradant l'ATP et les dNTP non incorporés du mélange réactionnel. La hauteur du pic est fonction de la quantité de nucléotides incorporés sur une position. **C.** Exemple d'un pyrogramme fourni après analyse d'une région F1 dont la séquence est listée en haut de la figure. La séquence en bas de la figure correspond à la distribution séquentielle de nucléotides. Cette région contient 7 CpG et un pourcentage de méthylation est calculé pour chacun d'entre eux. (E, enzymes; S, substrat)

2.1. Le bisulfite-pyroséquençage

C'est une méthode qui a été décrite pour la première fois pour l'analyse des SNP (Ronaghi et al. 1996), puis adaptée à la mesure de la méthylation de l'ADN (Dupont et al. 2004). En premier lieu l'ADNg doit être converti au bisulfite de sodium. L'ADNg traité au bisulfite sert ensuite de matrice pour une amplification par PCR des régions ciblées. Le dessin des amorces de PCR doit prendre en compte les changements de bases dus au traitement au bisulfite de sodium. Un grand nombre de cycles d'amplification est nécessaire (jusqu'à 50), et ce pour plusieurs raisons. L'amplification est techniquement délicate du fait de la présence de sels résiduels du traitement bisulfite, et ce malgré une étape de purification sur colonnes. Elle l'est également à cause de la perte de complexité de la matrice bisulfitée qui est majoritairement composée de 3 bases avec de nombreuses répétitions de T.

Un nombre élevé de cycles permet d'atteindre le plateau d'amplification et d'obtenir une quantité homogène d'amplicons entre échantillons, qui peuvent avoir subi des pertes variables au cours des différentes étapes de la conversion. Enfin, cela sert aussi à épuiser les amorces de PCR, dont l'une est biotinylée (généralement l'amorce antisens à son extrémité 5'). Ainsi lors de la purification de l'amplicon sur billes recouvertes de Streptavidine, il ne peut y avoir compétition avec l'amorce antisens biotinylée. Du fait du nombre important de cycles, une attention particulière doit être portée pour éviter les contaminations (notamment en dissociant dans le temps et l'espace les étapes précédant et suivant l'amplification).

La dernière étape du protocole est le pyroséquençage à proprement parler (Figure 61). Il s'agit d'une technique de séquençage direct ne nécessitant pas de sous-clonage. Comme les techniques de séquençage classiques, le pyroséquençage nécessite d'obtenir un brin d'ADN néosynthétisé à partir d'une matrice (le brin antisens du produit de PCR) et d'une amorce de pyroséquençage. Son originalité est de proposer séquentiellement le nucléotide à incorporer à chaque position, puis de doser cette incorporation de manière précise et quantitative. Au niveau des C situées dans un contexte CpG, l'appareil propose séquentiellement du T puis du C. La mesure de l'incorporation respective de ces deux nucléotides permet de calculer le taux de méthylation. En effet, la libération de pyrophosphate (PPi) lors de l'incorporation d'un nucléotide fournit le premier substrat d'un système enzymatique fournisseur d'énergie couplé à une production lumineuse. L'intensité de lumière émise est corrélée à la quantité de nucléotides incorporés. Des contrôles sont réalisés au niveau de C non méthylables (situés en dehors d'un contexte CpG) permettent de confirmer que la conversion au bisulfite est totale, et que le polymorphisme C/T des CpG renvoie bien au taux de méthylation.

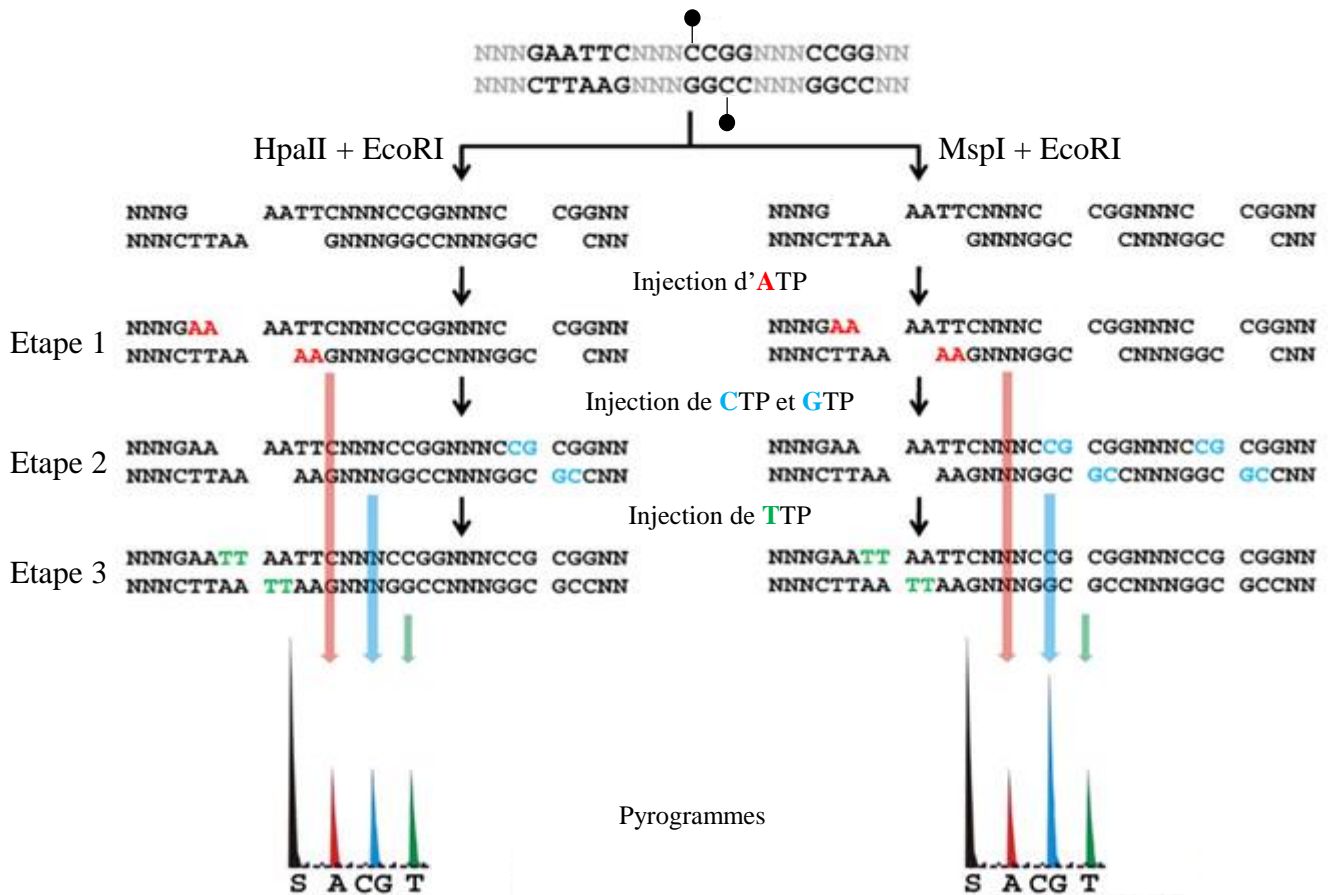


Figure 62. Principe de la mesure de la méthylation de l'ADN par LUMA

L'ADN génomique est digéré avec la combinaison d'enzymes MspI+EcoRI ou la combinaison HpaII+EcoRI. Cette digestion libère des bouts cohésifs TTAA (issus de la digestion par EcoRI) et CG (issus de la digestion par MspI ou HpaII). La procédure de pyroséquençage se déroule ensuite comme décrit précédemment. Les pics A (en rouge) et T (en vert) sur les pyrogrammes sont issus des étapes 1 et 3. Le pic de CG (en bleu) pendant l'étape 2 résulte de l'action de MspI ou HpaII. Le pic S correspond au pic de « substrat » injecté automatiquement par le programme en début de pyroséquençage. Le profil de méthylation est finalement déterminé par l'analyse du ratio de l'intensité du pic CG de HpaII par rapport à celui de MspI. Le ratio HpaII/MspI est proche de 1 pour un échantillon totalement déméthylé, et proche de 0 pour un échantillon totalement méthylé. La formule $\% \text{ méthylation} = 1 - (([CG] / [AT])_{\text{HpaII}} / ([CG] / [AT])_{\text{MspI}}) * 100$ permet alors de calculer le pourcentage global de méthylation des sites CCGG du génome, en prenant en compte la digestion par EcoRI pour la normalisation.

2.2. Le *LUMinometric Methylation Assay* (LUMA)

La méthode du LUMA a été développée par Karimi et ses associés (Karimi et al. 2006) (Karimi, Johansson, et Ekström 2006). Elle est basée sur la mesure de la méthylation globale par la comparaison de l'efficacité de coupure de l'ADN par les 2 enzymes isoschizomères MspI et HpaII (Figure 62). Toutes 2 clivent le site de restriction C[^]CGG, MspI indépendamment de l'état de méthylation du CG, et HpaII uniquement s'il n'est pas méthylé. Une fois l'ADNg digéré par l'une ou l'autre enzyme, les fragments générés servent de matrice pour le pyroséquençage. L'incorporation de nucléotides pour combler les bouts cohésifs générés par la digestion est faite selon la séquence programmée A/C+G/T/C+G/G/A/T. Ainsi, plus la méthylation globale du génome est faible, plus il y a de sites clivés par HpaII, et donc plus l'incorporation de nucléotides, de production d'énergie et d'intensité lumineuse seront élevées. Le calcul d'un ratio entre l'intensité de signal généré par l'ADN digéré par MspI ou HpaII permet d'établir la valeur de méthylation globale pour l'échantillon. La formule utilisée est la suivante :

$$\% \text{ méthylation} = 1 - (([\text{CG}] / [\text{AT}])_{\text{HpaII}} / ([\text{CG}] / [\text{AT}])_{\text{MspI}}) * 100$$

En plus des 2 isoschizomères, l'enzyme EcoRI est utilisée comme contrôle interne de la quantité d'ADNg présente dans les 2 digestions. Cette enzyme clive de manière équivalente les échantillons suivant le site G[^]AATTC. Son utilisation permet de s'affranchir des différences de concentration en ADNg de départ, qui peuvent avoir un impact sur les résultats. Bien qu'EcoRI ne possède pas de sites CpG au sein de son site de restriction, il a été montré que lorsque le site est suivi d'un G, les résultats peuvent être impactés du fait de la présence potentielle d'un groupement méthylé sur le C. De fait, d'autres normalisateurs ont été proposés dans la littérature (Lisanti et al. 2013). Au laboratoire des analyses *in silico* ont montré que cela ne concerne qu'un très petit nombre de sites, et que donc le biais est tout à fait négligeable (données non montrées).

Au laboratoire nous disposons d'un Pyromark Q24 qui permet de traiter jusqu'à 11 échantillons à la fois (MspI+EcoRI et HpaII+EcoRI par échantillon, plus un contrôle interne utilisant de l'ADNg de foie bovin démontrant la fiabilité entre les essais des mesures). Aussi, la durée d'un cycle de traitement est de seulement 15 minutes, ce qui permet de produire rapidement les données pour un grand nombre d'échantillons. Cette méthode n'est cependant pas sans présenter quelques inconvénients. Un d'entre eux est que cette méthode nécessite une quantité importante d'ADNg (1µg) ainsi que sa très bonne qualité : en effet, une mauvaise qualité est le reflet de clivages aléatoires de l'ADNg qui peut générer des sites cohésifs

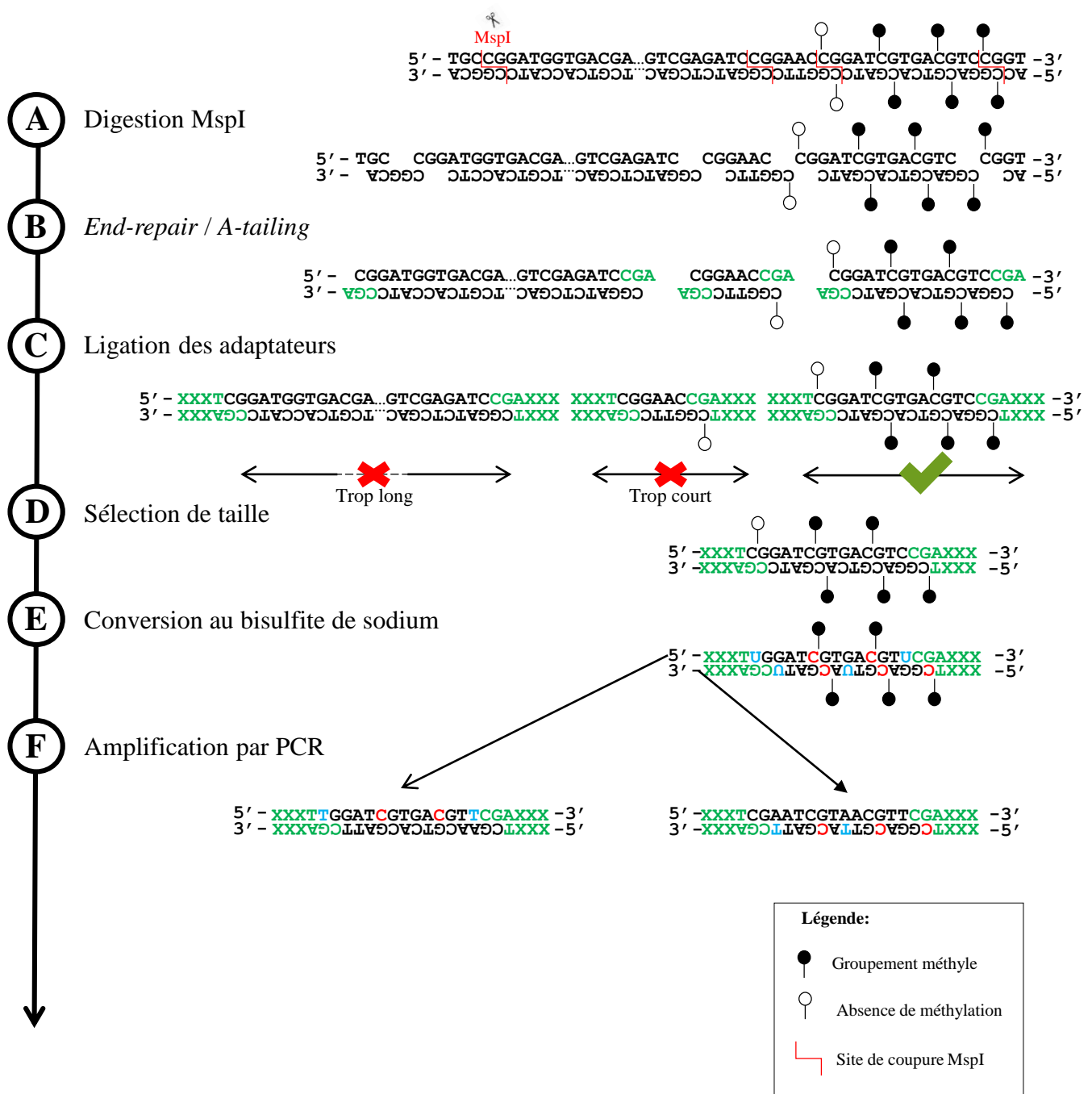


Figure 63. Etapes de préparation des banques RRBS

Pour ne pas encombrer la figure, sont représentées la présence ou l'absence de groupement méthyle uniquement sur les sites CpG et sur le fragment d'ADN conservé après la sélection de taille. **A.** Digestion enzymatique de l'ADNg avec l'enzyme de restriction MspI (C[^]CGG), indépendamment de l'état de méthylation du site CpG. **B.** End-repair/A-tailing. Les nucléotides ajoutés lors de ces étapes sont représentés en vert. Le end-repair permet d'ajouter les CG, et le A-tailing permet d'ajouter ensuite les A. **C.** Ligation des adaptateurs. Les adaptateurs sont représentés par les lettres XXX en vert. Nous avons utilisés des adaptateurs Illumina de 55 pb. **D.** Sélection de taille. Nous avons effectué une sélection de taille pour conserver les fragments de 150 à 400 pb (soit 40 à 290 pb si on ne prend pas en considération les 2 adaptateurs de 55 pb chacun). Les fragments trop longs ou trop courts sont par conséquent éliminés. **E.** Conversion de l'ADN au bisulfite de sodium. Les cytosines hors sites CpG et non-méthylées sont converties en uraciles (représentés en bleu). Les cytosines non-méthylées des sites CpG sont également converties en uraciles (représentés également en bleu). Les cytosines méthylées des sites CpG restent des cytosines. **F.** Amplification par PCR. Les uraciles sont alors remplacés par des thymines. L'information de méthylation est perdue au cours des cycles de PCR mais a préalablement été convertie en polymorphisme de séquence, qui pourra être révélé après le séquençage des fragments.

indépendants des digestions enzymatiques et augmenter le bruit de fond. Elle ne permet également de ne cibler que les CpG des sites CCGG, donc d'obtenir seulement une estimation du niveau global de méthylation car l'état de méthylation de ce site ne permet pas totalement d'extrapoler à l'état de méthylation de tous les CpG du génome. Cette méthode permet donc uniquement de comparer les conditions et les échantillons entre eux. De plus, les sites CCGG ne sont pas distribués de façon homogène dans le génome. Ils tendent en effet à être concentrés dans les régions denses en CpG, dont notamment certains éléments répétés (Fazzari et Greally 2004). Enfin, elle consomme une quantité assez importante d'ADN. Une optimisation du protocole au laboratoire a permis de partir de 1µg d'ADN de départ pour avoir de bons résultats.

2.3. Le *Reduced Representation Bisulfite Sequencing* (RRBS)

La méthode du RRBS a été développée par Meissner et son équipe en 2005 (Meissner 2005) (Gu et al. 2010) (Auclair et al. 2014). L'ensemble de la procédure est subdivisé en 3 étapes majeures : la construction et le séquençage des banques, ainsi que le traitement bio-informatique des données issues du séquençage. Elles peuvent optionnellement (ce qui a été notre cas) être précédées d'une étape d'analyse *in silico* pour optimiser la méthode en fonction de l'espèce étudiée.

2.3.1. Optimisation du protocole par analyses *in silico*

Les analyses *in silico* permettent de générer des jeux de données virtuels (appelés RRgenome pour *Reduced Representation genome*) qui peuvent être traités de la même façon qu'un jeu de donnée réel dans la procédure bio-informatique. Ce type d'analyse a déjà été développé et utilisé pour plusieurs espèces (Wang et al. 2012) (Chatterjee et al. 2013). A partir de la banque RRBS artificielle, des scripts permettent de calculer automatiquement des paramètres comme la taille du génome *in silico* (en pb), la proportion de génome total que ce génome *in silico* représente, le nombre de fragments générés et le nombre de sites CpG qu'ils contiennent, etc. Au début du projet SeQuaMol, et après avoir produit plusieurs banques selon le protocole RRBS classique, nous avons observé que les conditions ne permettaient pas d'éliminer totalement les séquences répétées. Nous avons alors utilisé l'approche *in silico* dans le but d'expérimenter de multiples conditions expérimentales, afin d'optimiser la procédure de production des banques RRBS au génome bovin. Nous avons également appliqué cette approche pour de multiples espèces dans le but d'en comparer les résultats. Les données issues de ces analyses se trouvent dans les articles n°1 et n°2 de ce document.

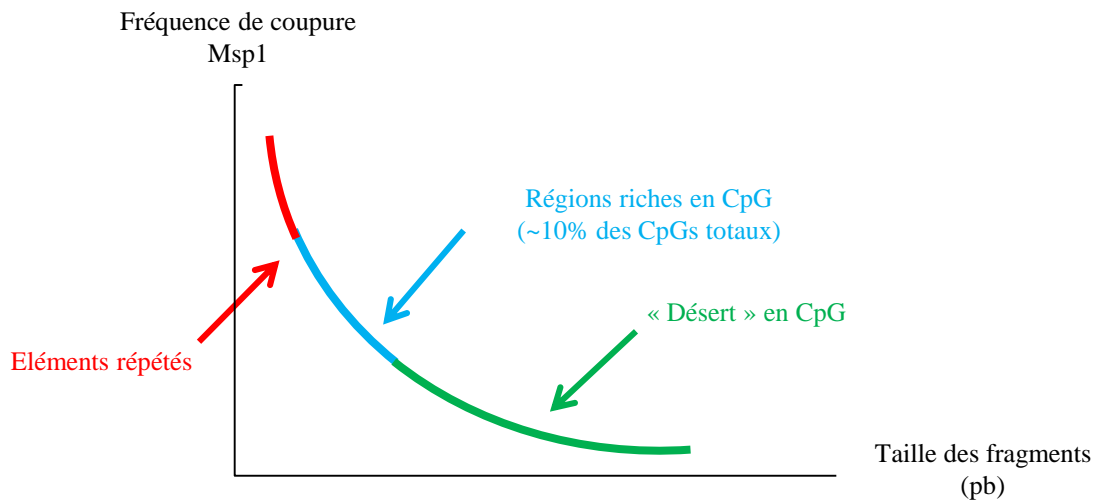
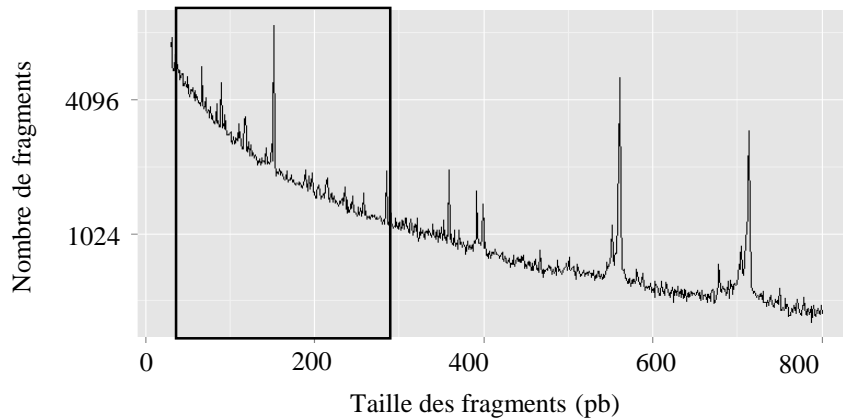
A**B**

Figure 64. Sélection de taille des fragments et densité en sites CpG

A. Représentation schématique de l'intérêt de l'étape de sélection de taille des fragments. L'axe des abscisses représente la taille des fragments générés par la digestion enzymatique du génome par MspI, et en ordonnée la fréquence de coupure de cette même enzyme. On observe que la fréquence de coupure est inversement proportionnelle à la taille des fragments générés. Ainsi, les petits fragments proviennent de régions très riches en CCGG, mais également très riches en CpG. Ces régions sont fréquemment retrouvées dans certains éléments répétés du génome. En parallèle, les très grands fragments proviennent de régions pauvres en sites CpG et CCGG (ici appelés « déserts » en CpG). Ces fragments ne sont pas de grand intérêt car leur séquençage ne fournirait de l'information que pour peu de sites CpG. En ne sélectionnant que des fragments de taille modérée, on se focalise sur des régions relativement riches en sites CpG. Ainsi, on optimise l'étape de séquençage. **B.** Distribution théorique de la taille des fragments générés par digestion *in silico* du génome bovin par MspI. Est encadrée la fenêtre de taille que nous avons utilisée pour nos analyses (40-290 pb). On peut observer la présence d'un certain nombre de pics, et notamment au sein de notre fenêtre de sélection. Ces pics correspondent à une sur-représentation de fragments d'une taille donnée. Ces pics sont potentiellement révélateurs de la présence d'éléments répétés.

2.3.2. La préparation des banques :

La préparation des banques de RRBS combine l'utilisation d'une enzyme de restriction et le traitement de l'ADN au bisulfite de sodium (Figure 63, page précédente). Elle débute par la digestion enzymatique de l'ADNg par l'enzyme MspI. Cette enzyme clive l'ADNg selon les sites 5'-C[^]CGG-3', indépendamment de l'état de méthylation du site CpG. La seconde étape est une combinatoire de ce qu'on appelle le « *end-repair* » et le « *A-tailing* ». Cette étape consiste à combler les extrémités des fragments d'ADN de type cohésif générés par le clivage enzymatique (*end-repair*). Ensuite, l'ajout dans le milieu réactionnel d'un excès d'adénine va provoquer l'ajout de ce nucléotide en 3' des fragments (*A-tailing*). Cette étape est en fait un prérequis à l'étape suivante de ligation des adaptateurs. En effet, l'adénine supplémentaire ajoutée en bout de fragment permet lors d'une nouvelle incubation de lier des adaptateurs de 55pb qui sont nécessaires aux étapes suivantes d'amplification PCR et de séquençage. Ces adaptateurs présentent en effet une thymine à une de leurs extrémités, permettant ainsi la liaison des adaptateurs aux fragments d'ADN.

Une des étapes clés du protocole de RRBS est la sélection de taille. Suite à la digestion enzymatique de l'ADNg par MspI, on obtient dans le milieu réactionnel un grand nombre de fragments d'ADN de tailles diverses. La taille des fragments est fonction de la densité en sites CpG des régions génomiques (Figure 64). En effet, plus une région est riche en sites CCGG, plus les fragments générés sont de courte taille. L'objectif de cette étape du protocole est de ne conserver que des fragments compris entre une certaine fenêtre de taille. L'élimination des fragments de trop courte ou de trop longue taille permet d'optimiser l'efficacité de l'étape de séquençage. Au laboratoire, cette étape a tout d'abord été effectuée par purification des fragments, après migration sur gel agarose et excision du gel. Cette méthode a été utilisée pour générer les banques de RRBS fournissant les données présentées dans l'article n°1. Pour générer les 58 banques de RRBS présentées dans l'article n°2, nous avons adapté la méthode afin d'automatiser l'ensemble des manipulations. Il a été nécessaire de modifier le protocole de l'étape de sélection des fragments selon la taille : cette étape utilise maintenant des billes magnétiques SPRIselect (Beckman Coulter). Les fragments d'ADN issus des étapes précédentes sont mis en contact de billes magnétiques sur lesquelles ils peuvent se fixer. Pour une quantité donnée d'ADN dans un volume fixe, augmenter la proportion de billes permet de fixer des fragments de plus en plus petits. En présence d'un aimant, les billes sont immobilisées et il est alors possible soit d'éliminer les fragments d'ADN libres et de conserver les fragments piégés, soit au contraire de prélever le surnageant et d'éliminer les fragments piégés. Pour la préparation des banques de RRBS cette sélection par billes

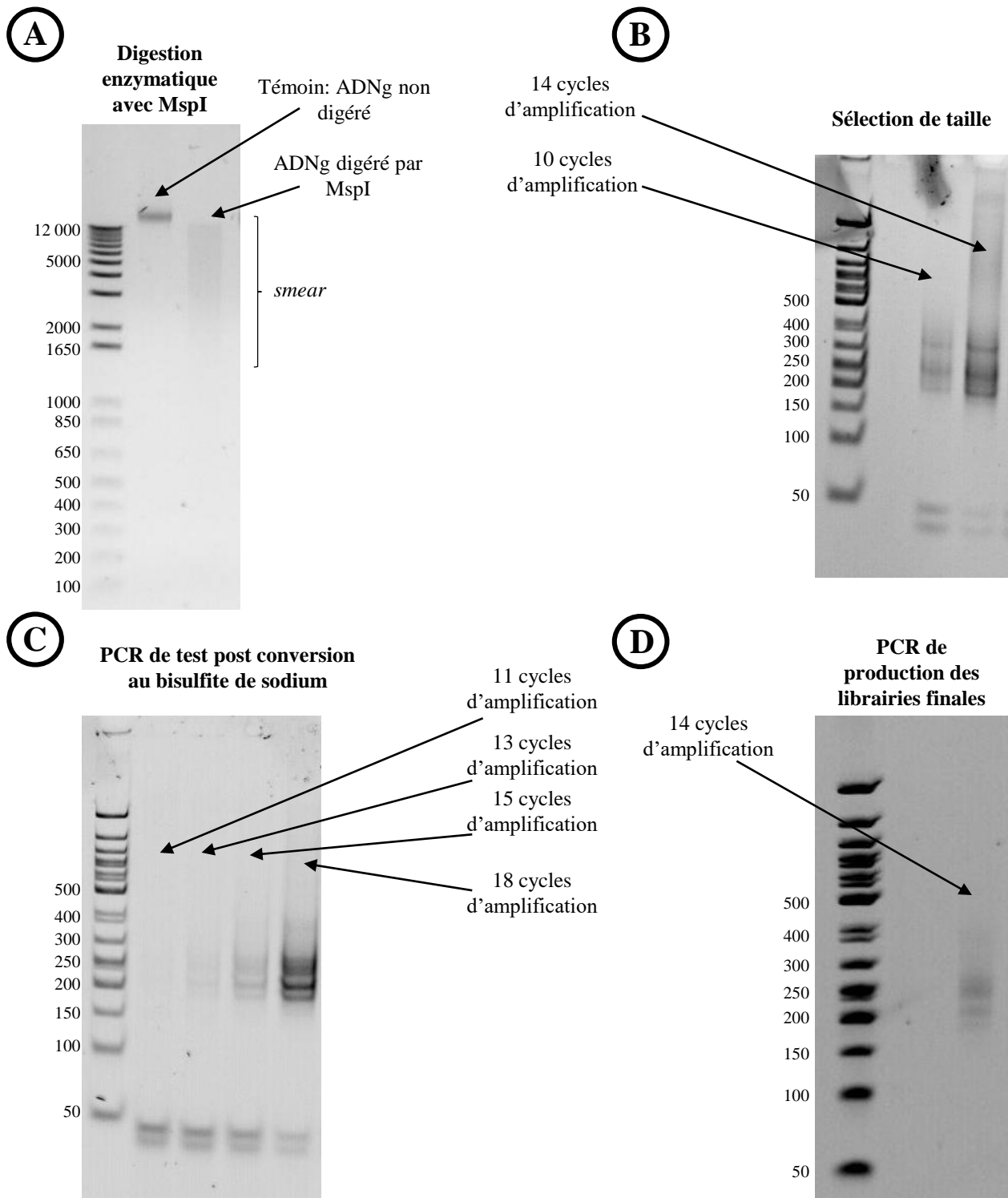


Figure 65. Contrôles qualité de production des banques RRBS

A. Electrophorèse sur gel à 1% d'agarose après digestion de l'ADNg par l'enzyme de restriction MspI. Une bande diffuse d'ADN signifie que la digestion a été efficace. **B.** Electrophorèse sur gel polyacrylamide 4-20% après les étapes de *end-repair*, *A-tailing* et de sélection de taille (fragments de 150 – 400 pb en comptant les adaptateurs). Avant la migration les fragments ont été amplifiés par PCR en utilisant 10 ou 14 cycles d'amplification. **C.** Electrophorèse sur gel polyacrylamide 4-20% après les étapes de conversion de l'ADN au bisulfite de sodium. Avant la migration, les fragments ont été amplifiés par PCR en utilisant 11, 13, 15 ou 18 cycles d'amplification. Cette étape permet de sélectionner le nombre d'amplifications nécessaires pour la production finale des banques de RRBS. **D.** Electrophorèse sur gel polyacrylamide 4-20% après les étapes d'amplification PCR et de purification finale des banques. (L'échelle située sur la gauche des images représente la taille des fragments en paires de bases)

magnétiques est effectuée en 2 temps. Dans un premier temps, on met en solution les billes magnétiques et l'ADN, et après passage sur l'aimant on conserve le surnageant. On élimine ainsi les gros fragments d'ADN qui sont en dehors de notre fenêtre de sélection. Dans un second temps, après une seconde mise en solution des billes et de l'ADN, on récupère cette fois les billes et on élimine le surnageant. A cette étape sont alors éliminés les fragments trop petits et qui ne rentrent pas dans la fenêtre de sélection. Au final, pour l'ensemble des banques RRBS préparées, nous avons sélectionné des fragments allant de 40 à 290 pb.

Les fragments d'ADN sont ensuite convertis au bisulfite de sodium. Nous avons pris la décision de réaliser 2 étapes de conversion successives. La raison est qu'une seule conversion ne permet d'obtenir en général qu'une efficacité d'environ 95%. Cette efficacité n'est pas suffisante pour assurer un calcul des taux de méthylation satisfaisant. La seconde étape de conversion permet de passer à une efficacité de plus de 99%.

Les fragments d'ADN convertis sont ensuite purifiés sur colonne et amplifiés par PCR. Le nombre de cycles d'amplification est contrôlé afin d'obtenir suffisamment de matériel biologique tout en limitant les biais d'amplification. Pour toutes les analyses présentées dans ce document, nous avons effectué 14 cycles d'amplification. Les banques RRBS obtenues après amplification sont enfin purifiées en utilisant à nouveau un système de billes magnétiques. La quantité d'ADN présente dans chaque banque est mesurée avec un fluoromètre Qubit, en utilisant le kit commercial *Quant-iT ds DNA HS Assay*, qui permet de mesurer à haute sensibilité de petites concentrations en ADN. Nous avons obtenu une moyenne de $16,2 \pm 12,4$ ng/ μ l d'ADN par banque, qui correspond à une concentration de $112,4 \pm 86,1$ nM.

Tout au long du processus de préparation des banques, des contrôles qualité sont réalisés (dont des exemples sont donnés Figure 65). Le premier est un contrôle de l'efficacité de la digestion enzymatique par électrophorèse sur gel d'agarose. D'autres sont effectués en faisant migrer les fragments sur un gel polyacrylamide et sont effectués aux étapes suivantes :

3. Après l'étape de sélection de taille
4. Après les 2 conversions successives au bisulfite de sodium
5. Après la purification des banques

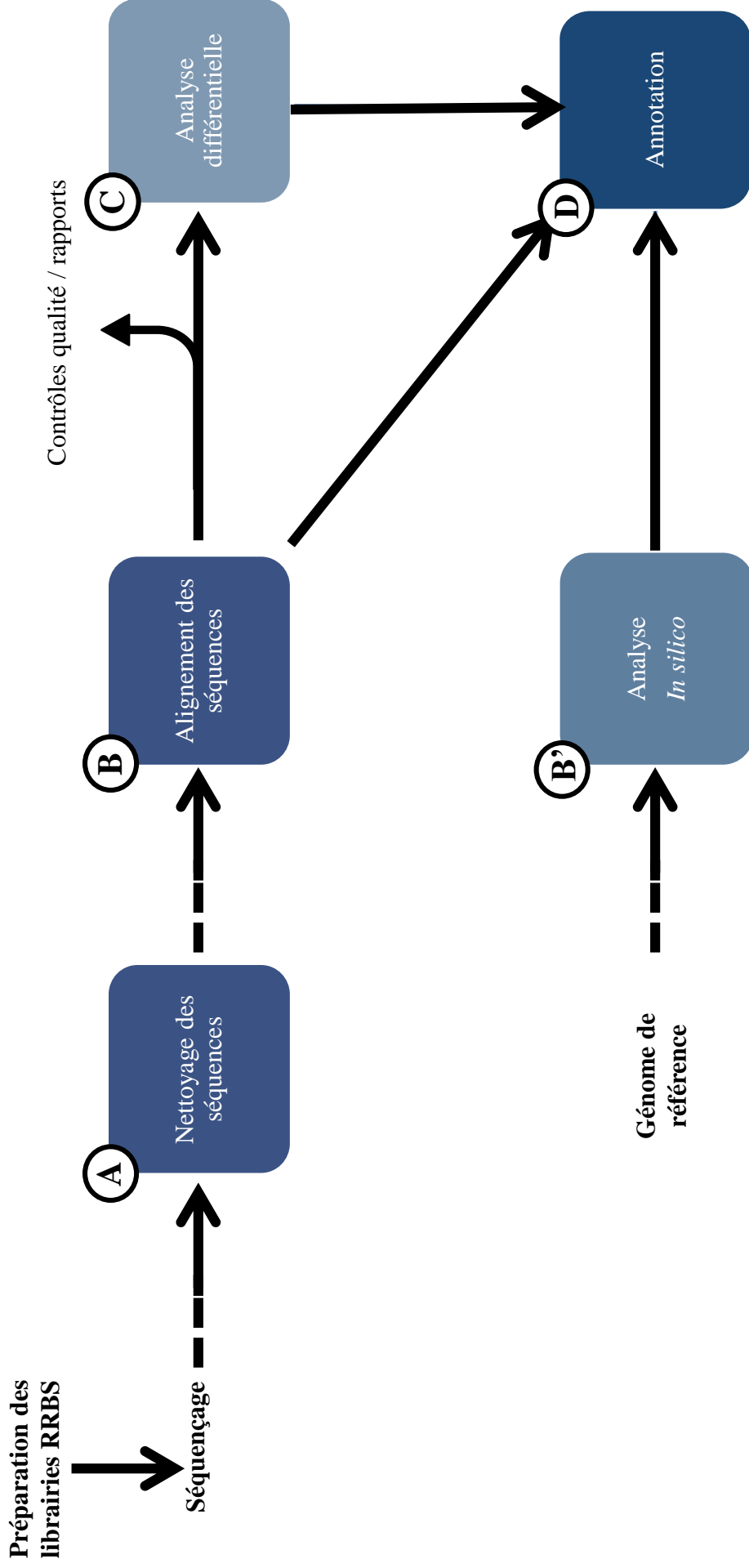


Figure 66. Vue d'ensemble de la procédure de traitement bio-informatique des données de RRBS

Suite à la préparation des librairies RRBS, celles-ci sont séquencées en NGS. Cela permet d'obtenir des fichiers au format .fastq. Ces fichiers contiennent les séquences de tous les fragments d'ADN. **A.** Nettoyage des séquences pour éliminer les bases et les fragments de mauvaise qualité ou trop courts. **B.** Alignement sur le génome de référence. Cette étape permet de positionner les séquences sur le génome, et de calculer un pourcentage de méthylation pour les sites CpG alignés. **C.** Analyse différentielle. Permet d'identifier les DMC et les DMR entre 2 groupes d'individus à partir des pourcentages de méthylation par CpG issus de l'alignement. **D.** Analyse d'annotation. Celle-ci peut-être effectuée sur les DMC et les DMR, mais également directement sur les CpG issus de l'alignement sur le génome. Permet de définir dans quel élément génomique le DMC/DMR/CpG se situe. **B'.** Analyse *in silico*. L'analyse *in silico* permet, à partir du génome de référence, de simuler la digestion par une enzyme de restriction ainsi que l'étape de sélection de taille des fragments. Les fragments ainsi générés peuvent être ensuite utilisés pour une analyse d'annotation.

2.3.3. Le séquençage des banques

Les banques préparées ont été envoyées à la société Integragen (Evry, France) pour effectuer le séquençage à haut-débit. Le séquençage a été réalisé sur un séquenceur Illumina HiSeq 4000, en *paired-end* (les fragments d'ADN sont séquencés dans les 2 sens de lecture, ce qui génère des fragments R1 et des fragments R2). La taille des séquences lues est de 75 pb. La profondeur de séquençage choisie est de 25 millions par échantillon (avec une marge $\pm 15\%$).

2.3.4. Le traitement bio-informatique des données

Une fois les données de séquençage obtenues (sous la forme de fichiers au format *.fastq*) il est nécessaire de les interpréter. A mon arrivée au laboratoire la procédure de traitement bio-informatique des données n'était pas développée. Un groupe de travail auquel j'ai activement participé, et réunissant les bio-informaticiens de l'unité, Luc Jouneau et François Piumi et les chercheurs de notre équipe a été mis en place pour développer cette procédure d'analyse. Ce travail a permis de mettre au point une procédure de traitement de données RRBS semi-automatisée, et relativement simple d'utilisation pour un « non bio-informaticien ».

La procédure a été rendue disponible pour la communauté scientifique et peut être téléchargée à l'adresse <https://github.com/FAANG/faang-methylation/tree/master/RRBS-toolkit/>. Un schéma représentatif de l'ensemble de la procédure de traitement bio-informatique est donné Figure 66. La procédure que nous avons mise en place intègre l'utilisation d'une succession de logiciels déjà existants ainsi que des éléments développés au laboratoire en langage Python, R et Shell.

Nettoyage des séquences et contrôles qualité

La première étape consiste à nettoyer les séquences. Cette étape, appelée *Trimming* (qu'on peut traduire en français par « élagage »), est réalisée en utilisant le logiciel TrimGalore (Babraham Institute). Le *trimming* consiste à éliminer les séquences des adaptateurs de séquençage ainsi que les fragments et les bases de trop mauvaise qualité. Le seuil de qualité (phred) est fixé à 20, ce qui correspond à une probabilité d'erreur d'identification d'une base de 1%. Le logiciel retire également 2 bases supplémentaires, dont la cytosine ajoutée artificiellement lors de l'étape de *end-repair* pour combler les bouts cohésifs. Nous avons également utilisé de manière automatique l'option `--trim1` qui permet de retirer encore une base supplémentaire en 3' de toutes les séquences, ce qui est nécessaire pour que les données

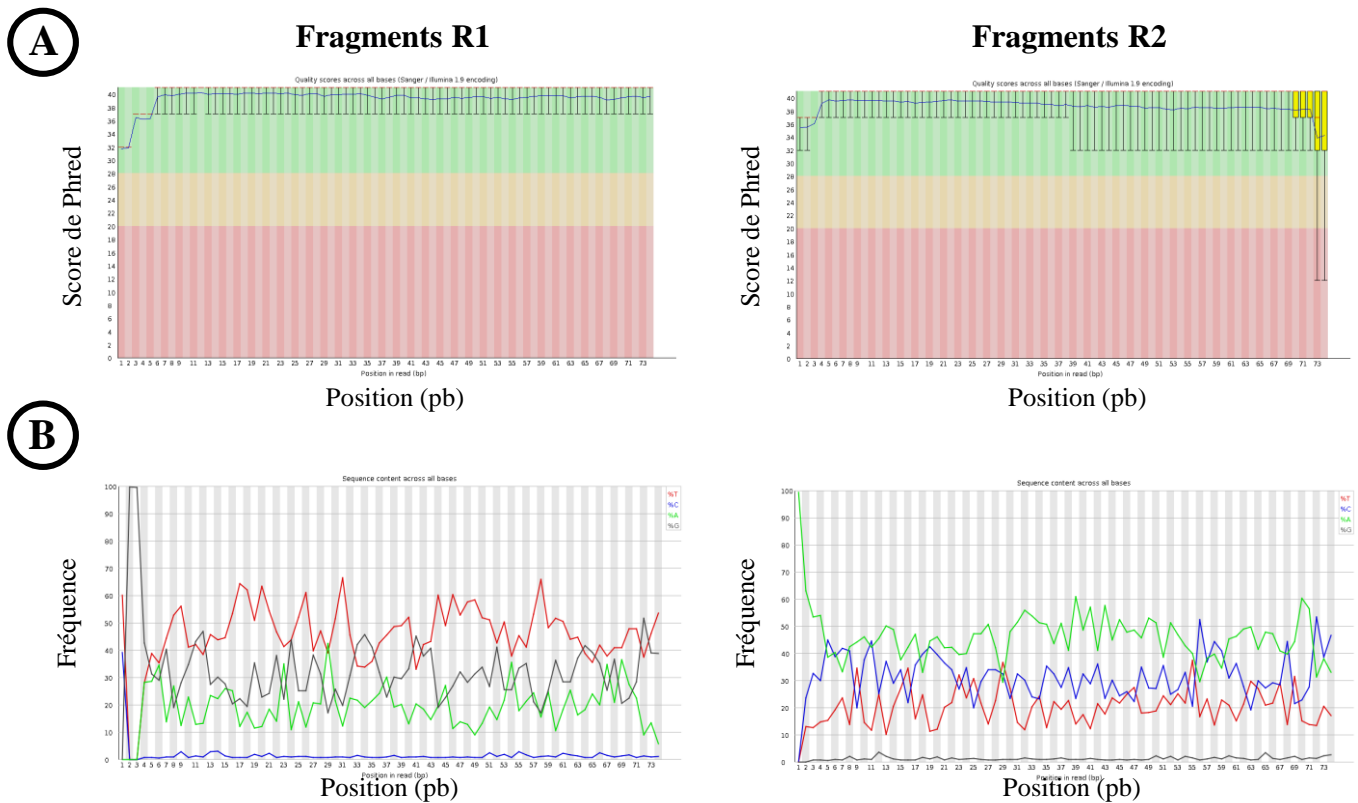


Figure 67. Rapports de contrôles qualité fournis par le logiciel FastQC

A. Cette représentation montre le degré d'incertitude du séquençage pour chaque base. Nous avons choisi pour les analyses un Phred de 20, ce qui signifie que la précision du séquençage est de 99% (limite basse du rectangle orange). Les bases dont le score est inférieur à 20 sont éliminées. **B.** Ce graphique représente la proportion en base à chaque position sur la séquence. En RRBS, chaque séquence R1 doit commencer par C ou T (en fonction de l'état de méthylation) puis GG. Le fragment R2 doit commencer par CAA. Ici il s'agit des FastQC fournis après l'étape de *trimming*, donc 2 nucléotides ont été retirés en R2, c'est pourquoi la séquence débute par un A à 100%. Ce graphique nous renseigne alors sur la bonne préparation des banques, ainsi que sur le succès de la conversion au bisulfite de sodium (taux de cytosine très bas en R1).

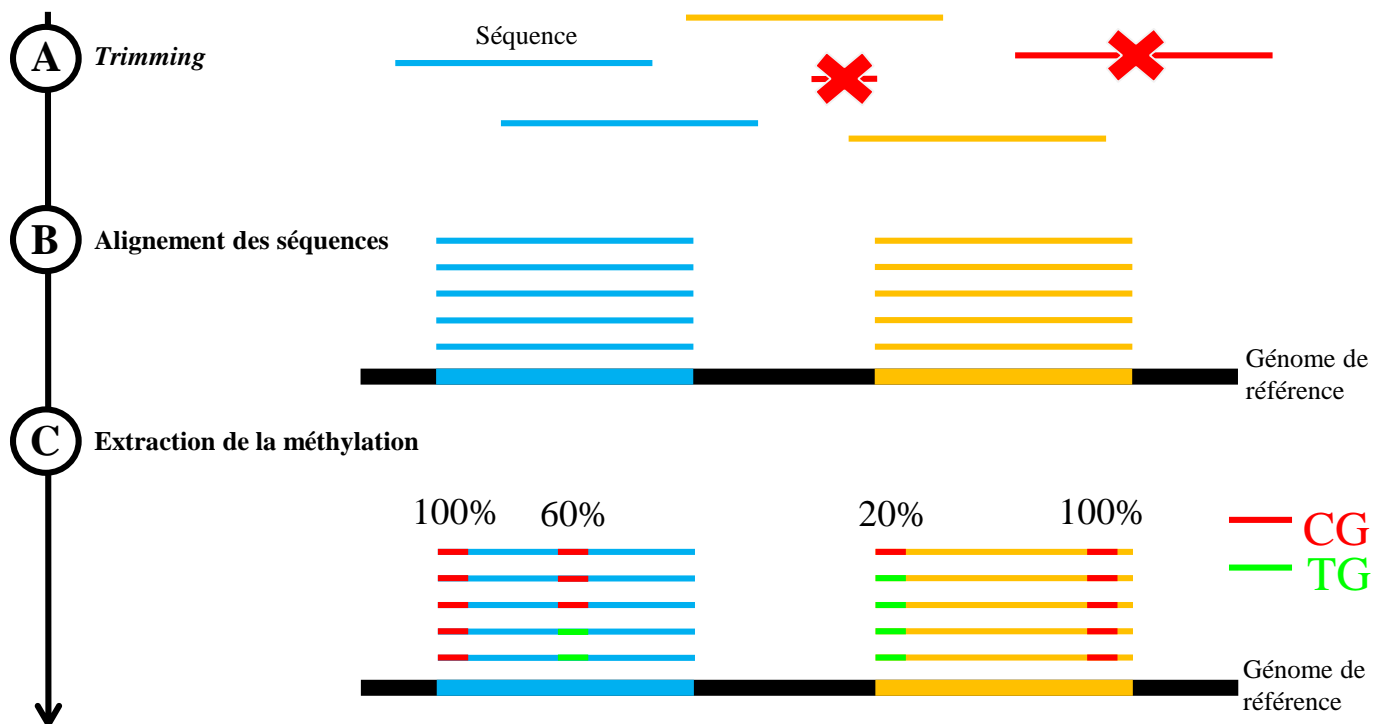


Figure 68. Contrôle qualité, nettoyage des séquences et alignement sur le génome de référence

A. Etape de *trimming*. Les séquences et les bases de mauvaise qualité, et les séquences trop courtes sont supprimées. **B.** Alignement des séquences sur le génome de référence. **C.** Extraction de la méthylation. Pour chaque CpG, un pourcentage de méthylation est calculé, en tenant compte du nombre de séquences alignées (couverture). Le nombre de C par rapport au nombre de T est révélateur du pourcentage de méthylation sur le site. La suite de la procédure d'analyse est présentée dans la Figure suivante.

soient compatibles avec le logiciel Bowtie1. A la suite de ce premier « élagage », TrimGalore élimine finalement les fragments devenus trop courts par rapport à un seuil fixé (dans notre cas à 20pb). Lorsque le séquençage est effectué en *paired-end*, le programme élimine la paire de séquences si un des deux fragments est trop court.

A la fin de cette étape et de manière automatisée, sont également effectués des contrôles qualité en utilisant le logiciel FastQC (Andrews et al. 2010). Ceci est nécessaire pour avoir des informations sur la bonne préparation des banques et sur les données de séquençage. Des exemples de rapports de qualité fournis par le logiciel FastQC sont donnés Figure 67.

Alignement sur le génome de référence

La seconde étape est celle de l'alignement des séquences sur le génome de référence (Figure 68). Nous avons intégré au pipeline le logiciel Bismark (Krueger et Andrews 2011), qui fait lui-même appel au logiciel d'alignement Bowtie 1 (Langmead et al. 2009). Le logiciel Bismark est spécialisé dans l'alignement de séquences bisulfitées. Ce type de logiciel spécialisé est nécessaire du fait de la baisse de complexité des séquences d'ADN après la conversion au bisulfite de sodium, qui passe de 4 à 3 bases. Ainsi, lors d'une première étape, deux génomes artificiels sont créés à partir du génome de référence : un dans lequel l'ensemble des C est remplacé par des T, l'autre dans lequel l'ensemble des G est remplacé par des A (correspondant au complémentaire de la conversion C en T). Chaque séquence d'ADN obtenue par séquençage est également convertie de cette façon. Le logiciel cherche ensuite le meilleur alignement possible des séquences sur les génomes ainsi modifiés, c'est-à-dire celui comportant le moins de mésappariements. Concernant les options utilisées pour l'alignement, nous avons opté pour n'autoriser qu'un seul mésappariement sur le génome. Si une séquence obtient un résultat maximal d'alignement pour une position sur le génome, on dit que l'alignement est unique. Dans certains cas, le meilleur résultat d'alignement est identique sur plusieurs positions génomiques. On parle alors d'alignements multiples. Pour nos analyses, nous avons décidé d'éliminer les séquences à alignements multiples car il est alors impossible de déterminer quelle est la position réelle de la séquence dans le génome. Dans le cas où les séquences ne s'alignent pas du tout sur le génome en fonction des critères fixés, elles sont éliminées.

Après l'alignement, des scripts développés au laboratoire permettent de générer automatiquement un fichier contenant l'information de la position, la couverture et le pourcentage de méthylation associé à chaque CpG. Le pourcentage de méthylation d'un CpG est calculé en comptant le nombre de C par rapport la couverture totale du CpG. Par exemple,

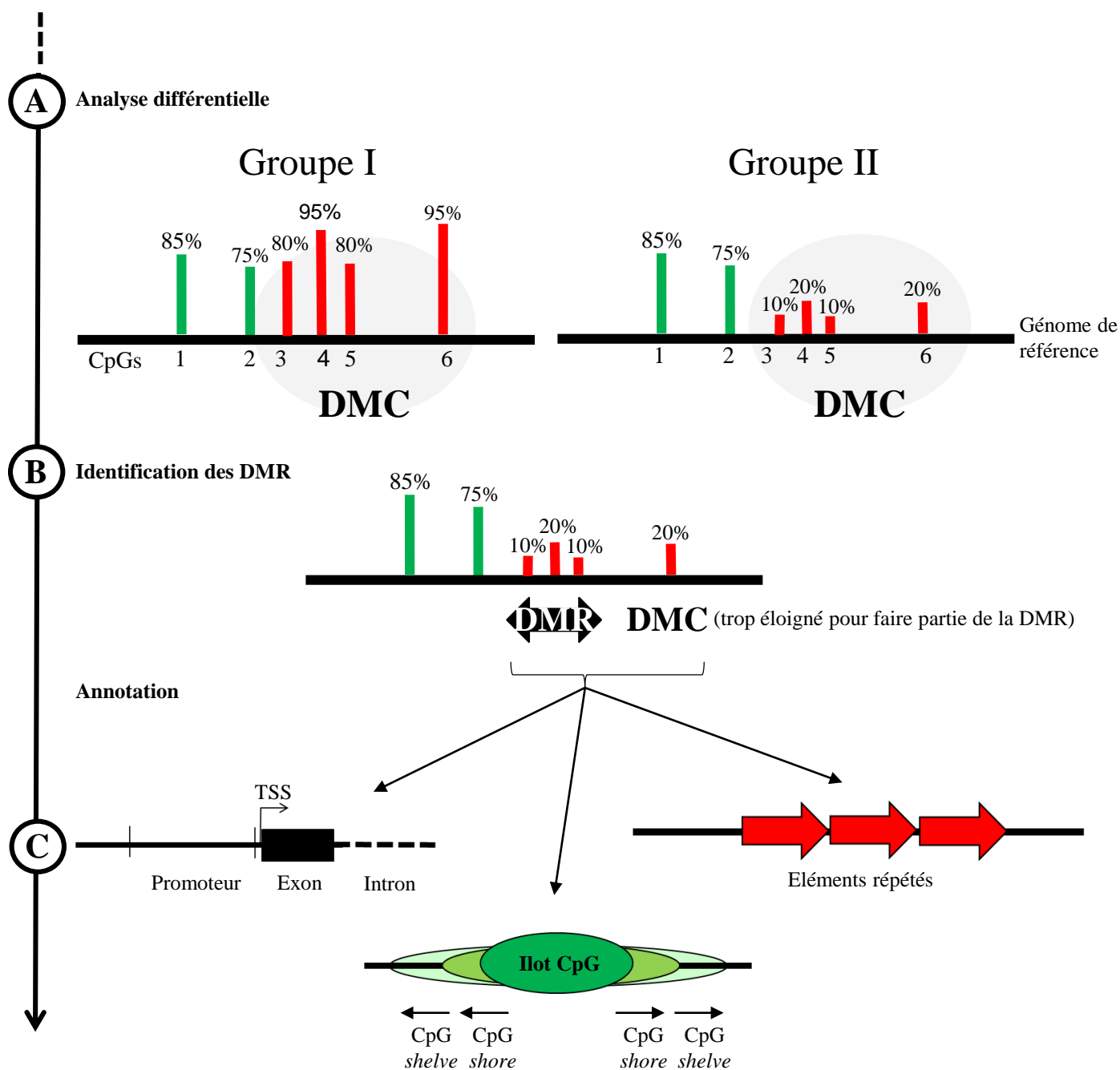


Figure 69. Identification et annotation des cytosines et régions différemment méthylées

A. Analyse différentielle. Permet de comparer l'état de méthylation des CpG pour 2 groupes d'individus donnés (ici appelés groupes I et II) et d'identifier quelles sont les cytosines pour lesquelles il existe un différentiel de méthylation statistiquement significatif. Ici, les CpG en vert ne présentent pas de différentiel de méthylation entre les conditions I et II, au contraire des CpG représentés en rouge présentent un différentiel de méthylation supérieur à 25% et sont alors considérés comme des DMC. **B.** Identification des DMR. Si plusieurs DMC sont localisés au sein d'une même région, les DMC peuvent être regroupées pour former une DMR. Nous avons défini comme DMR les régions contenant au moins 3 DMC dans une région de moins de 100pb. **C.** Analyse d'annotation. Cette analyse permet d'associer les DMC et les DMR avec des éléments génomiques d'intérêt, comme les différents éléments des gènes (promoteur-TSS, exon, intron, etc.), la densité en CpG ou encore les éléments répétés.

si un CpG est couvert par 10 séquences et que parmi elles on compte 3 C et 7 T, alors le pourcentage de méthylation du CpG est de 30%. Avec Bismark le pourcentage de méthylation sur les sites CHH et CHG est également calculé.

A l'étape d'alignement est également généré de manière automatique un rapport sur les résultats des étapes précédentes, ainsi que des statistiques générales sur la banque de RRBS. Nous obtenons par exemple des données sur le nombre de paires de séquences analysées, les taux d'alignements, le taux de conversion au bisulfite de sodium, des pourcentages globaux de méthylation en fonction de différents seuils de couverture des CpG, etc. Des graphiques sont également générés pour illustrer ces informations, présentant notamment la distance entre les fragments R1 et R2, la répartition de la couverture des CpG, la distance entre 2 CpG, etc.

Identification des cytosines différenciellement méthylées entre 2 conditions

L'étape suivante de la procédure est l'analyse différentielle de la méthylation (Figure 69). Cette analyse permet de mettre en évidence les DMC statistiquement significatives entre 2 jeux de données. Nous avons essayé plusieurs logiciels pour cette étape de l'analyse et nous avons retenu le logiciel methylKit (Akalin et al. 2012). Ce logiciel permet en effet d'identifier un nombre assez important de DMC, tout en permettant de travailler avec des p-values ajustées (qvalue), ce qui est un bon équilibre entre l'identification d'un nombre suffisant de DMC tout en évitant de générer trop de faux-positifs.

Pour effectuer cette analyse il faut en premier lieu filtrer les CpG en fonction de leur profondeur de séquençage, qu'on appelle aussi leur couverture. Il faut en effet que la profondeur de séquençage par CpG soit suffisante pour que le calcul du pourcentage de méthylation soit robuste. En général, on conserve les CpG avec une profondeur de séquençage de 5 à 10 séquences. Il est également possible de fixer un seuil maximal de couverture (que nous avons fixé à 500 pour certaines analyses). L'analyse statistique est effectuée à l'aide d'un test de régression logistique qui prend en compte le nombre de cytosines méthylées par rapport à la couverture des cytosines. Pour un individu et une position génomique donnée, il est parfois possible que les CpG ne passent pas les seuils de couverture choisis ou alors ne soient pas couverts du tout. Nous avons donc mis en place un critère de *minsample*, qui consiste à analyser un CpG s'il est couvert au seuil choisi pour une partie seulement des individus par groupe, et non pour l'ensemble des individus. Cela permet d'éviter de générer des faux-positifs si un nombre trop faible d'individus est couvert à une position donnée, tout en augmentant le potentiel d'identification de DMC. En effet, si l'on décidait de ne prendre en considération que les CpG passant effectivement le seuil de

couverture pour l'intégralité des échantillons, le nombre de CpG analysables chute drastiquement. Pour la plupart des analyses, le *minsamples* utilisé correspond à la moitié de l'effectif du groupe d'individu le plus petit.

A partir des DMC identifiées, un script développé au laboratoire permet de les assembler en régions dans le but d'identifier des régions différentiellement méthylées (*Differentially Methylated Regions*, DMR). En effet, si une région génomique donnée présente une succession de DMC, alors on considère que la région dans son ensemble peut être considérée comme différentiellement méthylée. Comme il n'existe pas de définition consensus des caractéristiques d'une DMR, nous avons considéré dans nos calculs qu'une région peut être définie comme DMR quand elle contient au moins 3 DMC dans une région de moins de 100pb. Dans le script nous avons également permis la possibilité d'étendre les DMR identifiées avec les CpG proches en utilisant un *qvalue* moins stricte.

Annotation sur le génome

La dernière étape clef du traitement bio-informatique est l'étape d'annotation des CpG, DMC ou DMR sur le génome. Cette annotation est faite en fonction du contexte génomique (promoteur-TSS, intron, exon, etc.), de la densité en CpG (CGI, CpG *shore*, CpG *shelve* ou *open sea*) ou de la présence d'éléments répétés (SINE, LINE, LTR, etc.). Cette analyse est faite en comparant la position du CpG ou DMC avec les données de positionnement des éléments génomiques en question. Les coordonnées des éléments génomiques proviennent de fichiers de référence :

- Un fichier Biomart provenant du site Ensembl. Il spécifie les coordonnées des gènes, leur identifiant Ensembl et leur nom (Aken et al. 2016)
- Un fichier GTF (*General Transfer Format*) qui renseigne sur les éléments des gènes comme la position de ses exons, intron, etc.
- Un fichier sur la position des îlots CpG du génome. Ce fichier provient du site *University of California Santa Cruz* (UCSC)
- Un fichier qui renseigne sur la position des éléments répétés et leur nature. Ce fichier provient de la base de données RepeatMasker et téléchargé sur la plateforme de l'UCSC

Lorsque l'on annote il faut définir un seuil de chevauchement pour considérer que la CpG/DMC/DMR et l'élément génomique co-localisent. Pour l'ensemble de nos analyses, nous avons pris la décision de choisir un seuil de chevauchement de 75% pour les CGI/*shore/shelve* et de 0% pour les éléments répétés.

Analyses complémentaires

Ont été incluses dans la procédure d'analyse, des analyses complémentaires. Il s'agit d'analyses descriptives utilisant l'ensemble de l'information de méthylation pour tous les CpG analysés, pour tous les échantillons inclus dans une étude. Elles consistent principalement à effectuer des Classifications hiérarchiques de corrélation de Pearson et des Analyses en Composantes Principales (ACP). Le but de ce type d'analyse est de comparer l'information de méthylation pour les CpG entre échantillons, et d'établir comment cette information permet de structurer des groupes d'individus. L'ACP permet dans un premier temps d'identifier la dimension qui permet de décrire au mieux notre jeu de variables (ici le niveau de méthylation). La seconde dimension fait de même mais est nécessairement orthogonale à la première. Pour chaque dimension, un pourcentage est calculé et nous informe sur la contribution de nos groupes à chaque dimension. La classification hiérarchique de corrélation permet quant à elle de structurer les données en prenant en compte l'ensemble des informations de contribution.

Résultats

Bull spermatozoa exhibit genome-wide DNA hypomethylation with both species peculiarities and conserved tissue-specific features

Article soumis pour publication

Jean-Philippe Perrier¹, Eli Sellem^{1,2}, Audrey Prézelin¹, Maxime Gasselin¹, Luc Jouneau¹, François Piumi^{1,3}, Hala Al Adhami^{1,4}, Michaël Weber⁴, Sébastien Fritz^{2,5}, Didier Boichard⁵, Chrystelle Le Danvic^{2,6}, Laurent Schibler², Hélène Jammes¹ and Hélène Kiefer^{1*}

¹UMR BDR, INRA, ENVA, Université Paris Saclay, 78350, Jouy en Josas, France

²ALLICE, 149 rue de Bercy, 75012, Paris, France

³Present affiliations : Institut Curie, PSL Research University, CNRS, UMR3664, 75005, Paris, France ; Sorbonne Universités, UPMC Univ Paris 06, CNRS, UMR3664, 75005, Paris, France

⁴CNRS, Université de Strasbourg, UMR7242 Biotechnologie et signalisation cellulaire, 300 bd Sébastien Brant, Illkirch cedex 67412, France

⁵UMR GABI, INRA, AgroParisTech, Université Paris Saclay, 78350, Jouy en Josas, France

⁶UMR CNRS/USTL 8576, UGSF, Villeneuve D'Ascq, France

*Corresponding author

Email addresses: jean-philippe.perrier@inra.fr (JPP); eli.sellem@alice.fr (ES); audrey.prezelin@inra.fr (AP); m.gasselin@hotmail.fr (MG); luc.jouneau@inra.fr (LJ); francois.piumi@curie.fr (FP); aladhami@unistra.fr (HAA); michael.weber@unistra.fr (MW); sebastien.fritz@inra.fr (SF); didier.boichard@inra.fr (DB); chrystelle.ledanvic@alice.fr (CLD); laurent.schibler@alice.fr (LS); helene.jammes@inra.fr (HJ); helene.kiefer@inra.fr (HK)

Introduction

Le spermatozoïde est une cellule hautement différenciée, qui requiert une programmation épigénétique issue d'un long processus associé à la gamétogenèse mâle (Carrell et al. 2012). Au cours de la gestation, la reprogrammation débute par l'effacement puis la réapposition de marques de méthylation de l'ADN dans les cellules germinales, un processus qui perdure après la naissance et jusqu'à l'âge adulte. La maintenance de la méthylation de l'ADN, l'accumulation d'ARN non-codants, la mise en place de modifications post-traductionnelles d'histones ou de variants sperme-spécifiques, et finalement le remplacement des histones par les protamines se mettent en place progressivement au cours des étapes de mitoses, méioses, différenciations et maturations au cours de la spermatogenèse.

La réorganisation de l'ensemble des marques épigénétiques apposées pendant la spermatogenèse permet une réorganisation de la chromatine aboutissant à une exceptionnelle compaction du noyau. Cette compaction permet une augmentation de la motilité et une protection de l'ADN. Elle joue aussi un rôle dans le développement embryonnaire. Une altération de la reprogrammation épigénétique dans la lignée germinale mâle peut potentiellement affecter les fonctions du spermatozoïde et l'efficacité de fécondation. De nombreuses études ont montré des associations entre un épigénome spermatique anormal et une concentration spermatique diminuée, une morphologie spermatique anormale, un échec de la fécondation, une mauvaise embryogenèse, des taux de naissance mauvais ou des désordres métaboliques chez les descendants. Des études de cohortes humaines, ou d'altérations génétiques et pharmacologiques de la méthylation de l'ADN chez la souris, ont souligné le rôle primordial de la méthylation de l'ADN dans la fertilité mâle.

Au contraire de cette abondante littérature (revue dans O'Doherty et McGettigan, 2015 ; Jenkins et al. 2017), les études sur la méthylation de l'ADN des spermatozoïdes bovins sont en nombre réduit, et relèvent d'une approche gène-candidat (Guo et al. 2013 ; Jena et al. 2014 ; Yao et al. 2015). Récemment, des études pangénomiques ont identifié des marques de méthylation de l'ADN spermatique associés avec la sub-fertilité chez le taureau ou le buffle, ainsi que des régions hyperméthylées dans le spermatozoïde comparé à l'embryon en utilisant une plateforme dédiée aux petits échantillons. Cependant, une vue générale du méthylome spermatique normal dans l'espèce bovine manque toujours malgré l'utilisation intensive de la semence bovine en insémination artificielle dans l'industrie. L'histoire de la domestication, l'apparition de races hautement spécialisées, et plus récemment des décennies de sélection

animale et d'amélioration génétique, ont façonné le génome bovin. Ceci a sans aucun doute eu un impact sur le méthylome, puisque la méthylation de l'ADN est directement affectée par le contenu en CpG du génome et ses altérations par le polymorphisme génétique. Alors que les mâles sont sélectionnés sur la base de leur haute valeur génétique, plus d'information sur ce que le spermatozoïde transmet à l'embryon en plus de l'héritage génétique paternel devient crucial dans le contexte de la sélection génomique.

Objectifs

Dans le but d'apporter des connaissances à ce sujet, nous avons établi une description du méthylome normal du spermatozoïde bovin en utilisant un large panel de méthodes, allant de la quantification globale aux analyses de la méthylation pangénomiques, et présentant différentes résolutions et couverture génomique, avec une validation finale par analyse gène-candidat. Nous rapportons ici sur la méthylation globale de la semence bovine par rapport aux cellules somatiques et à la semence d'autres espèces mammifères, et sur la comparaison des profils de méthylation pangénomique entre spermatozoïdes bovins et cellules somatiques.

Résultats et conclusions

La quantification globale de la méthylation de l'ADN de la semence de taureau, bélier, cheval, souris, bouc et humaine par LUMA a révélé une hypométhylation majeure de celle du taureau sur les sites CCGG (45,5% contre 57,1% à 76,1%). En revanche, les cellules somatiques présentaient un niveau de méthylation similaire pour les taureaux et les béliers, ce qui démontre que l'hypométhylation est spécifique de la semence bovine et n'est probablement pas due à une distribution biaisée des sites CCGG dans le génome bovin. Cela a également été confirmé par l'annotation des sites CCGG par rapports aux éléments génomiques disponibles pour les différentes espèces. Alors que le processus de cryoconservation n'a pas présenté d'effets sur la méthylation de l'ADN, de subtiles variations entre les races bovines ont été détectées, suggérant des facteurs génétiques sous-jacents dans la détermination des profils de méthylation. Dans le but de cartographier la méthylation de l'ADN à l'échelle pangénomique, nous avons mené une analyse par RRBS et MeDIP-chip. La semence bovine a été comparée avec des hépatocytes, fibroblastes et monocytes. Des séquences différenciellement méthylées ont été identifiées, la plupart hypométhylées dans les spermatozoïdes. En accord avec des résultats précédemment établis pour d'autres espèces, les séquences hypométhylées dans les spermatozoïdes bovins avaient trait au programme de différenciation de la lignée germinale, des fonctions spermatiques ainsi qu'aux séquences satellites.

BMC Genomics

Bull spermatozoa exhibit genome-wide DNA hypomethylation with both species peculiarities and conserved tissue-specific features

--Manuscript Draft--

Manuscript Number:	GICS-D-17-01344R2	
Full Title:	Bull spermatozoa exhibit genome-wide DNA hypomethylation with both species peculiarities and conserved tissue-specific features	
Article Type:	Research article	
Section/Category:	Non-human and non rodent vertebrate genomics	
Funding Information:	Agence Nationale de la Recherche (ANR-13-LAB3-0008-01)	Dr. H�el�ene Jammes
	Agence Nationale de la Recherche (ANR-11-INBS -0003)	Dr. H�el�ene Jammes
Abstract:	<p>Background: Spermatozoa have a remarkable epigenome in line with their degree of specialization, their unique nature and different requirements for successful fertilization. Accordingly, the incorrect establishment of DNA methylation patterns during male germ cell differentiation has been associated with male infertility in several species. While bull semen is widely used in artificial insemination, the literature describing DNA methylation in bovine spermatozoa is still scarce. The purpose of this study was therefore to characterize the bull sperm methylome relative to both bovine somatic cells and the sperm of other mammals through a multiscale analysis.</p> <p>Results: The global quantification of DNA methylation in bulls, rams, stallions, mice, goats and men using a luminometric methylation assay revealed the hypomethylation of bovine sperm at CCGG sites (45.5% vs. 57.1% to 76.1% in other mammals). In contrast, blood cells displayed a similar level of methylation in bulls and rams, demonstrating that bull sperm hypomethylation was tissue-specific and probably not due to a biased distribution of CCGG in the bovine genome. This was also confirmed by annotating CCGG sites relative to the genome features available in different species. While the cryopreservation process was not seen to exert any effect on sperm methylation, subtle breed-dependent variations were detected in cattle, suggesting either underlying genetic factors or the impact of contrasting environments on the determination of methylation patterns. To map DNA methylation at a genome-wide scale, we carried out simultaneous, reduced representation bisulfite sequencing and immunoprecipitation of methylated DNA followed by hybridization on a microarray. Bull sperm was compared with bovine liver, fibroblasts and monocytes, and differentially methylated sequences were identified and most were hypomethylated in sperm. Consistent with previous studies in other species, the sequences that were specifically hypomethylated in bull sperm were enriched for genes relevant to the germline differentiation program and sperm functions, as well as for satellite repeats.</p> <p>Conclusions: These results highlight the hypomethylation of bull spermatozoa when compared with both bovine somatic cells and the sperm of other mammals, and raise questions regarding its biological significance. Whether sperm-specific hypomethylation has potential interactions with structural variation in the cattle genome may deserve further attention.</p>	
Corresponding Author:	H�el�ene Kiefer INRA Centre de Jouy-en-Josas FRANCE	
Corresponding Author Secondary Information:		
Corresponding Author's Institution:	INRA Centre de Jouy-en-Josas	
Corresponding Author's Secondary Institution:		
First Author:	Jean-Philippe Perrier	
First Author Secondary Information:		

Order of Authors:	Jean-Philippe Perrier
	Eli Sellem
	Audrey Prézelin
	Maxime Gasselín
	Luc Jouneau
	François Piumi
	Hala Al Adhami
	Michael Weber
	Sébastien Fritz
	Didier Boichard
	Chrystelle Le Danvic
	Laurent Schibler
	Hélène Jammes
	Hélène Kiefer
Order of Authors Secondary Information:	
Response to Reviewers:	<p>All asked changes were applied</p> <ul style="list-style-type: none"> - We added a "Ethics approval paragraph in methods section - We confirm the list of authors and insure that the qualification for authorship is in good agreement with biomedcentral guidelines - we moved our "additional files section after the refernecs /figures legends

[Click here to view linked References](#)

1 **Bull spermatozoa exhibit genome-wide DNA hypomethylation with both**
2 **species peculiarities and conserved tissue-specific features**

3
4
5
6
7
8 4 Jean-Philippe Perrier¹, Eli Sellem^{1,2}, Audrey Prézélin¹, Maxime Gasselin¹, Luc Jouneau¹, François
9
10 5 Piumi^{1,3}, Hala Al Adhami^{1,4}, Michaël Weber⁴, Sébastien Fritz^{2,5}, Didier Boichard⁵, Chrystelle Le
11
12 6 Danvic^{2,6}, Laurent Schibler², Hélène Jammes¹ and Hélène Kiefer^{1*}
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1¹UMR BDR, INRA, ENVA, Université Paris Saclay, 78350, Jouy en Josas, France

2²ALLICE, 149 rue de Bercy, 75012, Paris, France

3³Present affiliations: Institut Curie, PSL Research University, CNRS, UMR3664, 75005, Paris, France;

Sorbonne Universités, UPMC Univ Paris 06, CNRS, UMR3664, 75005, Paris, France

4⁴CNRS, Université de Strasbourg, UMR7242 Biotechnologie et signalisation cellulaire, 300 bd

Sébastien Brant, Illkirch cedex 67412, France

5⁵UMR GABI, INRA, AgroParisTech, Université Paris Saclay, 78350, Jouy en Josas, France

6⁶UMR CNRS/USTL 8576, UGSF, Villeneuve D'Ascq, France

*Corresponding author

Email addresses: jean-philippe.perrier@inra.fr (JPP); eli.sellem@alice.fr (ES); audrey.prezelin@inra.fr

(AP); m.gasselin@hotmail.fr (MG); luc.jouneau@inra.fr (LJ); francois.piumi@curie.fr (FP);

aladhami@unistra.fr (HAA); michael.weber@unistra.fr (MW); sebastien.fritz@inra.fr (SF);

didier.boichard@inra.fr (DB); chrystelle.ledanvic@alice.fr (CLD); laurent.schibler@alice.fr (LS);

helene.jammes@inra.fr (HJ); helene.kiefer@inra.fr (HK)

23 **Abstract**

24

25 **Background:** Spermatozoa have a remarkable epigenome in line with their degree of
26 specialization, their unique nature and different requirements for successful fertilization.

27 Accordingly, the incorrect establishment of DNA methylation patterns during male germ cell
28 differentiation has been associated with male infertility in several species. While bull semen is widely
29 used in artificial insemination, the literature describing DNA methylation in bovine spermatozoa is
30 still scarce. The purpose of this study was therefore to characterize the bull sperm methylome
31 relative to both bovine somatic cells and the sperm of other mammals through a multiscale analysis.

32 **Results:** The global quantification of DNA methylation in bulls, rams, stallions, mice, goats
33 and men using a luminometric methylation assay revealed the hypomethylation of bovine sperm at
34 CCGG sites (45.5% vs. 57.1% to 76.1% in other mammals). In contrast, blood cells displayed a similar
35 level of methylation in bulls and rams, demonstrating that bull sperm hypomethylation was tissue-
36 specific and probably not due to a biased distribution of CCGG in the bovine genome. This was also
37 confirmed by annotating CCGG sites relative to the genome features available in different species.

38 While the cryopreservation process was not seen to exert any effect on sperm methylation, subtle
39 breed-dependent variations were detected in cattle, suggesting either underlying genetic factors or
40 the impact of contrasting environments on the determination of methylation patterns. To map DNA
41 methylation at a genome-wide scale, we carried out simultaneous, reduced representation bisulfite
42 sequencing and immunoprecipitation of methylated DNA followed by hybridization on a microarray.
43 Bull sperm was compared with bovine liver, fibroblasts and monocytes, and differentially methylated
44 sequences were identified and most were hypomethylated in sperm. Consistent with previous
45 studies in other species, the sequences that were specifically hypomethylated in bull sperm were
46 enriched for genes relevant to the germline differentiation program and sperm functions, as well as
47 for satellite repeats.

48 **Conclusions:** These results highlight the hypomethylation of bull spermatozoa when
49 compared with both bovine somatic cells and the sperm of other mammals, and raise questions
50 regarding its biological significance. Whether sperm-specific hypomethylation has potential
51 interactions with structural variation in the cattle genome may deserve further attention.

52
53 **Key words:** DNA methylation, sperm, cattle, satellite repeats

54 55 56 **Background**

57
58 The unique structural features of spermatozoa require dynamic epigenetic reprogramming of
59 the genome [1], which starts with the global erasure and reestablishment of DNA methylation marks
60 in fetal and post-natal germ cells and continues throughout adulthood. The maintenance of DNA
61 methylation, the accumulation of non-coding RNAs, the implementation of post-translational histone
62 modifications or sperm-specific variants and finally histone-to-protamine replacement then occur
63 progressively to regulate the sequential mitosis, meiosis, differentiation and maturation steps of
64 spermatogenesis. The profound reorganization of epigenetic marks during spermatogenesis enables
65 a dramatic compaction of the sperm nucleus, thus improving motility and DNA damage protection in
66 the female genital tract, and plays a fundamental role in subsequent development of the embryo [2].
67 Any alterations to the epigenetic reprogramming of the male germline may potentially affect sperm
68 functions and fertilization efficiency [3], and numerous studies have reported associations between
69 an abnormal sperm epigenome and a low sperm count or sperm dysmorphia, fertilization failures,
70 poor embryogenesis, low pregnancy outcomes and metabolic disorders affecting the offspring [4-14].
71 Accordingly, studies in human cohorts [9, 15, 16] and genetic or pharmacologic alterations to DNA
72 methylation in mice [17-19] have emphasized the prominent role of DNA methylation in male germ
73 cell differentiation and male fertility.

74 Unlike the abundant literature referred to above, studies on DNA methylation in bovine
1
2 75 spermatozoa are still scarce, and have often focused on candidate loci [20-22]. Recent genome-wide
3
4 76 studies have identified sperm DNA methylation marks associated with subfertility in buffalo and bulls
5
6
7 77 [23, 24], as well as regions that are hypermethylated in sperm relative to the embryo, studied using a
8
9 78 platform dedicated to small samples [25]. However, a comprehensive view of the normal sperm
10
11 79 methylome in bovine species is still lacking, even though this knowledge could enable promising
12
13
14 80 advances in the cattle industry. Indeed, domestication, the creation of highly specialized breeds and
15
16 81 decades of genetic improvement have shaped the bovine genome [26]. This undoubtedly has also
17
18 82 had a profound impact on the methylome, since DNA methylation is directly affected by the CpG
19
20
21 83 content of the genome and its alteration by DNA polymorphism [27]. Whether these changes are of
22
23 84 functional significance and contribute to the establishment of phenotypes needs to be ascertained.
24
25
26 85 In addition, in a context of genomic selection, more information on the epigenetic features
27
28 86 transferred to the embryo alongside the paternal genetic heritage is necessary in order to improve
29
30 87 semen quality control procedures as well as to guarantee semen fertility and proper embryo
31
32
33 88 development.

34
35 89 In order to contribute knowledge in this field, we established a thorough description of the
36
37 90 normal methylome of bull spermatozoa using a variety of methods, ranging from global methylation
38
39
40 91 quantification to genome-wide methylation assays achieving different resolutions and genome
41
42 92 coverage, with a final validation by candidate region analyses. We report here on the global
43
44 93 methylation of bull sperm relative to both bovine somatic cells and the sperm of other mammals,
45
46
47 94 and on a comparison of genome-wide methylation patterns between bovine sperm and somatic cells.
48
49 95 Interestingly, we found that bull sperm was largely hypomethylated at all the scales investigated.
50
51
52 96 This hypomethylation targeted genome features conserved with other species, but also revealed
53
54 97 some species peculiarities which could have important evolutionary implications.
55
56

57 98
58
59 99
60
61
62
63
64
65

100 **Methods**

101

102 **Animals and cell/tissue collection**

103 The bovine semen samples used in this study originated from bulls selected for artificial
104 insemination. In Figure 1A-B, semen and blood from the same individuals were collected using
105 standard procedures on bulls and rams maintained in semen production centers (Montbéliarde
106 breed, GEN'IATEST, France; Ile-de-France breed, OSON, France). Total blood was used for DNA
107 extraction. For both bulls and rams, collected semen was extended with Optidyl (Cryo-Vet) and either
108 underwent direct DNA extraction (fresh semen) or was subjected to standard techniques for semen
109 processing (straw conditioning, freezing and storage in liquid nitrogen; frozen semen). Other bull
110 semen samples were obtained from selection companies in the form of frozen straws stored in liquid
111 nitrogen. The straws came from the Holstein and Normande breeds (n=112 and n=18, respectively),
112 Montbéliarde breed (n=41) and Belgian White Blue breed (n=14) and were supplied by Evolution
113 (France), Umotest (France) and Awe (Belgium), respectively. Fresh stallion semen (n=2), fresh boar
114 semen (n=3), frozen human semen (n=7) and frozen goat semen (n=2) were supplied by Michèle
115 Magistrini (INRA, France), LNCR (France), Rachel Lévy (APHP, France) and Capgenes (France),
116 respectively. Fresh mouse semen was collected from the caudal epididymis after euthanasia by
117 cervical dislocation. For all samples except those from humans and stallions, the possible
118 contamination of spermatozoa by somatic cells was checked systematically under the microscope
119 and confirmed to be below detectable levels. The human and stallion semen samples contained
120 observable somatic cells and were therefore processed as previously described to ensure the
121 absence of potential contamination: the human semen samples were subjected to a stringent
122 somatic cell lysis protocol [15] and the sperm from stallions were purified by single layer
123 centrifugation using Androcoll-E-Large (SLU, Uppsala, Sweden) [28].

124 Other tissues were collected from Holstein cows maintained at the INRA experimental farm
125 (UCEA, INRA, France). Primary cultures of fibroblasts were derived from ear skin biopsies from three

126 separate adult heifers and cultured until passage 11 in Dulbecco's modified Eagle medium
127 supplemented with 10% fetal calf serum and 1% penicillin-streptavidin (Life Technologies) at 38°C
128 with 5% CO₂. Livers were obtained from adult cows slaughtered at the INRA experimental facilities.
129 Peripheral blood mononuclear cells (PBMCs) were isolated from blood collected from the jugular vein
130 and centrifuged using a Ficoll gradient. To obtain the monocyte fraction, PBMCs were incubated in
131 the presence of microbeads conjugated to monoclonal anti-CD14 antibodies (mouse IgG2a; Miltenyi
132 Biotec) in MACS BSA buffer, for 15 min. at 4°C under gentle agitation. Magnetic separation was then
133 performed using MS Columns following the manufacturer's instructions. The tissues and cells were
134 snap-frozen in liquid nitrogen and stored at -80°C until DNA extraction.

Ethics approval: all study methods were implemented in accordance with EU guidelines and
136 regulations (Directive 2010/63/UE). For animals maintained in INRA facilities, the experimental
137 protocols were approved by the INRA local Ethics Committee (COMETHEA, authorization numbers
138 12/160 and Méjusseaumes Animal Care committee 0162503). Mice semen samples were collected
139 from 7-week-old male C57Bl/6JOLAHsd mice supplied by Harlan Laboratory (Venray, Netherlands)
140 and euthanized by cervical dislocation. The bull samples were purchased by commercial companies:
141 Montbéliarde breed by GEN'IATEST (Roulans, France) and UMOTEST (Ceyzeriat, France), Holstein and
142 Normande breeds by EVOLUTION (Rennes, France) and Belgian White Blue Breed by AWE (Ciney,
143 Belgium); the ram, goat and boar semen were also supplied by commercial companies, ILE-DE
144 FRANCE breed (Oson, France), CAPGENES (Mignaloux-Beauvoir, France) and Laboratoire National de
145 Contrôle des Reproducteurs (LNCR, Maison Alfort, France), respectively. Stallion semen was supplied
146 by Dr. M. Magistrini (UMR INRA 085 PRC, Nouzilly, France).
147 Human sperm samples from patients included in a PHRC METASPERME study, coordinated by Dr. R.
148 Levy (Laboratoire d'Histologie Embryologie Cytogénétique CECOS, Hôpital Jean Verdier, Bondy,
149 France); this study received the approval from French ethics board (conseil d'évaluation éthique pour
150 les recherches en santé, CERES) and all the patients gave their informed written consent to
151 participate.

152

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

153 Genomic DNA extraction and genotyping

154 One straw of bull semen was used for DNA extraction (about 20 million spermatozoa). After
155 thawing at 37°C, the semen was washed with phosphate buffer saline (PBS) to remove the extender,
156 and incubated overnight at 55°C in 200 µl lysis buffer (10 mM Tris-HCl pH 7.5, 25 mM EDTA, 1% SDS,
157 75 mM NaCl, 50 mM dithiothreitol (DTT) and 0.5 µg glycogen) in the presence of 0.2 mg/ml
158 proteinase K. After incubation with 25 µg/ml RNase A for 1h at 37°C, genomic DNA was extracted
159 twice using phenol and phenol:chloroform (1:1), then ethanol precipitated and washed. The dried
160 pellet was re-suspended in TE buffer (10 mM Tris HCl pH7.5, 2 mM EDTA) and the DNA concentration
161 was measured using a Qubit 2.0 Fluorometer (Invitrogen). Fresh bull semen, total blood from bulls
162 and rams, fresh or frozen semen from rams, boars, mice, goats, processed semen from stallions and
163 men were treated in an identical manner as the frozen straws from bull semen.

164 DNA extraction from liver samples was performed as described elsewhere [29]. The same
165 procedure was used for fibroblasts and monocytes, except that the cells were lysed by the direct
166 addition of lysis buffer and proteinase K to the cell pellet.

167 The genotyping of six Montbéliarde bulls was performed by LABOGENA (France) on semen
168 and blood DNA from the same individuals using the commercially available BovineSNP50 v2
169 BeadChip (Illumina). Genotypes were determined using the Genotyping Module of GenomeStudio
170 software (Illumina). For each animal, copy number variations (CNVs) were searched for and
171 compared between tissues. The Log R Ratio (LRR, normalized measurement of total signal intensity)
172 and B Allele Frequency (BAF, measurement of the allelic intensity ratio) were used to infer copy
173 number changes in the genome. For example, in the presence of a deletion, LRR values increase and
174 BAF values cluster around 0 or 1 but are absent at around 0.5, due to a lack of heterozygotes. LRR
175 and BAF were then plotted along the genome and compared between tissues.

177 *In silico* analyses

178 A homemade script was used to extract the coordinates of all the CCGG sites present in the
1
2 179 genomes of different species (cattle, sheep, horse, pig, mouse and human; Figure 1D). The CCGG
3
4 180 sites were then annotated relative to different gene features, CpG density and repeats by means of a
5
6
7 181 homemade pipeline ([https://github.com/FAANG/faang-methylation/tree/master/RRBS-
10
11 183 toolkit/Annotation](https://github.com/FAANG/faang-methylation/tree/master/RRBS-
8
9 182 toolkit/Annotation)) and using the genome annotation features indicated in Supplementary Table S1.
12
13 184 The following criteria were applied: promoter-TSS, -1000 to +100 bp relative to the transcription start
14 184 site (TSS); TTS: -100 to +1000 bp relative to the transcription termination site (TTS); shore, up to 2000
15
16 185 bp from a CpG island (CGI); and shelve up to 2000 bp from a shore. A site/fragment was considered
17
18 186 to belong to a CGI (respective shore and shelve) if an overlap of at least 75% was observed between
19
20
21 187 the site/fragment and the CGI (respective shore and shelve). A site/fragment was considered as
22
23 188 being overlapped by a repetitive element whatever the extent of this overlapping.

26 189 In Supplementary Tables S2 and S3, *in silico* reduced representation (RR) genomes digested
27
28 190 by MspI restriction enzyme were produced for different species and using different size selection
29
30 191 criteria by means of another homemade pipeline ([https://github.com/FAANG/faang-
33 192 methylation/tree/master/RRBS-toolkit/RR_genome](https://github.com/FAANG/faang-
31
32 192 methylation/tree/master/RRBS-toolkit/RR_genome)). The RR genome fragments were then
34
35 193 annotated as explained above.

38 194

40 195 **Luminometric methylation assay (LUMA)**

42 196 Global DNA methylation levels were quantified using LUMA, as previously described [30, 31].
43
44
45 197 Briefly, 1 µg of genomic DNA was cleaved using the isochizomeres HpaII (methylation sensitive) and
46
47 198 MspI (non-methylation-sensitive) in two separate reactions and in the presence of EcoRI to
48
49
50 199 standardize for DNA amounts. The three enzymes were purchased from New England Biolabs. The
51
52 200 protruding ends were then used as templates for pyrosequencing with the Pyromark Q24 device and
53
54 201 Pyromark Gold Q96 reagents (Qiagen). The luminometric signals produced by either the sequential
55
56
57 202 incorporation of C and G nucleotides (reflecting the number of CCGG sites digested by HpaII or MspI)
58
59 203 or the sequential incorporation of A and T nucleotides (reflecting the number of AATT sites digested

204 by EcoRI), were then quantified using Pyromark Q24 software. Each sample was assayed in duplicate.

205 The global methylation percentage per sample was then calculated as follows:

$$206 \quad \text{Methylation \%} = 100 - \frac{\text{Average signal obtained with HpaII after EcoRI normalization}}{\text{Average signal obtained with MspI after EcoRI normalization}} * 100$$

207 The conditions were compared using non-parametric tests suited to small samples
208 (permutation tests for two independent samples or for two paired samples according to the
209 situation, with Monte-Carlo sampling of 100,000 permutations).

211 **Methylated DNA immunoprecipitation (MeDIP), microarray hybridization and data** 212 **analysis**

213 MeDIP and quality controls by PCR were performed as described elsewhere [29], on four liver
214 samples, three fibroblast samples and four sperm samples (Holstein). To prevent any technical bias,
215 the products of five independent MeDIP experiments were pooled for each sample. After moderate
216 genome amplification, the pooled MeDIP reactions and corresponding input DNA were labelled with
217 Cy3 and Cy5 and hybridized on a Roche-NimbleGen 3x720K microarray, with technical dye-swaps for
218 every sample. The microarray targeted the promoter region (-2000 to +1360 bp relative to the gene
219 start) of 21,296 bovine genes, according to an annotation file downloaded from the Johns Hopkins
220 University Center for Computational Biology FTP website
221 (ftp://ftp.ccb.jhu.edu/pub/data/assembly/Bos_taurus/Bos_taurus_UMD_3.0/annotation/); accessed
222 Aug. 2010). The microarray design and hybridization protocol, as well as more details on the data
223 analysis, can be found in [32].

224 Probes with signal enrichment in the MeDIP sample (“enriched probes”) were identified
225 using the ChIPmix R package [33] and could be visualized using Integrative Genomics Viewer software
226 (IGV) [34]. For Figure 2, a normalized enrichment factor NEpi was calculated for each promoter p and
227 each individual i as follows:

$$228 \quad \text{NEpi} = \text{Epi} \frac{G}{G_i}, \text{ where:}$$

229 Epi =Number of enriched probes for promoter p and sample i;

230 $G_i = \sum_{p=1}^{21296} E_{pi}$ = Total number of enriched probes at promoters for sample i;

231 $G = \frac{1}{11} \sum_{i=1}^{11} \sum_{p=1}^{21296} E_{pi}$ =Mean G_i for all samples.

232 Principal component analysis (PCA) and hierarchical clustering were then computed on the NEpi
233 matrix using the FactoMineR R package. For hierarchical clustering, the distance between samples
234 was calculated using Pearson correlation coefficients and Ward's method was applied as the linkage
235 function.

236 To identify regions of interest containing clusters of probes enriched in at least one tissue,
237 each promoter was screened for an anchor probe displaying enrichment in at least two samples from
238 the same tissue. Starting from this anchor, the region was then extended to upstream and
239 downstream probes until a probe displaying less than 10% enrichment in any tissue was
240 encountered. Two consecutive regions of interest separated by fewer than three probes were
241 coalesced into one larger region of interest. After coalescence, regions containing only the anchor
242 probe were eliminated. Indeed, the signal from a single probe could result from cross-hybridization
243 with another genomic region, whereas this probability might be reduced if several probes were
244 locally enriched. This strategy led to the identification of 27,684 regions of interest.

245 Among these regions of interest, those displaying differential methylation between tissues
246 were identified using an R package designed to analyze point patterns [35] which was therefore
247 appropriate for binary data (enriched vs. not enriched). For each pair of tissues, two models were
248 built and estimated using the ppm function of the spatstat R package (ANOVA for fitted point process
249 models): the full model took account of the origin of the tissue while the alternative model did not.

250 To identify differentially methylated regions (DMRs), the full model was compared with the
251 alternative model. All 27,684 regions of interest were tested and the resulting p-values were
252 corrected for multiple testing using the Benjamini-Hochberg procedure [36] with the multitest R
253 package. A given region was considered to be a DMR if the full model fitted the observations
254 significantly better than the alternative model (adjusted p-value <0.05).

255 The mean percentage of enriched probes Pr in each tissue was calculated as follows for each
1
2 256 region r (DMR or region of interest):
3
4
5 257 $Pr = \frac{1}{n} \sum_{i=1}^n \frac{Eri}{Tr} \times 100$, where n is the total number of biological replicates in the tissue considered
6
7 258 ($n=4$ for liver and sperm, $n=3$ for fibroblasts), Eri is the number of enriched probes for region r and
8
9
10 259 biological replicate i , and Tr is the number of probes included in region r . Three Pr were then
11
12 260 obtained per region: Pr_{sperm} , Pr_{liver} and $Pr_{fibroblasts}$. The scatterplot shown in Figure 3D illustrates the
13
14 261 $Pr_{liver} - Pr_{sperm}$ and $Pr_{fibroblasts} - Pr_{sperm}$ differences for the regions of interest (in black) and for the DMRs
15
16 262 specific to the comparison between sperm and somatic cells (in red). Positive values for both $Pr_{liver} -$
17
18
19 263 Pr_{sperm} and $Pr_{fibroblasts} - Pr_{sperm}$ indicated that the two somatic cell types were more methylated than
20
21 264 sperm in the region considered, while negative values for both $Pr_{liver} - Pr_{sperm}$ and $Pr_{fibroblasts} - Pr_{sperm}$
22
23
24 265 indicated the contrary. Similarly, an opposite sign for the values of $Pr_{liver} - Pr_{sperm}$ and $Pr_{fibroblasts} - Pr_{sperm}$
25
26 266 reflected the contrasting behaviors of liver and fibroblasts. Among the DMRs, those hypomethylated
27
28
29 267 in sperm were selected based on a positive value for both $Pr_{liver} - Pr_{sperm}$ and $Pr_{fibroblasts} - Pr_{sperm}$ and
30
31 268 were annotated as explained regarding the *in silico* analyses, together with the 27,684 regions of
32
33 269 interest. Genes containing DMRs were subjected to DAVID analysis (Database for Annotation,
34
35
36 270 Visualization and Integrated Discovery; [37]) using genes containing the 27,684 regions of interest as
37
38 271 the background.

40 272

43 273 **Reduced representation bisulfite sequencing (RRBS) and data analysis**

45 274 RRBS libraries were prepared as described elsewhere [38, 39] from two Holstein sperm
46
47
48 275 samples, two monocyte samples and two fibroblast samples. Briefly, 200 ng of genomic DNA were
49
50 276 digested by *MspI* (Thermo Scientific), end-repaired and ligated to 55 bp Illumina adapters for paired-
51
52 277 end sequencing. Size selection by gel excision was performed in order to select fragments ranging
53
54
55 278 from 150 to 400 bp (genomic fragments of 40-290 bp + adapters). The DNA was then purified using
56
57 279 the MinElute gel extraction kit (Qiagen) and then bisulfite-converted twice consecutively with the
58
59 280 EpiTect bisulfite kit (Qiagen), following the manufacturer's instructions for DNA extracted from FFPE

281 tissues. Converted DNA was amplified with Pfu Turbo Cx hotstart DNA polymerase (Agilent) using 14
1
2 282 PCR cycles for sperm and fibroblasts and 12 cycles for monocytes. The libraries were then purified
3
4 283 using Agencourt Ampure beads (Beckman-Coulter) and sequenced on an Illumina HiSeq2500
5
6
7 284 sequencer to produce 75 bp paired-end reads (Integrage SA, France).

8
9 285 RRBS sequences were analyzed using an integrated pipeline combining homemade scripts
10
11 286 developed in Python, R and Shell, together with external tools ([https://github.com/FAANG/faang-](https://github.com/FAANG/faang-methylation/tree/master/RRBS-toolkit/)
12
13 287 [methylation/tree/master/RRBS-toolkit/](https://github.com/FAANG/faang-methylation/tree/master/RRBS-toolkit/)). The bisulfite conversion rate was estimated from the
14
15
16 288 unmethylated cytosine added artificially during the end-repair step. The sequences displayed the
17
18 289 expected nucleotide composition based on MspI digestion and bisulfite conversion according to
19
20
21 290 FastQC quality control (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Subsequent
22
23 291 quality checks and trimming were carried out using Trim Galore v0.3.7
24
25
26 292 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) which removed adapter
27
28 293 sequences, poor quality bases and reads (Phred score below 20) and reads shorter than 20
29
30 294 nucleotides. High quality reads were aligned on the bovine reference genome (UMD 3.1 assembly)
31
32
33 295 using Bismark v0.14.3 in the default mode with Bowtie 1 [40, 41].

34
35 296 Only CpGs covered by 5 to 500 uniquely mapped reads for each sample (termed as CpGs 5-
36
37 297 500) were retained for subsequent analyses. Each CpG 5-500 was assigned a methylation percentage
38
39 298 per sample calculated from Bismark methylation calling
40
41
42 299 $((\text{reads with "C"} \times 100) / (\text{reads with "C"} + \text{reads with "T"}))$, which could be visualized using
43
44 300 IGV. For Figure 2, PCA and hierarchical clustering were then computed on the matrix of methylation
45
46 301 percentages for each CpG 5-500 and each sample as explained for the MeDIP data.

47
48
49 302 For each pair of tissues, differentially methylated CpGs (DMCs) were identified using
50
51 303 methylKit [42]. A CpG 5-500 was considered as a DMC when the associated q-value was weaker than
52
53 304 0.001 and the methylation difference between the two tissues was at least 25% (according to the
54
55 305 methylKit calculation mode, which takes account of the coverage per sample; see Additional Data
56
57 306 File 3, column K). In parallel with this differential analysis, some CpGs 5-500 were selected based on a
58
59
60
61
62
63
64
65

307 stringent methylation threshold criterion (methylation percentage weaker than 20% for the two
1
2 308 biological replicates in the first cell type and higher than 80% for the two biological replicates in the
3
4 309 second cell type) and added to the DMC list in case they had not previously been detected by
5
6
7 310 methylkit (“obvious DMCs”).
8

9 311 For each tissue, the mean methylation percentage was calculated (mean of the methylation
10
11 312 percentages obtained in the two biological replicates) as well as the difference between two tissues
12
13 313 (Additional Data File 3, column L). The scatterplot shown in Figure 3E illustrates the differences
14
15
16 314 between monocytes and sperm and between fibroblasts and sperm for the 1,580,644 CpGs 5-500
17
18 315 covered in all six samples (in black) and for the DMCs specific to the comparison between sperm and
19
20
21 316 somatic cells (in red). Among these DMCs, those hypomethylated in sperm were selected based on a
22
23 317 positive value for both differences. Together with the CpGs 5-500, they were then annotated relative
24
25
26 318 to gene features, CGIs and repeats as explained for the *in silico* analyses. Genes containing DMCs
27
28 319 were subjected to DAVID analysis using genes containing the 1,580,644 CpGs 5-500 as the
29
30
31 320 background.
32

33 321 To better characterize repetitive elements, an artificial genome containing the consensus
34
35 322 sequence of each bovine repeat was constituted from the Replibase database [43]. Reads were
36
37
38 323 aligned on this artificial genome as explained above, and the average methylation percentage was
39
40 324 calculated for each repeat and each sample (average methylation percentage for all CpGs included in
41
42 325 one genomic repeat and covered by either 5-500 reads or by >500 reads).
43
44
45 326

47 327 **Pyrosequencing**

48
49 328 Bisulfite conversion was performed on 1 µg genomic DNA as described elsewhere [44]. After
50
51
52 329 ethanol precipitation, the DNA pellet was re-suspended in 20 µl H₂O.
53

54 330 For *LSM4* and *BTSAT4*, primers were designed using the MethPrimer program [45] and
55
56
57 331 amplifications were carried out from 1 µl treated DNA with Platinum Taq DNA polymerase
58
59 332 (Invitrogen), according to the manufacturer’s instructions with variable MgCl₂ concentrations. The
60
61
62
63
64
65

333 following program was used: 3 min. at 94°C followed by 50 cycles of 30 sec. at 94°C, 1 min. at
1
2 334 variable hybridization temperatures, 1 min. at 72°C, and finally 10 min. at 72°C. For *DDX4* and *SYCP3*,
3
4 335 primers were designed using the Pyromark assay design software (Qiagen) and amplifications were
5
6
7 336 performed using the Pyromark PCR kit (Qiagen) according to the manufacturer's instructions. The
8
9 337 primers used to amplify each region are listed in Supplementary Table S4, together with the
10
11 338 hybridization temperatures and MgCl₂ concentrations, and the reverse primers were 5'-biotinylated.

13
14 339 After denaturation and purification, the biotinylated antisense strand of PCR product was
15
16 340 used as a template for pyrosequencing with 0.3 μM pyrosequencing primer, using the Pyromark Q24
17
18 341 device and Pyromark Gold Q96 reagents (Qiagen). The pyrosequencing primers are listed in
19
20 342 Supplementary Table S5. Each CpG was assayed in duplicate, and inconsistent duplicates (more than
21
22 343 5% difference) were repeated. The methylation percentage per CpG was then obtained by calculating
23
24 344 the mean of all replicates that passed quality control by the Pyromark Q24 software. The statistical
25
26 345 analysis was performed as explained for LUMA on the mean percentage per CpG.

33 347 **Results**

39 349 **Global DNA methylation is low in bovine sperm**

41 350 We first assessed the global level of methylation in bovine sperm relative to somatic cells
42
43 351 (PBMCs) using LUMA [30, 31] on a large sample size. The average methylation at CCGG sites was
44
45 352 dramatically lower in sperm from different breeds (45.5%, n=185) than in Holstein PBMCs (74.9%,
46
47 353 n=74; data not shown). To determine whether the global hypomethylation of sperm was specific to
48
49 354 the bovine species, we collected paired semen and blood samples from bulls and rams. Differences
50
51 355 between species were observed in both cell types, and sperm was hypomethylated compared to
52
53 356 blood in both species (Figure 1A). However the degree of sperm hypomethylation was much greater
54
55 357 in bulls than in rams, and was independent of the cryopreservation process (Figure 1B). Because the
56
57 358 presence of DNA polymorphism could influence the global CCGG content and methylation, we then
58
59
60
61
62
63
64
65

359 compared sperm methylation in four bovine breeds. Methylation was significantly lower in Belgian
1
2 360 White Blue than in any other breed and significantly lower in Holstein than in Normande,
3
4 361 Montbéliarde displaying an intermediate status (Figure 1C). A bootstrap analysis confirmed that
5
6
7 362 these breed-related differences were not due to an unbalanced number of bulls from each breed
8
9 363 (data not shown). These results demonstrated that global methylation varied in different bovine
10
11 364 breeds. However, the range of variation (from 42.6% in Belgian White Blue to 47.1% in Normande)
12
13
14 365 was weak relative to the 30% difference we observed between bovine sperm and blood, suggesting
15
16 366 that sperm hypomethylation was not breed-dependent.

17
18
19 367 This sperm-specific hypomethylation may have resulted from a biased representation of the
20
21 368 regions present in our sperm genomic DNA. Indeed, the high level of sperm chromatin compaction
22
23 369 impeded the complete extraction of genomic DNA using standard procedures, which were therefore
24
25
26 370 optimized by the addition of reducing agents such as DTT (see for instance [46]). The protocol we
27
28 371 used during this study performed well for DNA extraction from human sperm (50 mM DTT, [4]). To
29
30
31 372 investigate whether these conditions were also suited to bovine sperm, we used genomic DNA from
32
33 373 six paired sperm and blood samples (the same samples as in Figure 1A), both of which were purified
34
35 374 in the presence of 50 mM DTT as a template for genotyping. As expected, the same genotype was
36
37
38 375 obtained from blood and sperm DNA in each bull. Likewise, for each animal, CNV profiles were
39
40 376 compared between tissues because any difference might be indicative of preferential extraction. The
41
42 377 plots were similar and no gross discrepancies could be observed between tissues (data not shown).
43
44
45 378 This result therefore ruled out the possibility that some specific regions of the bovine genome failed
46
47 379 to be extracted from sperm chromatin under our experimental conditions.

48
49 380 Because the genomic distribution of CCGG sites may display species-specific variations that
50
51
52 381 might influence the methylation results, we compared sperm CCGG methylation in several species
53
54 382 relative to the genomic features available. In the horse, pig, mouse, human and goat, sperm
55
56
57 383 methylation was much higher than that measured in bovine and closer to that determined in sheep
58
59 384 (Figure 1D). The CCGG distribution in the bovine genome was within the range observed in other
60
61
62
63
64
65

385 species for most of the features examined, except for CGIs that accounted for 14.2% of total CCGG
1
2 386 sites in bovine vs. 8.0% to 12.2% in other species. The important representation of bovine CCGG in
3
4 387 CGIs was balanced by a low representation in open sea (regions with a low CpG density). Most
5
6
7 388 mammalian CGIs are unmethylated in somatic cells and can become methylated during development
8
9 389 and disease [47]. Because blood cells (in which most CGIs are supposed to be unmethylated)
10
11 390 displayed a roughly similar methylation level in bovine and sheep, it is unlikely that the massive
12
13 391 hypomethylation of bull sperm resulted solely from the higher percentage of bovine CCGG sites in
14
15 392 CGIs. Another remarkable feature of the bovine CCGG sites was their strong enrichment in satellites,
16
17 393 which represented 1.6% of all CCGGs vs. 0.1% to 0.5% in other species. However, whatever the
18
19 394 species examined, the <2% difference in CCGG sites present in satellites could not account for the
20
21 395 >10% difference in sperm methylation (but see Discussion).

22
23
24
25
26 396 Taken together, these data demonstrated that compared to somatic cells, bull sperm
27
28 397 displayed dramatic CCGG hypomethylation which was specific to the bovine species. This
29
30 398 hypomethylation was neither related to the process of semen cryopreservation nor to a technical
31
32 399 artefact, and could not be fully explained by the genomic distribution of CCGG sites in the bovine
33
34 400 species.

401

402 **MeDIP-chip and RRBS are complementary approaches for whole-genome DNA methylation**

403 **analyses in cattle**

404 Because high-throughput analyses were necessary to identify regions that were
405 hypomethylated in bovine sperm, we decided to assess two cost-effective approaches which are
406 widely used to study DNA methylation: MeDIP-chip [48] and RRBS [38]. MeDIP-chip enables precise
407 targeting of specific regions in the genome through custom design of the microarray (in our case,
408 3360 bp spanning the promoter and upstream region of each of the 21,296 bovine genes [32]), while
409 RRBS offers a base-resolution analysis of CpG-rich regions through the combined use of enzymatic
410 digestion (MspI) and the size selection of restriction fragments.

1
2 412
3
4 413
5
6
7 414
8
9 415
10
11 416
12
13
14 417
15
16 418
17
18
19 419
20
21 420
22
23 421
24
25
26 422
27
28 423
29
30
31 424
32
33 425
34
35 426
36
37
38 427
39
40 428
41
42 429
43
44

Because no data were available regarding the optimal size window for RRBS in cattle, we conducted an *in silico* prediction of the genome coverage by RRBS and compared the results with those of MeDIP-chip. We produced five RR genomes digested *in silico* by MspI and displaying size windows of 250 bp with a 40 bp increment (from 0-250 bp to 160-410 bp). The results are summarized in Supplementary Table S2. Consistent with the microarray design, the coverage in terms of promoter-TSS and genes was higher for the MeDIP-chip than for any RR genome, illustrating the complementarity of these two approaches. While the percentage of the whole genome covered in RR genomes increased from 2.7% to 3.2% with size window increments, the CpG sites covered decreased from 13% to 9.7%. This decrease paralleled the fall in CGI and intragenic percentages. The 0-250 bp RR genome was the most cost-effective in terms of the CpGs covered versus the fraction of genome sequenced. This RR genome was also interesting in terms of the gene features covered, with the highest percentage of CpGs included in genes, promoter-TSS, exons and CGIs, and the lowest percentage of CpGs included in repeats. However, we selected the 40-290 bp RR genome because very short fragments have a high probability of mapping on multiple locations [49, 50] and are little informative after sequencing. This RR genome was also close to the 40-220 bp size window which is widely used in other species [38, 49, 51]. We then compared the bovine 40-290 bp RR genome with 40-290 bp RR genomes from other species, and noticed that a roughly similar CpG enrichment ratio was obtained in all species (ratio between the number of CpGs per Mb RR genome and the number of CpGs per Mb whole genome; Supplementary Table S3).

45 430
46
47 431
48
49 432
50
51
52 433
53

These *in silico* analyses suggested that the RRBS procedure could be successfully adapted to the bovine genome using MspI with a size window of 40-290 bp, and would lead to a base-resolution map of the methylome with coverage that would be complement that of the MeDIP-chip.

54 434 **The tissue/cell type is a massive determinant of DNA methylation landscapes in cattle**

55
56
57 435
58
59 436
60
61
62
63
64
65

In parallel with a MeDIP-chip analysis, which was conducted as described elsewhere [32] on bovine sperm, liver and fibroblasts, we generated RRBS libraries from bovine sperm, monocytes and

437 fibroblasts in order to determine cell type-dependent variations of the methylome using these two
1
2 438 complementary technologies. Thirty million pairs of reads per RRBS library were obtained on
3
4 439 average, among which 89% could be aligned to the reference genome with 36.6% unique mapping
5
6
7 440 and 52.4% multi-mapping (ambiguous reads; Supplementary Table S6). The rate of unique mapping
8
9 441 was lower than that reported for other mammals including sheep [50] and pig [51], but higher than
10
11 442 that observed in zebrafish [49] and within the same range as another RRBS study conducted in cattle
12
13
14 443 [52]. The bisulfite conversion rate was $\geq 99.5\%$ for all samples (see Methods) but did not reach 100%.
15
16 444 Under these conditions, it was impossible to distinguish non-CpG methylation (which represents only
17
18 445 0.02% of total methylation in somatic cells [53]) from residual cytosines due to incomplete bisulfite
19
20
21 446 conversion. We therefore focused our analysis exclusively on CpG methylation. The CpGs were then
22
23 447 selected based on their coverage by uniquely mapped reads. We conserved the CpGs covered by at
24
25
26 448 least 5 reads, which is the recommended threshold when the differences between conditions are
27
28 449 expected to be important [54], and filtered out the CpGs covered by >500 reads since excessive
29
30
31 450 coverage might be related to a biased amplification of repetitive elements during library preparation.
32
33 451 These filtering steps led to the selection of approximately 2.5 million CpGs per sample (CpGs 5-500),
34
35 452 from which an average methylation rate was calculated. In contrast with the results obtained by
36
37
38 453 LUMA, the average methylation in sperm (51.8%) was higher than in fibroblasts (48%) and lower than
39
40 454 in monocytes (57.6%). We next categorized the CpGs 5-500 into hypo- ($<20\%$ methylation),
41
42 455 intermediate (20–80% methylation) and hypermethylated CpGs ($>80\%$ methylation) and observed a
43
44
45 456 larger proportion of hypo- and a smaller proportion of intermediate CpGs in sperm than in somatic
46
47 457 cells, which was counterbalanced by a large proportion of hypermethylated CpGs. This bimodal
48
49
50 458 distribution of methylation in sperm probably explained the intermediate level of average
51
52 459 methylation for CpGs 5-500. It was noted that when only CpGs >500 were considered, average
53
54 460 methylation increased in monocytes (81.4%) and in fibroblasts (66.6%), but fell dramatically in sperm
55
56
57 461 (22.5%), which is reminiscent of the sperm hypomethylation we observed using LUMA. It is possible
58
59
60
61
62
63
64
65

462 that the CpGs >500, which represented less than 1% of all the CpGs analyzed by RRBS, corresponded
1
2 463 to CpGs that are more widely represented in LUMA.
3

4 464 We next conducted descriptive analyses of the MeDIP-chip and RRBS data. For MeDIP-chip, a
5
6
7 465 normalized factor NE_p, representing the number of enriched probes at promoter p for sample i, was
8
9 466 calculated for all promoters and samples, and hierarchical clustering and PCA were run on the
10
11 467 resulting matrix (Figure 2A). For RRBS, PCA and hierarchical clustering were computed from the
12
13 468 matrix of methylation percentages obtained for each CpG 5-500 and each sample (Figure 2B). With
14
15 469 both types of descriptive analysis, the samples were clearly clustered according to the tissue/cell
16
17 470 type using both MeDIP-chip and RRBS. Interestingly, hierarchical clustering revealed that the distance
18
19 471 between sperm and other cell types was more important than the distance between liver and
20
21 472 fibroblasts or monocytes and fibroblasts, highlighting the methylation specificities of germinal cells
22
23 473 compared to somatic cells. This could also be seen in PCA, where dimension 1 opposed sperm to one
24
25 474 or both somatic cells/tissue, while the difference between the two somatic cell types was more
26
27 475 apparent along dimension 2.
28
29
30
31

32
33 476 In conclusion, our results showed that the tissue/cell type represented the main source of
34
35 477 variation in methylation, and that sperm-specific methylation profiles could emerge from our MeDIP
36
37 478 and RRBS data.
38
39
40
41

42 480 **Identification of regions and CpGs hypomethylated in bull sperm**

43
44

45 481 We therefore ran a differential analysis on each pair of tissue/cell types. For MeDIP, we
46
47 482 compared the proportion of enriched probes between two tissues/cell types in 27,684 regions of
48
49 483 interest containing clusters of probes enriched in at least one tissue/cell type (see Methods). This led
50
51 484 to the identification of 4,329 DMRs between sperm and liver; 3,780 DMRs between sperm and
52
53 485 fibroblasts and 2,803 DMRs between fibroblasts and liver. The features of each DMR and the
54
55 486 corresponding promoter regions are summarized in Additional Data File 2. To analyze the RRBS data,
56
57 487 we used methylKit software [42] which is widely used for differential analysis using bisulfite
58
59
60
61
62
63
64
65

488 sequencing data. Because our experimental design included only two biological replicates per cell
1
2 489 type, we set the q-value threshold at 0.001 in order to prevent the generation of false positives. In
3
4 490 parallel with this methylKit analysis, some CpGs 5-500 displaying extreme methylation percentages in
5
6
7 491 both biological replicates of each cell type were added to the DMC list in case they had not previously
8
9 492 been detected by methylKit (“obvious DMCs”, see Methods). Using these stringent criteria, we
10
11 493 identified 298,901 DMCs between monocytes and sperm; 450,971 DMCs between fibroblasts and
12
13
14 494 sperm, and 239,036 DMCs between monocytes and fibroblasts (Supplementary Table S7 and
15
16 495 Additional Data File 3).

18 496 Consistent with the results of clustering and PCA, the number of DMRs/DMCs was higher
19
20
21 497 between sperm and any somatic cell type/tissue than between two somatic cell types, for both
22
23 498 MeDIP and RRBS. Figure 3A-B shows the distribution of DMRs and DMCs in different Venn territories
24
25
26 499 corresponding to pairwise comparisons under MeDIP (Figure 3A) and RRBS (Figure 3B). DMRs and
27
28 500 DMCs specific to the comparison between sperm and somatic cells could be deduced from these
29
30 501 territories (in red). These 1,678 DMRs and 174,103 DMCs were located at the intersection between
31
32
33 502 the “sperm vs. somatic cell type 1” and “sperm vs. somatic cell type 2” territories (yellow and orange)
34
35 503 excluding the DMRs/DMCs also shared by the “somatic cell type 1 vs. somatic cell type 2” comparison
36
37
38 504 (three-color territory). Because the comparison between sperm and fibroblasts was performed using
39
40 505 both MeDIP and RRBS, the actual complementarity of the two technologies could be assessed using
41
42 506 real data. As shown in Figure 3C, only a limited subset of CpGs was shared by DMRs identified using
43
44
45 507 MeDIP and DMCs identified using RRBS. This result, together with the *in silico* analysis indicating that
46
47 508 the targeted regions were largely different under MeDIP-chip and RRBS, clearly demonstrated that
48
49 509 we were able to identify two distinct subsets of sperm-specific DMRs/DMCs.

51
52 510 We next investigated whether these two subsets of sperm-specific DMRs/DMCs might
53
54 511 display similar features in terms of variations in methylation. For both MeDIP (Figure 3D) and RRBS
55
56
57 512 (Figure 3E), we plotted the differences in methylation between each somatic cell type and sperm for
58
59 513 sperm-specific DMRs/DMCs. The appropriate background was used for each technology: the 27,684
60
61
62
63
64
65

514 regions of interest subjected to differential analysis in MeDIP and the 1,580,644 CpGs 5-500 covered
1
2 515 in all six samples analyzed using RRBS. With both MeDIP and RRBS, background CpGs (in black) were
3
4 516 particularly concentrated around the center of the plot, illustrating that most of the CpGs analyzed
5
6
7 517 did not display cell type-dependent variations. By contrast, sperm-specific DMRs/DMCs were located
8
9 518 along a diagonal that ran from the lower left-hand corner to the upper right-hand corner, meaning
10
11 519 that these DMRs/DMCs behaved similarly in the two somatic cell types when compared to sperm.
12
13
14 520 Most strikingly, a great majority of sperm-specific DMRs/DMCs (68% for MeDIP and 79% for RRBS)
15
16 521 were grouped in the upper right-hand corner, indicating that they were hypomethylated in sperm
17
18
19 522 when compared to the two somatic cell types examined.
20

21 523 Taken together, these results demonstrated that using two complementary technologies we
22
23 524 were able to identify two distinct subsets of DMRs/DMCs specific to the comparison between sperm
24
25
26 525 and somatic cells, most of them being hypomethylated in sperm. These hypomethylated sperm-
27
28 526 specific DMRs/DMCs (hypo-DMRs/DMCs) could therefore partly explain the global hypomethylation
29
30
31 527 of sperm observed in LUMA which was particularly marked in cattle.
32

33 528

35 529 **Hypomethylation in bull sperm targets specific genomic features and functions**

38 530 To determine whether specific gene ontology (GO) terms were enriched in the sperm
39
40 531 hypomethylated regions, we next annotated the hypo-DMRs identified by MeDIP relative to genes.
41
42
43 532 Consistent with the microarray design, most of the 1144 hypo-DMRs were located in or close to
44
45 533 genes according to the criteria described in the Methods, resulting in a list of 701 unique genes which
46
47 534 were then subjected to DAVID analysis. Significant enrichments were found for biological processes
48
49
50 535 such as sexual reproduction (Benjamini-adjusted p-value: 0.0013; 36 genes), fertilization (adjusted p-
51
52 536 value: 0.003; 15 genes) and RNA transport (adjusted p-value: 0.025; 11 genes). Further analysis using
53
54
55 537 a more restrictive list of GO terms (those designated as DIRECT by DAVID) led to the identification of
56
57 538 functional clusters involved in mRNA processing (enrichment score: 1.53; Figure 4A) and
58
59 539 meiosis/spermatogenesis (enrichment score: 1.49; Figure 4B).
60
61
62
63
64
65

540 Because CpGs targeted by RRBS are scattered along the genome, we then started to
1
2 541 characterize the hypo-DMCs identified by RRBS relative to different genomic features (genes, CpG
3
4 542 density and overlapping repeats; Figure 5A). The most remarkable observation was a dramatic
5
6
7 543 enrichment of hypo-DMCs for repeats (24.5% vs. 13.2% in background; right panel), and particularly
8
9 544 for satellites (64.7% vs. 17.8%). We suspected that only a small fraction of the repetitive elements
10
11 545 was represented in the uniquely mapped reads which were analyzed here, and hypothesized that
12
13
14 546 more information on the repeats hypomethylated in sperm could be extracted from the ambiguous
15
16 547 reads. We therefore built an artificial genome containing one copy of each bovine repeat as defined
17
18
19 548 in the Repbase database [43], and aligned the totality of the reads on this artificial genome. Under
20
21 549 these conditions, reads mapping at a unique location reached 24% on average, which was quite high
22
23 550 considering the reduced size of the artificial genome (Supplementary Table S8). By contrast with the
24
25
26 551 results obtained using the bovine reference genome (Supplementary Table S6), the number of CpGs
27
28 552 >500 was not dramatically lower than that of CpGs 5-500, which was in line with the stacking of
29
30
31 553 numerous reads at the same location on the artificial genome. The average methylation was the
32
33 554 lowest in sperm, especially for CpGs >500, and the proportion of hypomethylated CpGs (average
34
35 555 methylation <20%) was consistently the highest in sperm. Additional Data File 4 shows the average
36
37
38 556 percentage of methylation calculated individually for each Repbase consensus repeat, from either
39
40 557 CpGs 5-500 or CpGs >500, and Supplementary Figure S1 recapitulates the average percentage per
41
42 558 family of repeats in each cell type. The hypomethylation of sperm was clear in satellites and also in
43
44
45 559 rDNA repeats encoding ribosomal RNAs. We therefore concluded from both CpGs targeted by
46
47 560 uniquely mapped reads (that had been used to identify hypo-DMC) and CpGs targeted by ambiguous
48
49
50 561 reads, that specific families of repetitive elements were massively hypomethylated in bovine sperm.

51
52 562 Regarding gene features with respect to RRBS data, the proportion of genes containing hypo-
53
54 563 DMCs was relatively unchanged compared with background (Figure 5A, left panel), but interestingly,
55
56
57 564 exons were more frequently represented (26.8% vs. 13.8% in background) while promoter-TSS were
58
59 565 represented less than in background (4.8% vs. 13.1%). The distribution of hypo-DMCs in CGIs, shores

566 and shelves was identical to that observed with the background (middle panel). We investigated
1
2 567 whether specific GO terms were enriched in gene features displaying a different representation in
3
4 568 hypo-DMCs and background. For exonic hypo-DMCs, which accounted for 2713 unique genes,
5
6
7 569 significant enrichments were found for biological processes such as cell adhesion (Benjamini-
8
9 570 adjusted p-value: 5.3e-5; 213 genes), the regulation of signaling (adjusted p-value: 3.9e-2; 382 genes)
10
11 571 and cell migration (adjusted p-value: 3.8e-2; 177 genes), and one main functional cluster related to
12
13 572 cell adhesion could also be identified (Figure 5B, enrichment score: 5.2). For hypo-DMCs located in
14
15 573 promoter-TSS, which accounted for 1200 unique genes, the most enriched biological process was
16
17 574 sexual reproduction (adjusted p-value: 7.2e-2; 44 genes), and two functional clusters were identified
18
19 575 (Figure 5C) as being related to piRNA metabolism (enrichment score: 2.45; left panel) and to meiosis
20
21 576 and spermatogenesis (enrichment score: 2.06; right panel).

22
23
24
25 577 From an analysis of both hypo-DMRs obtained in MeDIP and hypo-DMCs identified through
26
27 578 RRBS, we therefore concluded that hypomethylation in sperm essentially targeted repeats and the
28
29 579 promoters of genes important to spermatogenesis (which is the differentiation process that
30
31 580 eventually leads to the mature sperm we analyzed), but also to genes involved in cell
32
33 581 communication, signaling and migration that may be essential to both sperm functions and post-
34
35 582 fertilization steps.
36
37
38
39
40
41

42 584 **Validation by pyrosequencing**

43
44
45 585 Four regions were selected for validation, based on their hypomethylation in sperm, their
46
47 586 content in DMRs/DMCs and their position relative to genes involved in sperm functions. Figure 6A
48
49 587 shows the detailed localization of these regions together with their coverage and individual
50
51 588 methylation in both MeDIP and RRBS. *LSM4*, which contained a hypo-DMR, is a gene involved in RNA
52
53 589 processing [55]. Genes *SYCP3* and *DDX4*, which contained both hypo-DMRs and hypo-DMCs in our
54
55 590 study, play a major role in spermatogenesis insofar as either mutation or aberrant methylation of
56
57 591 these genes can lead to male infertility [56, 57]. *BTSAT4* was the most frequently represented bovine
58
59
60
61
62
63
64
65

592 satellite in our RRBS data and displayed important hypomethylation in sperm (Additional Data File 4).

1
2 593 Overall, the four regions represented 42 analyzed CpGs. We used the pyrosequencing of bisulfite-
3
4 594 converted DNA [44] to quantify the absolute methylation percentage of individual CpGs in sperm,
5
6
7 595 liver and fibroblasts (Figure 6B; genomic DNA from monocytes was in limited amounts and was saved
8
9 596 for separate investigations). Consistent with the MeDIP data that revealed an important enrichment
10
11 597 of probes along the *LSM4* promoter in liver and fibroblasts, the CpGs analyzed by pyrosequencing
12
13 598 were strongly hypomethylated in sperm when compared to somatic cells. In this region, the CpGs
14
15 599 assessed by pyrosequencing were not covered by RRBS, but those in *DDX4*, *SYCP3* and *BTSAT4* were
16
17 600 covered by the three techniques and displayed important hypomethylation in sperm whatever the
18
19 601 technique used. By pyrosequencing four additional CpGs, we also checked that one region specifically
20
21 602 hypermethylated in sperm compared to somatic cells validated (data not shown).
22
23
24

25
26 603 In conclusion, the results obtained using MeDIP, RRBS and pyrosequencing were in excellent
27
28 604 agreement, which validated the high-throughput data and led to the characterization of gene regions
29
30 605 with methylation patterns specific to sperm in cattle.
31
32

33 606

34 35 607 **Discussion**

36 608

37
38
39
40
41 609 During this study, we performed DNA methylation analyses at different genome scales in
42
43 610 order to exhaustively characterize the bovine sperm methylome. Our main findings were that global
44
45 611 methylation was dramatically low in bull sperm compared with other species, and that bull sperm
46
47 612 was hypomethylated when compared to bovine somatic cells in the context of two genome-wide
48
49 613 methylation assays targeting distinct genomic regions, namely RRBS and MeDIP-chip, with a focus on
50
51 614 gene promoters.
52

53
54
55 615 The hypomethylation of bull sperm compared to somatic cells agreed well with findings in
56
57 616 other mammals and provides additional evidence of features of the male germline differentiation
58
59 617 program being conserved across species. In line with our observations, a full-genome analysis of
60
61

1
2 618 sperm DNA methylation conducted in human and chimp highlighted the hypomethylation of a
3
4 619 majority of promoters in both species [58]. In addition, studies of the human sperm methylome
5
6 620 performed using the Infinium 450K methylation platform revealed that hypomethylated CpGs were
7
8 621 more abundant in sperm than in somatic cells [16, 59]. Together with our work, these reports
9
10 622 demonstrate that the hypomethylated status of sperm is independent of both the genome coverage
11
12 623 and resolution of the technology used to map DNA methylation, and suggest that sperm
13
14 624 hypomethylation is conserved among mammals. We found that promoters and exons of genes
15
16 625 hypomethylated in bull sperm were enriched for biological processes essential to sperm functions,
17
18 626 such as sexual reproduction, fertilization, cell adhesion and migration, meiosis, RNA transport and
19
20 627 processing (including piRNA metabolism), and the regulation of signaling. Interestingly, genes
21
22 628 involved in these processes displayed highly dynamic expression in post-natal mouse spermatogonial
23
24 629 stem cells, with an initial prominence of cell adhesion and piRNA categories at birth that coincides
25
26 630 with gonad colonization and piRNA expression; this is followed by histone synthesis and RNA
27
28 631 splicing/translation when the germ cells proliferate and expand within the seminiferous tubules; the
29
30 632 signaling pathways are then modified and finally meiotic genes are activated after puberty [60]. In
31
32 633 human sperm, some of these processes (cell adhesion, sexual reproduction, meiosis and piRNA
33
34 634 metabolism) are enriched in the hypomethylated promoters [58]. Hypermethylation of the piRNA
35
36 635 machinery in testes has also been associated with human spermatogenic disorders [61]. In several
37
38 636 species including bovine, the hypomethylation of mature spermatozoa could therefore reflect a
39
40 637 dynamic sequence of past transcriptional events in the male germline differentiation program which
41
42 638 are essential to sperm functions.

43
44 639 Another striking finding revealed by our RRBS data was the dramatic hypomethylation of
45
46 640 repetitive elements in bull sperm, which particularly targeted satellites. While the approach used to
47
48 641 consider repetitive sequences is still a matter of debate due to the mapping artefacts they potentially
49
50 642 generate [62, 63], we decided initially to disregard ambiguous reads. Because a high proportion of
51
52 643 repeats are represented by ambiguous reads, this strategy probably led us to underestimate the total
53
54
55
56
57
58
59
60
61
62
63
64
65

1 644 contribution of repeats to the hypomethylated CpGs we identified. In order to rescue a proportion of
2 645 these discarded repeats, we then aligned the sequences on an artificial genome that contained one
3
4 646 copy of each bovine repetitive element, and were able to confirm the drastic hypomethylation of
5
6
7 647 satellites in bull sperm using this alternative approach. The hypomethylation of satellites in sperm
8
9 648 has long been described in several species, including bovine, by analyses of candidate sequences [64-
10
11 649 66]. More recently, the hypomethylated status of satellites in human and chimp sperm has been
12
13
14 650 generalized to the whole genome by means of bisulfite sequencing [58]. Hypermethylation of a
15
16 651 human satellite sequence was shown to be associated with sperm abnormalities [67], suggesting that
17
18
19 652 methylation erasure at satellites is an important process in the differentiation of male germ cells.
20
21 653 Satellites are essential components of the constitutive heterochromatin in mammals, and this
22
23 654 function is partly mediated by DNA methylation [68]. Satellites play key roles in chromosome
24
25
26 655 structure, stability and segregation. Through their high molecular dynamics and ability to drive
27
28 656 chromosome rearrangements, they are considered to be major actors in diseases such as cancer, but
29
30
31 657 also in genome evolution and speciation [69]. The significance of satellite hypomethylation in sperm
32
33 658 could be related to the transcriptional burst that arises from paternal satellites in early mouse
34
35 659 development, which is necessary for normal formation of the heterochromatin in embryos and for
36
37
38 660 developmental progression [70]. Consistent with this important transcriptional activity, satellites
39
40 661 remain hypomethylated after fertilization in normal preimplantation embryos [66, 71]. In contrast,
41
42 662 embryos resulting from somatic cell nuclear transfer (SCNT) and have a reduced development
43
44
45 663 potential display somatic-like hypermethylated satellites in the mouse [66] and bovine [72-74]. Of
46
47 664 note, the hypermethylation of satellites seems to persist in the sperm of adult SCNT-derived bulls
48
49
50 665 [75], suggesting that it has resisted the two waves of epigenetic reprogramming that occur during
51
52 666 early development and germ cell differentiation. As well as satellites, the results we obtained using
53
54 667 the artificial genome of repetitive elements also highlighted the hypomethylation of rDNA repeats
55
56
57 668 encoding ribosomal RNAs. A study of mouse sperm reported that the average methylation of rDNA
58
59 669 was relatively weak (which is consistent with our observation) but highly variable across individuals,
60
61
62
63
64
65

670 which we did not investigate here due to the small sample size [76]. Methylation was correlated to
1
2 671 the rDNA copy number in each mouse sire and was transmitted to the somatic cells of the offspring,
3
4 672 suggesting a strong relationship between genetic and epigenetic inheritance.
5
6

7 673 Strikingly, the sperm-specific hypomethylated sequences identified throughout this study,
8
9 674 and which were particularly enriched in genes related to germline differentiation and in satellite and
10
11 675 rDNA repeats, displayed several common features. Firstly, these hypomethylated sequences have
12
13 676 previously been described as targets of *DNMT3B de novo* DNA methyltransferase in both germ cells
14
15 677 and somatic cells. The methylation of satellites, which is limited to 40% in the germline of newborn
16
17 678 male mice, is half-reduced in mutant deficient for *DNMT3B* [77]. In humans, germline *DNMT3B*
18
19 679 mutations lead to the ICF syndrome (Immunodeficiency Centromeric instability Facial anomalies). In
20
21 680 somatic cells, this disease is associated with a hypomethylated status of germline genes and
22
23 681 centromeric satellites that closely mimics that of gametes, affecting nuclear organization and
24
25 682 chromosome stability [78]. From the data accumulated in other species, it is therefore tempting to
26
27 683 speculate that the establishment of methylation is governed by common mechanisms for the
28
29 684 sequences we found to be hypomethylated in bull sperm. Another common feature shared by
30
31 685 sperm-specific hypomethylated sequences is that they partly remain associated to nucleosomes in
32
33 686 mature spermatozoa, as reported for genes involved in RNA processing and for repetitive elements
34
35 687 (including centromeric satellites) in humans and bovine [79], and for satellites and rDNA repeats in
36
37 688 bovine [80]. In addition, sperm histone retention particularly affects CpGs that lack methylation in
38
39 689 humans [59] and the mouse [81]. The association with retained nucleosomes in sperm, together with
40
41 690 the important function of paternal satellite transcripts in early embryos and the hypermethylation of
42
43 691 satellites in SCNT embryos, support the hypothesis that regions which are hypomethylated in sperm
44
45 692 play a fundamental role not only in germline differentiation but also in post-fertilization epigenetic
46
47 693 reprogramming.
48
49
50
51
52
53
54

55
56 694 The final question arising from our study originates from the intriguing finding that global
57
58 695 methylation at CCGG sites was more than 10% lower in bull sperm than in any other species
59
60
61
62
63
64
65

696 investigated (sheep, horse, pig, mouse, goat and human). Because satellites are hypomethylated in
1
2 697 the sperm of many species, the global hypomethylation of bull sperm could partly be explained by
3
4 698 the larger amount of satellite sequences present in the bovine genome (eight satellite components
5
6
7 699 representing 23% of the total genomic content; [82]). We annotated the CCGG sites relative to the
8
9 700 genomic features available in different species, and indeed observed that CCGGs overlapping
10
11 701 satellites represented 1.6% of all CCGGs in bovine vs. 0.1 to 0.5% in other species. From a purely
12
13 702 mathematical point of view, the low percentage of CCGG sites present in satellites whatever the
14
15 703 species probably did not accounted for the >10% difference in sperm methylation. However, it
16
17 704 should be kept in mind that the repetitive nature of satellites precludes their correct integration in
18
19 705 genome assemblies, which probably led to an underestimation of their contribution to global CCGG
20
21 706 methylation. Although we cannot rule out that the weaker global methylation in bull sperm is due to
22
23 707 a higher representation of bovine satellites in CCGGs, an alternative explanation might be that
24
25 708 satellite methylation is quantitatively lower in bull sperm than in the sperm of other species. This is
26
27 709 supported by an old report which demonstrated that relative to somatic tissues, most satellites are
28
29 710 largely undermethylated in the sperm of cattle while they are only slightly undermethylated in
30
31 711 mouse sperm [64]. The abundance of satellites in the bovine genome, together with their low
32
33 712 methylation content in male germ cells, may contribute to explaining some of the bovine-specific
34
35 713 features of meiotic recombination. Crossing-over events are more frequently observed in the
36
37 714 spermatocytes of *Bos taurus* than in those of related Bovidae species (wildebeests; [83]). Moreover,
38
39 715 the meiotic recombination rate in cattle is particularly elevated in males, while in most species it is
40
41 716 higher in females [84, 85]. The frequency of crossing-overs is usually low in repeat-rich domains
42
43 717 associated with heterochromatin [86]; however, the weak methylation of satellites in bull germ cells
44
45 718 probably reflects a particular chromatin structure that may promote crossing-overs. The global
46
47 719 hypomethylation of bull spermatozoa may also have contributed to shaping the bovine genome.
48
49 720 Indeed, segmental duplications in the bovine reference genome are particularly enriched for satellite
50
51 721 repeats that are hypomethylated in bull sperm, including *BTSAT4* [87]. Segmental duplications
52
53
54
55
56
57
58
59
60
61
62
63
64
65

722 promote both chromosome rearrangements that drive bovine genome evolution (as indicated by
1
2 723 their enrichment in the vicinity of cattle-specific evolutionary breakpoints [26]), and inter-individual
3
4 724 variability through their ability to promote CNVs [88]. The role of segmental duplications in structural
5
6
7 725 variations of the bovine genome might be mediated by global hypomethylation of the bull germline.
8
9 726 In support of this hypothesis, human-specific evolutionary rearrangements and CNVs associate not
10
11 727 only with low copy repeats and the deletions/duplications they generate, but also with
12
13 728 hypomethylated regions of human sperm [89], thus providing a potential link between
14
15 729 hypomethylation in the germline and genome structural variation.
16
17
18
19 730

20

21 731 **Conclusions**

22
23 732 By means of a thorough characterization of bull sperm DNA methylation at different genome
24
25 733 scales, this study has provided evidence that bull spermatozoa are hypomethylated when compared
26
27 734 to not only bovine somatic cells but also the sperm of other mammals. The sequences
28
29 735 hypomethylated in bull sperm are conserved across species, which may denote an important role in
30
31 736 germline differentiation and in post-fertilization epigenetic reprogramming. The cattle-specific
32
33 737 hypomethylation at CCGG sites may be partly related to the abundance of satellites in the bovine
34
35 738 genome and to their hypomethylated status in the male germline. Given the potential evolutionary
36
37 739 implications of these findings, it would be of considerable interest to quantify DNA methylation at
38
39 740 different stages of bovine male germline differentiation in order to understand when and how this
40
41 741 hypomethylation takes place.
42
43
44
45
46

47 742

48
49 743

50 744 **List of abbreviations**

51
52 745

53
54
55 746 PBMCs, peripheral blood mononuclear cells; PBS, phosphate buffer saline; DTT, dithiothreitol; TSS,
56
57 747 transcription start site; TTS, transcription termination site; CGI, CpG island; RR genome, reduced
58
59
60
61

748 representation genome; LUMA, luminometric methylation assay; MeDIP, methylated DNA
1
2 749 immunoprecipitation; IGV, Integrative Genomics Viewer; PCA, principal component analysis; DMR,
3
4 750 differentially methylated region; DAVID, Database for Annotation, Visualization and Integrated
5
6
7 751 Discovery; RRBS, reduced representation bisulfite sequencing; DMC, differentially methylated CpG;
8
9 752 GO, gene ontology; SCNT, somatic cell nuclear transfer; CNV, copy number variation.

10
11 753

12
13
14 754

15 16 755 **Declarations**

17
18
19 756

20
21
22 757 **Ethics approval and consent to participate:** all study methods were implemented in accordance with
23
24 758 EU guidelines and regulations (Directive 2010/63/UE). For animals maintained in INRA facilities, the
25
26 759 experimental protocols were approved by the INRA local Ethics Committee (COMETHEA,
27
28
29 760 authorization numbers 12/160 and Méjusseaumes Animal Care committee 0162503). Mice semen
30
31 761 samples were collected from 7-week-old male C57Bl/6J0laHsd mice supplied by Harlan Laboratory
32
33
34 762 (Venray, Netherlands) and euthanized by cervical dislocation. The bull samples were purchased by
35
36 763 commercial companies: Montbéliarde breed by GEN'IATEST (Roulans, France) and UMOTEST
37
38 764 (Ceyzeriat, France), Holstein and Normande breeds by EVOLUTION (Rennes, France) and Belgian
39
40
41 765 White Blue Breed by AWE (Ciney, Belgium); the ram, goat and boar semen were also supplied by
42
43 766 commercial companies, ILE-DE FRANCE breed (Oson, France), CAPGENES (Mignaloux-Beauvoir,
44
45 767 France) and Laboratoire National de Contrôle des Reproducteurs (LNCR, Maison Alfort, France),
46
47
48 768 respectively. Stallion semen was supplied by Dr. M. Magistrini (UMR INRA 085 PRC, Nouzilly, France).
49
50 769 Human sperm samples from patients included in a PHRC METASPERME study, coordinated by Dr. R.
51
52 770 Levy (Laboratoire d'Histologie Embryologie Cytogénétique CECOS, Hôpital Jean Verdier, Bondy,
53
54
55 771 France); this study received the approval from French ethics board (conseil d'évaluation éthique pour
56
57 772 les recherches en santé, CERES) and all the patients gave their informed written consent to
58
59
60 773 participate.

774 **Consent for publication:** not applicable

775 **Availability of data and material:** additional data files are provided (see below). The MeDIP-chip and
776 RRBS datasets supporting the results of this article are available in the NCBI Gene Expression
777 Omnibus database under accession numbers GSE102960 (MeDIP) and GSE102169 (RRBS).

778 **Competing interests:** the authors declare that they have no competing interests.

779 **Funding:** this work was supported by grants from the French National Research Agency (grant ANR-
780 13-LAB3-0008-01 'SeQuaMol' and grant ANR-11-INBS -0003 in the framework of the 'Investing for
781 the Future' program). SeQuaMol funding allowed the collection of large number of animal semen
782 samples, molecular analysis (MeDIP/Chip, RRBS, Bisulfite conversion, pyrosequencing). The grant
783 ANR-11-INBS -0003 supported molecular analysis such as LUMA. The funders had no role in the
784 design of the study and collection, analysis, and interpretation of data and in writing the manuscript.
785 JPP (pHd student) was supported by the French Ministry of Higher Education and Research and was a
786 fellow of the ABIES doctoral school.

787 **Authors' contributions:** JPP carried out the experiments, data analysis and drafting of the
788 manuscript. ES participated in conception of the study, collected the samples and performed the
789 experiments. AP performed the LUMA experiments. LJ performed the bioinformatics and statistical
790 analyses of RRBS and MedIP data and FP performed the bioinformatics analyses of RRBS data. HAA
791 and MG participated to the RRBS bioinformatics analyses. MW supervised the RRBS library
792 preparations. CLD, SF and DB contributed to sample collection and to critical revision of the
793 manuscript. LS and HJ participated in the conception of the study, obtained funding and editing of
794 the manuscript. HK coordinated the study and carried out the experiments, data analysis and drafting
795 of the manuscript. All authors have read and approved the final manuscript.

796 **Acknowledgements:** we would like to thank Evolution, Umotest, Awe, GEN'IATEST, OSON, Capgenes,
797 LNCR, Michèle Magistrini and Rachel Lévy for providing semen samples; Véronique Duranthon and
798 Delphine Dubé for providing fibroblasts and Chaneze Mehalla for help with the pyrosequencing of
799 *DDX4* and *SYCP3*.

800 References

801

- 802 1. Carrell DT: **Epigenetics of the male gamete.** *Fertility and sterility* 2012, **97**(2):267-274.
- 803 2. Gannon JR, Emery BR, Jenkins TG, Carrell DT: **The sperm epigenome: implications for the embryo.** *Advances in experimental medicine and biology* 2014, **791**:53-66.
- 804 3. Boissonnas CC, Jouannet P, Jammes H: **Epigenetic disorders and male subfertility.** *Fertility and sterility* 2013, **99**(3):624-631.
- 805 4. 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.** *European journal of human genetics : EJHG* 2010, **18**(1):73-80.
- 806 5. Nanassy L, Carrell DT: **Abnormal methylation of the promoter of CREM is broadly associated with male factor infertility and poor sperm quality but is improved in sperm selected by density gradient centrifugation.** *Fertility and sterility* 2011, **95**(7):2310-2314.
- 807 6. Lambrot R, Xu C, Saint-Phar S, Chountalos G, Cohen T, Paquet M, Suderman M, Hallett M, Kimmins S: **Low paternal dietary folate alters the mouse sperm epigenome and is associated with negative pregnancy outcomes.** *Nature communications* 2013, **4**:2889.
- 808 7. Martinez D, Pentinat T, Ribo S, Daviaud C, Bloks VW, Cebria J, Villalmanzo N, Kalko SG, Ramon-Krauel M, Diaz R *et al*: **In utero undernutrition in male mice programs liver lipid metabolism in the second-generation offspring involving altered Lxra DNA methylation.** *Cell metabolism* 2014, **19**(6):941-951.
- 809 8. Radford EJ, Ito M, Shi H, Corish JA, Yamazawa K, Isganaitis E, Seisenberger S, Hore TA, Reik W, Erkek S *et al*: **In utero effects. In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism.** *Science* 2014, **345**(6198):1255903.
- 810 9. Aston KI, Uren PJ, Jenkins TG, Horsager A, Cairns BR, Smith AD, Carrell DT: **Aberrant sperm DNA methylation predicts male fertility status and embryo quality.** *Fertility and sterility* 2015.
- 811 10. Salas-Huetos A, Blanco J, Vidal F, Godo A, Grossmann M, Pons MC, S FF, Garrido N, Anton E: **Spermatozoa from patients with seminal alterations exhibit a differential micro-ribonucleic acid profile.** *Fertility and sterility* 2015, **104**(3):591-601.
- 812 11. Ni K, Spiess AN, Schuppe HC, Steger K: **The impact of sperm protamine deficiency and sperm DNA damage on human male fertility: a systematic review and meta-analysis.** *Andrology* 2016, **4**(5):789-799.
- 813 12. de Castro Barbosa T, Ingerslev LR, Alm PS, Versteijhe S, Massart J, Rasmussen M, Donkin I, Sjogren R, Mudry JM, Vetterli L *et al*: **High-fat diet reprograms the epigenome of rat spermatozoa and transgenerationally affects metabolism of the offspring.** *Molecular metabolism* 2016, **5**(3):184-197.
- 814 13. Sharma U, Conine CC, Shea JM, Boskovic A, Derr AG, Bing XY, Belleannee C, Kucukural A, Serra RW, Sun F *et al*: **Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals.** *Science* 2016, **351**(6271):391-396.
- 815 14. Kobayashi N, Miyauchi N, Tatsuta N, Kitamura A, Okae H, Hiura H, Sato A, Utsunomiya T, Yaegashi N, Nakai K *et al*: **Factors associated with aberrant imprint methylation and oligozoospermia.** *Scientific reports* 2017, **7**:42336.
- 816 15. Jenkins TG, Aston KI, Meyer TD, Hotaling JM, Shamsi MB, Johnstone EB, Cox KJ, Stanford JB, Porucznik CA, Carrell DT: **Decreased fecundity and sperm DNA methylation patterns.** *Fertility and sterility* 2015.
- 817 16. Urdinguio RG, Bayon GF, Dmitrijeva M, Torano EG, Bravo C, Fraga MF, Bassas L, Larriba S, Fernandez AF: **Aberrant DNA methylation patterns of spermatozoa in men with unexplained infertility.** *Human reproduction* 2015, **30**(5):1014-1028.

- 848 17. Bourc'his D, Bestor TH: **Meiotic catastrophe and retrotransposon reactivation in male germ**
1 849 **cells lacking Dnmt3L.** *Nature* 2004, **431**(7004):96-99.
- 2 850 18. Song N, Endo D, Song B, Shibata Y, Koji T: **5-aza-2'-deoxycytidine impairs mouse**
3 851 **spermatogenesis at multiple stages through different usage of DNA methyltransferases.**
4 852 *Toxicology* 2016, **361-362**:62-72.
- 5 853 19. Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Herault Y, Guillou F, Bourc'his D: **The**
6 854 **DNA methyltransferase DNMT3C protects male germ cells from transposon activity.** *Science*
7 855 2016, **354**(6314):909-912.
- 8 856 20. Guo F, Yang B, Ju ZH, Wang XG, Qi C, Zhang Y, Wang CF, Liu HD, Feng MY, Chen Y *et al*:
9 857 **Alternative splicing, promoter methylation, and functional SNPs of sperm flagella 2 gene in**
10 858 **testis and mature spermatozoa of Holstein bulls.** *Reproduction* 2014, **147**(2):241-252.
- 11 859 21. Jena SC, Kumar S, Rajput S, Roy B, Verma A, Kumaresan A, Mohanty TK, De S, Kumar R, Datta
12 860 **TK: Differential methylation status of IGF2-H19 locus does not affect the fertility of**
13 861 **crossbred bulls but some of the CTCF binding sites could be potentially important.**
14 862 *Molecular reproduction and development* 2014, **81**(4):350-362.
- 15 863 22. Yao W, Li Y, Li B, Luo H, Xu H, Pan Z, Xie Z, Li Q: **Epigenetic regulation of bovine**
16 864 **spermatogenic cell-specific gene boule.** *PloS one* 2015, **10**(6):e0128250.
- 17 865 23. Verma A, Rajput S, De S, Kumar R, Chakravarty AK, Datta TK: **Genome-wide profiling of**
18 866 **sperm DNA methylation in relation to buffalo (*Bubalus bubalis*) bull fertility.**
19 867 *Theriogenology* 2014, **82**(5):750-759 e751.
- 20 868 24. Kropp J, Carrillo JA, Namous H, Daniels A, Salih SM, Song J, Khatib H: **Male fertility status is**
21 869 **associated with DNA methylation signatures in sperm and transcriptomic profiles of bovine**
22 870 **preimplantation embryos.** *BMC genomics* 2017, **18**(1):280.
- 23 871 25. Shojaei Saadi HA, O'Doherty AM, Gagne D, Fournier E, Grant JR, Sirard MA, Robert C: **An**
24 872 **integrated platform for bovine DNA methylome analysis suitable for small samples.** *BMC*
25 873 *genomics* 2014, **15**:451.
- 26 874 26. Bovine Genome S, Analysis C, Elsik CG, Tellam RL, Worley KC, Gibbs RA, Muzny DM,
27 875 Weinstock GM, Adelson DL, Eichler EE *et al*: **The genome sequence of taurine cattle: a**
28 876 **window to ruminant biology and evolution.** *Science* 2009, **324**(5926):522-528.
- 29 877 27. Kader F, Ghai M: **DNA methylation-based variation between human populations.** *Molecular*
30 878 *genetics and genomics : MGG* 2017, **292**(1):5-35.
- 31 879 28. Ortiz I, Dorado J, Ramirez L, Morrell JM, Acha D, Urbano M, Galvez MJ, Carrasco JJ, Gomez-
32 880 Arrones V, Calero-Carretero R *et al*: **Effect of single layer centrifugation using Androcoll-E-**
33 881 **Large on the sperm quality parameters of cooled-stored donkey semen doses.** *Animal : an*
34 882 *international journal of animal bioscience* 2014, **8**(2):308-315.
- 35 883 29. Kiefer H: **Genome-Wide Analysis of Methylation in Bovine Clones by Methylated DNA**
36 884 **Immunoprecipitation (MeDIP).** *Methods in molecular biology* 2015, **1222**:267-280.
- 37 885 30. Karimi M, Johansson S, Stach D, Corcoran M, Grander D, Schalling M, Bakalkin G, Lyko F,
38 886 Larsson C, Ekstrom TJ: **LUMA (LUMinometric Methylation Assay)--a high throughput**
39 887 **method to the analysis of genomic DNA methylation.** *Experimental cell research* 2006,
40 888 **312**(11):1989-1995.
- 41 889 31. Attig L, Vige A, Gabory A, Karimi M, Beauger A, Gross MS, Athias A, Gallou-Kabani C, Gambert
42 890 P, Ekstrom TJ *et al*: **Dietary alleviation of maternal obesity and diabetes: increased**
43 891 **resistance to diet-induced obesity transcriptional and epigenetic signatures.** *PloS one* 2013,
44 892 **8**(6):e66816.
- 45 893 32. Kiefer H, Jouneau L, Champion E, Rousseau-Ralliard D, Larcher T, Martin-Magniette ML,
46 894 Balzergue S, Ledevin M, Prezelin A, Chavatte-Palmer P *et al*: **Altered DNA methylation**
47 895 **associated with an abnormal liver phenotype in a cattle model with a high incidence of**
48 896 **perinatal pathologies.** *Scientific reports* 2016, **6**:38869.
- 49 897 33. Martin-Magniette ML, Mary-Huard T, Berard C, Robin S: **ChIPmix: mixture model of**
50 898 **regressions for two-color ChIP-chip analysis.** *Bioinformatics* 2008, **24**(16):i181-186.

- 899 34. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP:
 1 900 **Integrative genomics viewer**. *Nature biotechnology* 2011, **29**(1):24-26.
- 2 901 35. Baddeley A, Turner R: **spatstat: An R Package for Analyzing Spatial Point Patterns**. *Journal of*
 3 902 *Statistical Software* 2005, **12**(6).
- 4 903 36. Benjamini Y, Hochberg Y: **Controlling the False Discovery Rate: A Practical and Powerful**
 5 904 **Approach to Multiple Testing**. *Journal of the Royal Statistical Society Series B*
 6 905 *(Methodological)* 1995, **57**(1):289-300.
- 7 906 37. Huang da W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene**
 8 907 **lists using DAVID bioinformatics resources**. *Nature protocols* 2009, **4**(1):44-57.
- 9 908 38. Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A: **Preparation of reduced**
 10 909 **representation bisulfite sequencing libraries for genome-scale DNA methylation profiling**.
 11 910 *Nature protocols* 2011, **6**(4):468-481.
- 12 911 39. Auclair G, Guibert S, Bender A, Weber M: **Ontogeny of CpG island methylation and**
 13 912 **specificity of DNMT3 methyltransferases during embryonic development in the mouse**.
 14 913 *Genome biology* 2014, **15**(12):545.
- 15 914 40. Krueger F, Andrews SR: **Bismark: a flexible aligner and methylation caller for Bisulfite-Seq**
 16 915 **applications**. *Bioinformatics* 2011, **27**(11):1571-1572.
- 17 916 41. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of**
 18 917 **short DNA sequences to the human genome**. *Genome biology* 2009, **10**(3):R25.
- 19 918 42. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE:
 20 919 **methyKit: a comprehensive R package for the analysis of genome-wide DNA methylation**
 21 920 **profiles**. *Genome biology* 2012, **13**(10):R87.
- 22 921 43. Bao W, Kojima KK, Kohany O: **Rebase Update, a database of repetitive elements in**
 23 922 **eukaryotic genomes**. *Mobile DNA* 2015, **6**:11.
- 24 923 44. Dupont JM, Tost J, Jammes H, Gut IG: **De novo quantitative bisulfite sequencing using the**
 25 924 **pyrosequencing technology**. *Analytical biochemistry* 2004, **333**(1):119-127.
- 26 925 45. Li LC, Dahiya R: **MethPrimer: designing primers for methylation PCRs**. *Bioinformatics* 2002,
 27 926 **18**(11):1427-1431.
- 28 927 46. Weyrich A: **Preparation of genomic DNA from mammalian sperm**. *Current protocols in*
 29 928 *molecular biology* 2012, **Chapter 2**:Unit 2 13 11-13.
- 30 929 47. Schubeler D: **ESCI award lecture: regulation, function and biomarker potential of DNA**
 31 930 **methylation**. *European journal of clinical investigation* 2015, **45**(3):288-293.
- 32 931 48. Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schubeler D: **Chromosome-wide**
 33 932 **and promoter-specific analyses identify sites of differential DNA methylation in normal and**
 34 933 **transformed human cells**. *Nature genetics* 2005, **37**(8):853-862.
- 35 934 49. Chatterjee A, Ozaki Y, Stockwell PA, Horsfield JA, Morison IM, Nakagawa S: **Mapping the**
 36 935 **zebrafish brain methylome using reduced representation bisulfite sequencing**. *Epigenetics*
 37 936 2013, **8**(9):979-989.
- 38 937 50. Doherty R, Couldrey C: **Exploring genome wide bisulfite sequencing for DNA methylation**
 39 938 **analysis in livestock: a technical assessment**. *Frontiers in genetics* 2014, **5**:126.
- 40 939 51. Choi M, Lee J, Le MT, Nguyen DT, Park S, Soundrarajan N, Schachtschneider KM, Kim J, Park
 41 940 JK, Kim JH *et al*: **Genome-wide analysis of DNA methylation in pigs using reduced**
 42 941 **representation bisulfite sequencing**. *DNA research : an international journal for rapid*
 43 942 *publication of reports on genes and genomes* 2015, **22**(5):343-355.
- 44 943 52. Doherty R, Whiston R, Cormican P, Finlay EK, Couldrey C, Brady C, O'Farrelly C, Meade KG:
 45 944 **The CD4(+) T cell methylome contributes to a distinct CD4(+) T cell transcriptional signature**
 46 945 **in Mycobacterium bovis-infected cattle**. *Scientific reports* 2016, **6**:31014.
- 47 946 53. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z,
 48 947 Ngo QM *et al*: **Human DNA methylomes at base resolution show widespread epigenomic**
 49 948 **differences**. *Nature* 2009, **462**(7271):315-322.

- 949 54. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE,
1 950 Nusbaum C, Jaffe DB *et al*: **Genome-scale DNA methylation maps of pluripotent and**
2 951 **differentiated cells.** *Nature* 2008, **454**(7205):766-770.
- 3 952 55. Arribas-Layton M, Dennis J, Bennett EJ, Damgaard CK, Lykke-Andersen J: **The C-Terminal RGG**
4 953 **Domain of Human Lsm4 Promotes Processing Body Formation Stimulated by Arginine**
5 954 **Dimethylation.** *Molecular and cellular biology* 2016, **36**(17):2226-2235.
- 7 955 56. Miyamoto T, Minase G, Okabe K, Ueda H, Sengoku K: **Male infertility and its genetic causes.**
8 956 *The journal of obstetrics and gynaecology research* 2015, **41**(10):1501-1505.
- 9 957 57. Cheng YS, Lu CW, Lin TY, Lin PY, Lin YM: **Causes and Clinical Features of Infertile Men With**
10 958 **Nonobstructive Azoospermia and Histopathologic Diagnosis of Hypospermatogenesis.**
11 959 *Urology* 2017.
- 13 960 58. Molaro A, Hodges E, Fang F, Song Q, McCombie WR, Hannon GJ, Smith AD: **Sperm**
14 961 **methylation profiles reveal features of epigenetic inheritance and evolution in primates.**
15 962 *Cell* 2011, **146**(6):1029-1041.
- 16 963 59. Krausz C, Sandoval J, Sayols S, Chianese C, Giachini C, Heyn H, Esteller M: **Novel insights into**
17 964 **DNA methylation features in spermatozoa: stability and peculiarities.** *PloS one* 2012,
18 965 **7**(10):e44479.
- 20 966 60. Hammoud SS, Low DH, Yi C, Lee CL, Oatley JM, Payne CJ, Carrell DT, Guccione E, Cairns BR:
21 967 **Transcription and imprinting dynamics in developing postnatal male germline stem cells.**
22 968 *Genes & development* 2015, **29**(21):2312-2324.
- 23 969 61. Heyn H, Ferreira HJ, Bassas L, Bonache S, Sayols S, Sandoval J, Esteller M, Larriba S:
24 970 **Epigenetic disruption of the PIWI pathway in human spermatogenic disorders.** *PloS one*
25 971 2012, **7**(10):e47892.
- 27 972 62. Royo H, Stadler MB, Peters AH: **Alternative Computational Analysis Shows No Evidence for**
28 973 **Nucleosome Enrichment at Repetitive Sequences in Mammalian Spermatozoa.**
29 974 *Developmental cell* 2016, **37**(1):98-104.
- 30 975 63. Dansranjav T, Schagdarsurengin U: **The Rationale of the Inevitable, or Why Is the**
31 976 **Consideration of Repetitive DNA Elements Indispensable in Studies of Sperm Nucleosomes.**
32 977 *Developmental cell* 2016, **37**(1):13-14.
- 34 978 64. Adams RL, Burdon RH, Fulton J: **Methylation of satellite DNA.** *Biochemical and biophysical*
35 979 *research communications* 1983, **113**(2):695-702.
- 36 980 65. Feinstein SI, Racaniello VR, Ehrlich M, Gehrke CW, Miller DA, Miller OJ: **Pattern of**
37 981 **undermethylation of the major satellite DNA of mouse sperm.** *Nucleic acids research* 1985,
38 982 **13**(11):3969-3978.
- 40 983 66. Yamagata K, Yamazaki T, Miki H, Ogonuki N, Inoue K, Ogura A, Baba T: **Centromeric DNA**
41 984 **hypomethylation as an epigenetic signature discriminates between germ and somatic cell**
42 985 **lineages.** *Developmental biology* 2007, **312**(1):419-426.
- 43 986 67. Houshdaran S, Cortessis VK, Siegmund K, Yang A, Laird PW, Sokol RZ: **Widespread epigenetic**
44 987 **abnormalities suggest a broad DNA methylation erasure defect in abnormal human sperm.**
45 988 *PloS one* 2007, **2**(12):e1289.
- 47 989 68. Nishibuchi G, Dejardin J: **The molecular basis of the organization of repetitive DNA-**
48 990 **containing constitutive heterochromatin in mammals.** *Chromosome research : an*
49 991 *international journal on the molecular, supramolecular and evolutionary aspects of*
50 992 *chromosome biology* 2017, **25**(1):77-87.
- 52 993 69. Biscotti MA, Olmo E, Heslop-Harrison JS: **Repetitive DNA in eukaryotic genomes.**
53 994 *Chromosome research : an international journal on the molecular, supramolecular and*
54 995 *evolutionary aspects of chromosome biology* 2015, **23**(3):415-420.
- 55 996 70. Probst AV, Okamoto I, Casanova M, El Marjou F, Le Baccon P, Almouzni G: **A strand-specific**
56 997 **burst in transcription of pericentric satellites is required for chromocenter formation and**
57 998 **early mouse development.** *Developmental cell* 2010, **19**(4):625-638.
- 59 999 71. Kaneda M, Akagi S, Watanabe S, Nagai T: **Comparison of DNA methylation levels of**
60 1000 **repetitive loci during bovine development.** *BMC proceedings* 2011, **5 Suppl 4**:S3.

- 1001 72. Kang YK, Lee HJ, Shim JJ, Yeo S, Kim SH, Koo DB, Lee KK, Beyhan Z, First NL, Han YM: **Varied patterns of DNA methylation change between different satellite regions in bovine preimplantation development.** *Molecular reproduction and development* 2005, **71**(1):29-35.
- 11002
- 21003 73. Yamanaka K, Kaneda M, Inaba Y, Saito K, Kubota K, Sakatani M, Sugimura S, Imai K, Watanabe S, Takahashi M: **DNA methylation analysis on satellite I region in blastocysts obtained from somatic cell cloned cattle.** *Animal science journal = Nihon chikusan Gakkaiho* 2011, **82**(4):523-530.
- 31004
- 41005
- 51006
- 61007
- 71008 74. Yamanaka K, Sakatani M, Kubota K, Balboula AZ, Sawai K, Takahashi M: **Effects of downregulating DNA methyltransferase 1 transcript by RNA interference on DNA methylation status of the satellite I region and in vitro development of bovine somatic cell nuclear transfer embryos.** *The Journal of reproduction and development* 2011, **57**(3):393-402.
- 81009
- 91010
- 101011
- 111012
- 121013 75. Couldrey C, Wells DN: **DNA methylation at a bovine alpha satellite I repeat CpG site during development following fertilization and somatic cell nuclear transfer.** *PloS one* 2013, **8**(2):e55153.
- 131014
- 141015
- 151016 76. Shea JM, Serra RW, Carone BR, Shulha HP, Kucukural A, Ziller MJ, Vallaster MP, Gu H, Tapper AR, Gardner PD *et al*: **Genetic and Epigenetic Variation, but Not Diet, Shape the Sperm Methylome.** *Developmental cell* 2015, **35**(6):750-758.
- 161017
- 171018
- 181019 77. Kato Y, Kaneda M, Hata K, Kumaki K, Hisano M, Kohara Y, Okano M, Li E, Nozaki M, Sasaki H: **Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse.** *Human molecular genetics* 2007, **16**(19):2272-2280.
- 201020
- 211021
- 221022
- 231023 78. Walton EL, Francastel C, Velasco G: **Dnmt3b Prefers Germ Line Genes and Centromeric Regions: Lessons from the ICF Syndrome and Cancer and Implications for Diseases.** *Biology* 2014, **3**(3):578-605.
- 241024
- 251025
- 261026 79. Samans B, Yang Y, Krebs S, Sarode GV, Blum H, Reichenbach M, Wolf E, Steger K, Dansranjavin T, Schagdarsurengin U: **Uniformity of nucleosome preservation pattern in Mammalian sperm and its connection to repetitive DNA elements.** *Developmental cell* 2014, **30**(1):23-35.
- 271027
- 281028
- 291029
- 301030 80. Sillaste G, Kaplinski L, Meier R, Jaakma U, Eriste E, Salumets A: **A novel hypothesis for histone-to-protamine transition in Bos taurus spermatozoa.** *Reproduction* 2017, **153**(3):241-251.
- 311031
- 321032
- 331033 81. Erkek S, Hisano M, Liang CY, Gill M, Murr R, Dieker J, Schubeler D, van der Vlag J, Stadler MB, Peters AH: **Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa.** *Nature structural & molecular biology* 2013, **20**(7):868-875.
- 341034
- 351035
- 361036 82. Macaya G, Cortadas J, Bernardi G: **An analysis of the bovine genome by density-gradient centrifugation. Preparation of the dG+dC-rich DNA components.** *European journal of biochemistry* 1978, **84**(1):179-188.
- 371037
- 381038
- 391039 83. Vozdova M, Sebestova H, Kubickova S, Cernohorska H, Vahala J, Rubes J: **A comparative study of meiotic recombination in cattle (Bos taurus) and three wildebeest species (Connochaetes gnou, C. taurinus taurinus and C. t. albojubatus).** *Cytogenetic and genome research* 2013, **140**(1):36-45.
- 401040
- 411041
- 421042
- 431043 84. Ma L, O'Connell JR, VanRaden PM, Shen B, Padhi A, Sun C, Bickhart DM, Cole JB, Null DJ, Liu GE *et al*: **Cattle Sex-Specific Recombination and Genetic Control from a Large Pedigree Analysis.** *PLoS genetics* 2015, **11**(11):e1005387.
- 441044
- 451045
- 461046 85. Kadri NK, Harland C, Faux P, Cambisano N, Karim L, Coppieters W, Fritz S, Mullaart E, Baurain D, Boichard D *et al*: **Coding and noncoding variants in HFM1, MLH3, MSH4, MSH5, RNF212, and RNF212B affect recombination rate in cattle.** *Genome research* 2016, **26**(10):1323-1332.
- 471047
- 481048
- 491049 86. Termolino P, Cremona G, Consiglio MF, Conicella C: **Insights into epigenetic landscape of recombination-free regions.** *Chromosoma* 2016, **125**(2):301-308.
- 501050
- 511051 87. Liu GE, Ventura M, Cellamare A, Chen L, Cheng Z, Zhu B, Li C, Song J, Eichler EE: **Analysis of recent segmental duplications in the bovine genome.** *BMC genomics* 2009, **10**:571.
- 521052
- 531053
- 541054
- 551055
- 561056
- 571057
- 581058
- 591059
- 601059
- 61
- 62
- 63
- 64
- 65

1053 88. Fadista J, Thomsen B, Holm LE, Bendixen C: **Copy number variation in the bovine genome.**
11054 *BMC genomics* 2010, **11**:284.
21055 89. Li J, Harris RA, Cheung SW, Coarfa C, Jeong M, Goodell MA, White LD, Patel A, Kang SH, Shaw
31056 *C et al: Genomic hypomethylation in the human germline associates with selective*
41057 **structural mutability in the human genome.** *PLoS genetics* 2012, **8**(5):e1002692.

6
71058

8
91059 **Figure legends**

10
11
121060

13
141061 **Figure 1. Global methylation measured by LUMA is low in bovine sperm.** (A, B) The y-axis
151062 represents global CCGG methylation in bull and ram samples. Significant differences between
16
171063 independent samples are indicated by asterisks ($p < 0.05$, permutation test), while paired samples are
18
191064 connected by plain lines. (A) Global methylation in blood cells and semen for bulls ($n=6$) and rams
20
211065 ($n=5$). The p-values are indicated in red and green for comparisons involving samples collected from
221066 the same individuals (permutation test for paired samples). (B) Global methylation in bull ($n=6$) and
23
241067 ram ($n=4$) sperm cells from fresh and frozen semen. The difference between fresh and frozen semen
25
261068 is not significant. (C) Global methylation of sperm in four bovine breeds. For each box, the middle line
27
281069 indicates the median and the edges the 25th/75th percentiles. Significant differences between
291070 breeds are indicated by different letters ($p < 0.05$, permutation test). (D) Global sperm methylation
30
311071 and CCGG distribution in several mammalian species. The CCGG sites were annotated relative to
32
331072 gene features, CpG density and overlapping repeats. For each genomic feature examined, species
34
351073 with extreme values are indicated (min. and max.). In the bovine genome, CCGG sites are particularly
361074 enriched in CpG islands and satellites, and are within the ranges of other species for other genomic
37
381075 features. SD: standard deviation.

39
40
411076
42
43
44
45
46
47
48
49
50
511077

52
53
541078 **Figure 2. Cell type is a massive determinant of DNA methylation landscapes in cattle.** (A). MeDIP-
55
561079 chip on sperm ($n=4$, red), fibroblasts ($n=3$, green) and liver ($n=4$, blue). For each sample i and each
57
58 promoter p , a normalized number of enriched probed NEpi was computed (see Methods).
591079

60
61
62
63
64
65

1080 Correlation clustering and PCA were then performed on the matrix of the normalized number of
1
21081 enriched probes. (B) RRBS on sperm (n=2, red), fibroblasts (n=2, green) and monocytes (n=2, blue).
3
41082 Correlation clustering and PCA were run on the totality of CpGs covered between 5 and 500 reads in
5
6
71083 the six samples.
8
9
101084
11
12
131085 **Figure 3. Two distinct sets of differentially methylated regions and differentially methylated CpGs**
14
15 **display hypomethylation in sperm.** (A) Venn diagram showing the DMRs identified using MeDIP-chip
161086 in sperm, fibroblasts and liver. A total of 1,678 DMRs specific to the comparison between sperm and
17
181087 somatic cells was obtained (in red). (B). Venn diagram showing the DMCs identified using RRBS in
19
201088 sperm, fibroblasts and monocytes. A total of 174,103 DMCs specific to the comparison between
21
221089 sperm and somatic cells was obtained (in red). (C) The CpG positions included in the 3,780 DMRs
23
241090 identified between sperm and fibroblasts using MeDIP were extracted. The Venn diagram shows the
25
261091 intersection between these CpGs and the 450,971 DMCs identified between sperm and fibroblasts
27
281092 using RRBS. (D, E) Scatterplot showing the methylation differences between two somatic tissues and
29
301093 sperm, for all regions or CpGs used during differential analysis (in black; background) and for
31
321094 DMRs/DMCs specific to the comparison between sperm and somatic cells (in red). The proportions of
33
341095 background regions/CpGs hypermethylated (upper right edge) and hypomethylated (lower left edge)
35
361096 in both somatic tissues compared to sperm are indicated in black. The proportions of DMRs/DMCs
37
381097 hypermethylated and hypomethylated in both somatic tissues compared to sperm are indicated in
39
401098 red. (D) Differences in methylation between liver and sperm ($Pr_{liver} - Pr_{sperm}$; x-axis) and between
41
421099 fibroblasts and sperm ($Pr_{fibroblasts} - Pr_{sperm}$; y-axis) for regions and DMRs identified using MeDIP. See
43
441100 Methods for the definition of Pr_{sperm} , Pr_{liver} and $Pr_{fibroblasts}$. (E) Differences in methylation between
45
461101 monocytes and sperm (x-axis) and between fibroblasts and sperm (y-axis) for CpGs and DMCs
47
481102 identified using RRBS.
49
501103
51
52
53
54
55
561104
57
58
59
60
61
62
63
64
65

1105 **Figure 4. Hypo-DMRs identified by MeDIP target genes involved in mRNA processing and**
1
2 1106 **spermatogenesis.** Genes containing the 1,144 hypo-DMRs were subjected to DAVID analysis, with
3
4 1107 the regions of interest used as the background. Terms of gene ontology, pathways or Uniprot
5
6
7 1108 keywords enriched among the DMRs and their corresponding p-values are indicated, as are the genes
8
9 1109 present in each category. The green color on the heatmap represents a correspondence between a
10
11 1110 gene and a category. To limit the size of the heatmaps, only GO terms designated as DIRECT by
12
13 1111 DAVID were used for cluster generation. (A) Functional cluster related to mRNA processing. (B)
14
15 1112 Functional cluster related to meiosis and spermatogenesis.

16
17
18
19 1113
20
21
22
23 1114 **Figure 5. Hypo-DMCs identified by RRBS target specific genomic features and functions.** The
24
25 1115 137,861 hypo-DMCs and 1,580,644 CpGs (background) were annotated relative to gene features,
26
27 1116 CpG density and overlapping repeats. (A) Distribution of hypo-DMCs and background CpGs among
28
29 1117 these genomic features. (B) Genes with hypo-DMCs located in exons were subjected to DAVID
30
31 1118 analysis, with genes from some of the 1,580,644 CpGs in exons used as the background. The
32
33 1119 heatmap represents a functional cluster related to cell adhesion. (C) Genes with hypo-DMCs located
34
35 1120 in promoter-TSS were subjected to DAVID analysis, with genes containing some of the 1,580,644
36
37 1121 CpGs in promoter-TSS used as the background. The heatmaps represent functional clusters related to
38
39 1122 piRNA metabolism (left panel) and to meiosis and spermatogenesis (right panel). To limit the size of
40
41 1123 the heatmaps, only GO terms designated as DIRECT by DAVID were used for cluster generation.
42
43
44
45
46
47
48 1124

49
50
51 1125 **Figure 6. Validation by pyrosequencing.** (A) IGV browser views of the gene regions targeted for
52
53 1126 pyrosequencing. In the MeDIP panels, the “Probes” track indicates the probe positions on the
54
55 1127 microarray. The blue, green, and red bar charts represent probes with signal enrichment in the
56
57 1128 MeDIP fraction for liver, fibroblast and sperm samples, respectively. An absence of chart at a given
58
59 1129 probe position denotes that signal was not enriched in the MeDIP fraction. In the RRBS panels, the
60
61
62
63
64
65

1130 blue, green, and red bar charts represent the methylation percentages at each CpG 5-500 position
1
21131 for monocyte, fibroblast and sperm samples, respectively. Individual CpGs are shown, as are the
3
41132 MeDIP probe classes based on CpG frequency (the upper, middle and lower bands represent high,
5
6
71133 intermediate and low class probes, respectively). The orange boxes delineate the regions analyzed by
8
91134 pyrosequencing. (B) Methylation percentages of the CpGs assayed by pyrosequencing in sperm (n=6),
10
11
121135 fibroblasts (n=3) and liver (n=4). The difference between sperm and somatic cells is significant at
13
141136 every position ($p < 0.05$, permutation test).
15
16

171137

18

19

20

211138

Description of additional data files

22

23

241139

25

261140

The following additional data are available with the online version of this paper:

27

281141

Additional Data File 1 is a Microsoft Word file (.docx) containing eight Supplementary Tables and one

29

30

311142

supplementary Figure. **Supplementary Table S1:** reference genomes used for *in silico* analyses and

32

331143

origin of the files used for annotation. **Supplementary Table S2:** *in silico* characterization of bovine

34

351144

reduced restriction (RR) genomes generated using different size selection criteria. **Supplementary**

36

37

381145

Table S3: comparison of RR genomes obtained with a 40-290 bp selection size window in different

39

401146

species. **Supplementary Table S4:** primers and PCR conditions used to generate the pyrosequencing

41

421147

templates. **Supplementary Table S5:** pyrosequencing primers. **Supplementary Table S6:** library

43

44

451148

characterization, mapping efficiency on the bovine genome (UMD3.1), coverage and average

46

471149

methylation in RRBS libraries. **Supplementary Table S7:** results of comparisons between tissues by

48

49

501150

RRBS. **Supplementary Table S8:** mapping efficiency on a Rebase artificial bovine genome, coverage

51

521151

and average methylation in RRBS libraries. **Supplementary Figure S1:** Average methylation

53

541152

percentages for CpGs 5-500 and CpGs >500 in each cell type, in reads uniquely aligned on a Rebase

55

56

571153

bovine artificial genome.

58

59

60

61

62

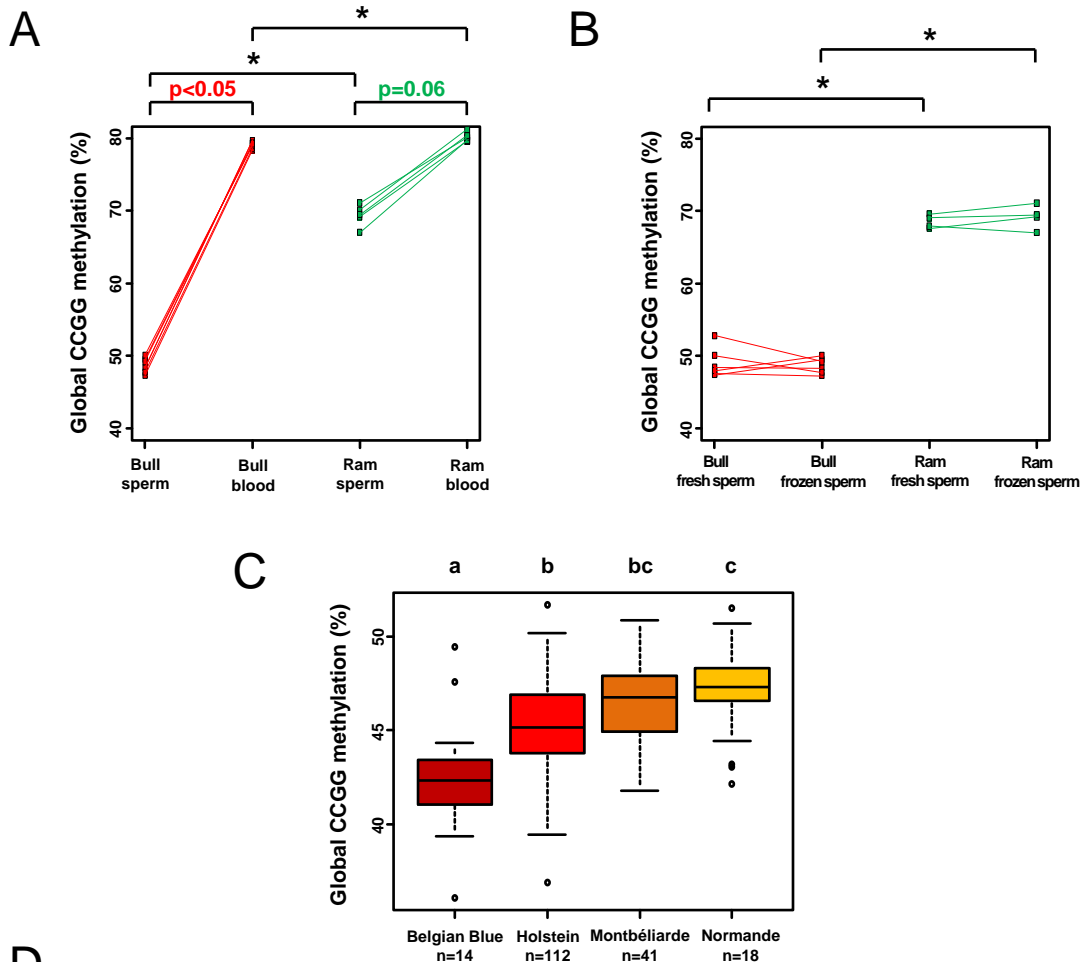
63

64

65

1154 **Additional Data File 2** is a Microsoft Excel file (.xlsx) listing the DMRs identified using MeDIP. This file
1
21155 includes three datasheets corresponding to the pairwise comparisons between sperm, fibroblasts
3
41156 and liver. One DMR may be present in several lanes if shared by several promoters.
5
6
71157 **Additional Data File 3** is a Microsoft Excel file (.xlsx) listing the DMCs identified using RRBS. This file
8
91158 includes three datasheets corresponding to the pairwise comparisons between sperm, fibroblasts
10
111159 and monocytes.
12
13
141160 **Additional Data File 4** is a Microsoft Excel file (.xlsx) listing the consensus sequences of each bovine
15
161161 repetitive element as defined in Repbase and the average methylation percentages for CpGs 5-500
17
181162 and CpGs >500 in each RRBS sample.
19
20
211163
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Figure 1

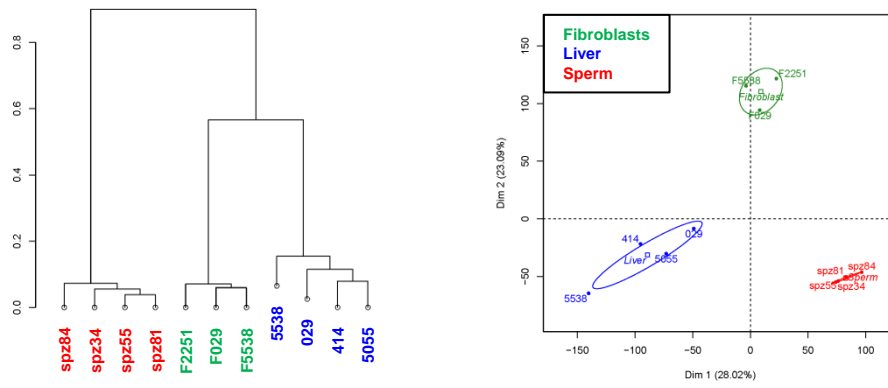


D

	Cattle	Sheep	Horse	Pig	Mouse	Human	Goat
Global CCGG methylation in sperm (mean % ± SD)	45.5 ± 2.6 (min.) n=185	69.4 n=5	76.1 (max.) n=2	72.6 n=3	59.2 n=2	57.1 n=7	67.2 n=2
Number of CCGG sites	1,987,520	1,855,703	2,602,517 (max.)	2,366,897	1,548,665 (min.)	2,196,730	NA
Gene features (%)							
Promoter-TSS	4.9	3.9	3.0 (min.)	3.4	6.1	7.2 (max.)	NA
Intron	27.0	30.4	20.8 (min.)	25.2	30.4	39.1 (max.)	NA
Exon	5.2	5.4	3.0 (min.)	4.1	7.1	7.3 (max.)	NA
TTS	2.0	2.4	1.5 (min.)	1.9	2.1	2.7 (max.)	NA
3'UTR	0.4 (min.)	0.4 (min.)	1.1	0.8	1.1	1.8 (max.)	NA
5'UTR	2.7	1.8 (min.)	1.9	3.2	7.1	8.6 (max.)	NA
Intergenic	57.8	55.7	68.7 (max.)	61.4	46.1	33.3 (min.)	NA
CpG density (%)							
Island	14.8 (max.)	11.3	NA	12.8	8.2 (min.)	12.1	NA
Shore	12.8	13.0	NA	13.2 (max.)	5.3 (min.)	9.1	NA
Shelve	6.0	6.2 (max.)	NA	6.2 (max.)	2.8 (min.)	5.0	NA
Open sea	66.4 (min.)	69.5	NA	67.8	83.7 (max.)	73.8	NA
Repeats (%)							
LINE	16.9	15.3	6.3	5.9 (min.)	20.3 (max.)	7.0	NA
SINE	11.5	12.7	13.6	26.7	10.3 (min.)	33.6 (max.)	NA
LTR	4.4	2.9 (min.)	3.8	3.1	12.6 (max.)	7.0	NA
Satellite	1.6 (max.)	0.1 (min.)	0.3	0.2	0.1 (min.)	0.6	NA
Other	2.5	2.3	2.7	2.7	1.8 (min.)	4.8 (max.)	NA
No overlapping repeat	63.1	66.7	73.3 (max.)	61.4	54.9	47.0 (min.)	NA

Figure 2

A



B

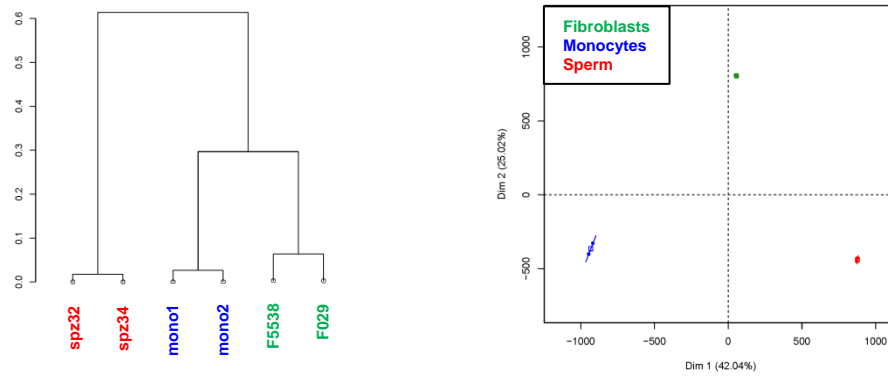


Figure 3

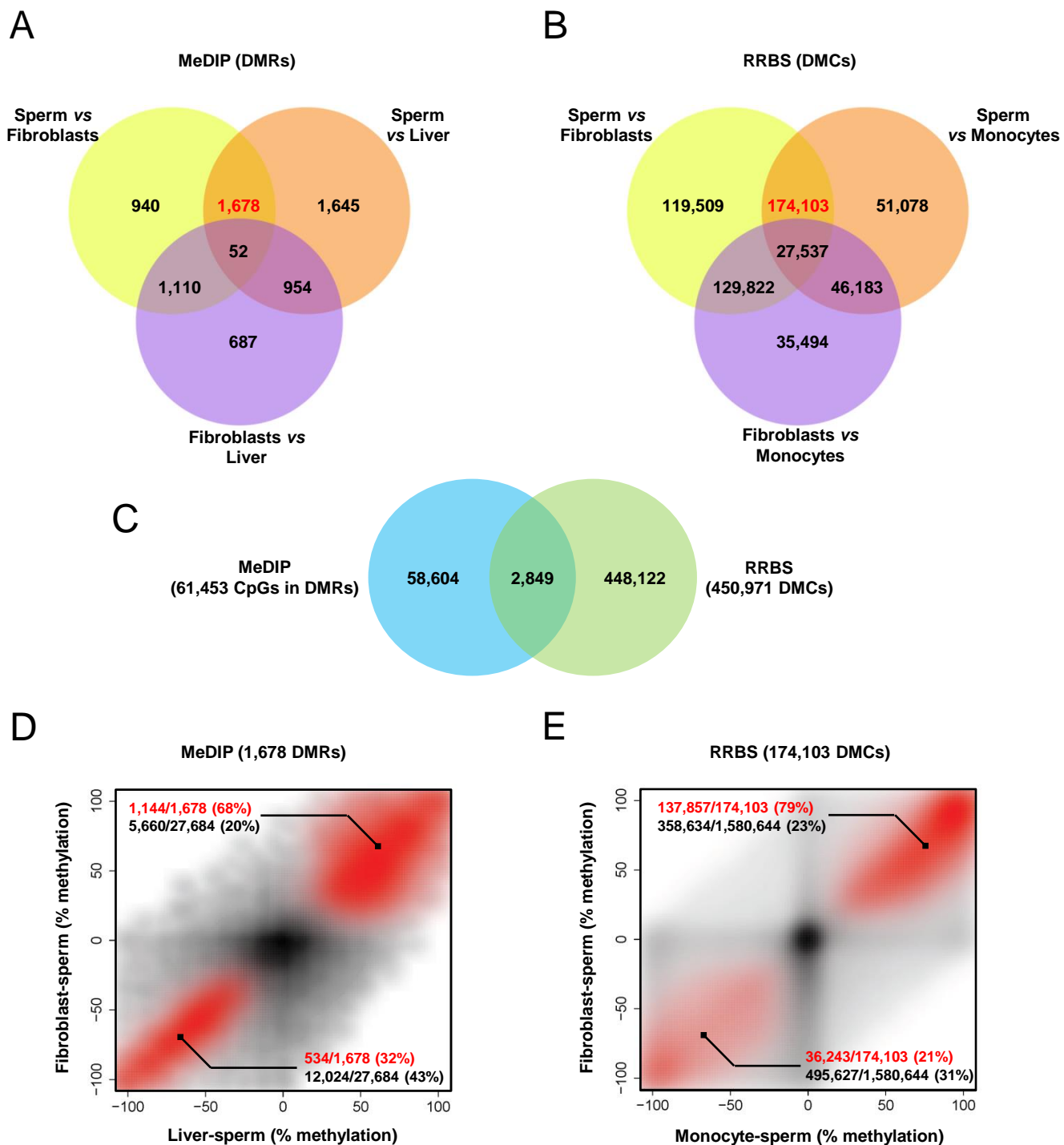
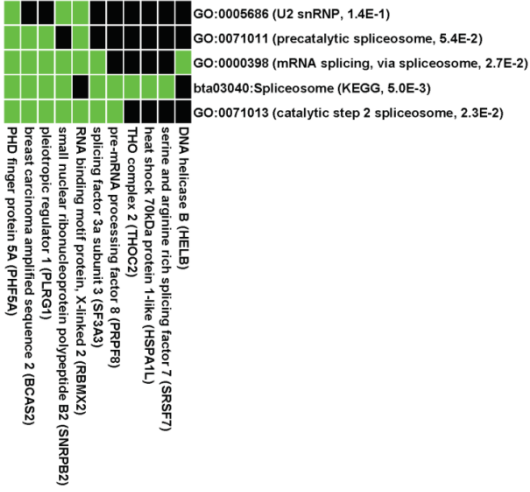


Figure 4

A



B

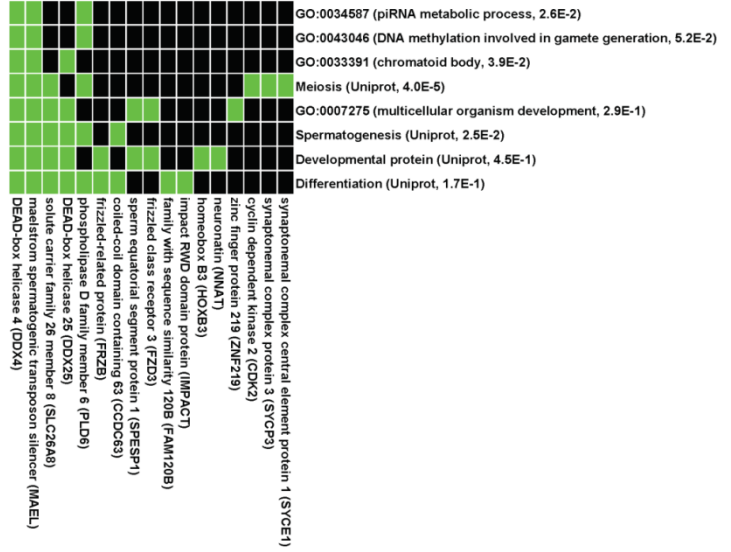
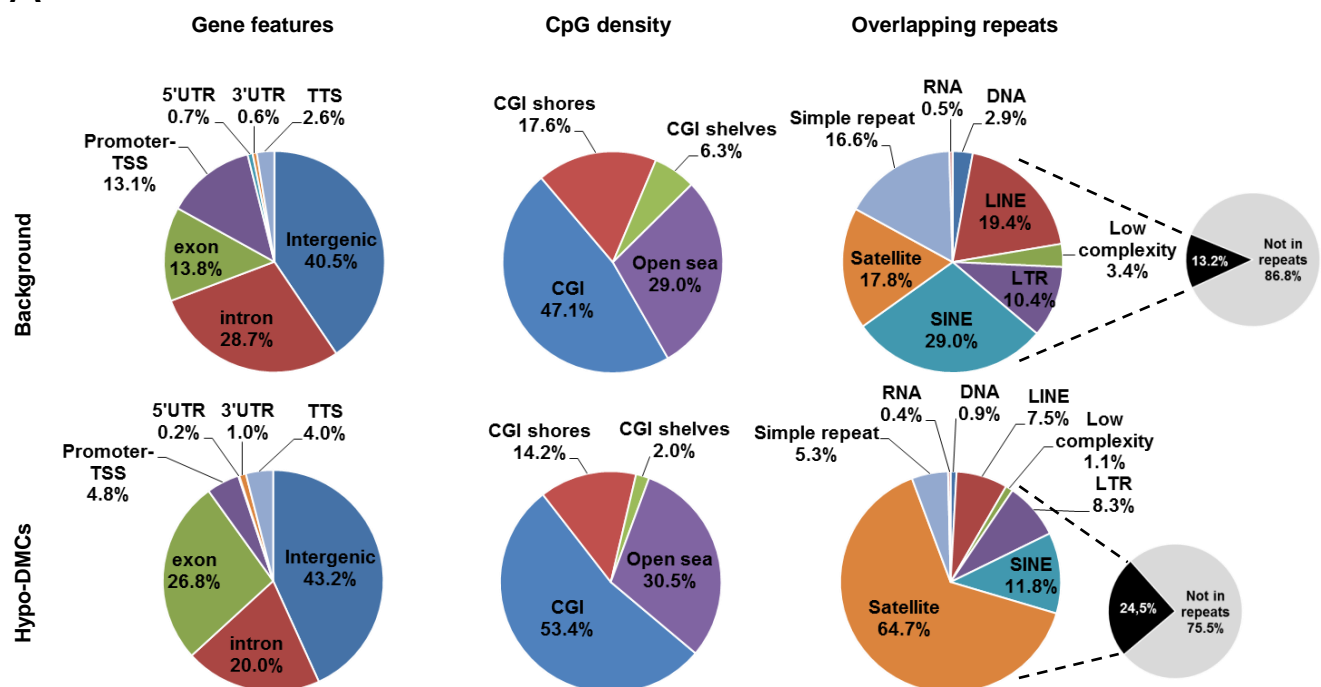
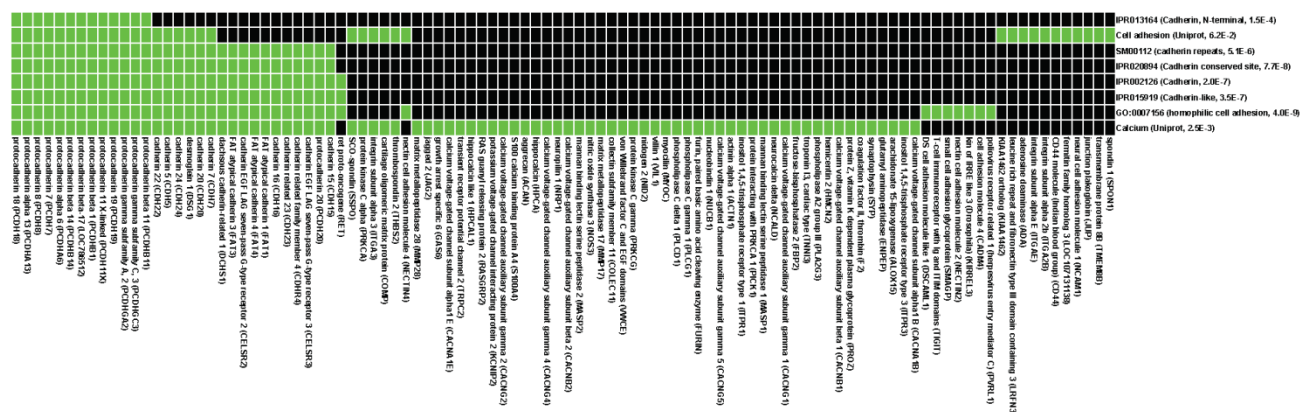


Figure 5

A



B



C

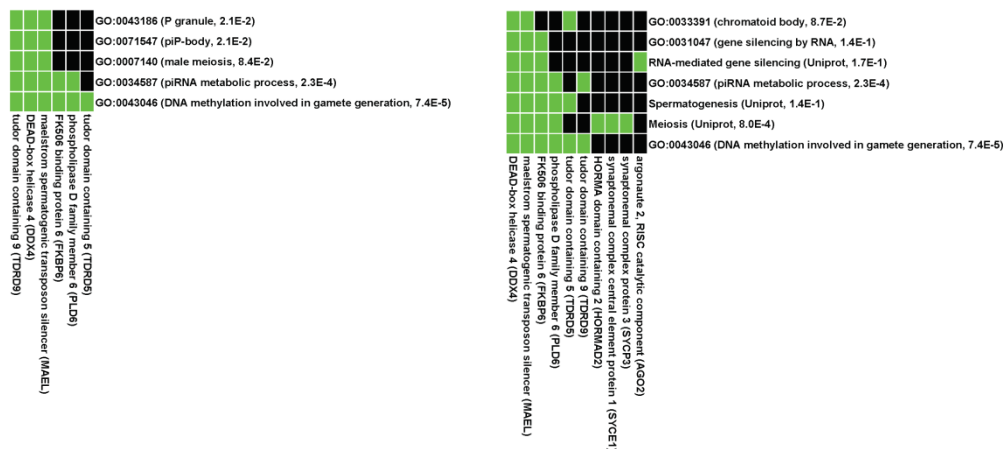
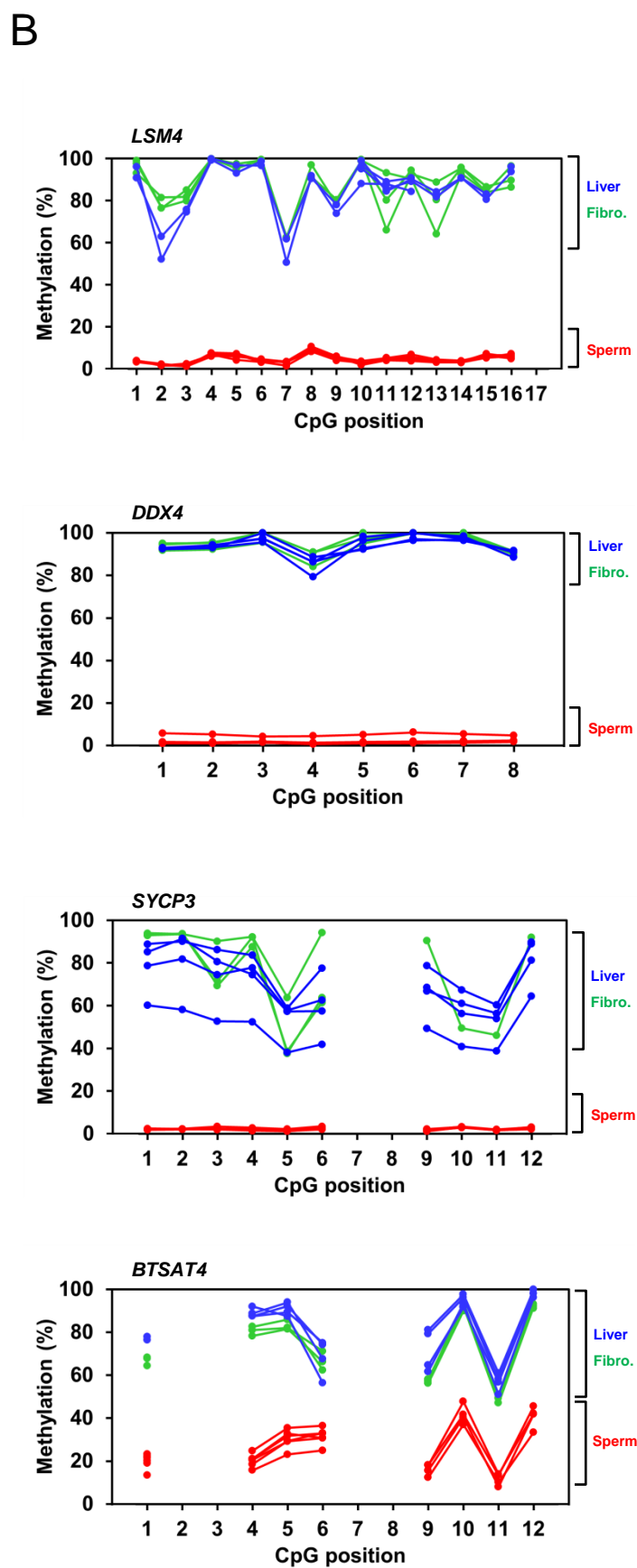
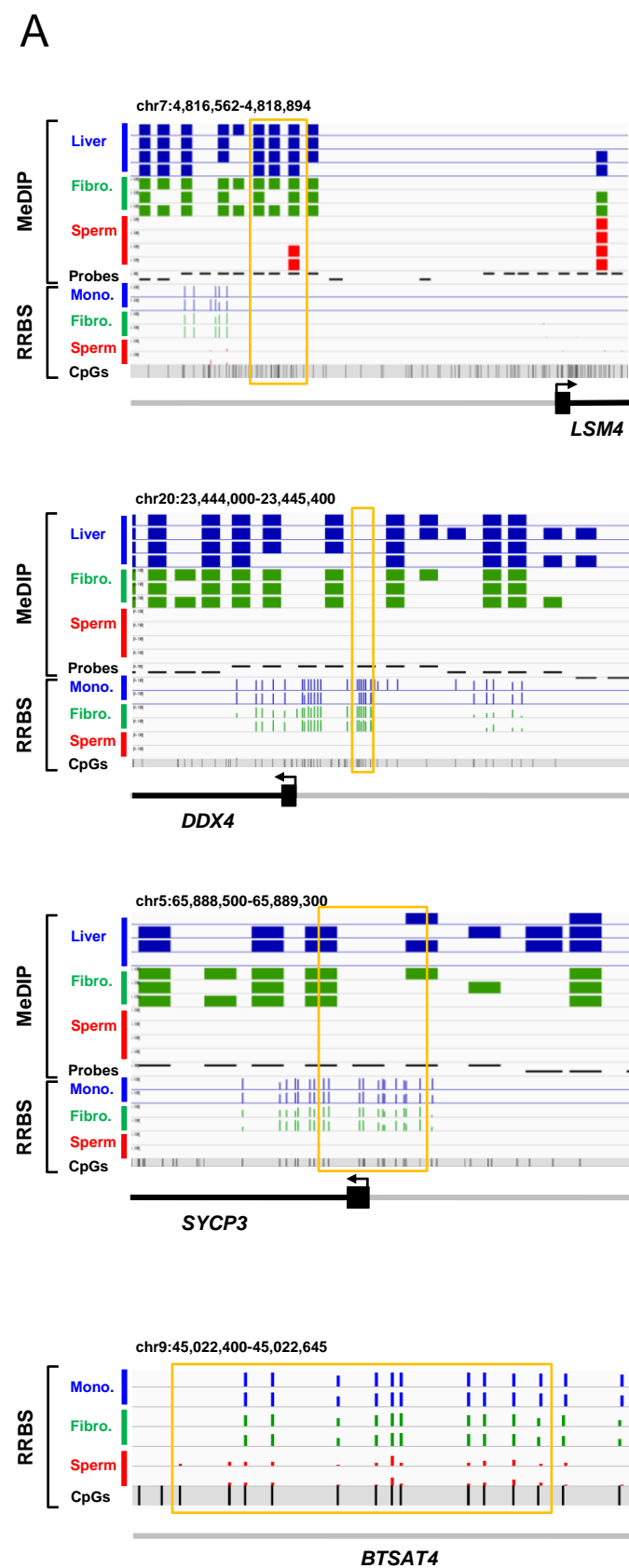


Figure 6



Identification de nouveaux biomarqueurs de la fertilité mâle dans le méthylome de la semence bovine : une analyse par RRBS

Article en cours de préparation

Introduction

L'amélioration génétique des ruminants exploite principalement la diffusion de la semence des mâles comme vecteur du progrès génétique. Pendant de nombreuses années, les mâles reproducteurs ont ainsi été sélectionnés sur les performances de leur descendance, selon une procédure de « testage » d'une durée de 3 à 5 ans. La sélection génomique (Meuwissen, Hayes, et Goddard 2001) permet désormais de connaître la valeur génétique d'un individu par l'analyse de son génome. Les reproducteurs mâles de haute valeur génétique ainsi sélectionnés produisent, dès leur maturité sexuelle, de la semence utilisée à grande échelle par les éleveurs pour inséminer leurs vaches. D'éventuels défauts qualitatifs impactant la fécondance de cette semence et la fertilité des taureaux ne sont donc détectés qu'*a posteriori*.

Au cours de la production de semence des contrôles qualité sont toutefois effectués en routine. Ces contrôles ont pour objectif de s'assurer de la bonne conformité de la semence avant son utilisation. Les contrôles sont basés sur des critères phénotypiques, comme le nombre de spermatozoïdes, leur concentration ou leur motilité. Néanmoins, la semence de certains taureaux satisfaisant l'ensemble des critères phénotypiques s'avère parfois peu fertile (Sellem et al. 2015), conduisant à des taux de conception en IA très inférieurs à ceux attendus. Les conséquences de l'utilisation de taureaux sub-fertiles en IA sont nombreuses, que ce soit pour les entreprises de sélection (ES) qui élèvent ces taureaux, les entreprises de mise en place qui conduisent les IA ou encore les éleveurs (achat de la paille, retards dans les gestations, désynchronisation des lactations, etc.).

De nombreuses études ont été menées pour identifier des marqueurs génétiques de la fertilité mâle (Fritz et al. 2013) (Hering, Olenski, et Kaminski 2014) (Hering et al. 2014) (Fortes et al. 2013). Un grand nombre de marqueurs ont été identifiés, notamment des polymorphismes nucléotidiques (*Single Nucleotide Polymorphism*, SNP) létaux au stade embryonnaire ou fœtal. D'autres facteurs entrent indubitablement en jeu, qui ne sont pas encore identifiés. Il est possible que certains de ces facteurs soient épigénétiques.

L'épigénétique est définie comme « l'étude des processus moléculaires impliqués dans des changements de l'expression des gènes, héréditaires à travers la mitose voire la méiose, sans modification de la séquence d'ADN » (Bird 2002). Les principales marques épigénétiques décrites sont les modifications post-traductionnelles des histones, l'intervention des ARN non-codants, et la méthylation de l'ADN. Cette dernière consiste en l'addition d'un groupement méthyle sur le carbone 5 des cytosines lorsqu'elles sont suivies d'une guanine

(CpG). Au cours de la spermatogenèse et du développement embryonnaire, on assiste à de profondes restructurations des marques épigénétiques, phénomène appelé reprogrammation épigénétique. Des altérations durant ce processus de reprogrammation peuvent conduire à des altérations du fonctionnement des spermatozoïdes, ainsi que du développement embryonnaire (Jenkins et Carrell 2012) (Dada et al. 2012).

A l'initiation de ce travail, nous disposions de peu d'informations concernant le méthylome du spermatozoïde bovin. Néanmoins, une étude par MeDIP-chip a montré que 151 gènes sont différenciellement méthylés entre les spermatozoïdes de 3 buffles fertiles et 3 buffles sub-fertiles (Verma et al. 2014). Parmi ces gènes, certains sont impliqués dans le développement des cellules germinales, la spermatogenèse ou le développement embryonnaire. En 2017, Kropp a identifié par MBD-seq 76 régions différenciellement méthylées dans les spermatozoïdes de 2 groupes de taureaux, classés selon leur taux de conception (très bon ou très mauvais) (Kropp et al. 2017). Parmi ces régions, 25 sont situées dans des gènes et 20 d'entre elles sont hypométhylées chez les taureaux sub-fertiles. De nombreuses autres études ont montré des résultats similaires sur plusieurs gènes candidats, comme *SPEF2*, *IGF2-H19*, ou *Boule* (Guo et al. 2013) (Jena et al. 2014) (Yao et al. 2015).

Il n'est observé aucun gène commun entre ces différentes études. Cette observation pourrait être le reflet d'utilisation de méthodes d'analyse pangénomiques différentes, ne renseignant pas les mêmes régions génomiques. Afin de qualifier au mieux la semence, notamment par l'identification de nouveaux marqueurs moléculaires de qualité, il était nécessaire de mettre en place un dispositif expérimental basé sur un nombre important de taureaux, classés en fonction de leur fertilité et des caractéristiques de leur semence. Notre objectif est ici de générer un catalogue de biomarqueurs en utilisant la méthylation de l'ADN. Cette étude a été réalisée sur 58 taureaux issus des 2 principales races laitières françaises, les races Holstein et Montbéliarde. Les taureaux de chacune des races sont répartis en 2 catégories en fonction de leur fertilité. Un troisième groupe est constitué de taureaux prédits comme fertiles sur la base de leur profil génétique, mais qui se sont révélés sub-fertiles sur le terrain. L'identification des biomarqueurs de méthylation a été effectuée par une approche pangénomique en utilisant le *Reduced Representation Bisulfite Sequencing* (RRBS), dont les résultats ont été validés par bisulfite-pyroséquençage.

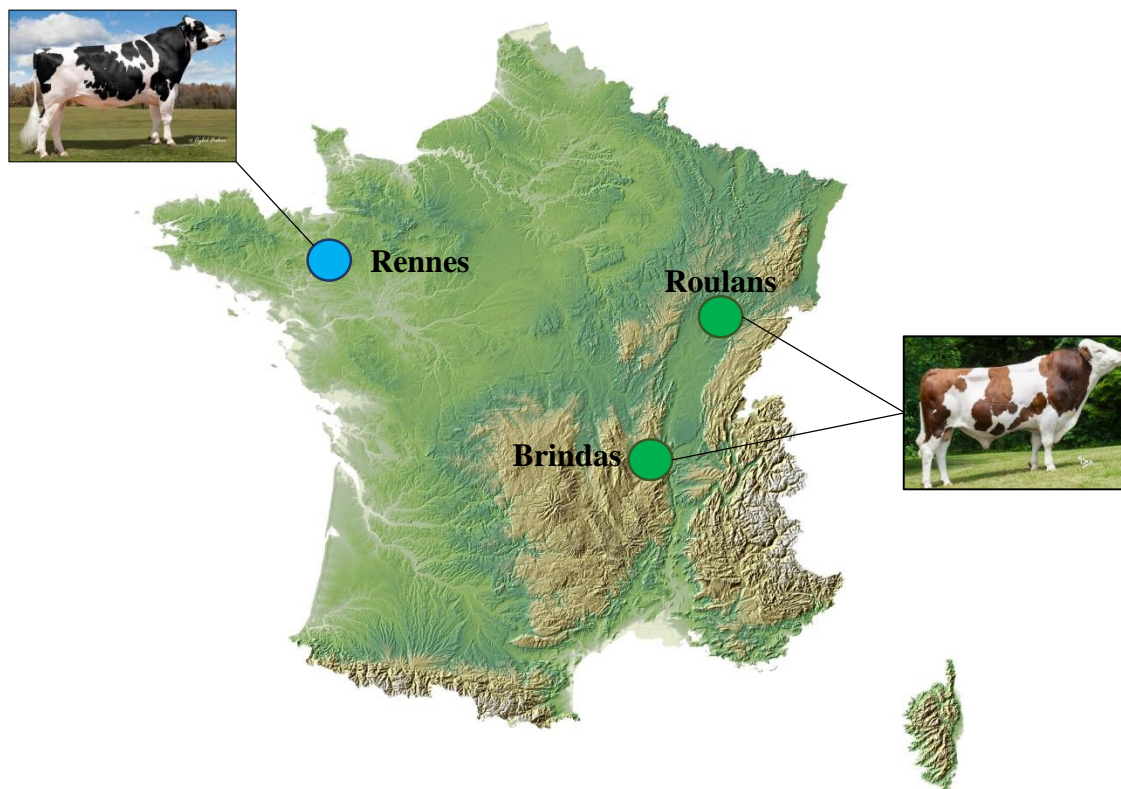


Figure 1. Dispositif expérimental

Origine des taureaux en fonction de leur race. Tous les taureaux de race Holstein proviennent du centre de production de semence de l'ES Evolution, située à Rennes (Ille-et-Vilaine). Les taureaux de race Montbéliard proviennent de 2 centres de production, soit du centre situé à Roulans (Doubs), soit du centre situé à Brindas (Rhône), tous 2 sous contrôle de l'ES Umotest.

Matériels et méthodes

A. Dispositif expérimental et collecte des échantillons

Le dispositif expérimental est composé de 58 taureaux de races Holstein (n=28) et Montbéliard (n=30). Les taureaux Holstein proviennent du centre de production de semence de l'ES Evolution situé à Saint-Aubin. Les Montbéliard proviennent des centres de production de l'ES Umotest situés à Brindas ou à Roulans (Figure 1).

Les taureaux et les éjaculats ont été sélectionnés à partir d'une large population, sur la base des taux de non-retour en chaleur (TNR) de chaque éjaculat, de façon à constituer des groupes de fertilité contrastée. Pour la race Holstein le TNR disponible est le TNR 90 jours (TNR₉₀). Pour la race Montbéliarde, nous disposons également des TNR à 24 et 56 jours (TNR₂₄ et TNR₅₆). Pour chacune des 2 races, une 3^{ème} catégorie de fertilité « Déviants » a pu être constituée en mettant en évidence des taureaux pour lesquels la prédiction génétique de fertilité n'était pas en cohérence avec les données de terrain. En effet, un prédicteur génomique de la fertilité mâle a été conçu pour ce projet (Sébastien Fritz, Alice). Il est basé sur l'utilisation de l'indicateur de fertilité taureau (ensemble des TNR₂₈₂ des différents éjaculats d'un même mâle, corrigé de tous les effets environnementaux : conduite d'élevage, jour/mois de l'IA, parité...), et des données de génotypage des taureaux en question. Ce modèle de prédiction a été construit en utilisant un ensemble de mâles nés avant 2010 et a été appliqué aux mâles nés après 2010. Si la prédiction génomique et l'indicateur de fertilité d'un mâle diffèrent de plus de 3 écarts-types, le taureau est considéré comme déviant. Dans la cohorte étudiée, tous les taureaux Déviants identifiés avaient une prédiction génomique élevée, mais un indicateur de fertilité faible ($IF_{[Holstein-Déviants]} = -11,4$; $IF_{[Holstein-Fertiles+]} = 2,79$; $IF_{[Holstein-Fertiles-]} = 2,68$; $IF_{[Montbéliard-Déviants]} = -12,7$; $IF_{[Montbéliard -Fertiles+]} = 4,53$; $IF_{[Montbéliard -Fertiles-]} = 2,28$; Figure 2.B, page 165).

Toutes les analyses de méthylation ont été réalisées à partir d'une paillette de semence produite à partir d'un éjaculat collecté selon les méthodes standards, et conservée dans de l'azote liquide avant utilisation. Pour chacun des taureaux, lors de la décongélation de la paillette servant pour la préparation d'ADN génomique, nous avons aussi effectué des mesures d'analyses fonctionnelles au CASA (mesure de la concentration spermatique, motilité, potentiel mitochondrial et intégrité membranaire ; Annexe 1).

L'ensemble des individus a été génotypé sur une puce BovineSNP50 v2 BeadChip (Illumina). Au vu des données de Fritz (Fritz et al. 2013), nous avons recherché si certains individus étaient porteurs des haplotypes délétères pour le développement embryonnaire. Les individus porteurs ont été éliminés du dispositif. Une seconde analyse des données concernant ces taureaux (IDELE) a révélé qu'un taureau Holstein Fertile+ est porteur de *Complex Vertebral Malformation* (CVM, chromosome 3, gène *SLC35A3*), un taureau Déviant porteur de *Brachyspina* (chromosome 21, gène *FANCI*), et un taureau Montbéliard Déviant est quant à lui porteur du Syndrome d'Hypoplasie Généralisé Capréoliforme (SHGC, chromosome 13, gène *CEP250*).

B. Extraction d'ADN génomique

La décongélation des paillettes est effectuée au bain-marie à 37°C - 30 secondes. A cette étape, 2µl du contenu de la paillette sont prélevés et sont utilisés pour réaliser les analyses de la motilité au CASA. Le contenu de la paillette est dilué dans 1ml de PBS 1X et centrifugé à 2400g - 5 minutes. Le surnageant retiré, le culot est repris dans 200µl PBS 1X. A cette étape, 20µl utilisés pour réaliser les analyses morphologiques au CASA et la mesure de la concentration spermatique par cytométrie en flux. La contamination de l'éjaculat par des cellules somatiques a aussi été systématiquement contrôlée. Nous n'avons observé aucune contamination importante pour l'ensemble des individus (<5%). Les 180µl restants sont lavés une seconde fois avec 820µl PBS 1X pour effectuer une seconde centrifugation à 2400g - 5 minutes. Le surnageant retiré, le culot est repris dans 187µl de tampon de lyse (10mM Tris HCl pH 7,5, 25mM EDTA, 1% SDS, 75mM NaCl, qsp H₂O). Environ 20 millions de spermatozoïdes, dans un volume de 187 µl sont utilisés pour poursuivre l'extraction, en présence de 50mM de DTT, 200µg/ml de protéinase K, et 0,5µg de glycogène. Le mélange réactionnel est incubé à 55°C sur la nuit. Le lysat est ensuite extrait en utilisant du phénol puis du phénol-chloroforme (1 :1, v/v) en utilisant des tubes Phase Lock Gel (5Prime), puis précipité avec de l'éthanol 100% à -20°C sur la nuit. Après centrifugation (15000g, 4°C, 20 minutes), le culot d'ADN est rincé à l'éthanol 70%, puis séché et repris dans 200µl de tampon TE (10mM Tris HCl pH7,5, 2mM EDTA). La mesure de la quantité d'ADN est faite par un fluoromètre Qubit 2.0 (Invitrogen), et la qualité de l'ADN contrôlée par électrophorèse sur un gel d'agarose à 0,5%. L'ADNg est conservé à -20°C avant son utilisation.

C. Analyse de la méthylation de l'ADN par RRBS

Les analyses pangénomiques de la méthylation de l'ADN ont été réalisées en utilisant la méthode du *Reduced Representation Bisulfite Sequencing* (RRBS), en suivant la méthode précédemment décrite (Meissner 2005) (Gu et al. 2010) (Auclair et al. 2014). L'ensemble du protocole de préparation des banques sur automate, le séquençage des banques ainsi que leur traitement bio-informatique sont également détaillés dans la partie Méthodologie de ce document. Brièvement, 200ng d'ADN génomique est digéré en utilisant l'enzyme de restriction MspI (Thermo Scientific, site de restriction C[^]CGG). Après les étapes de « *end-repair* », de « *A-tailing* » et de ligation des adaptateurs de séquençage Illumina (55pb), les fragments d'ADN sont sélectionnés en fonction de la taille (comprise entre 150 et 400 pb, soit de 40 à 290 + les adaptateurs). L'étape de sélection de taille est réalisée en utilisant des billes magnétiques SPRIselect (Beckman Coulter), en suivant le protocole de double sélection de taille décrit par le fabricant. L'ADN est ensuite converti 2 fois au bisulfite de sodium avec le *EpiTect bisulfite kit* (Qiagen), en suivant les instructions du fabricant (*protocol for DNA extracted from FFPE tissues*). L'ADN converti est amplifié par la Pfu Turbo Cx hotstart DNA polymerase (Agilent) avec 14 cycles de PCR. Les banques sont purifiées avec des billes magnétiques *Agencourt Ampure beads* (Beckman Coulter), et séquencées avec un séquenceur Illumina HiSeq 4000 pour produire des séquences de 75 paires de bases en *paired-end* (Integrage, France).

L'enchaînement des étapes d'analyse bio-informatique des données, ainsi que les paramètres de contrôle qualité et de *trimming* des séquences sont identiques à ce qui a été décrit dans la section Méthodologie. La version du programme de *trimming* des données est TrimGalore.0.4.0 avec Cutadapt.1.8.3. En revanche, concernant leur alignement sur le génome de référence avec le programme Bismark v0.14.3 (Krueger et Andrews 2011), une particularité de notre approche est d'utiliser un génome composite comprenant l'assemblage bovin UMD 3.1, auquel a été ajouté le chromosome Y provenant du génome bovin Btau5.0.1.

Seuls les CpG couverts par plus de 10 séquences alignées de façon unique sur le génome ont été conservés pour les analyses (appelés CpG₁₀). Nous avons aussi décidé de nous intéresser en parallèle aux CpG dont la couverture est supérieure à 500 (appelés CpG₅₀₀). Ces CpG sont souvent retirés des analyses car ils sont potentiellement dus à des biais d'amplification par PCR. Il n'est cependant pas possible d'écarter le fait qu'ils puissent aussi correspondre à des séquences répétées. En effet, lors du contrôle qualité des banques, sur gel d'acrylamide, nous

avons toujours observé des bandes indiquant la présence de séquences de même taille en quantité importante. Il était donc important de suivre leur profil de méthylation.

Pour chaque échantillon, un pourcentage de méthylation globale pour les CpG₁₀ et pour les CpG₅₀₀ sont calculés à partir des résultats fournis par le programme Bismark. Des analyses en Composantes Principales (ACP) et en classification hiérarchique de corrélation de Pearson ont été réalisées en utilisant le package R FactoMineR, en utilisant les données de méthylation des CpG₁₀. Ont été conservés les CpG₁₀ présents pour au moins la moitié des individus de la catégorie comportant le moins d'individus. Par exemple, lors de la comparaison Fertiles+/Fertiles- chez les Holstein, les CpG sélectionnés présentaient une couverture de 10 pour au moins 4 individus de chaque groupe. Les analyses ont été effectuées sans échantillonnage des données, et en utilisant des données centrées-réduites.

Pour chaque comparaison 2 à 2 entre catégories de fertilité, les cytosines différentiellement méthylées (*Differentially Methylated Cytosine*, DMC) ont été identifiées avec le logiciel methylkit (Akalin et al. 2012) et en utilisant les paramètres suivants : une couverture minimale 10 CpG pour au moins 4 échantillons par catégorie, avec un différentiel de méthylation supérieur à 25% et une pvalue ajustée (qvalue) de 0,01. Une région génomique est considérée comme différentiellement méthylée (*Differentially Methylated Region*, DMR) lorsqu'elle présente une taille d'au moins 100 pb et contient un minimum de 3 DMC.

Une analyse en *Partial Least Squares - Discriminant Analysis* (PLS-DA) a été effectuée sur les DMC. Elle permet d'identifier les dimensions qui expliquent le mieux leur distribution. Le modèle établi à partir du jeu de donnée des DMC est ensuite évalué pour 2 indicateurs de qualité, le R²Y (indicateur de l'adéquation entre le modèle et le jeu de donnée) et le Q²Y (valeur prédictive du modèle). Les analyses en PLS-DA ont été effectuées en utilisant le package ropls du logiciel R, en utilisant 100 permutations aléatoires entre les groupes.

Les DMC ont également été annotées en fonction de leur association aux différents éléments génomiques, à la densité en CpG et aux éléments répétés du génome. Les critères d'annotation suivants ont été utilisés : les promoteurs-TSS sont situés de -2000 à +100 pb par rapport au site d'initiation de la transcription (*Transcription Start Site*, TSS) ; les sites de terminaison de la transcription (*Transcription Termination Site*, TTS) à ± 100 pb autour du TTS ; les CpG *shores* encadrent les îlots CpG (CGI) à ± 2000 pb ; les CpG *shelves* encadrent les CpG *shores* à ± 2000 pb. Une DMR est considérée comme participant au CGI/CpG *shore*/CpG *shelve*, si elle se superpose à hauteur d'au moins 75%. Une DMR est considérée comme comportant un élément répété dès lors qu'il y a un chevauchement. Lorsque les

DMC/DMR ne sont pas directement situées dans un gène, on considère qu'elles peuvent encore lui être associées jusqu'à une distance de 10kb.

Pour les analyses d'enrichissement, nous avons utilisé le programme *DAVID Functional Annotation Clustering (Database for Annotation, Visualization and Integrated Discovery)*, (Huang, Sherman, et Lempicki 2008) (Huang, Sherman, et Lempicki 2009). Ce programme, en plus d'identifier des fonctions biologiques significativement enrichies en DMC à partir d'un jeu de données, permet aussi de regrouper en domaines plus larges les fonctions biologiques qui sont en relation entre elles. L'analyse d'enrichissement a été effectuée selon les termes de Gene Ontology (GO) : composants cellulaires (*cellular components*), fonctions moléculaires (*molecular functions*) et processus biologiques (*biological processes*), en utilisant les catégories FAT (dont les termes GO très larges ont été filtrés). La note d'enrichissement des groupements de termes GO correspond à la moyenne géométrique de toutes les p-values (*EASE score*; (Hosack et al. 2003)) de tous les termes GO associés au groupe. Une note supérieure à 1,3 signifie que le regroupement associé est significativement enrichi en DMC. Pour chaque groupement de terme GO est donnée une p-value corrigée par Benjamini-Hocheberg. Cette p-value corrigée correspond à celle du terme GO le plus centré autour de 0,05. Enfin, pour chaque comparaison les enrichissements sont effectués par rapport à un *background*, c'est-à-dire l'ensemble des gènes couverts par au moins un CpG, analysés pour les échantillons inclus dans la comparaison.

D. Evaluation des apports de la double digestion enzymatique dans la description du méthylome

Des banques RRBS ont été préparées en utilisant un couple d'enzymes de restriction. En effet, après les analyses de premières banques de RRBS, nous avons été surpris par le fort taux d'alignements multiples des séquences sur le génome de référence (51,1% en moyenne). L'analyse de la section *overrepresented sequences* des données de FastQC, ainsi que la présence de fortes bandes d'amplifications sur les gels d'électrophorèse ont révélé que ce phénomène était dû à la présence en nombre important de séquences répétées dans la fenêtre de sélection de taille. Ce problème a déjà été rencontré par des équipes de recherche travaillant sur l'espèce ovine pour la construction des banques RRBS (Couldrey et al. 2014). Nous avons alors tenté une approche par double digestion enzymatique, une approche couramment employée en RRBS (Wang et al. 2013) (Martinez-Arguelles, Lee, et Papadopoulos 2014). Il a fallu tout d'abord sélectionner un couple d'enzymes et une fenêtre

de sélection de taille, permettant de cliver d'avantage les séquences répétées suramplifiées pour ainsi les éliminer de la fenêtre de sélection. Les enzymes qui ont été évaluées ont été sélectionnées sur la base des critères suivants :

- L'enzyme doit être non-méthylsensible
- Elle permet l'analyse du meilleur ratio CpG analysés / génome analysé. En d'autres termes, après utilisation de l'enzyme et sélection de taille des fragments, il est intéressant que les fragments générés soient le plus enrichis possible en CpG. Ainsi, on peut couvrir le maximum de CpG possible tout en réduisant au maximum la profondeur de séquençage requise pour pouvoir ensuite les analyser
- Elle permet dans la mesure du possible, l'analyse de CpG différents de ceux analysés par MspI seul, pour optimiser la couverture en CpG du génome
- Pour faciliter la préparation des banques et l'analyse bio-informatique des données, l'enzyme n'implique peu/pas de mises au point par rapport au protocole utilisé avec MspI

Cette analyse a dans un premier temps été réalisée *in silico*. Parmi les 52 couples d'enzymes et les 10 fenêtres de sélection de taille testées par couple (dont les résultats sont présentés en Annexe n°2), notre choix s'est porté sur les enzymes DraI (TTT[^]AAA), TaqI (T[^]CGA), XmnI (GAANN[^]NNTTC) et HaeIII (GG[^]CC), toutes conjointement avec l'utilisation de MspI. Après avoir préparé et séquencé des banques avec ces couples d'enzymes, il s'est avéré que celui donnant les meilleurs résultats en termes de taux d'alignement est HaeIII+MspI, en utilisant la fenêtre de sélection de taille usuelle (40 à 290pb).

La construction des banques de double digestion a été réalisée pour 15 des taureaux de race Holstein de notre dispositif (5 de chaque catégorie de fertilité). Les taureaux ont été sélectionnés pour être les plus représentatifs de leur catégorie de leur fertilité, selon les critères ci-dessous :

- Ils présentent une différence en méthylation globale la plus importante possible entre les DMC Fertiles + et Déviants
- Les taureaux sont homogènes en analyse descriptive (ACP et classification hiérarchique), ainsi qu'en analyse discriminante (voir ci-après pour le détail concernant les analyses descriptives et discriminantes). Cela signifie qu'ils sont les plus représentatifs de leur catégorie de fertilité respective
- Les taureaux présentent une bonne corrélation linéaire entre leur taux de méthylation globale et leurs données de fertilité

- Le nombre de paillettes de semence encore disponibles est suffisant

Contrairement au séquençage des banques MspI, dont la profondeur de séquençage choisie est de 25 M ($\pm 15\%$) de séquences, pour les doubles digestions la profondeur de séquençage choisie est de 40 M ($\pm 15\%$) pour optimiser la couverture des CpG. En effet, il s'est avéré que le nombre CpG à couvrir avec la double digestion était plus important qu'en simple digestion (voir Figure 19). Pour conserver une couverture par CpG en adéquation avec celle des banques préparées par simple digestion, il a fallu augmenter la profondeur de séquençage pour ces banques.

E. Validations de DMR par pyroséquençage

Pour réaliser le pyroséquençage, 1 μ g d'ADN est dénaturé à 95°C pendant 5 minutes puis incubé avec 0,2 M de NaOH à 37°C pendant 15 minutes. Une solution de 500 μ l de bisulfite de sodium est préparée (5M de métabisulfite de sodium (Sigma), 130mM d'hydroquinone (Sigma), 0,36M de NaOH) puis ajoutée à l'ADN en solution. L'échantillon est incubé à 55°C pendant 4h dans le noir, sous une couche d'huile minérale pour éviter l'oxydation de la solution. L'échantillon est ensuite purifié en utilisant le kit Wizard DNA Clean-Up System (Promega). L'ADN bisulfité est élué dans 40 μ l d'eau. Pour arrêter la réaction de conversion au bisulfite, NaOH est ajouté à une concentration finale de 0,3 M, et les échantillons sont incubés à 37°C pendant 15 minutes. La solution est neutralisée avec l'ajout de NH₄OAc 3M, et l'ADN est précipité en ajoutant 250 μ l d'éthanol 100% et 1 μ l de glycogène (-20°C, toute la nuit). Après centrifugation, le culot d'ADN est repris dans 20 μ l d'eau et conservé à -20°C jusqu'à utilisation.

Pour les gènes *TLE3*, *SNIP1* et *RARB*, le dessein des amorces de PCR a été réalisé en utilisant MethPrimer (Li et Dahiya 2002). De manière à éviter l'amplification préférentielle d'un sous-ensemble de molécules en fonction de l'état de méthylation du CpG, les amorces ont été sélectionnées pour ne pas chevaucher de CpG. Les amplifications par PCR ont été réalisées à partir de 1 μ l d'ADN bisulfité, dans un volume réactionnel de 50 μ l contenant une concentration finale de 0,2 μ M des amorces sens et antisens, 0,2mM de dNTP, 2mM de MgCl₂, 2U de Platinum Taq Polymerase (Invitrogen), et 1X de tampon de PCR. L'amplification par PCR a été effectuée dans un thermocycleur Eppendorf selon les conditions suivantes : 94°C – 3 minutes ; 50 cycles d'amplification (94°C – 30 secondes ; 60°C – 1 minute ; 72°C – 1 minute) ; 72°C – 10 minutes. Le dessein des amorces de

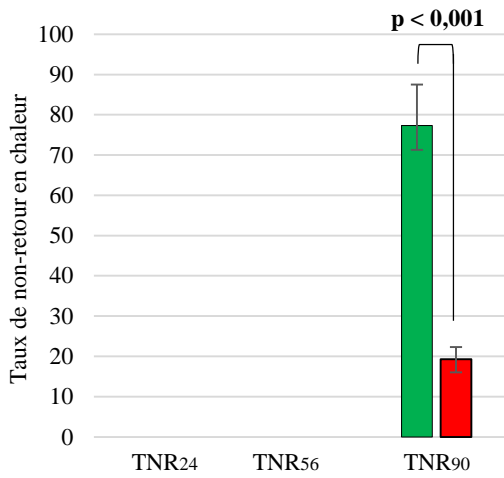
pyroséquençage a été fait manuellement : ce sont des oligomères d'environ 15pb, ne chevauchent pas de CpG et présentant la meilleure diversité en nucléotides possible.

Pour les gènes *WWC2*, *LAMB2* et *SMARCD3*, en raison de difficultés rencontrées pour l'amplification du gène *WWC2* avec le système décrit ci-dessus, nous avons opté pour l'utilisation d'un kit commercialisé par la société Qiagen et adapté au pyroséquençage. Le programme *Pyromark Assay Design* a été utilisé pour sélectionner les amorces de PCR et de pyroséquençage et le *Pyromark PCR kit* a permis de réaliser les amplifications. Les amplifications par PCR ont été réalisées à partir de 1µl d'ADN bisulfité, dans un volume réactionnel de 25µl contenant une concentration finale de 0,2µM des amorces sens et antisens, 1X de *Pyromark PCR Master Mix*, 1X de colorant de migration, et une concentration variable de MgCl₂ en fonction du gène (voir Annexe 3 et 4). L'amplification par PCR a été effectuée dans un thermocycleur Eppendorf selon les conditions suivantes : 95°C – 15 minutes ; 45 cycles d'amplification (94°C – 30 secondes ; température d'hybridation variable en fonction du gène - 1 minute ; 72°C – 1 minute) ; 72°C – 10 minutes. Dans le cas de ces gènes, certaines amorces chevauchent des CpG. Pour éviter une amplification préférentielle en fonction de l'état de méthylation du CpG, un A remplace le C ou le T dans la séquence de l'amorce. Ainsi, le biais est équivalent quel que soit l'état de méthylation du CpG.

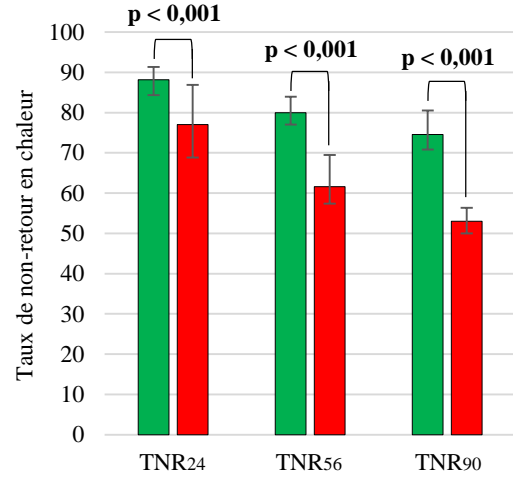
Enfin, quelle que soit la méthode utilisée, les produits de PCR ont été systématiquement séquencés afin de vérifier qu'ils correspondent bien à la séquence attendue, de contrôler s'il n'existe pas de SNP sur les sites de fixation des amorces de pyroséquençage, et de vérifier le rendement de la conversion au bisulfite de sodium (~100%). Une électrophorèse sur gel d'agarose 1,5% a également été réalisée pour chaque produit de PCR pour contrôler la qualité de l'amplification, la quantité de produits de PCR obtenue, ainsi que l'absence d'amplification parasite et/ou de contamination.

F. Intégration des données de génotypage aux données de méthylation

Pour chaque race prise séparément a été effectuée une analyse en *Sparse Partial Least Squares* (SPLS) en mode régression. Cette analyse a été réalisée en utilisant le package R *mixOmics*. Cette méthode permet d'intégrer 2 jeux de données pour en identifier les variables avec la plus forte covariance. Le premier jeu de données concerne les DMC, préalablement identifiées avec les paramètres suivants : une couverture minimale de 10 séquences par CpG pour tous les individus, un différentiel de méthylation de 25% entre les catégories de fertilité et une pvalue de 0,01. Le second jeu de données concerne les SNP issus des données de génotypage des individus. Pour que l'analyse puisse fonctionner avec ce type de données, l'information génétique en un *locus* donné s'est vue attribuée un poids en fonction de sa fréquence allélique dans la population. Le poids attribué est de 0 si l'individu est homozygote pour l'allèle majoritaire, 1 si l'individu est hétérozygote, et 2 si l'individu est homozygote pour l'allèle minoritaire. Les génotypes ayant la même valeur pour tous les individus ont été éliminés. Parmi l'ensemble des relations de covariance calculées, seuls les 2000 DMC et les 2000 SNP présentant la covariance la plus importante ont été conservés pour effectuer d'autres analyses. Les résultats sont alors représentés sur un graphe des individus montrant les 2 dimensions expliquant au mieux la covariance entre les 2 jeux de données pour chaque individu. La contribution des variables (DMC et SNP) pour les 2 premières dimensions sont également représentées sur un graphe des variables.

A

■ Fertiles+ (N=10)
 ■ Fertiles- (N=8)



■ Fertiles+ (N=10)
 ■ Fertiles- (N=10)

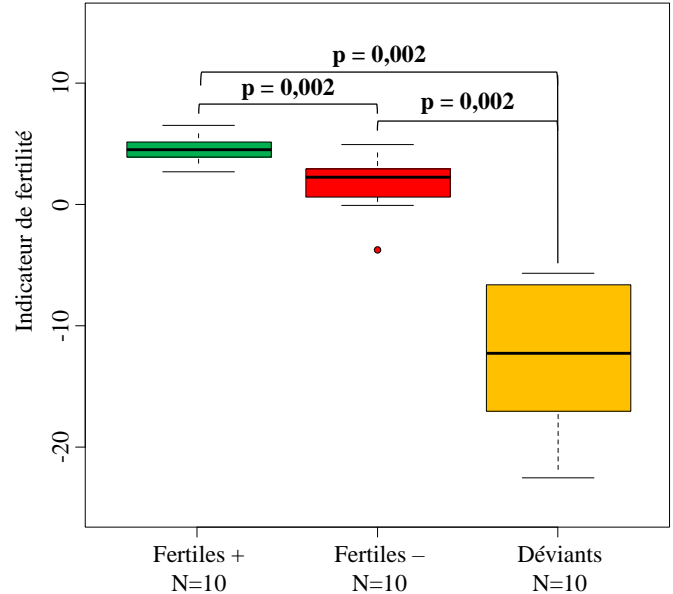
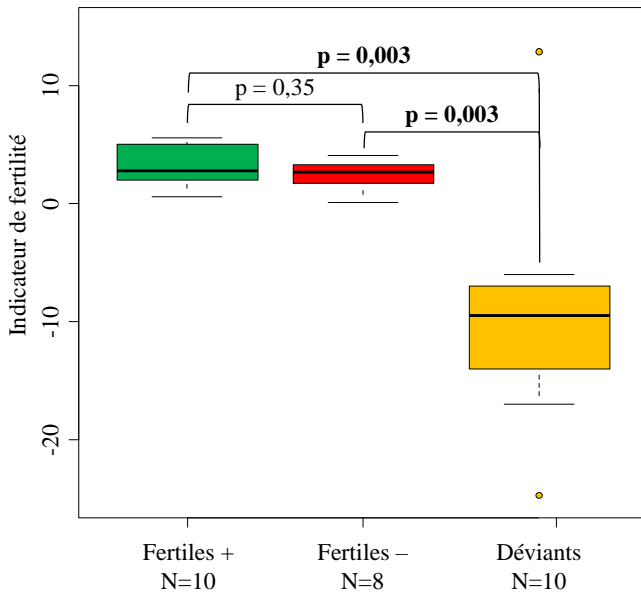
B

Figure 2. Analyses phénotypiques

A. Taux de non-retour (TNR) pour les catégories de Fertiles+ et Fertiles-. Pour les taureaux Montbéliard, nous disposons des données de TNR à 24, 56 et 90 jours. Pour les taureaux Holstein nous ne disposons que des données de TNR à 90 jours. Les TNR à 90 jours sont similaires entre les 2 races pour la catégorie Fertiles+. En revanche, on observe que les taureaux Holstein Fertiles- ont un TNR significativement différent de celui des Montbéliard (médianes respectivement de 19,8% et 53,4%, $p < 0,001$). L'ensemble des analyses statistiques ont été effectuées avec des tests de permutation pour séries indépendantes avec échantillonnage Monte Carlo (100 000 tirages), suivis de comparaisons 2 à 2 (1000 tirages) avec correction de Benjamini-Hochberg pour tests multiples. **B.** Comparaison des indicateurs de fertilité pour les 3 catégories. Pour les taureaux de race Holstein, les Déviants présentent un indicateur de fertilité significativement plus faible (médiane de -11,38) que celui des Fertiles+ et Fertile- (médianes de 2,79 et 2,68, respectivement). Pour les taureaux de race Montbéliard, les indicateurs de fertilité sont significativement différents entre les 3 catégories, avec les Déviants présentant l'indicateur le plus faible (médianes respectivement de -12,27, 4,54 et 2,28 pour les Déviants, Fertiles+ et Fertiles-).

Résultats

A. Constitution des groupes de taureaux en fonction de leurs caractéristiques de fertilité

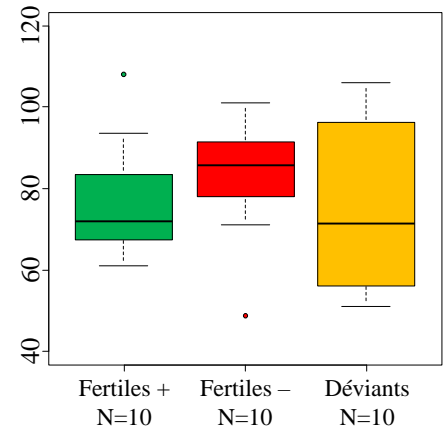
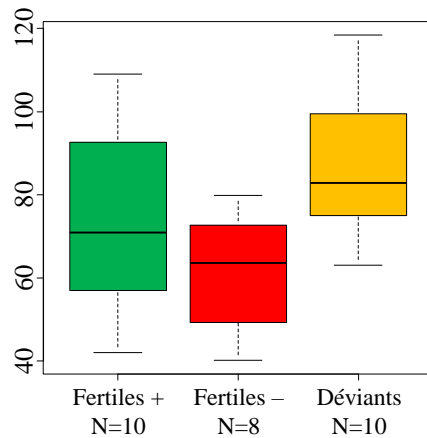
Les mesures de TNR ont été obtenues pour l'ensemble des taureaux afin de caractériser phénotypiquement leur fertilité (Figure 2.A, page précédente). Nous disposons des TNR₂₄, TNR₅₆ et TNR₉₀ pour les taureaux Montbéliard, et des TNR₉₀ pour les Holstein. Les éjaculats retenus pour constituer les groupes présentent des niveaux de TNR contrastés, obtenus sur la base d'un nombre minimum d'IA. Ainsi, en race Holstein les taureaux Fertiles+ présentent une moyenne de TNR₉₀ de $77,3 \pm 5,4\%$, alors que la moyenne des taureaux Fertiles- est de $19,3 \pm 2,6\%$ ($p < 0.0001$; Figure 2.A). Ces TNR ont été évalués en moyenne sur 300 IA réalisées (minimum 120, maximum 567). En race Montbéliarde, les taureaux Fertiles+ présentent une moyenne de TNR₉₀ de $74,5 \pm 3,5\%$ alors que la moyenne des Fertiles- est de $53 \pm 2,4\%$ ($p < 0.0001$). Le nombre d'IA moyen est de 115 (minimum 61, maximum 245).

De façon générale, une diminution progressive des TNR est observée avec le temps de gestation (TNR₂₄ > TNR₅₆ > TNR₉₀), pour les Fertiles+ comme les Fertiles-. La comparaison des TNR₉₀ entre races met en évidence une forte différence de fertilité entre Fertiles- Holstein et Montbéliard (19,3 et 53%, respectivement), alors que les niveaux de fertilité des catégories Fertiles+ Holstein et Montbéliard sont similaires (77,3 et 74,3%, respectivement). Il est à noter que les paillettes utilisées n'ont pas été produites au même âge selon les catégories de taureaux Holstein. En effet, les paillettes des taureaux Holstein Déviants ont été produites dès leur maturité sexuelle (âge en production médian de 15 ± 6 mois), alors que les paillettes des taureaux Fertiles+ et les Fertiles- ont été produites plus tardivement (54 ± 24 et 47 ± 26 mois pour les Fertiles+ et Fertiles-, respectivement). Cet aspect sera traité en détail plus tard dans l'analyse (partie K. page 184). Aucune différence significative d'âges de production des différentes catégories de fertilité n'est observée en race Montbéliarde.

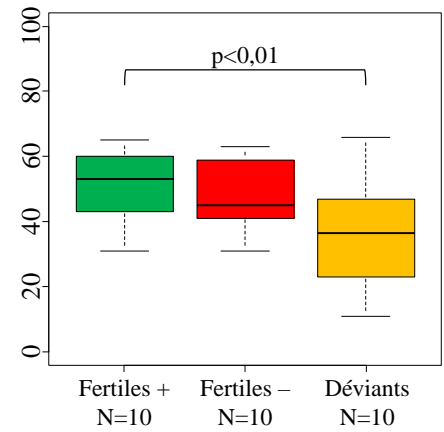
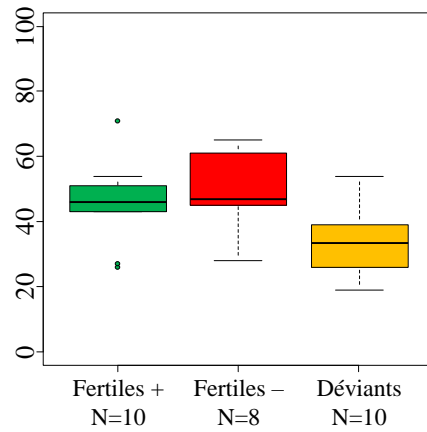
Enfin, pour la quasi-totalité des paramètres spermatiques mesurés par CASA, aucune différence significative n'est observée entre catégories. Ces analyses concernent notamment la concentration en spermatozoïdes, leur type de déplacement, le potentiel mitochondrial ou encore l'oxydation cellulaire (l'ensemble des résultats sont répertoriés en Annexe 1, Figure 3). L'unique différence significative (pvalue inférieure < 0,01) réside dans le pourcentage de spermatozoïdes motiles entre les taureaux Fertiles+ et Déviants pour la race Montbéliarde.



Ⓐ
Concentration
Spermatique



Ⓑ
Motilité (%)



Ⓒ
phmmp

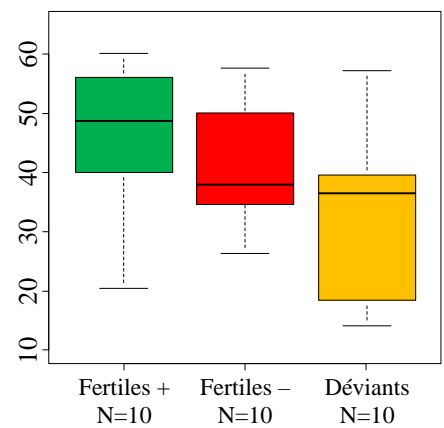
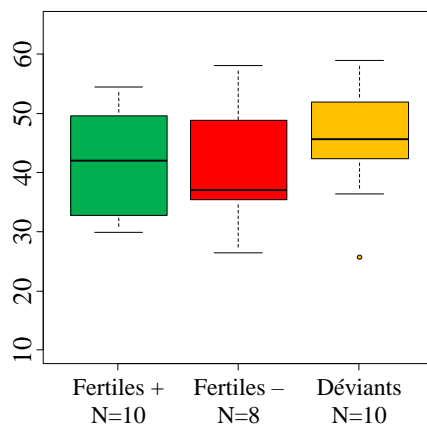


Figure 3. Analyses fonctionnelles

Quelques exemples de résultats d'analyses fonctionnelles sont représentées ici. L'ensemble des données est disponible en Annexe 1. **A.** Analyse de la concentration spermatique en millions de spermatozoïdes par millilitre. **B.** Mesure de la proportion de spermatozoïdes motiles dans l'échantillon. **C.** Mesure du potentiel mitochondrial. Cette mesure est révélatrice de la viabilité cellulaire.

B. Bonne qualité des séquences de RRBS

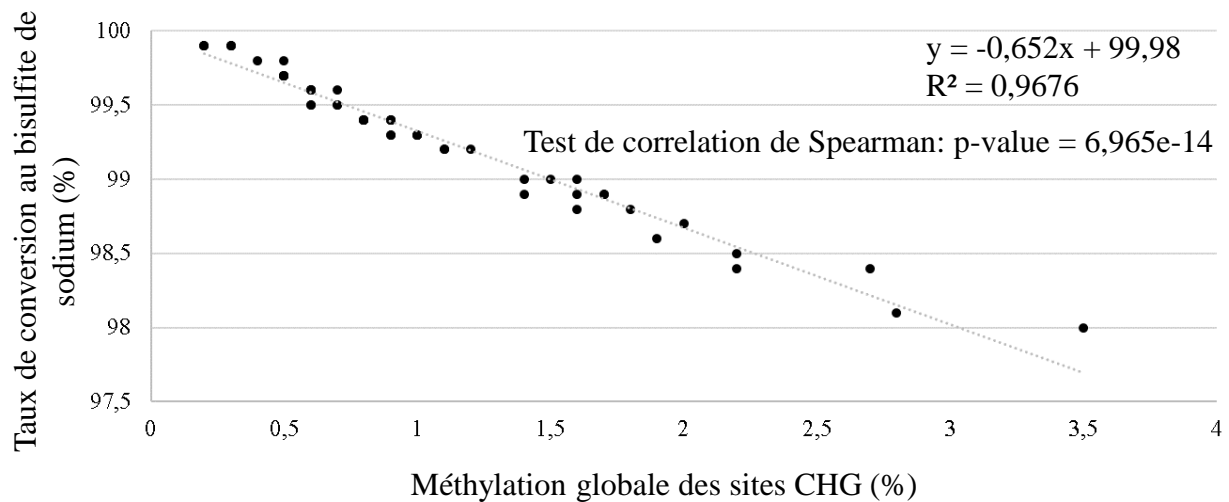
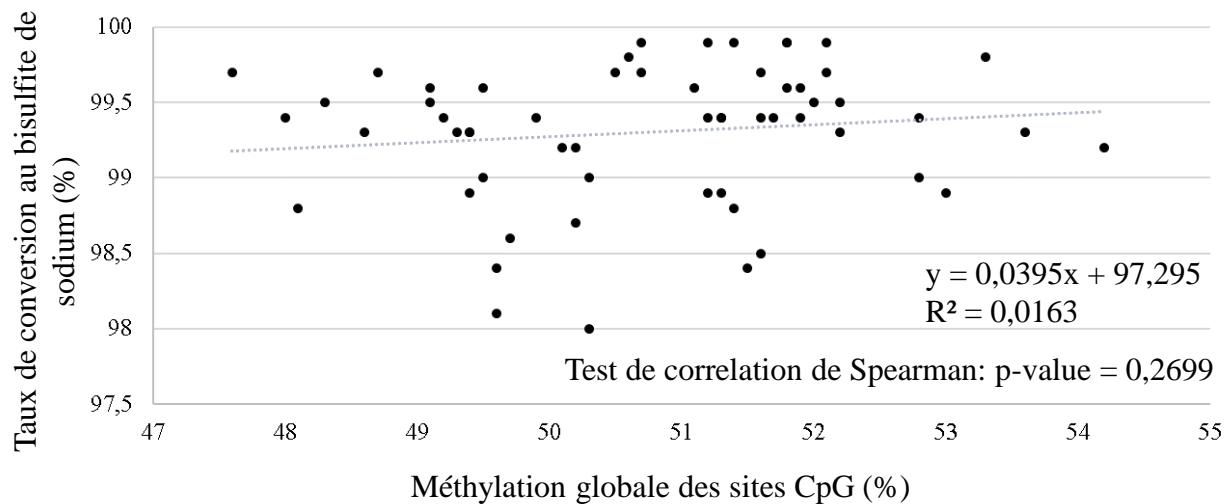
Suite à la réception des données de séquençage des banques RRBS, à leur *trimming* et à leur alignement sur le génome, des contrôles qualité de ces séquences sont systématiquement réalisées (un récapitulatif est donné Figure 4.A, et l'ensemble des données est présenté en Annexe 5). A partir des 58 banques incluses dans cette étude, il peut être observé que la profondeur de séquençage moyen est de $35\,223\,758 \pm 4\,359\,844$. L'alignement sur le génome permet de distinguer 3 types de séquences : celles s'alignant sur un unique *locus* du génome (alignement unique), celles pouvant s'aligner en de multiples *loci* (alignement multiple), et celles ne s'alignant jamais sur le génome (non-alignées). Le taux d'alignement des séquences sur le génome est de $39,3 \pm 2,4\%$ pour les alignements uniques, $51,1 \pm 2,3\%$ pour les alignements multiples et $9,6 \pm 0,3\%$ pour les séquences non-alignées. Les séquences non-alignées sont éliminées, et les séquences alignées de façon multiple sur le génome ne sont pas prises en considération dans nos analyses. Seules les séquences alignées de façon unique sont conservées pour les analyses ultérieures. Le nombre de CpG alignés de façon unique est de $3\,216\,778 \pm 122\,139$, ce qui représente $11,5\%$ des CpG totaux du génome (le génome bovin UMD 3.1 comporte $27\,540\,276$ CpG, et le chromosome Y du génome Btau5.0.1 comporte $341\,967$ CpG, soit un total de $27\,882\,243$ CpG). Il est aussi possible de comparer les données générées avec celles provenant de la digestion *in silico* du génome (RRgenome, voir le premier article). Le génome *in silico* produit selon les mêmes conditions expérimentales (sites de restriction C^ACGG, fenêtre de sélection de taille 40-290pb) permet d'obtenir des fragments du génome couvrant 79Mb (3% du génome) et contenant $3\,454\,028$ CpG. En condition réelle, nous avons obtenu une moyenne de $3\,216\,778$ CpG, soit $90,5\%$ des CpG attendus d'après les prédictions *in silico*.

Parmi ces CpG, $61,0 \pm 2,6\%$ sont couverts par plus de 10 séquences (CpG₁₀). Finalement, après avoir pris en considération les aspects d'alignement unique et de couverture supérieure à 10 séquences par CpG, le nombre de CpG que nous avons sélectionné pour les analyses est d'en moyenne $1\,962\,235$, soit $7,0\%$ du nombre total de CpG du génome bovin.

La conversion de l'ADN au bisulfite de sodium est quasi-totale, de l'ordre de $99,3 \pm 0,5\%$. Cette donnée est importante car le pourcentage de méthylation calculé est fortement dépendant du taux de conversion. Dans notre cas, la corrélation entre taux de conversion au bisulfite de sodium et méthylation globale (moyenne de méthylation des CpG uniques et couverts plus de 10 fois) présente un R^2 de $0,0163$ (Figure 4.B, page précédente). Le taux de méthylation calculé pour nos échantillons n'est donc pas biaisé par le taux de conversion au

A

	Nombre de paires de séquences analysées	Taux de conversion bisulfite (%)	Alignements uniques (%)	Alignements multiples (%)	Pas d'alignement (%)	Nombre de CpG couverts	Nombre de CpG sur le chromosome Y	Nombre de CpG avec couverture ≥ 10 (%)	Nombre de CpG avec couverture ≥ 500 (%)
Moyenne	35 223 758	99,3	39,3	51,1	9,6	3 216 778	3896	61,0	0,4
Minimum	26 370 072	98,0	33,5	44,8	8,9	2 956 237	3010	54,8	0,3
Maximum	45 869 353	99,9	45,5	56,7	10,5	3 458 347	4738	66,4	0,4

B**Figure 4. Contrôles qualité des séquences de RRBS**

A. Tableau récapitulatif des données de contrôle qualité issu du séquençage des 58 banques de RRBS. **B.** Graphes de corrélation entre le taux de conversion au bisulfite de sodium et la méthylation globale des sites CpG et CHG (H étant n'importe quelle base nucléotidique excepté un G). Chaque point représente la corrélation pour un des individus du dispositif expérimental. Alors qu'aucune corrélation n'existe concernant les sites CpG ($R^2=0,0163$, p-value= 0,2699, graphe du haut), une forte corrélation est présente pour les sites CHG ($R^2=0,9676$, p-value=6,965e-14, graphe du bas). Le taux de méthylation CHG est donc fortement influencé par le taux de conversion de l'ADN au bisulfite de sodium.

bisulfite de sodium. En revanche, une forte corrélation entre le taux de méthylation des sites CHG et CHH est observée avec le taux de conversion au bisulfite de sodium (R^2 respectivement de 0,9676 (Figure 4.B, page précédente) et 0,955 (non montré, p -value= $8,045 \times 10^{-14}$)). Ceci suggère que le taux de méthylation calculé est très fortement influencé par le taux de conversion, et représente plutôt un artéfact technique. Dans ces conditions, ces données de méthylation sur les sites CHG et CHH n'ont pas été prises en considération dans les analyses ultérieures.

Aucun des différents critères de contrôle qualité ne présente de différence significative entre les banques construites pour les 2 races et pour les 3 groupes de fertilité. Ainsi, nous pouvons conclure que l'ensemble des banques peut être analysé sans biais technique afférant à la préparation.

C. La race influence de manière significative le taux de méthylation

En prenant en considération l'ensemble des CpG₁₀, le taux de méthylation moyen est respectivement de $51,3 \pm 1,5$ et $50,7 \pm 1,4$ % pour les races Holstein et Montbéliarde (Figure 5.A, page suivante), cette légère tendance à l'hypométhylation chez la race Montbéliarde n'étant pas significative (p -value=0,061). Cependant, en considérant uniquement les CpG₅₀₀, le taux de méthylation moyen est respectivement de $18,5 \pm 3,4$ et $16,9 \pm 2,6$ % pour les races Holstein et Montbéliarde (Figure 5.B, page suivante). L'hypométhylation globale des CpG₅₀₀ des individus de race Montbéliarde par rapport aux individus de race Holstein est significative (p -value=0,035).

La seconde partie du processus d'analyse de la méthylation est celle des analyses descriptives. Le premier type d'analyse est une classification hiérarchique de corrélation de Pearson (Figure 6, page suivante). Elle permet de classer très nettement les individus en fonction de leur race : 2 groupes apparaissent distinctement, eux-mêmes séparés en 2. Le second type de représentation est celui donné par l'ACP. La première dimension explique 5,51% de la variabilité, et la seconde 3,24%. Ce sont des valeurs très faibles. Il faut cependant prendre en considération le nombre de variables mises en jeu (ici 1 403 732 CpG), ce qui tend à faire diminuer ces chiffres. En prenant en considération les 2 premières dimensions, une séparation entre les 2 races est suggérée mais n'est pas totale du fait de l'importante dispersion des individus : Ceci sous-tend une importante part de variabilité inter-individuelle dans le niveau de méthylation des CpG, en particulier chez les animaux Holstein.

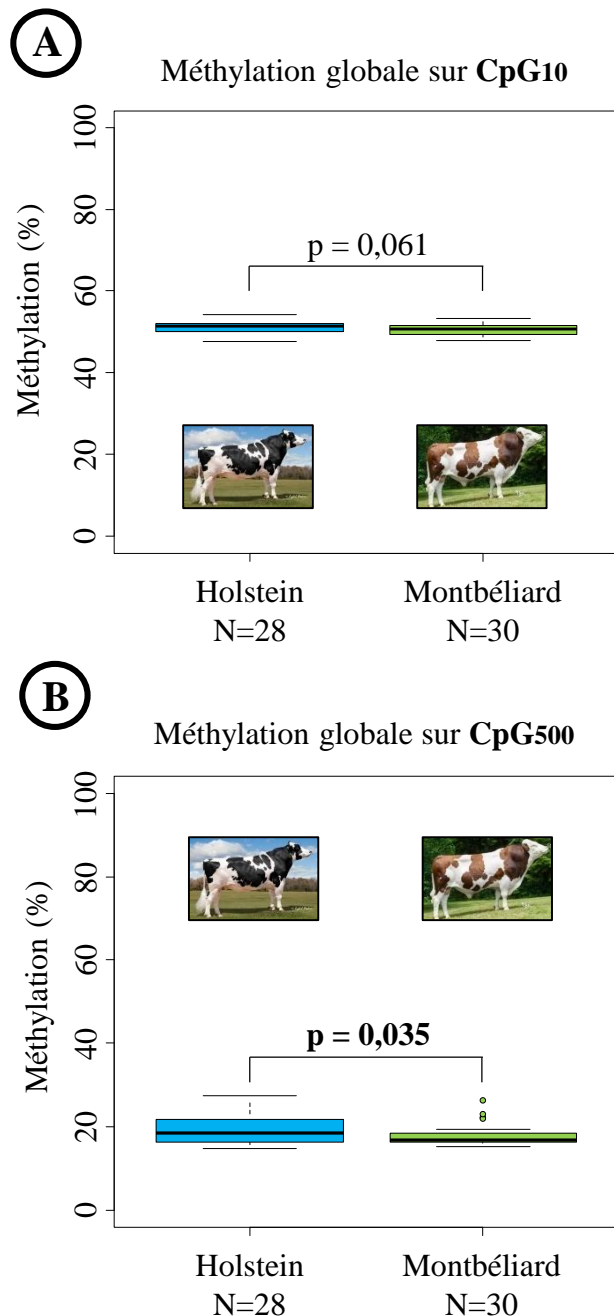


Figure 5. Comparaisons de la méthylation entre les races Holstein et Montbéliarde

A. Méthylation globale des CpG10. Dans un premier temps, nous avons analysé la méthylation de tous les CpG ayant une couverture d'au moins 10 séquences (CpG10). L'analyse porte en moyenne sur 1 962 235 CpG par individu. La méthylation médiane pour les taureaux de race Holstein est de 51,3% (minimum= 47,6 %; maximum= 54,2%), et pour les taureaux de race Montbéliard de 50,7% (minimum= 48,0%; maximum= 53,3%). On observe une tendance à l'hypométhylation chez les taureaux de race Montbéliard comparé aux taureaux de race Holstein (pvalue=0,061, test de permutation pour séries indépendantes avec échantillonnage Monte Carlo de 100 000 tirages). **B.** La même comparaison est conduite en ne considérant que les CpG500. Au total, l'analyse porte en moyenne sur 12 867 CpG par individu. La méthylation médiane pour les taureaux de race Holstein est de 18,5% (minimum= 14,8 %; maximum= 27,5%), et pour les taureaux de race Montbéliard de 16,9% (minimum= 15,2%; maximum= 26,4%). La méthylation des CpG500 est significativement inférieure chez les Montbéliard comparé à celle chez les Holstein (pvalue=0,035, test de permutation pour séries indépendantes avec échantillonnage Monte Carlo de 100 000 tirages).

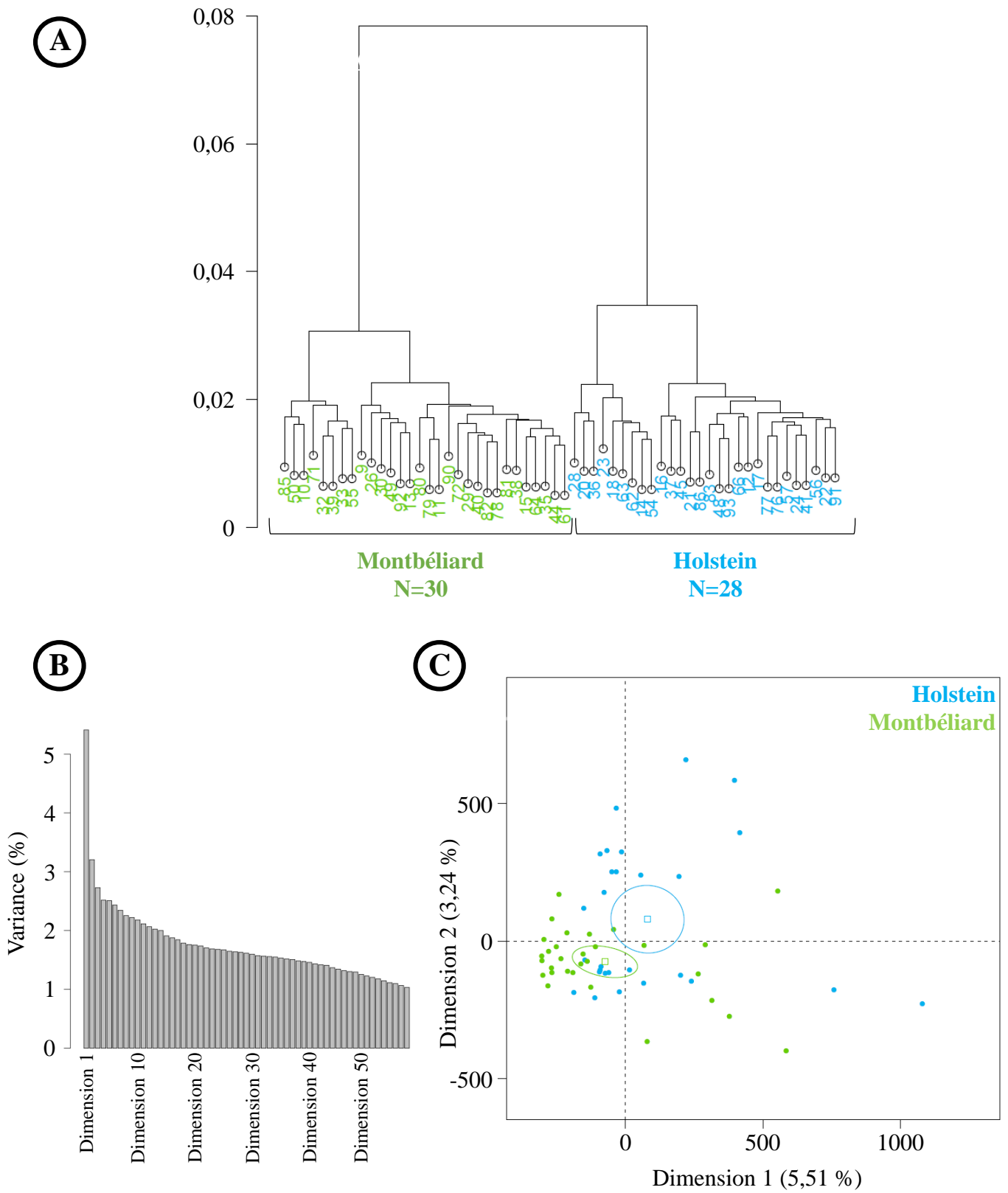


Figure 6. Analyse descriptive comparant les races Holstein et Montbéliarde

Les CpG utilisés pour cette analyse descriptive sont couverts par un minimum de 10 séquences pour au moins 14 individus par race. Ainsi, l'analyse porte sur un total de 1 403 732 CpG. **A.** Classification hiérarchique de corrélation de Pearson. On observe une nette séparation des individus en fonction de leur race. **B.** Digramme en éboulis pour l'Analyse en Composantes Principales (ACP). Cette représentation permet de montrer la contribution de chaque axe à la variabilité totale. Le premier axe est celui qui explique le plus de variabilité. Dans notre cas, le premier axe n'explique que 5,51% de la variabilité. **C.** L'ACP permet de visualiser la séparation entre les 2 races selon la dimension 1. La séparation n'est pas totale, et est certainement due à une importante part de variabilité inter-individuelle en terme de méthylation de l'ADN, en particulier chez les individus de race Holstein.

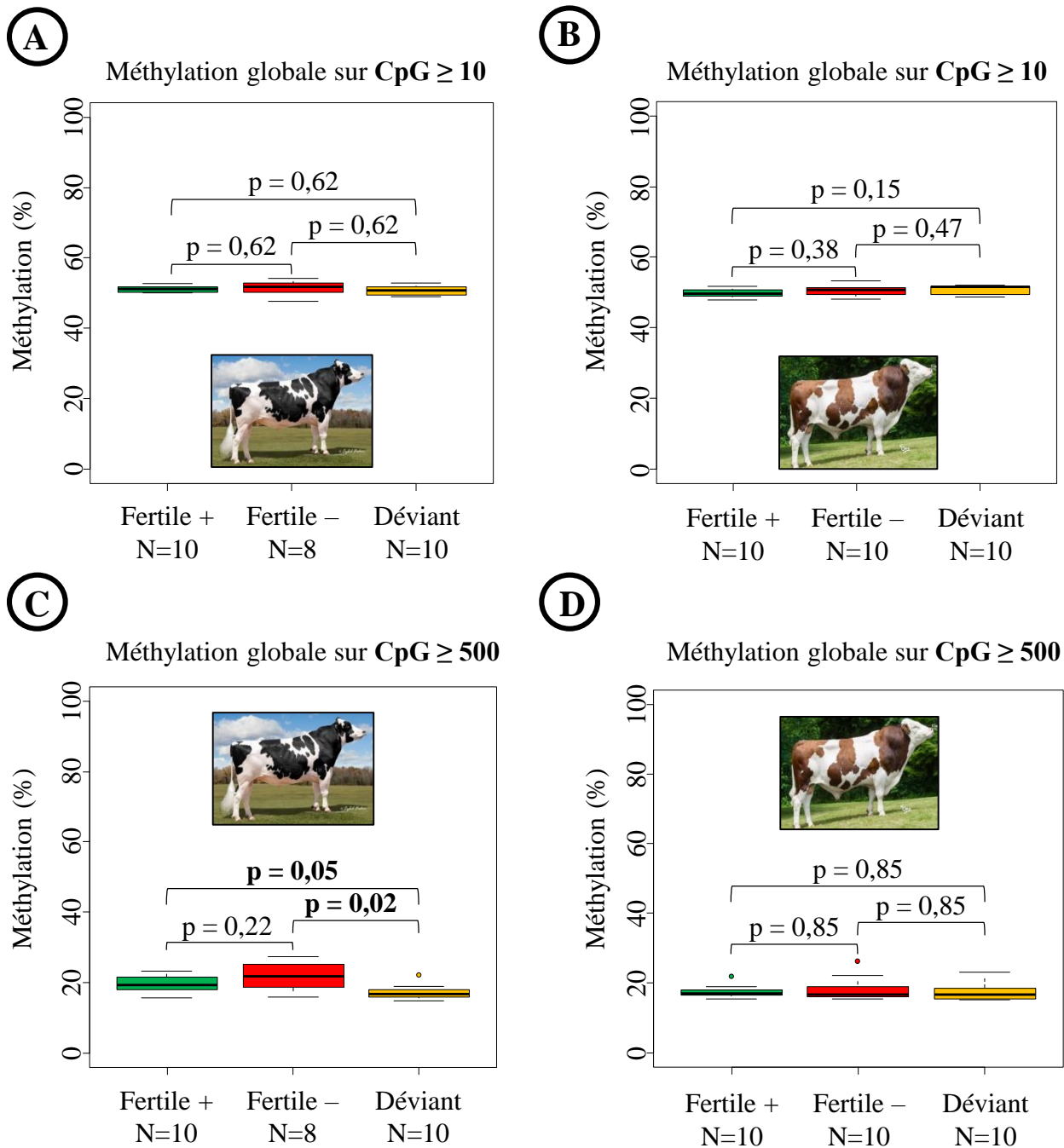


Figure 7. Méthylation globale des taureaux Holstein et Montbéliard, avec distinction entre les catégories de fertilité

A. Méthylation globale des CpG10. Au total, l'analyse porte sur 1 905 833 CpG pour la race Holstein. La méthylation médiane pour les Fertiles + est de 51,3% (minimum= 50,1%; maximum= 52,8%), 51,9% pour les Fertiles- (minimum= 47,6%; maximum= 54,2%) et 50,8% pour les Déviants (minimum= 49,1%; maximum= 53,0%). Ces différences ne sont pas significatives (pvalue=0,62). **B.** Pour la race Montbéliard, l'analyse porte sur 1 885 260 CpG. La méthylation médiane pour les Fertiles + est de 49,6% (minimum= 48,0%; maximum= 51,8%), 50,7% pour les Fertiles- (minimum= 48,1%; maximum= 53,3%) et 51,6% pour les Déviants (minimum= 48,7%; maximum= 52,1%). Ces différences ne sont pas significatives (pvalues comprises entre 0,15 et 0,47). **C.** Méthylation globale des CpG500 pour les taureaux de race Holstein (moyenne de 13 044 CpG par individu). La méthylation médiane pour les Fertiles + est de 19,3% (minimum= 15,7%; maximum= 23,4%), 21,9% pour les Fertiles- (minimum= 15,9%; maximum= 27,5%) et 16,9% pour les Déviants (minimum= 14,8%; maximum= 22,2%). On observe une différence statistiquement significative entre les Déviants et les Fertiles + (pvalue=0,05), et les Déviants et les Fertiles - (pvalue=0,02). **D.** Analyse de la méthylation globale des CpG500 pour la race Montbéliard (moyenne de 12 702 CpG par individu). La méthylation médiane pour les Fertiles + est de 17,1% (minimum= 15,4%; maximum= 22,0%), 16,9% pour les Fertiles- (minimum= 15,4%; maximum= 26,4%) et 16,9% pour les Déviants (minimum= 15,2%; maximum= 23,2%). On observe aucune différence significative entre les groupes de fertilité (pvalue=0,85 pour toutes les combinaisons). L'ensemble des analyses statistiques ont été effectuées avec des tests de permutation pour séries indépendantes avec échantillonnage Monte Carlo (100 000 tirages), suivis de comparaisons 2 à 2 (1000 tirages) avec correction de Benjamini-Hochberg pour tests multiples.

D. Des différentiels de méthylation entre les catégories de fertilité sont mis en évidence pour les 2 races

Après avoir comparé les races entre elles, le second axe d'analyse consiste à comparer les catégories de fertilité au sein de chacune des races. Les résultats en termes de méthylation globale des CpG₁₀ et CpG₅₀₀ sont représentés Figure 7. Pour les 2 races, aucune différence significative n'est observée entre les catégories de fertilité pour les CpG₁₀, quels que soient les groupes comparés 2 à 2 (p-values détaillées dans la Figure 7). De même, il n'y a pas de différence significative pour les CpG₅₀₀ entre les catégories de fertilité en race Montbéliarde. En revanche, pour la race Holstein, les taureaux de la catégorie Déviants présentent une hypométhylation significative des CpG₅₀₀ : méthylation médiane de $16,9 \pm 2,1\%$, par rapport aux 2 autres catégories ($19,3 \pm 2,4\%$ et $21,9 \pm 4,2\%$ pour les Fertiles+ et Fertiles-, respectivement).

Plus précisément, nous avons catégorisé les CpG en 3 groupes en fonction de leur état de méthylation : les CpG hypométhylés dont le pourcentage de méthylation est inférieur à 20%, les CpG hyperméthylés dont le pourcentage de méthylation est supérieur à 80%, et les CpG intermédiaires dont le pourcentage de méthylation est compris entre 20 et 80 %. La répartition des CpG₁₀ entre les 3 groupes est équivalente quelle que soit la race considérée (Figure 8). Pour les CpG₅₀₀, on observe qu'en moyenne 59,2% et 53,8% sont hypométhylés chez les catégories Fertiles+ et Fertiles- de la race Holstein, respectivement, et 65,9% sont hypométhylés chez les Déviants Holstein. Pour la race Montbéliarde, 64,2%, 63,3% et 65,7% sont hypométhylés chez les catégories Fertiles+, Fertiles- et Déviants, respectivement. Il semble alors que les individus Déviants de la race Holstein présentent plus de CpG₅₀₀ hypométhylés que les 2 autres catégories de fertilité. De plus, il semblerait que l'hypométhylation globale constatée entre la race Holstein et Montbéliard soit liée à une augmentation du nombre de CpG₅₀₀ hypométhylés pour la race Montbéliarde.

Une analyse descriptive a été réalisée pour chacune des races, basée sur des classifications hiérarchiques et des Analyses en Composantes Principales (ACP) utilisant l'ensemble des CpG (1 905 833 CpG en race Holstein et 1 885 260 CpG en race Montbéliarde. Les résultats sont présentés Figure 9). De façon générale, la fertilité ne semble pas en lien direct avec le profil de méthylation globale. En effet, ces approches ne permettent pas de regrouper les individus en fonction des catégories de fertilité. Par exemple en race Holstein, la classification hiérarchique de corrélation montre une séparation entre 2 groupes d'individus qui ne se superpose pas à la distribution entre les catégories de fertilité, puisque chaque branche

		CpG10			CpG500		
		Couverture ≥ 10 hypométhylé (<20%)	Couverture ≥ 10 intermédiaire (>20% et <80%)	Couverture ≥ 10 hyperméthylé (>80%)	Couverture ≥ 500 hypométhylé (<20%)	Couverture ≥ 500 intermédiaire (>20% et <80%)	Couverture ≥ 500 hyperméthylé (>80%)
Holstein	Fertile +	45,1	6,8	48,1	59,2	40,4	0,5
	Fertile -	44,5	7,2	48,3	53,8	45,6	0,6
	Déviant	45,2	7,2	47,7	65,9	33,7	0,4
Montbéliard	Fertile +	46,5	6,9	46,6	64,2	35,4	0,4
	Fertile -	45,6	7,0	47,4	63,3	36,2	0,5
	Déviant	45,1	6,9	48,1	65,7	33,8	0,5

Tableau 1. Comparaisons de la méthylation entre les races Holstein et Montbéliarde

Comparaison entre le taux de CpG hypométhylés (méthylation < 20%), les CpG hyperméthylés (méthylation > 80%), et les CpG intermédiaires (20% ≤ méthylation ≤ 80%). Les Déviants de race Holstein semblent présenter plus de CpG500 hypométhylés que les 2 autres catégories de fertilité. De plus, les catégories Fertiles+ et Fertiles- de race Holstein présentent moins de CpG500 hypométhylés que les 3 catégories de fertilité de la race Montbéliard. En conclusion, il existe une différence importante de méthylation des CpG500 entre les Holstein et Montbéliard, et ce différentiel est le fait des taureaux Déviants.

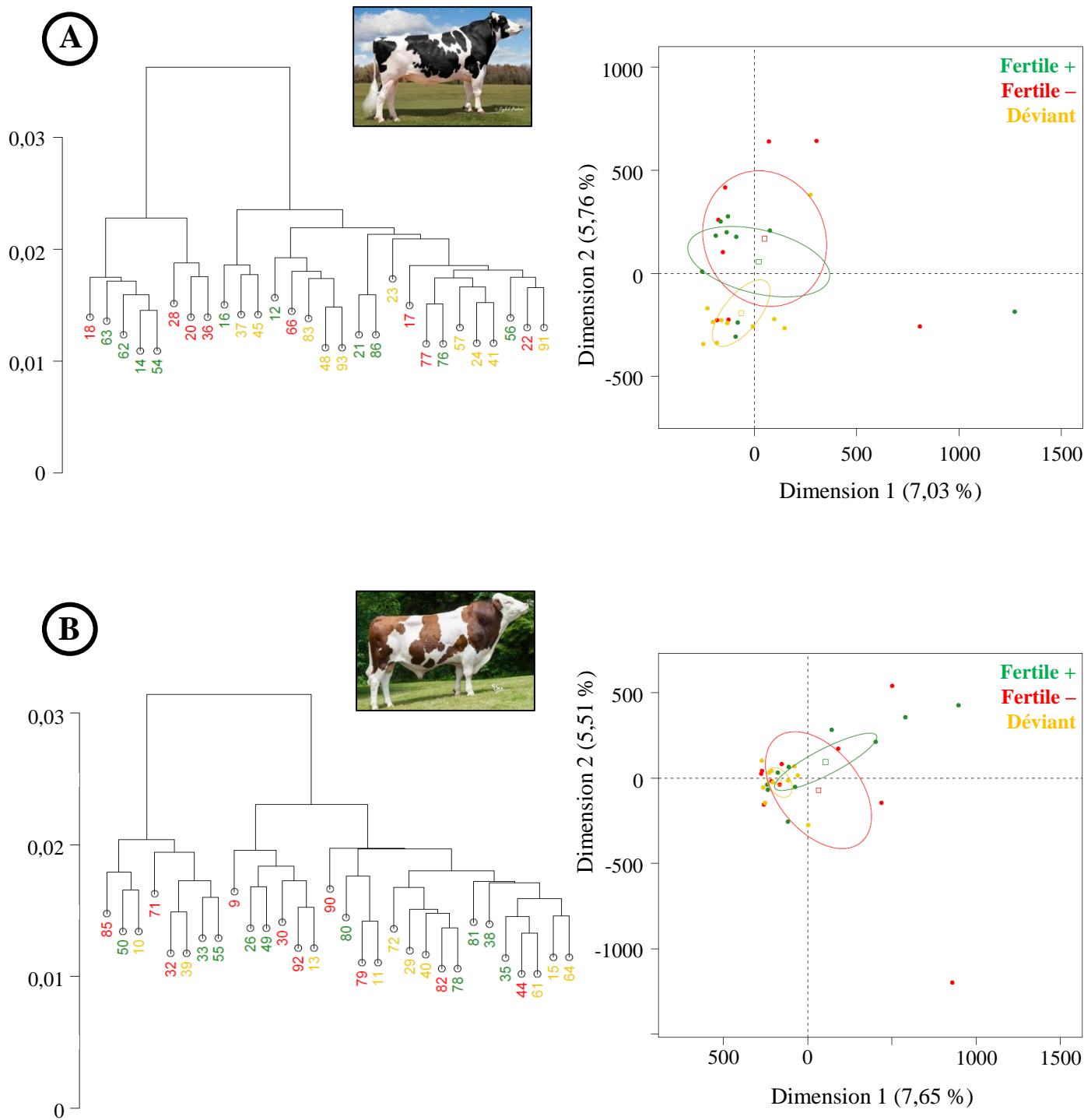
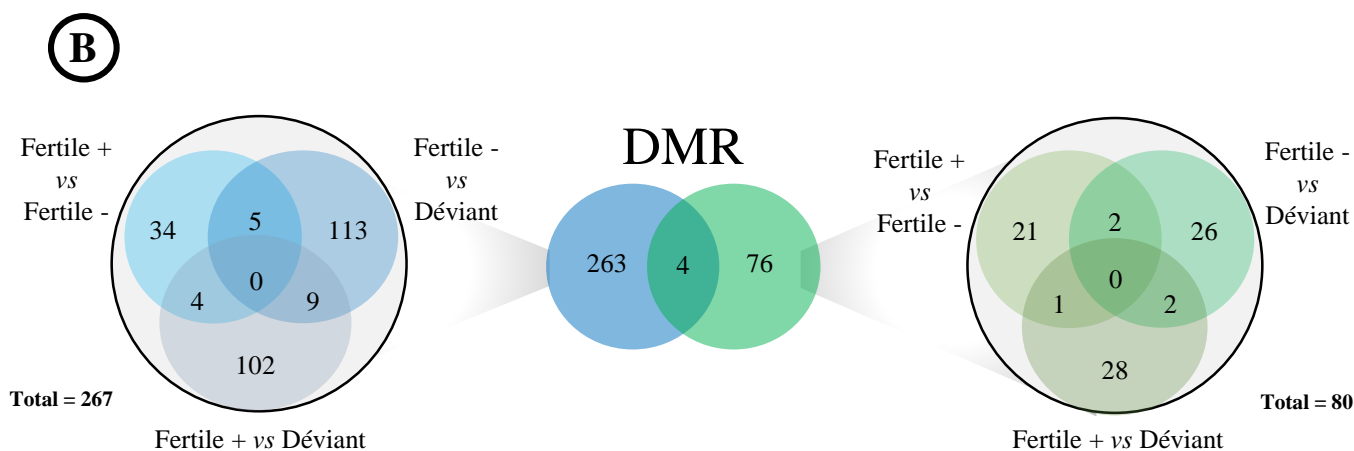
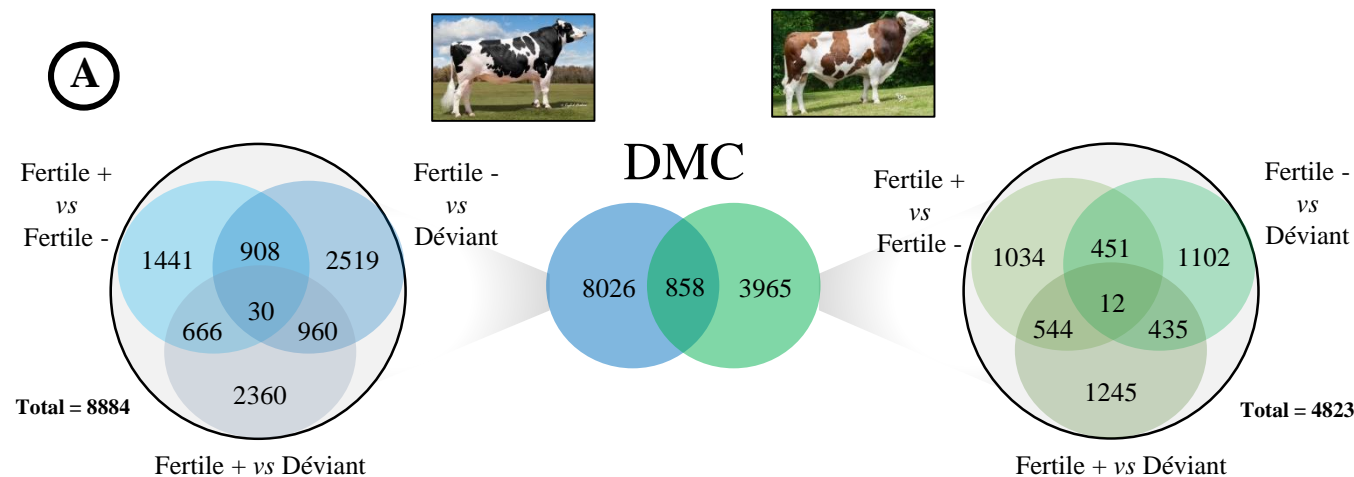


Figure 8. Analyse descriptive comparant les catégories de fertilité pour les races Holstein et Montbéliarde

A. Classification hiérarchique de corrélation de Pearson et Analyse en Composantes Principales (ACP) pour la race Holstein. L'analyse porte sur un total de 1 905 833 CpG. On observe peu séparation entre les catégories de fertilité que ce soit sur la classification hiérarchique ou l'ACP. On observe cependant un début de regroupement des Déviants sur la classification hiérarchique, et un début de séparation de ceux-ci sur l'ACP par rapport aux autres catégories. **B.** Analyse similaire conduite sur la race Montbéliarde. L'analyse repose ici sur un total de 1 885 260 CpG différents. En classification hiérarchique comme en ACP, les 3 catégories de fertilité sont indissociables.



C

		DMC			
		Hypométhylés contre Fertiles-	Hypométhylés contre Déviant	Hyperméthylés contre Fertiles-	Hyperméthylés contre Déviant
Holstein	Fertiles +	54,1%	32,7%	45,9%	67,3%
	Fertiles -		37,6%		62,4%
	Déviant				
Montbéliard	Fertiles +	49,5%	51,9%	50,5%	48,1%
	Fertiles -		47,5%		52,5%
	Déviant				

		DMR			
		Hypométhylés contre Fertiles-	Hypométhylés contre Déviant	Hyperméthylés contre Fertiles-	Hyperméthylés contre Déviant
Holstein	Fertiles +	32,6%	13,0%	67,4%	87,0%
	Fertiles -		31,5%		68,5%
	Déviant				
Montbéliard	Fertiles +	54,2%	61,3%	45,8%	38,7%
	Fertiles -		60,0%		40,0%
	Déviant				

Figure 9. Analyse différentielle entre les catégories de fertilité, pour les races Holstein et Montbéliarde

A. Les DMC sont identifiées en menant les analyses différentielles en utilisant les catégories 2 à 2. Dans les diagrammes de Venn à gauche et à droite sont respectivement représentés les DMC pour la race Holstein et Montbéliarde. Au centre sont représentés les recouvrements qui existent entre les DMC identifiées chez les Holstein et Montbéliarde. Chaque diagramme de Venn présente les DMC spécifiques d'une comparaison, les DMC communs à 2 comparaisons et aux 3 comparaisons. Ainsi, un total de 8884 DMC sont observés pour la race Holstein, et 4823 pour la race Montbéliarde, et 858 sont communs entre les 2 races. **C.** Le même type de représentation est obtenu en prenant en compte les DMR. **C.** Répartition des DMC et DMR pour les 2 races en prenant en considération l'état de méthylation relatif pour chacune des comparaisons 2 à 2 entre les groupes de fertilité.

comporte à la fois des individus Fertiles+ et Fertiles-. De même, l'ACP ne fait pas apparaître de séparation nette des groupes d'individus. Il est à noter que plusieurs individus Fertiles+ et Fertiles- présentent des profils extrêmes, et contribuent ainsi à la variabilité inter-individuelle. La contribution des composantes principales à l'inertie totale est faible. Les deux premières composantes expliquent ainsi 7,03% et 5,76% de la variabilité de méthylation. Néanmoins, la seconde composante principale semble permettre de distinguer le groupe des Déviants des groupes Fertiles+ et Fertiles-. Pour les taureaux de race Montbéliarde, ni la classification hiérarchique de corrélation ni l'ACP ne permettent de distinguer les groupes d'individus. Le groupe des taureaux Déviants semble plus homogène que les 2 autres groupes.

L'analyse différentielle a été menée afin d'identifier les cytosines différentiellement méthylées (*Differentially Methylated Cytosine*, DMC) entre les catégories de fertilité comparées 2 à 2. Les résultats sont présentés sous forme de diagramme de Venn (Figure 10.A, page suivante). Nous avons identifié, toutes catégories confondues, 8884 DMC pour la race Holstein et 4823 DMC pour la race Montbéliarde (soit 45,7% de moins que pour la race Holstein). Entre les 2 races 858 DMC sont communes, ce qui représente 9,66% des DMC de la race Holstein et 17,8% des DMC de la race Montbéliarde.

Pour la race Holstein, la comparaison entre Déviants et autres catégories de fertilité met en évidence le plus grand nombre de DMC. En effet, la comparaison entre Fertiles+ et Fertiles- génère 3045 DMC, contre 4417 DMC pour la comparaison Fertiles- contre Déviants (soit une augmentation de 45%) et 4016 DMC pour la comparaison entre Fertiles+ et Déviants (augmentation de 32%). Cette observation est en accord avec les données de l'ACP, qui suggère que le groupe Déviant Holstein se comporte différemment des 2 autres groupes.

Cette observation est en accord avec les données de l'ACP, qui montraient que le groupe Déviant Holstein semble se comporter différemment des 2 autres groupes. Pour les taureaux de race Montbéliarde, les DMC sont réparties de façon assez uniforme entre les territoires, ce qui corrobore le fait que la catégorie Déviant ne semble pas particulièrement se distinguer des autres catégories de fertilité. Il y a 2041 DMC pour la comparaison Fertiles+ contre Fertiles-, 2000 DMC pour la comparaison Fertiles- contre Déviants, et 2236 DMC pour la comparaison entre Fertiles+ contre Déviants.

Selon les territoires du diagramme de Venn, les territoires comportant 960 DMC chez les Holstein et 435 DMC chez les Montbéliard seront appelées « spécifiques des Déviants » dans la suite du document. Ces DMC sont différentiellement méthylées chez les Déviants par rapport aux Fertiles+ et Fertiles-, mais ne le sont pas quand on compare Fertiles + et Fertiles -.

Pour la race Holstein, on note qu'une majorité des DMC est hypométhylée chez les taureaux Déviants (Figure 10.B). En effet, alors que dans la comparaison entre Fertiles+ et Fertiles- les DMC se répartissent de manière assez équitable en termes de méthylation (54,1% sont hypométhylés chez les Fertiles- et 45,9% sont hypométhylés chez les Fertiles+), dans la comparaison entre Fertiles- et Déviants 62,4% sont hypométhylés chez les Déviants. De façon similaire, dans la comparaison entre Fertiles+ et Déviants, 67,3% des DMC sont hypométhylés chez les taureaux Déviants. Alors qu'on n'observe pas de différence significative pour la méthylation globale des CpG₁₀ entre les 3 catégories, il existe une tendance à l'hypométhylation pour les DMC de la catégorie de fertilité Déviant pour les taureaux de race Holstein. Cette tendance à l'hypométhylation des DMC des taureaux Déviants n'est pas retrouvée pour la race Montbéliarde, les DMC hypométhylées et hyperméthylées de chaque comparaison se répartissant assez équitablement entre chaque catégorie de fertilité.

L'ensemble des DMC mises en évidence peut être utilisé pour définir des DMR (Figure 10.C). Assez peu de DMR ont été obtenues pour les 2 races. Toutes catégories confondues, nous avons identifié 267 DMR pour la race Holstein et 80 DMR pour la race Montbéliarde (soit 70,0% de moins que pour la race Holstein). Parmi elles, 4 DMR sont communes, ce qui représente 1,5% des DMR de la race Holstein et 5% des DMR de la race Montbéliarde. De plus, toutes comparaisons confondues, le nombre de moyen de DMC par DMR est de 5 pour les races Holstein comme Montbéliard (minimum de 3 et maximum de 21, données non montrées). Ces résultats tendent à montrer que la plupart des DMC sont espacées les unes et des autres, donc qu'elles seraient distribuées ponctuellement sur le génome sans former d'importants regroupements. Il ne faut cependant pas oublier que cela peut aussi être dû à nos conditions expérimentales (notamment la fenêtre de sélection de taille) qui implique que les régions les plus denses en CpG ont été éliminées.

Chez les Holstein, les tendances observées pour les DMC s'accroissent pour les DMR concernant l'état de méthylation. Ainsi, les DMR sont majoritairement hypométhylées chez les Déviants comparés aux 2 autres catégories de fertilité, en particulier pour la comparaison entre Fertiles+ et Déviants (Figure 10.B). Par contre, chez les taureaux Montbéliard, les DMR sont en majorité hyperméthylées chez les Déviants par rapport aux 2 autres catégories.

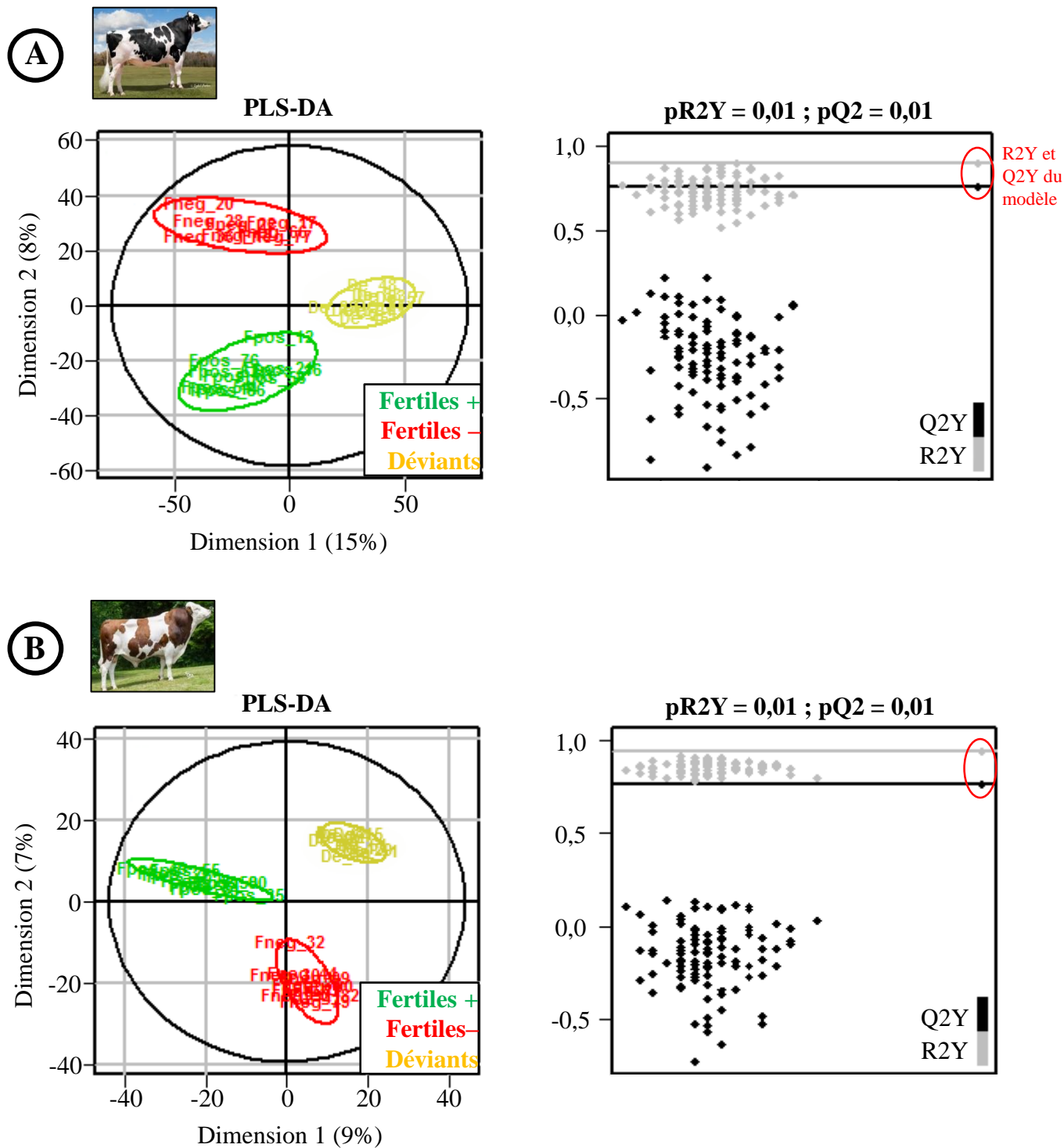


Figure 10. Analyse discriminante pour les DMC identifiées

A. Analyse pour la race Holstein, en utilisant les 8884 DMC. Les 2 premières dimensions de la PLS-DA expliquent 15 et 8% de la variance, respectivement. La dimension 1 permet de séparer très nettement les Déviants par rapport aux 2 autres catégories; la dimension 2 sépare surtout les 2 autres catégories entre elles. La représentation de droite renseigne sur la qualité du modèle de jeu de données. Les indicateurs de qualité de notre modèle (Q2Y: valeur prédictive du modèle; R2Y: adéquation entre le modèle et le jeu de donnée) sont représentés par les points noirs et gris en haut à droite de la figure. On observe qu'ils sont proches de 1 sur l'axe des ordonnées (la valeur maximale possible). Les nuages de points à gauche représentent la valeur des indicateurs calculée pour des jeux de données factices générés à partir de notre jeu de donnée, en permutant aléatoirement les individus entre les groupes (100 permutations). On n'obtient jamais un modèle d'aussi bonne qualité que notre jeu de donnée initial. **B.** Analyse similaire en utilisant les 4823 DMC des taureaux Montbéliard. Les résultats sont très similaires à ceux obtenus pour les Holstein; mis à part le fait que la variance expliquée par les 2 premières dimensions est inférieure que ce qui a été identifié pour la race Holstein (9 et 7% respectivement). Aussi, la séparation entre les catégories de fertilité sur le PLS-DA est moins nette que pour les Holstein. Les Déviants sont séparés uniquement des Fertiles+ sur la dimension1, puis des Fertiles- sur la dimension 2. La représentation de droite montre cependant que le modèle de PLS-DA utilisant notre jeu de donnée de DMC est d'aussi bonne qualité que pour la race Holstein.

E. Les DMC représentent de bons biomarqueurs de la fertilité

Pour s'assurer que les DMC identifiées permettent de caractériser les catégories de fertilité, nous avons réalisé une analyse de type discriminante par PLS-DA. Cette analyse a été réalisée sur l'ensemble des DMC identifiées provenant de toutes les comparaisons, soit les 8884 DMC pour la race Holstein et les 4823 DMC pour la race Montbéliarde.

Pour la race Holstein comme pour la race Montbéliarde, on observe que la PLS-DA sépare distinctement les individus en fonction de leur catégorie de fertilité (Figure 12). Pour la race Holstein, la première dimension explique 15% de la variance entre les groupes, et on observe que cette dimension sépare les Déviants des 2 autres catégories de fertilité. La seconde dimension explique 8% de la variance, et sépare les Fertiles+ des Fertiles-.

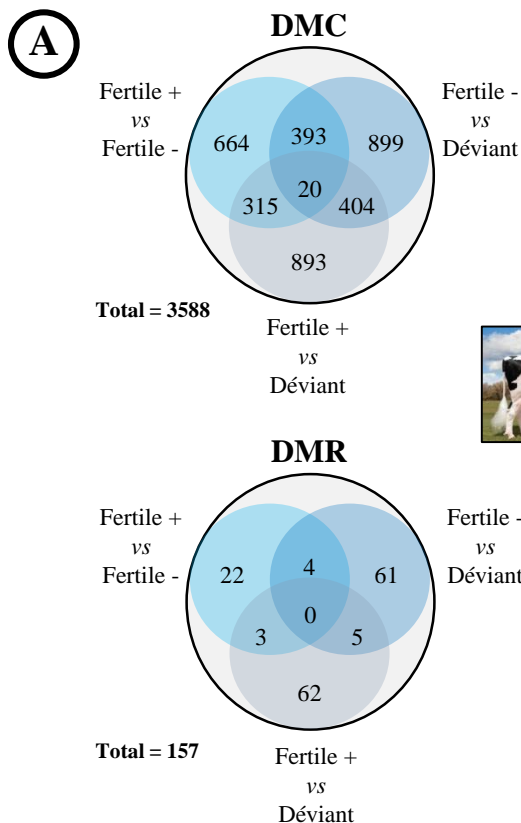
Pour la race Montbéliarde, la première dimension explique 9% de la variance entre les individus, et sépare les Déviants des Fertiles+, mais pas totalement des Fertiles-. La seconde dimension (7%) permet alors de séparer les Déviants des Fertiles-. Les 3 catégories de fertilité sont donc séparées selon les 2 premières dimensions, mais les résultats sont moins nets que pour les taureaux Holstein. L'évaluation de la qualité de notre modèle est assurée par l'observation de coefficients R2Y et Q2Y proches de 1, avec une probabilité de l'ordre de 0,01. Comparés aux coefficients obtenus pour cent permutations aléatoires, ces valeurs attestent donc de la qualité du modèle basé sur une répartition des DMC en 3 catégories.

En conclusion, les DMC identifiées par l'analyse différentielle sont de bons biomarqueurs pour discriminer les catégories de fertilité dans les deux races.

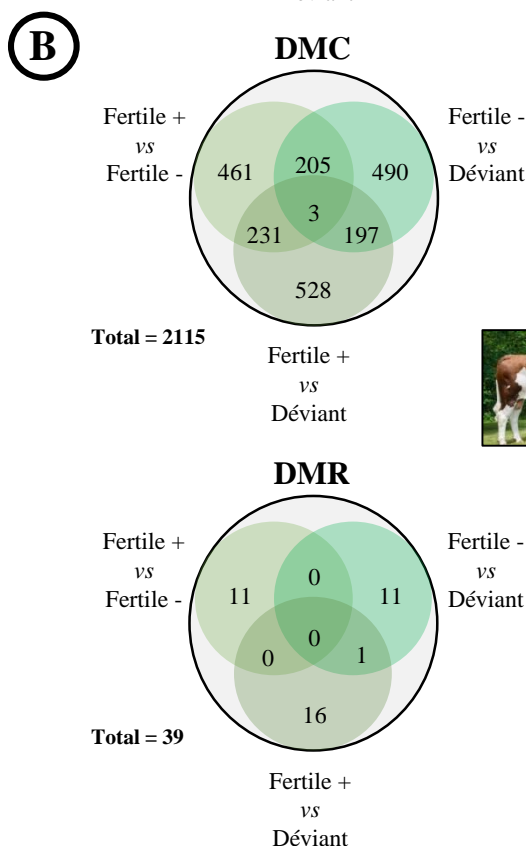
F. Répartition génomique des DMC identifiées

Comparée à l'annotation des 1 905 333 CpG₁₀ analysés (ou 1 885 260 pour la race Montbéliarde, qu'on nomme également les *backgrounds*), l'annotation des 8884 DMC pour les Holstein et des 4823 DMC pour les Montbéliard présente une répartition plus importante dans les régions intergéniques, dans des régions éloignées des îlots CpG, et sur des régions répétées (Figure 13, page suivante).

Pour les 2 races, les DMC sont principalement observées dans les introns des gènes (à 32,4% pour la race Holstein et 36,1% pour la race Montbéliarde, contre 26,2 et 26,0% pour les *backgrounds*). Au contraire, une plus faible proportion de DMC est associée aux promoteurs-TSS et exons (respectivement à 2,7 et 2,4% pour les races Holstein et Montbéliard, contre



Holstein	Background (%)	DMC (%)	DMR (%)
Nombre de CpG	1 905 833	8884	267
Eléments génomiques			
Promoteur-TSS	13,8	2,7	3,4
Intron	26,2	32,4	25,5
Exon	13,1	9,6	20,2
TTS	0,7	0,7	0,7
3'UTR	0,8	1,2	0,7
5'UTR	3,5	1,6	1,1
Intergénique	41,8	51,7	48,3
Densité en CpG			
Island	42,8	19,6	38,6
Shore	18,0	18,3	11,2
Shelve	6,4	8,2	5,6
Open Sea	32,8	53,9	44,5
Eléments répétés			
LINE	3,2	7,0	6,4
SINE	4,2	7,1	3,7
LTR	1,6	2,9	1,5
Satellite	3,8	2,9	0,4
Autre	3,0	1,8	3,7
Sans élément répété	84,2	78,2	84,3



Montbéliard	Background (%)	DMC (%)	DMR (%)
Nombre de CpG	1 885 260	4823	80
Eléments génomiques			
Promoteur-TSS	14,0	2,4	3,8
Intron	26,0	36,1	33,8
Exon	13,1	6,8	7,5
TTS	0,7	0,6	0,0
3'UTR	0,8	0,8	0,0
5'UTR	3,5	1,6	1,3
Intergénique	41,9	51,7	53,8
Densité en CpG			
Island	43,3	17,1	30,0
Shore	17,8	21,2	23,8
Shelve	6,3	9,9	8,8
Open Sea	32,7	51,8	37,4
Eléments répétés			
LINE	3,0	5,2	7,5
SINE	4,1	7,7	11,3
LTR	1,6	2,9	6,3
Satellite	3,8	1,3	0,0
Autre	3,1	1,7	7,5
Sans élément répété	84,5	81,2	67,4

Figure 11. Annotation des DMC et DMR

A. Identification du nombre de gènes associés aux DMC et DMR pour chaque comparaison entre les catégories de fertilité pour la race Holstein. Dans le tableau de droite sont comparés les résultats d'annotation entre le *background* et ceux des DMC et DMR. **B.** Même type de représentation des résultats pour la race Montbéliard.

13,8 et 14% dans leur *background*). Le résultat est différent pour les DMR. Pour la race Holstein, on les retrouve indistinctement dans les introns et les exons (respectivement 25,5 et 20,2%). Pour la race Montbéliarde, on les retrouve principalement dans les introns (33,8 contre 7,5% dans les exons).

Le même type d'analyse est effectué pour les CGI, CpG *shores* et CpG *shelves*. Pour les 2 races, les DMC sont majoritairement observées en *open sea*. Par exemple pour la race Holstein, on retrouve les DMC à 19,6% dans les CGI, 18,3% dans les CpG *shores*, 8,2% dans les CpG *shelves* et 53,9% en *open sea*. Des résultats similaires sont observés pour la race Montbéliarde. En revanche, pour les 2 races les DMR sont retrouvées majoritairement associées aux CGI et dans la même proportion que leur *background* respectif.

Pour l'analyse d'annotation ciblant les éléments répétés du génome, nous nous sommes principalement focalisés sur l'analyse des LINE, SINE, LTR et séquences satellites. Dans les 2 races, une grande majorité des DMC et DMR ne ciblent pas des éléments répétés (78,2% des DMC et 84,3% des DMR pour la race Holstein ; 81,5% pour les DMC et 67,4% pour les DMR pour la race Montbéliarde). Les éléments répétés les plus représentés pour leur association avec une DMC sont les LINE et les SINE, avec respectivement 7,0 et 7,1% en race Holstein, et 7,5 et 11,3% en race Montbéliarde. Alors que les tendances restent similaires pour les DMR en race Holstein (avec cependant une baisse de la proportion de SINE), on observe une augmentation de la part des LINE, SINE et LTR dans les DMR en race Montbéliarde.

Pour la race Holstein, sur les 8884 DMC identifiées par l'analyse différentielle, 3588 peuvent être associées à un gène. Concernant les DMR, sur les 267 identifiées, 157 peuvent être associées à un gène. Pour la race Montbéliarde, les chiffres sont de 2115 DMC pouvant être associées à un gène, et 39 DMR. La liste des gènes associés à une DMR pour les races Holstein et Montbéliard sont données respectivement en Annexes n°7 et n°8. Seul un gène présent sur le chromosome Y comporte des DMC, le gène *LOC107132102* (ou gène 32436) : une DMC est spécifique aux Déviants par rapport aux Fertiles+ et Fertiles-, et 2 autres DMC correspondent à la comparaison entre Fertiles- et Déviants.

A

Fertiles+ contre Fertiles-

Note	Nombre de gènes	Significativité	Fonction biologique (Gene Ontology term)
2,94	86	4,00E-02	Neuron differentiation (GO:0030182)
2,85	65	2,30E-02	Phosphotransferase activity, alcohol group as acceptor (GO:0016773)
2,84	43	1,80E-02	Protein serine/threonine kinase activity (GO:0004674)
2,57	101	4,90E-02	Locomotion (GO:0040011)
2,57	38	1,50E-01	Cell-substrate junction (GO:0030055)
2,44	74	4,70E-02	Neuron part (GO:0097458)
1,87	22	1,20E-01	Second-messenger-mediated signaling (GO:0019932)
1,86	4	1,70E-01	Protein-arginine deiminase activity (GO:0004668)
1,82	45	1,90E-01	Synapse (GO:0045202)
1,79	12	1,60E-01	Stress fiber assembly (GO:0043149)
1,72	24	8,30E-02	Modulation of synaptic transmission (GO:0050804)
1,68	14	3,00E-01	Regulation of phosphatase activity (GO:0010921)
1,58	31	2,70E-01	Somatodendritic compartment (GO:0036477)
1,55	54	1,50E-01	Regulation of cell motility (GO:2000145)
1,55	66	2,10E-01	Cardiovascular system development (GO:0072358)
1,53	15	4,10E-01	Sensory perception of mechanical stimulus (GO:0050954)
1,44	36	4,20E-01	Skeletal system development (GO:0001501)
1,43	158	3,20E-01	Regulation of cell communication (GO:0010646)
1,42	9	2,50E-01	Transmembrane receptor protein tyrosine kinase activity (GO:0004714)
1,35	32	7,60E-02	Signal release (GO:0023061)
1,32	38	5,50E-01	Sensory organ development (GO:0007423)
1,31	63	4,10E-01	Cellular response to endogenous stimulus (GO:0071495)

B

Déviants contre Fertiles+

Note	Nombre de gènes	Significativité	Fonction biologique (Gene Ontology term)
4,13	40	2,60E-02	Enzyme activator activity (GO:0008047)
3,48	68	3,00E-02	Phosphotransferase activity, alcohol group as acceptor (GO:0016773)
3,48	98	2,90E-02	Cell projection organization (GO:0030030)
3,22	81	4,50E-02	Neuron part (GO:0097458)
2,64	25	4,80E-02	Guanyl-nucleotide exchange factor activity (GO:0005085)
2,33	160	3,00E-02	Regulation of cellular component organization (GO:0051128)
2,31	188	4,00E-02	Regulation of signaling (GO:0023051)
2,06	72	1,90E-01	Cardiovascular system development (GO:0072358)
2	14	1,70E-01	Receptor signaling protein serine/threonine kinase activity (GO:0004702)
1,88	121	3,90E-02	Movement of cell or subcellular component (GO:0006928)
1,87	17	5,50E-02	Axon extension (GO:0048675)
1,85	24	1,30E-01	Modulation of synaptic transmission (GO:0050804)
1,78	8	3,80E-02	Intracellular ligand-gated ion channel activity (GO:0005217)
1,77	16	2,20E-01	Protein tyrosine kinase activity (GO:0004713)
1,63	19	4,90E-01	Regulation of peptidyl-tyrosine phosphorylation (GO:0050730)
1,61	15	2,00E-01	Cerebral cortex development (GO:0021987)
1,61	6	1,30E-01	Semaphorin-plexin signaling pathway involved in neuron projection guidance (GO:1902285)
1,59	23	2,30E-01	Cognition (GO:0050890)
1,46	235	4,80E-01	Ion binding (GO:0043167)
1,46	33	4,60E-01	Somatodendritic compartment (GO:0036477)
1,4	17	5,00E-01	Dendrite development (GO:0016358)
1,39	166	3,00E-02	Cellular localization (GO:0051641)
1,35	22	1,90E-01	Regulation of synapse structure or activity (GO:0050803)
1,34	28	5,10E-01	Anatomical structure homeostasis (GO:0060249)
1,33	18	5,00E-01	Regulation of protein binding (GO:0043393)

G. Identification des fonctions biologiques enrichies en gènes associés aux DMC/DMR identifiées

A partir des gènes identifiés comme étant associés aux DMC et/ou DMR, il a été possible d'effectuer des analyses d'enrichissement en utilisant la fonction *Functional Annotation Clustering* du programme DAVID. Pour chaque race prise séparément, l'analyse d'enrichissement a été réalisée sur l'ensemble des gènes associés à au moins une DMC et identifiée pour chaque comparaison 2 à 2 entre les catégories de fertilité. L'enrichissement est évalué en prenant en compte l'ensemble des gènes associés à l'ensemble des CpG analysés (*background* annoté). Nous avons déterminé les fonctions biologiques enrichies en gènes associés à au moins une DMC identifiée dans la comparaison Fertiles+/Fertiles-, qui nous sert de référentiel. Nous avons ensuite déterminé les voies enrichies pour les 2 autres comparaisons, à savoir Déviants/Fertiles+ et Déviants/Fertiles-. Les fonctions sont classées en fonction d'un EASE score, pour « *Expression Analysis Systematic Explorer* », qui correspond à une pvalue corrigée après test de Fisher. Il est considéré que les scores supérieurs à 1,3 sont des scores avec pvalue significative voire hautement significative.

Pour la race Holstein, entre les catégories **Fertiles+** et **Fertiles-**, nous avons pu identifier **22 groupements de termes GO** avec un EASE score de 1,3 à 2,94 (Figure 14.A). Six groupements de termes GO ont trait au système nerveux et à sa formation : « *Neuron differentiation* », « *Neuron part* », « *Synapse* », « *Modulation of symaptic transmission* », « *Somatodendritic compartment* », « *Sensory perception of mechanical stimulus* ». Il est aussi intéressant de remarquer la présence de groupements de termes GO concernant le développement du système cardiovasculaire, du squelette ou des organes sensoriels. Ceci suggère que les DMC sont associées à des gènes ayant un rôle dans différents processus biologiques reliés au développement embryonnaire. Quatre groupements de termes GO concernent des activités enzymatiques de la machinerie cellulaire : « *Phosphotransferase activity* », « *Protein serine/threonine kinase activity* », « *Protein-arginine deiminase activity* », « *Regulation of phosphatase activity* ». Par ailleurs, 5 groupements de termes GO concernent les interactions cellule-cellule et les mécanismes membranaires de transmission du stimulus : « *Cell-substrate junction* », « *Second messenger mediated signaling* », « *Regulation of cell communication* », « *Transmembrane receptor protein tyrosine kinase activity* », « *Signal release* », « *Cellular response to endogenous stimulus* ». De plus, des processus comme la *Locomotion* et la régulation de la motilité cellulaire (« *regulation of cell motility* ») arrivent respectivement en 4^{ème} et 15^{ème} position dans le classement. En comparant les

C

Note	Nombre de gènes	Significativité	Fonction biologique (Gene Ontology term)
5,38	65	3,00E-03	Regulation of cell migration (GO:0030334)
5,16	48	1,20E-03	Cell-substrate adherens junction (GO:0005924)
5,11	41	1,80E-02	Angiogenesis (GO:0001525)
5,09	198	3,20E-03	Small molecule binding (GO:0036094)
4,04	51	4,80E-02	Negative regulation of cell proliferation (GO:0008285)
3,88	22	4,90E-02	Positive regulation of neuron projection development (GO:0010976)
3,30	68	2,80E-02	Regulation of cellular component biogenesis (GO:0044087)
2,96	79	4,90E-02	Organ morphogenesis (GO:0009887)
2,80	34	4,80E-02	Positive regulation of cell adhesion (GO:0045785)
2,65	30	8,40E-02	Membrane region (GO:0098589)
2,50	87	2,20E-02	Embryo development (GO:0009790)
2,36	48	2,50E-03	Protein serine/threonine kinase activity (GO:0004674)
2,22	34	2,90E-02	Cell-substrate adhesion (GO:0031589)
2,22	14	7,40E-02	Basement membrane (GO:0005604)
2,14	24	1,40E-01	Respiratory system development (GO:0060541)
2,06	12	1,70E-01	Odontogenesis of dentin-containing tooth (GO:0042475)
2,04	30	1,10E-01	Plasma membrane organization (GO:0007009)
1,95	17	1,60E-01	Protein tyrosine kinase activity (GO:0004713)
1,85	23	1,60E-01	Modulation of synaptic transmission (GO:0050804)
1,78	12	1,30E-01	Stress fiber assembly (GO:0043149)
1,78	65	6,00E-02	Central nervous system development (GO:0007417)
1,74	24	2,90E-01	Negative regulation of cellular component movement (GO:0051271)
1,71	174	3,20E-02	Intracellular signal transduction (GO:0035556)
1,70	34	7,80E-02	Ameboidal-type cell migration (GO:0001667)
1,67	16	3,60E-01	Positive regulation of epithelial cell proliferation (GO:0050679)
1,66	19	2,70E-01	Regulation of protein binding (GO:0043393)
1,62	15	2,80E-01	Columnar/cuboidal epithelial cell differentiation (GO:0002065)
1,56	11	4,10E-01	Antiporter activity (GO:0015297)
1,48	24	1,20E-02	Regulation of cell size (GO:0008361)
1,45	24	2,40E-01	GTPase activator activity (GO:0005096)
1,42	4	1,90E-01	Protein-arginine deiminase activity (GO:0004668)
1,41	10	2,30E-01	Positive regulation of extrinsic apoptotic signaling pathway (GO:2001238)
1,38	8	2,20E-01	T cell selection (GO:0045058)
1,34	21	2,40E-01	Regulation of synapse structure or activity (GO:0050803)
1,31	10	1,90E-01	Regulation of peptidyl-lysine acetylation (GO:2000756)
1,31	21	1,10E-01	Central nervous system neuron differentiation (GO:0021953)
1,30	20	1,80E-01	Regulated exocytosis (GO:0045055)

Figure 12. Processus biologiques enrichis en DMC – enrichissements pour les comparaisons Fertiles+ contre Fertiles-, Déviants/Fertiles+, Déviants/Fertiles- – taureaux Holstein

L'analyse a été réalisée avec l'outil DAVID *Functional Annotation Clustering*, en utilisant les termes Gene Ontology *Biological Processes*, *Cellular Component* et *Molecular Function*. Les enrichissements sont faits par rapport aux gènes annotés sur le *background* de la race Holstein, soit 18 931 gènes. La comparaison Fertiles+ contre Fertiles- sert de témoin de comparaison pour étudier les processus biologiques spécifiquement enrichis entre Déviants/Fertiles+, et Déviants/Fertiles-. Pour la comparaison entre Fertiles+ contre Fertiles-, l'analyse est réalisée sur 1192 gènes différents comportant au moins une DMC entre ces catégories. Pour la comparaison entre Déviants/Fertiles+, l'analyse est réalisée sur 1357 gènes différents comportant au moins une DMC entre ces catégories. Pour la comparaison entre Déviants-/Fertiles-, l'analyse est réalisée sur 1419 gènes différents comportant au moins une DMC entre ces catégories. Les groupements de termes GO surlignés en vert correspondent aux processus biologiques qui interviennent pendant le développement embryonnaire ; en bleu ceux qui sont relatifs au système nerveux ; en rouge ceux relatifs à la motilité cellulaire, et en jaune ceux relatifs aux activités enzymatiques de la machinerie cellulaire.

catégories **Déviants** et **Fertiles+**, on obtient **25 regroupements de termes GO** significatifs avec un EASE score de 1,33 à 4,13 (Figure 14.B). Certaines fonctions biologiques décrites précédemment sont conservées, voire mieux représentées : Concernant le système nerveux, 9 groupements de termes GO sont identifiés dont 3 ont été mentionnés pour la comparaison Fertile+/Fertiles- (« *Neurone part* », « *Modulation of synaptic transmission* », « *somatodendritic compartment* »). Le regroupement de gènes touchant au développement du système cardiovasculaire est non seulement conservé, mais aussi plus significativement enrichi dans cette comparaison (note de 2,06 contre 1,55 pour le témoin de comparaison). Comme dans la précédente analyse d'enrichissement, on dénombre aussi plusieurs groupements de termes GO concernant les activités enzymatiques de la machinerie cellulaire : parmi les 6 regroupements, un seul est commun aux deux comparaisons (« *Phosphotransferase activity, alcohol group as acceptor* »). Le processus biologique le plus significatif est « *enzyme activator activity* », qui regroupe des gènes impliqués dans l'augmentation de l'activité des enzymes. Quand on compare les **Déviants** et les **Fertiles-**, on obtient cette fois **37 regroupements de termes GO** significatifs, avec un EASE score de 1,3 à 5,38 soit plus de 50% de plus qu'avec la comparaison témoin (Figure 14.C). Parmi ces regroupements, 8/37 sont en commun avec ceux précédemment mis en évidence : « *Modulation of synaptic transmission* », commun aux 3 comparaisons), « *Cell-substrate adherens junction* », « *Protein serine/threonine kinase activity* », « *Stress fiber assembly* », et « *Protein-arginine deiminase activity* », communs aux comparaisons Fertiles+/Fertiles- et Déviants/Fertiles- et enfin « *Protein tyrosine kinase activity* », « *Regulation of synapse structure or activity* », « *Regulation of protein binding* », communs aux comparaisons Déviants/Fertiles+ et Déviants/Fertiles-. Il est à noter une forte proportion de regroupements de termes GO concernant le développement embryonnaire (« *Embryo development* », « *Organ morphogenesis* », « *Angiogenesis* », « *Odontogenesis of dentin-containing tooth* », « *Respiratory system development* », « *Central nervous system development* », « *Central nervous system neuron differentiation* »), la prolifération/différenciation/apoptose cellulaire (« *Negative regulation of cell proliferation* », « *Regulation of cellular component biogenesis* », « *Positive regulation of epithelial cell proliferation* », « *Columnar/cuboidal epithelial cell differentiation* », « *Positive regulation of extrinsic apoptotic signaling pathway* », « *Regulation of cell size* ») et les interactions cellule-cellule incluant la migration cellulaire avec un EASE score extrêmement significatif de 5,31 (« *Regulation of cell migration* », « *Cell-substrate adherens junction* », « *Positive regulation of cell adhesion* », « *Cell-substrate adhesion* », « *Ameboidal-type cell migration* »).

A

Fertiles+ contre Fertiles-

Note	Nombre de gènes	Significativité	Fonction biologique (Gene Ontology term)
3,24	11	3,90E-02	Negative regulation of angiogenesis (GO:0016525)
3,13	32	5,10E-02	Response to growth factor (GO:0070848)
2,65	25	4,20E-02	Proteinaceous extracellular matrix (GO:0005578)
2,40	46	2,30E-02	Regulation of cellular component biogenesis (GO:0044087)
2,15	64	5,00E-02	Cell adhesion (GO:0007155)
2,06	78	3,30E-02	Regulation of multicellular organismal development (GO:2000026)
2,00	15	4,50E-01	Cell cortex (GO:0005938)
1,96	28	5,10E-02	Cell surface receptor signaling pathway involved in cell-cell signaling (GO:1905114)
1,95	41	1,60E-01	Regulation of cell development (GO:0060284)
1,71	57	6,90E-01	Molecular function regulator (GO:0098772)
1,67	24	4,50E-01	Cell-substrate adherens junction (GO:0005924)
1,62	98	2,60E-02	Cell surface receptor signaling pathway (GO:0007166)
1,60	65	1,20E-01	Locomotion (GO:0040011)
1,59	96	1,60E-01	Cellular localization (GO:0051641)
1,53	13	5,60E-01	Regulation of Ras protein signal transduction (GO:0046578)
1,46	24	3,90E-02	Cell-substrate adhesion (GO:0031589)
1,41	16	3,20E-01	Extrinsic apoptotic signaling pathway (GO:0097191)
1,40	94	2,70E-01	Regulation of protein metabolic process (GO:0051246)
1,38	9	4,70E-01	Cardiocyte differentiation (GO:0035051)

B

Déviants contre Fertiles+

Note	Nombre de gènes	Significativité	Fonction biologique (Gene Ontology term)
3,04	123	1,80E-01	Regulation of signaling (GO:0023051)
2,28	60	2,20E-01	Molecular function regulator (GO:0098772)
1,99	11	3,20E-01	Rho guanyl-nucleotide exchange factor activity (GO:0005089)
1,99	23	2,20E-01	Phospholipid binding (GO:0005543)
1,66	41	5,30E-01	Regulation of cellular component biogenesis (GO:0044087)
1,43	94	1,10E-01	Nervous system development (GO:0007399)
1,39	11	7,10E-01	Inositol lipid-mediated signaling (GO:0048017)
1,32	69	5,30E-01	ATP binding (GO:0005524)
1,31	60	2,30E-01	Neuron differentiation (GO:0030182)

C

Déviants contre Fertiles-

Note	Nombre de gènes	Significativité	Fonction biologique (Gene Ontology term)
4,35	22	1,10E-03	Guanyl-nucleotide exchange factor activity (GO:0005085)
1,65	66	6,70E-01	Cell proliferation (GO:0008283)
1,62	7	7,00E-01	Basement membrane (GO:0005604)
1,46	52	6,50E-01	Response to endogenous stimulus (GO:0009719)
1,41	63	6,40E-01	Locomotion (GO:0040011)
1,41	17	6,10E-01	GTPase regulator activity (GO:0030695)
1,34	114	3,90E-01	Regulation of signaling (GO:0023051)
1,34	27	6,90E-01	Response to growth factor (GO:0070848)
1,31	15	7,70E-01	Epithelial cell development (GO:0002064)
1,30	12	7,20E-01	Cell cortex (GO:0005938)

Figure 13. Processus biologiques enrichis en DMC – enrichissements pour les comparaisons Fertiles+ contre Fertiles-, Déviants/Fertiles+ et Déviants/Fertiles- – taureaux Montbéliard

L'analyse a été réalisée avec l'outil DAVID *Functional Annotation Clustering*, en utilisant les termes Gene Ontology *Biological Processes*, *Cellular Component* et *Molecular Function*. Les enrichissements sont faits par rapport aux gènes annotés sur le *background* de la race Montbéliard, soit 18 802 gènes. La comparaison Fertiles+ contre Fertiles- sert de témoin de comparaison pour étudier les processus biologiques spécifiquement enrichis entre Déviants/Fertiles+, ainsi que Déviants/Fertiles-. Pour la comparaison entre Fertiles+ contre Fertiles-, l'analyse est réalisée sur 783 gènes différents comportant au moins une DMC entre ces catégories. Pour les comparaisons entre Déviants/Fertiles+ et Déviants/Fertiles-, l'analyse est réalisée sur respectivement 832 et 788 gènes différents comportant au moins une DMC. Les groupements de termes GO surlignés en vert correspondent aux processus biologiques qui interviennent pendant le développement embryonnaire ; en bleu ceux qui sont relatifs au système nerveux ; en rouge ceux relatifs à la motilité cellulaire, et en jaune ceux relatifs aux activités enzymatiques de la machinerie cellulaire.

Pour les taureaux Montbéliard, la comparaison **Fertiles+ contre Fertiles-**, permet d'identifier **19 groupements de termes GO** significativement enrichis en gènes associés aux DMC, avec un EASE score de 1,38 à 3,24 (Figure 15.A). Un seul regroupement est commun avec ceux observés pour les taureaux Holstein dans le cadre du même type de comparaison : le regroupement « *locomotion* ». Le regroupement le plus significatif concerne la régulation négative de l'angiogenèse (avec un EASE score de 3,24). On peut observer des regroupements de termes GO concernant les interactions cellule-cellule et cellule-matrice extracellulaire (« *Proteinaceous extracellular matrix* », « *Cell adhesion* », « *Cell surface receptor signaling pathway involved in cell-cell signaling* », « *Cell-substrate adherens junction* », « *Cell surface receptor signaling pathway* », « *Cell-substrate adhesion* »). Pour la comparaison entre **Déviants et Fertiles+**, seuls **9 regroupements de termes GO** présentent un EASE score de 1,3 à 3,04 (Figure 15.B). Néanmoins, aucune fonction ne présente une p-value significative. Deux regroupements concernent le système nerveux (« *Nervous system developmen* », « *Neuron differentiation* ») et trois la signalisation intracellulaire (« *Regulation of signaling* », « *Rho guanyl-nucleotide exchange factor activity* », « *Inositol lipid-mediated signaling* »). Enfin, quand on s'intéresse à la comparaison entre **Déviants et Fertiles-**, seuls **10 regroupements de termes GO** ont un score entre 1,3 et 4,35 (Figure 14.C). Deux sont communs à ceux observés pour la comparaison Fertiles+/Fertiles- (« *Response to growth factor* », « *locomotion* ») et un à ceux issus de la comparaison Déviants/Fertiles+ (« *Regulation of signaling* »). Cependant, le premier regroupement de termes GO, « *Guanyl-nucleotide exchange factor activity* » est très significatif avec un score de 4,35 et une p-value ajustée de 0,0011. Ce terme GO correspond aux gènes codants pour des protéines dont la fonction est de stimuler l'échange de nucléotides guanine des enzymes GTPases. Ce regroupement peut être associé à celui intitulé « *GTPase regulator activity* ».

Pour la race Holstein, les modifications du méthylome ciblent un nombre de fonctions plus important chez les déviants que chez les Fertiles-. L'ensemble des données obtenues par ces analyses de fonctions biologiques enrichies en gènes associés aux DMC suggère que les fonctions les plus touchées sont des fonctions ayant trait au développement embryonnaire, mettant en jeu des processus de prolifération/différenciation cellulaire et d'interactions cellule-cellule. Peu de fonctions semblent être impliquées dans le fonctionnement du spermatozoïde. Ainsi les altérations du méthylome des taureaux Déviants pourraient jouer un rôle important dans la transmission à la descendance d'informations délétères et dans l'arrêt du développement à divers stades.

Gène	Description
<i>Retinoic Acid Receptor Beta</i> (<i>RARB</i>)	- Activation de médiateurs de la signalisation cellulaire dans la morphogénèse embryonnaire - Rôle dans la spermiogénèse et la différenciation des spermatoocytes
<i>Smad Nuclear Interacting Protein 1</i> (<i>SNIP1</i>)	- Impliqué dans la voie de signalisation Notch qui régule la différenciation cellulaire au cours du développement
<i>Transducin Like Enhancer Of Split 3</i> (<i>TLE3</i>)	- Régule la stabilité de l'ARNm de la cycline D1 (cycle cellulaire) - Répresseur de la transcription pour les voies de signalisation TGF- β impliquées dans la croissance et la différenciation cellulaire
<i>WW And C2 Domain Containing 2</i> (<i>WWC2</i>)	- Module la voie de signalisation Hippo, qui contrôle la croissance des organes par la régulation de la prolifération cellulaire et l'apoptose
<i>Laminin Subunit Beta 2</i> (<i>LAMB2</i>)	- Impliqué dans les processus d'adhésion cellulaire, de différenciation, de migration et de signalisation cellulaire

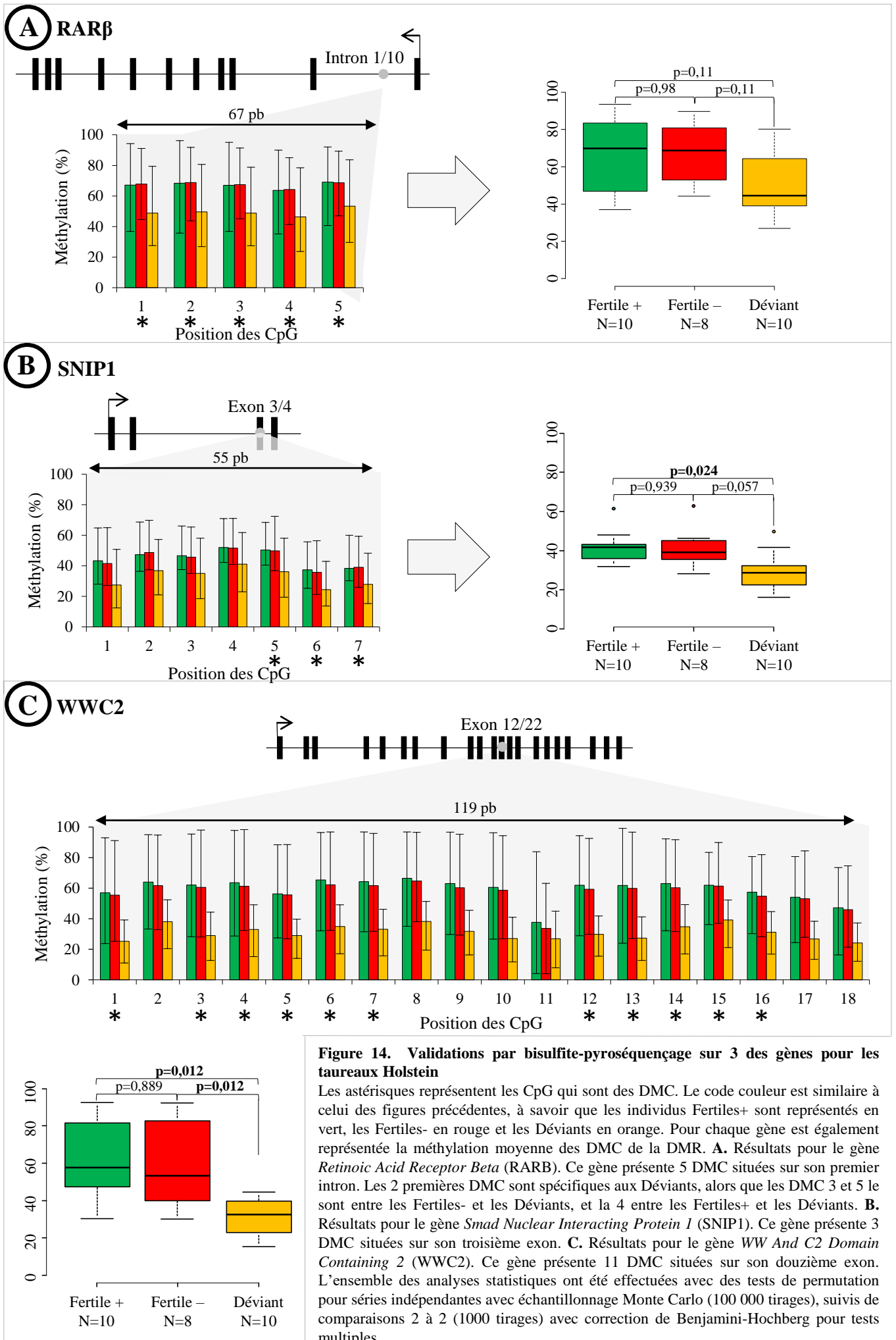
Tableau 2. Implications biologiques des gènes présentant une DMR et validés par bisulfite-pyrosequencing

H. Validation des DMR par bisulfite-pyroséquence

Toute analyse haut-débit nécessite une validation en utilisant une autre technologie. Nous avons donc décidé de procéder à la validation de plusieurs DMR en utilisant une approche séquence-spécifique, par bisulfite-pyroséquence. Pour la race Holstein, 4 DMR ont été sélectionnées. Celles des gènes *WW And C2 Domain Containing 2 (WWC2)* et *Retinoic Acid Receptor Beta (RARβ)* ont été sélectionnées car elles sont observées spécifiquement hypométhylées chez les taureaux Déviants par rapport aux taureaux Fertiles+ et Fertiles-. Les gènes *Smad Nuclear Interacting Protein 1 (SNIP1)* et *Transducin Like Enhancer Of Split 3 (TLE3)* présentent quant à eux une DMR comportant un mélange de DMC présentant un taux de méthylation soit spécifique aux Déviants, soit différent pour chaque catégorie (Déviants, Fertiles+, Fertiles-). Ces 4 gènes jouent un rôle important dans le développement embryonnaire précoce et la spermatogenèse (voir Tableau 1). Au cours des analyses RRBS qui ont conduit à la sélection de ces DMR, 2 seuils différents de couverture ont été utilisés. Les CpG de la DMR de *WWC2* sont couverts par plus de 10 séquences chez 8/10 individus au minimum pour chacun des groupes de fertilité (avec 8/8 individus chez les Fertiles-). Les CpG des DMR des gènes *RARβ*, *SNIP1* et *TLE3* sont quant à eux couverts par plus de 10 séquences pour au minimum 4 individus parmi chaque catégorie de fertilité. Deux autres gènes ont été sélectionnés car ils présentaient au moins une DMC commune pour les races Holstein et Montbéliard, pour la comparaison entre Fertiles+ contre Déviants. Ces gènes sont les *Laminin Subunit Beta 2 (LAMB2)* et *SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily D, Member 3 (SMARCD3)*.

Les résultats des analyses CpG par CpG sont représentés Figures 16 et 17. La première chose que l'on observe est la variabilité inter-individuelle très importante pour chacun des gènes et chacune des catégories de fertilité. Pour un CpG donné et quel que soit le gène, les moyennes entre catégories sont toujours différentes mais les écarts-types sont d'une telle amplitude qu'il n'est pas possible de conclure à une différence significative. En utilisant le logiciel methylKit, le calcul du différentiel de méthylation est basé sur la moyenne par catégories sans tenir compte de la variabilité intra-groupe. Ce mode de calcul masque donc toute variabilité inter-individus.

Si on s'intéresse à la méthylation moyenne des CpG de la DMR par catégories, il en ressort que les taureaux Déviants présentent systématiquement une hypométhylation. Seul *WWC2* conserve un différentiel de méthylation significatif entre les taureaux Déviants et les 2 autres catégories de fertilité (Figure 16). Un différentiel de méthylation statistiquement significatif



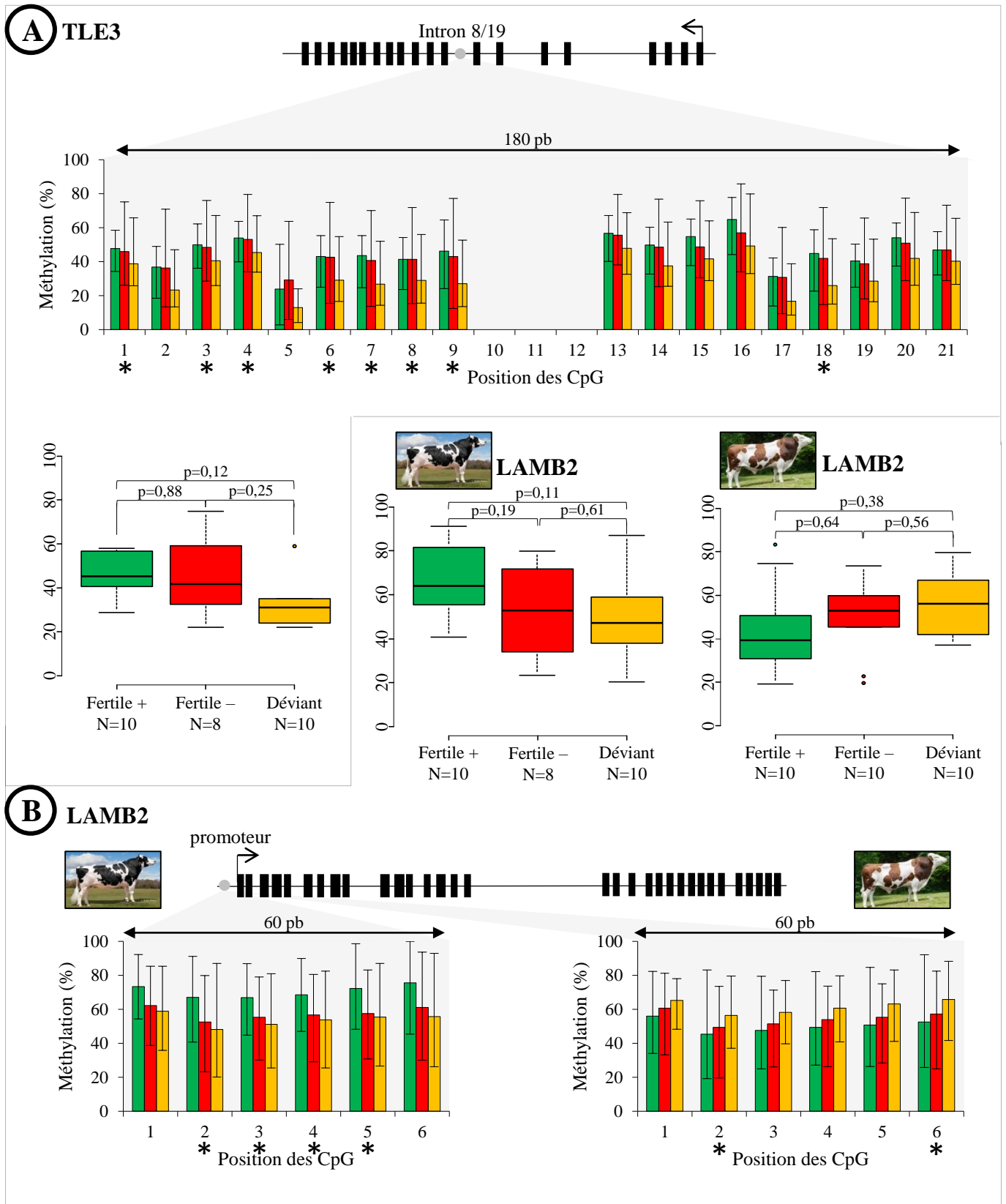


Figure 15. Validations par bisulfite-pyrosequençage sur 1 autre gène pour les taureaux Holstein, et le gène LAMB2 pour les taureaux Holstein et Montbéliard

Les astérisques représentent les CpG qui sont des DMC. Le code couleur est similaire à celui des figures précédentes, à savoir que les individus Fertiles+ sont représentés en vert, les Fertiles- en rouge et les Déviants en orange. Pour chaque gène est également représentée la méthylation moyenne des DMC de la DMR. **A.** Résultats pour le gène *Transducin Like Enhancer Of Split 3* (TLE3). Ce gène présente 8 DMC situées sur son huitième intron. **B.** Résultats pour le gène *Laminin Subunit Beta 2* (LAMB2). Ce gène présente 4 DMC pour la race Holstein et 2 DMC pour la race Montbéliard. Ces DMC sont situées dans le promoteur du gène. La deuxième DMC est celle qui est commune entre les 2 races. L'ensemble des analyses statistiques ont été effectuées avec des tests de permutation pour séries indépendantes avec échantillonnage Monte Carlo (100 000 tirages), suivis de comparaisons 2 à 2 (1000 tirages) avec correction de Benjamini-Hochberg pour tests multiples.

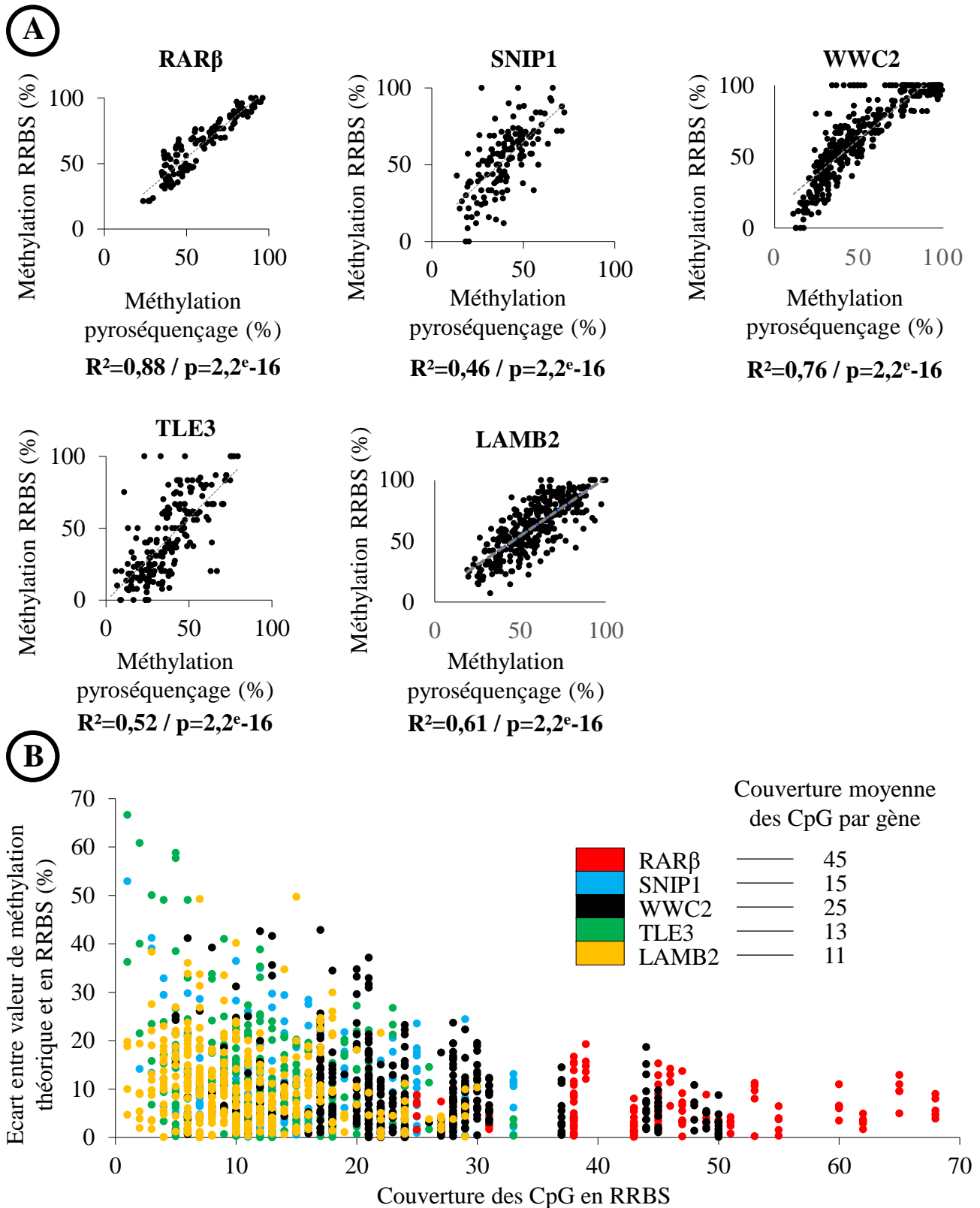


Figure 16. Corrélation entre les résultats de RRBS et de bisulfite-pyroséquenceage

A. Graphes de corrélation entre la valeur de méthylation calculée par RRBS par rapport à celle calculée par pyroséquenceage. Chaque point représente un CpG analysé pour un individu donné. En fonction du gène, le R^2 varie de 0,88 à 0,52. Les tests de corrélation de rang de Spearman donnent systématiquement la valeur maximale possible pour chacun des tests. **B.** Graphe représentant la précision de la mesure de la méthylation en RRBS par rapport à la couverture des CpG. Les valeurs en ordonnées correspondent à l'écart entre le pourcentage de méthylation calculé en RRBS et un pourcentage de méthylation théorique calculé à partir des valeurs de pyroséquenceage. Ainsi, plus la valeur est faible, plus la mesure en RRBS est satisfaisante. La plupart des pourcentages des méthylation calculés en RRBS sont à moins de 20% d'écart avec la valeur théorique. Cependant, avec une couverture des CpG inférieure à 30, une partie des mesures présente un écart de valeur de plus de 30%, pouvant aller jusqu'à près de 70% dans quelques cas. L'écart de valeur se stabilise après une couverture de 30 à 40, pour se fixer en dessous de 20%.

entre les Déviants et les Fertiles+ pour le gène *SNIP1* est aussi observé. Ces tendances à l'hypométhylation chez les taureaux Déviants sont similaires à celles observées en RRBS, bien que le différentiel de méthylation mis en évidence par pyroséquençage, soit moins important. Les résultats pour le gène *SMARCD3* ne sont pas représentés ici car ce gène présente certaines particularités. Celui est donc traité séparément, dans le chapitre suivant.

Les résultats pour le gène *LAMB2* sont représentés séparément pour les races Holstein et Montbéliarde. Alors que l'on observe une hypométhylation chez les Déviants Holstein par rapport aux autres catégories de fertilité, on note une hyperméthylation chez les Déviants Montbéliard par rapport aux autres catégories. Il semble donc que sur certains gènes, les profils de méthylation divergent entre les races, bien qu'il y ait un différentiel entre les catégories de fertilité.

Afin de contrôler la bonne adéquation entre les pourcentages de méthylation obtenus en RRBS et ceux obtenus en pyroséquençage, nous avons pour chaque CpG analysé, chaque individu et chacun des gènes, corrélé les résultats obtenus par les 2 méthodes d'analyse. Les résultats sont présentés Figure 18.A. Quel que soit le gène étudié, en utilisant un test de corrélation de Spearman les résultats de corrélation sont statistiquement significatifs. Cela signifie que le pourcentage de méthylation calculé par RRBS est très proche de celui calculé par le pyroséquençage, L'importance de la couverture des CpG pour le calcul du pourcentage de méthylation est illustrée par la Figure 18.B. Pour cela, nous avons calculé l'écart entre le taux de méthylation des CpG observé par RRBS et sa valeur théorique déduite de la courbe de corrélation. L'écart pour chaque CpG des cinq gènes analysés a été reporté en fonction de la couverture obtenue après séquençage. Il apparaît nettement que plus la couverture des CpG augmente, plus l'écart diminue, suggérant que l'erreur sur le calcul du taux de méthylation par RRBS diminue aussi. La grande majorité des CpG analysés présentent une erreur de calcul inférieure à 25%. En dessous d'une couverture de 10, l'erreur peut être très importante pour quelques CpG. L'erreur se stabilise en dessous de 20% au-delà d'une couverture d'environ 30. Pour toutes nos analyses nous avons conservé les CpG couverts plus de 10 fois ; cela paraît un compromis acceptable pour avoir à la fois un nombre important de CpG analysés et une couverture pour laquelle l'erreur de mesure n'est pas trop importante.

Pris dans leur ensemble, ces résultats montrent que les DMC/DMR identifiées par RRBS ont potentiellement des profils de méthylation différents entre catégories de fertilité. Néanmoins dans le cadre d'identification de DMC comme biomarqueurs, il serait certainement pertinent de refaire des analyses en augmentant le seuil de différentiel de méthylation afin d'éviter les risques de faux positifs.

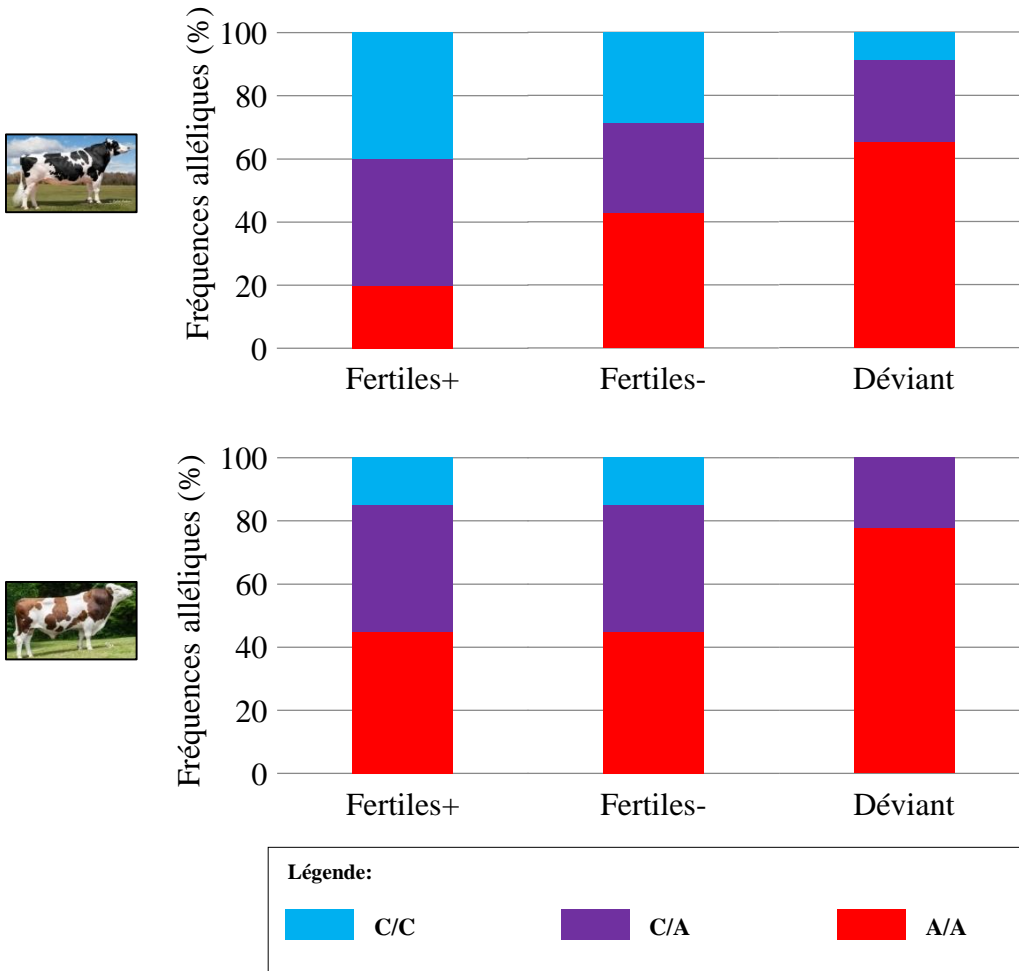


Figure 17 Répartition des polymorphismes de séquence sur la DMC du gène SMARCD3

Pour les taureaux de race Holstein, après séquençage de la région du gène SMARCD3 nous avons observé 6 individus C/C, 3 individus C/A et un individu A/A pour la catégorie Fertiles+ ; et respectivement 4 individus A/A, 2 C/A et 2 A/A pour la catégorie Fertiles- ; pour la catégorie des Déviants les proportions sont inversées avec 2 individus C/C, 3 individus C/A et 5 individus A/A. Pour les taureaux de race Montbéliard, nous avons noté pour les catégories Fertiles+ et Fertiles-, 4 individus A/A, 4 individus C/A et 3 individu C/C. Pour les Déviants, on observe 7 individus A/A et 3 individus C/A. Ces polymorphismes en fonction de la catégorie de fertilité, situés directement sur la cytosine participant à la DMC, peuvent biaiser les calculs du pourcentage de méthylation en RRBS. Ces résultats peuvent expliquer le différentiel de méthylation identifié par RRBS alors qu'il ne l'est pas en pyroséquençage.

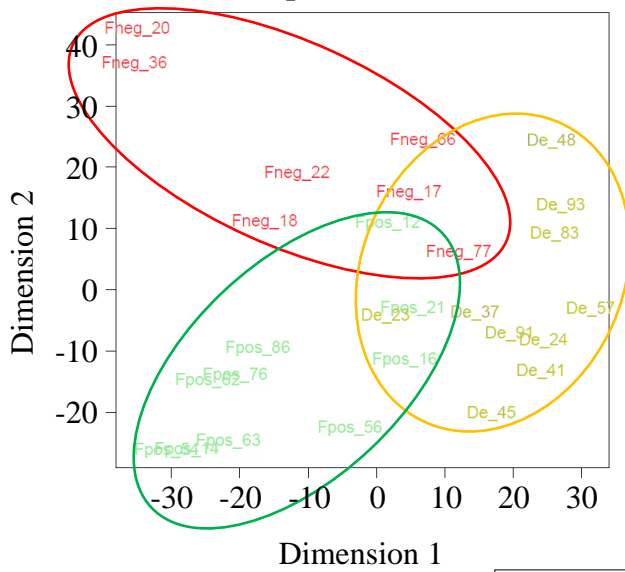
I. Intégration des données de SNP et de DMC

Pour le gène *SMARCD3*, l'analyse des pyrogrammes suggère la présence d'un SNP sur la cytosine participant à la DMC. En amont de la position C/T à lire par pyroséquençage, une adénine, n'existant pas dans la séquence prévue, est distribuée et doit servir de témoin négatif d'incorporation. Pour certains individus, les profils d'incorporation obtenus présentaient soit une absence d'incorporation de A et une incorporation variable de C et de T (en fonction de l'état de méthylation initial), soit une incorporation équivalente de A et de C/T, soit une incorporation de A et une absence complète d'incorporation de C/T. Le séquençage de la région selon la méthode de Sanger nous a permis de valider définitivement la présence d'un SNP expliquant les différentiels de méthylation entre les catégories de fertilité. Les résultats concernant la distribution des polymorphismes en fonction de la catégorie de fertilité est donné Figure 19. On observe clairement que les Déviants sont majoritairement homozygotes A/A pour les deux races. Au contraire, les individus Fertiles+ et Fertiles- sont plutôt homozygotes C/C et hétérozygotes C/A. La distribution allélique est significativement différente entre catégories de fertilité pour les deux races. Le différentiel de méthylation sur cette position entre catégories de fertilité est donc certainement dû au polymorphisme de séquence. Ces données montrent la nécessité fondamentale de prendre en considération l'information génétique dans l'interprétation des différentiels de méthylation observés. Ainsi, en comparant les données de génotypage des taureaux et les positions des DMC, il s'est avéré que 42 SNP sont situés directement sur des DMC (sur le C ou le G) pour la race Holstein, et 25 pour les Montbéliard. Parmi ces SNP, 10 sont communs entre les 2 races.

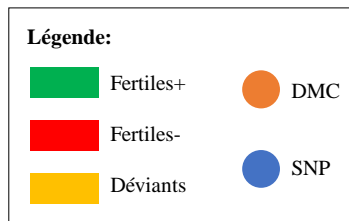
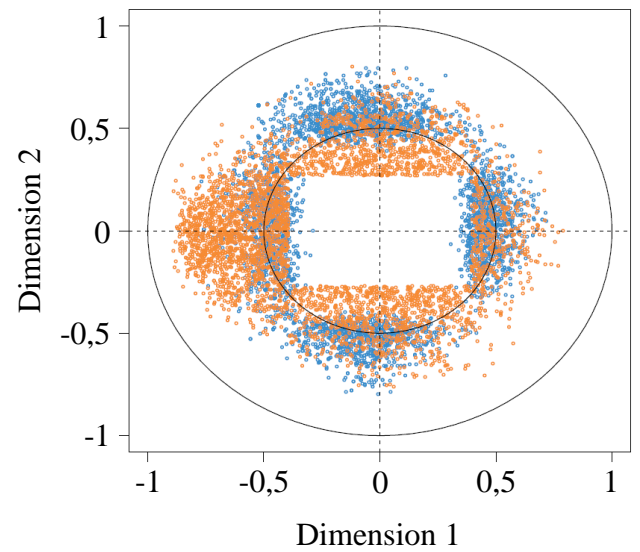
Suite à ces résultats, afin de mieux évaluer les interactions entre haplotypes et DMC, Hélène Kiefer et Luc Jouneau ont mis en œuvre une approche intégrative de type « *Sparse Partial Least Squares Regression* » (SPLS). En corrélant l'information de méthylation des DMC et les polymorphismes, ils ont identifié pour la race Holstein 296 964 couples DMC/SNP avec une corrélation supérieure à $\pm 0,5$. Les 2000 DMC et les 2000 SNP qui contribuent le plus à maximiser la covariance ont été sélectionnés et utilisés pour effectuer une SPLS (représentée Figure 20). On observe que pour la race Holstein, la structure en groupe de fertilité est relativement conservée. Cela montre qu'il existe alors potentiellement un lien entre le taux de méthylation de ces DMC et le polymorphisme génétique pour ces individus. Pour la race Montbéliarde, on observe sur la PLS que la structure des groupes n'est pas préservée. Les DMC et SNP corrélés entre eux ne séparent donc pas les catégories de fertilité. Le lien entre



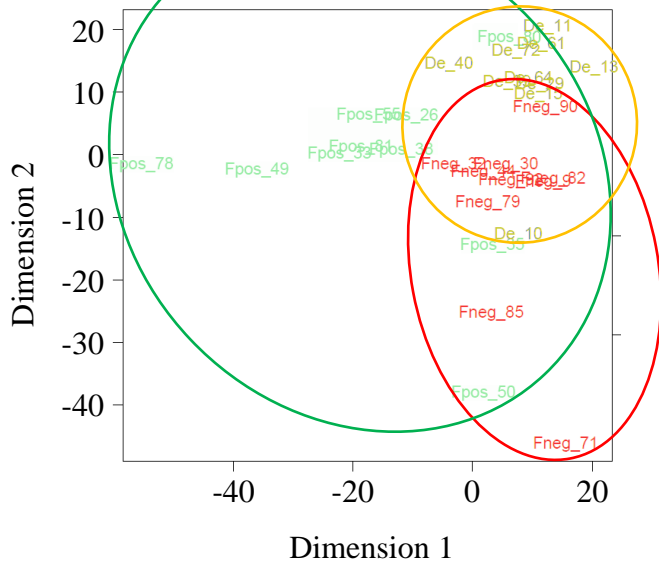
Graphe des individus



Graphe des variables



Graphe des individus



Graphe des variables

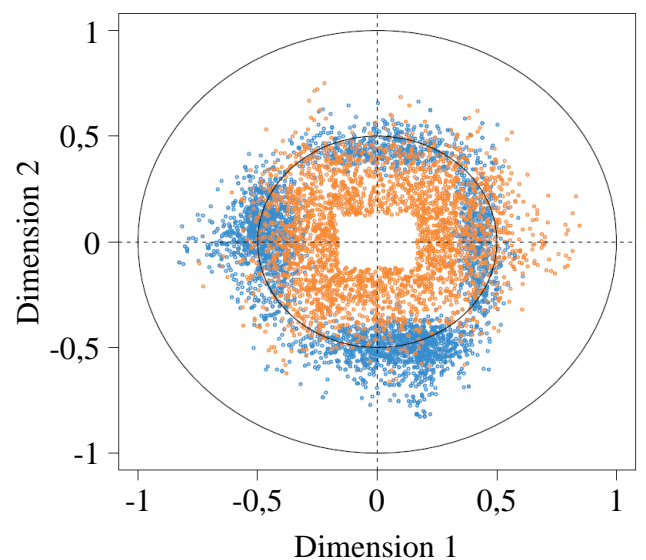


Figure 18 Intégration des données de génotypage aux DMC

Sur le graphe des individus le code couleur est similaire à celui des figures précédentes, à savoir que les individus Fertiles+ sont représentés en vert, les Fertiles- en rouge et les Déviants en orange. Pour les taureaux de race Holstein, en intégrant les données de génotypage aux données de DMC on conserve la structure en groupes de fertilité (sur le graphe des individus). Ce n'est pas le cas pour la race Montbéliard, où les catégories de fertilité se recoupent totalement. Sur le graphe des variables pour la race Holstein, les DMC sont majoritairement regroupées sur la gauche sur la dimension 1. Ce n'est pas le cas pour la race Montbéliard.

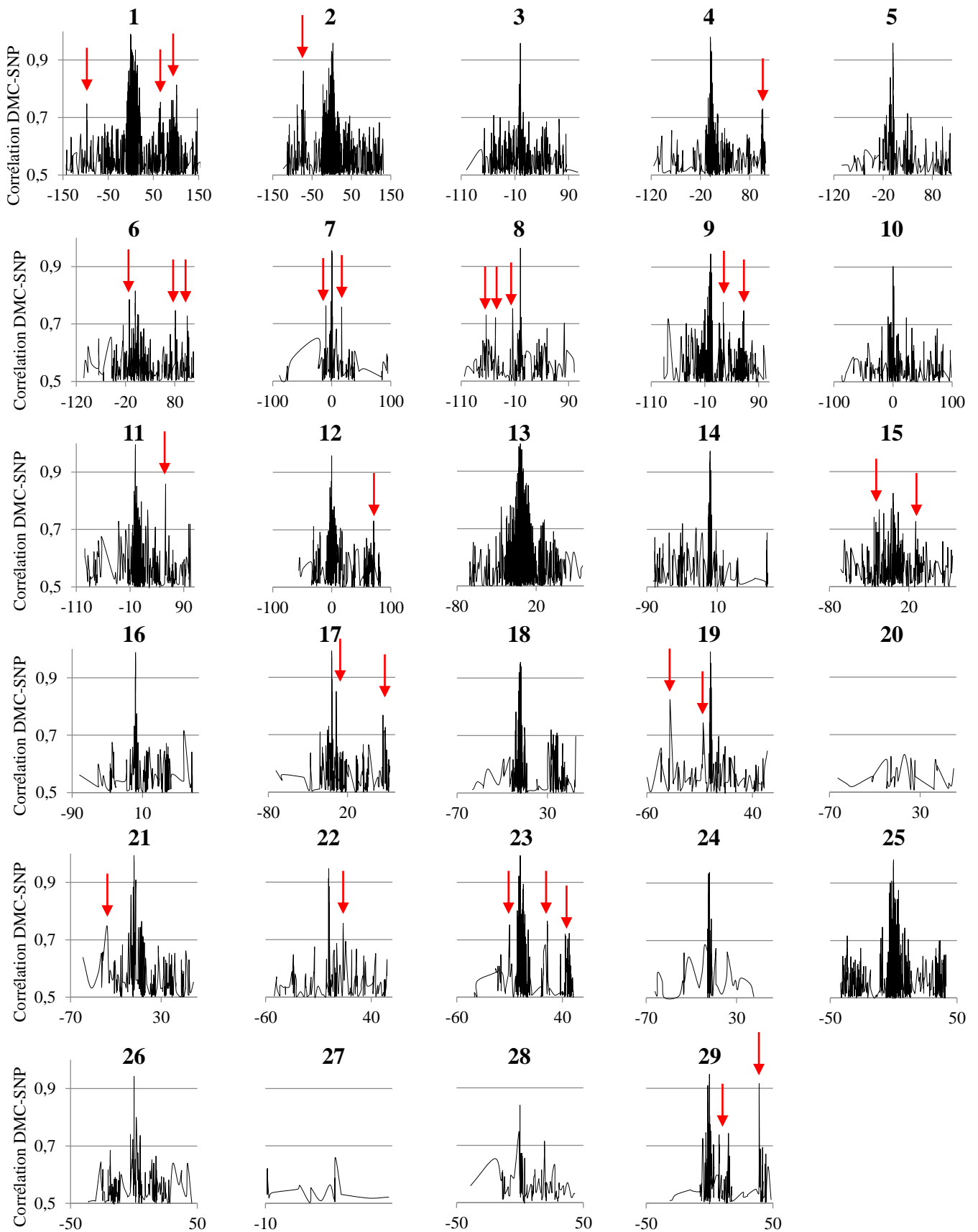


Figure 19. Relation spatiale entre les DMC et les SNP pour la race Holstein

Les données sont représentées chromosome par chromosome, en utilisant les 2000 DMC et les 2000 SNP les plus corrélés. L'axe des ordonnées correspond à la distance entre les DMC et SNP en millions de paires de bases. On observe systématiquement un maximum de corrélation quand le SNP est sur la DMC (distance de 0). On note également des effets à distance, symbolisés par les flèches rouges.

l'état de méthylation et le polymorphisme génétique n'est alors pas identifiable. Sur le graphe des variables pour la race Holstein, la dimension 1 représente la contribution des variables à la dimension 1 du graphe des individus. On peut observer que la majorité des DMC se situent sur la partie de gauche du graphe. Il semble alors qu'une part importante des DMC corrélés aux SNP se comportent de la même manière, et doivent donc présenter un différentiel de méthylation allant dans le même sens. Il apparaît alors que ces DMC sont celles qui sont majoritairement hypométhylées chez les taureaux Déviants. Sur le graphe des variables pour la race Montbéliarde, on observe que les DMC sont peu contributeurs sur la dimension 1.

Pour la race Holstein, les liens qui unissent les 2000 DMC et SNP les plus corrélés, peuvent être représentés chromosome par chromosome (Figure 21). On peut observer que pour chaque chromosome, la corrélation est maximale quand le SNP est directement sur la DMC (position « 0 »). De plus, une très grande majorité des 2000 SNP identifiés par l'approche PLS se situent à quelques Mb des DMC avec lesquelles ils sont corrélés. Cependant, on peut également observer de façon ponctuelle de fortes corrélations à distance entre DMC et SNP. Notamment, 170 SNP semblent contrôler la méthylation de DMC situées sur un autre chromosome (effet *trans*). Un d'entre eux semble particulièrement intéressant. Situé dans le gène *Aldehyde Oxidase 1-like (AOX4 ; position chr2 :89752875)*, ce SNP est fortement corrélé à 31 DMC qui semblent se regrouper en *clusters*. On peut alors émettre l'hypothèse que certains SNP peuvent contrôler l'état de méthylation de plusieurs régions du génome.

J. Intérêts d'utiliser une double digestion enzymatique HaeIII + MspI, pour construire une banque RRBS chez le bovin

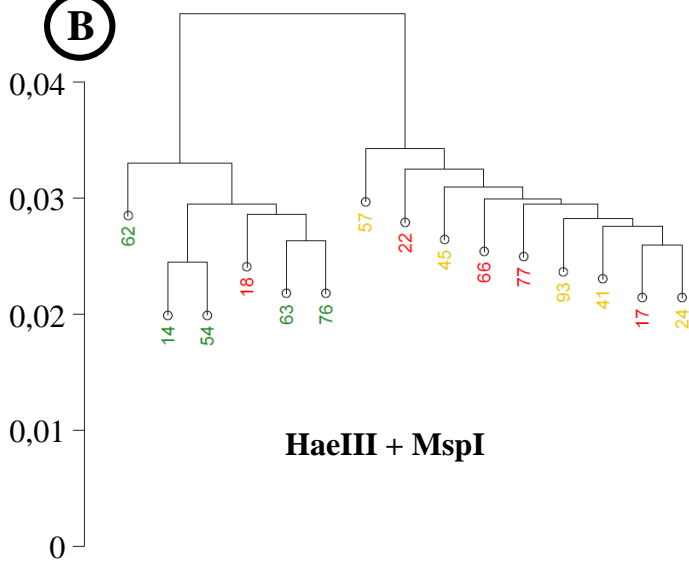
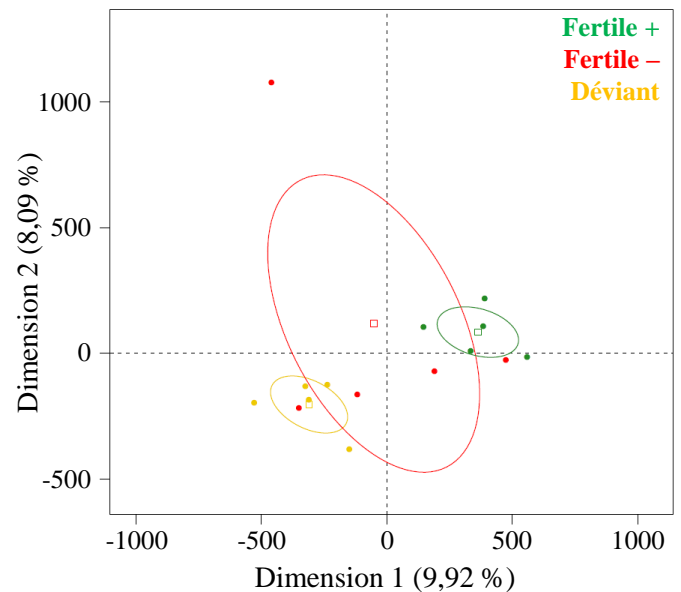
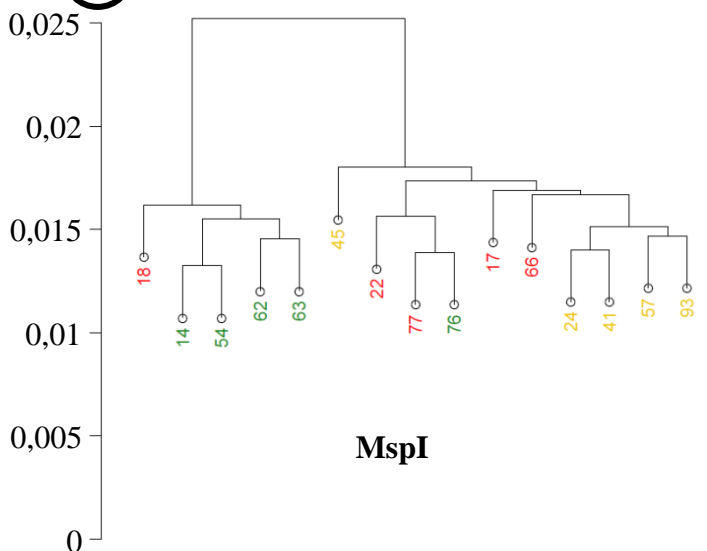
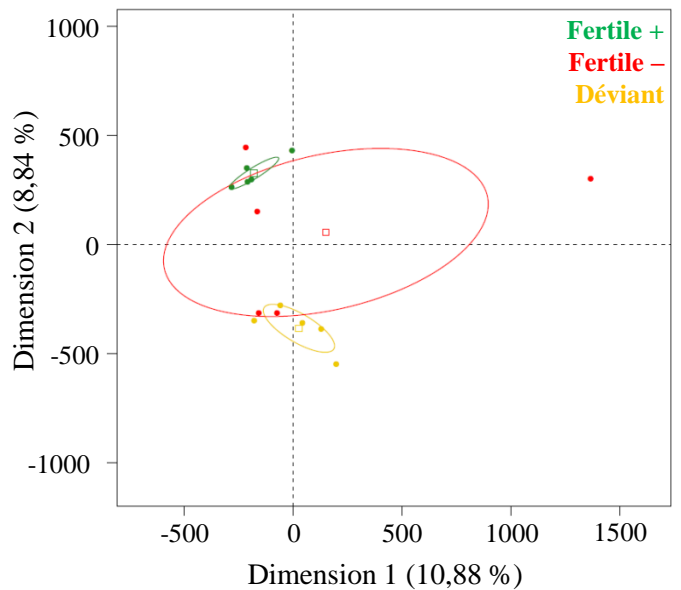
Différentes publications suggéraient l'utilisation d'une double digestion enzymatique afin de réduire la part de séquences répétées non analysables (Wang et al. 2013 ; Chatterjee et al. 2013 ; Martinez-Arguelles et al. 2014). Suite à l'analyse *in silico* d'une banque réalisée avec une double digestion enzymatique HaeIII en combinaison avec MspI, nous avons observé que celle-ci ne permet pas d'éliminer efficacement les séquences répétées chez les bovins, notamment BTSAT4 qui est représentée de façon importante (voir Résultats article n°1). En revanche, la double digestion enzymatique HaeIII+MspI permet d'enrichir considérablement les banques en CpG analysables (Annexe 1), c'est pourquoi nous l'avons sélectionnée pour effectuer des analyses complémentaires.

A**HaeIII
+ MspI
N = 15**

	Nombre de paires de séquences analysées	Taux de conversion bisulfite (%)	Alignements uniques (%)	Alignements multiples (%)	Pas d'alignement (%)	Nombre de CpG couverts	Nombre de CpG sur le chromosome Y	Nombre de CpG avec couverture ≥ 10 (%)	Nombre de CpG avec couverture ≥ 500 (%)
Moyenne	47 135 519	99,3	50,4	42,3	7,3	6 499 180	8722	40,1	0,2
Minimum	34 811 154	98,4	48,5	39,3	6,5	6 304 876	7805	27,1	0,1
Maximum	55 490 243	99,7	53,4	44,7	7,8	6 726 566	9692	45,7	0,3

**MspI
N = 15**

	Nombre de paires de séquences analysées	Taux de conversion bisulfite (%)	Alignements uniques (%)	Alignements multiples (%)	Pas d'alignement (%)	Nombre de CpG couverts	Nombre de CpG sur le chromosome Y	Nombre de CpG avec couverture ≥ 10 (%)	Nombre de CpG avec couverture ≥ 500 (%)
Moyenne	35 827 486	99,2	40,7	49,7	9,6	3 287 154	4175	62,2	0,4
Minimum	29 070 976	98,4	37,5	44,8	9,1	3 103 676	3729	58,1	0,3
Maximum	42 920 692	99,7	45,5	52,4	10,5	3 458 347	4738	66,4	0,4

B**HaeIII + MspI****C****MspI****Figure 20. Résultats obtenus par la double digestion enzymatique HaeIII + MspI**

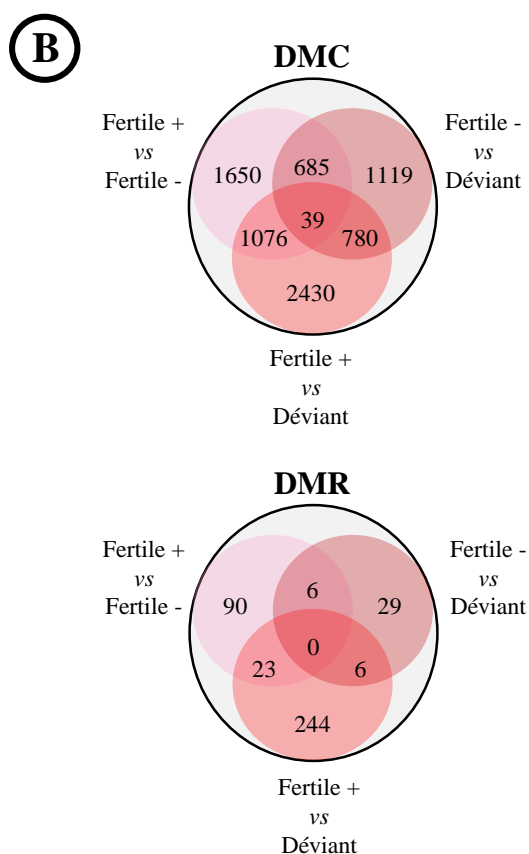
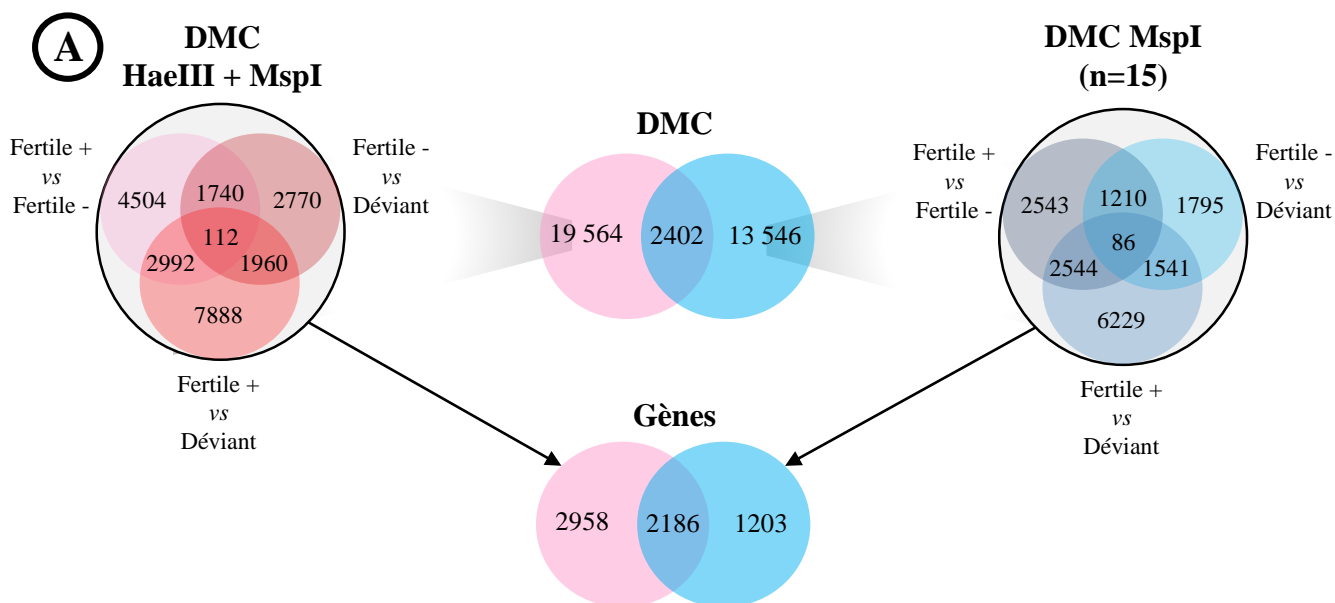
A. Tableau récapitulatif des données de contrôle qualité issu du séquençage des bibliothèques de RRBS en utilisant la double digestion enzymatique (tableau du haut), comparé aux données obtenues avec MspI utilisé seul (tableau du bas). Pour le tableau concernant MspI seul, les calculs sont faits uniquement sur les mêmes échantillons que ceux utilisés pour la double digestion enzymatique (n=15). **B.** ACP et classification hiérarchique de corrélation. La couverture minimale des CpG10 utilisée est de 4 individus par catégorie de fertilité. L'analyse porte alors sur un total de 1 321 826 CpG. **C.** ACP et classification hiérarchique de corrélation pour les individus traités par MspI seul (n=15, les mêmes individus que ceux analysés en utilisant HaeIII + MspI). La couverture minimale des CpG10 utilisée est de 4 individus par catégorie de fertilité. L'analyse porte alors sur un total de 1 403 469 CpG.

Nous avons donc préparé 15 nouvelles banques correspondant à 5 taureaux Holstein choisis dans chacune des trois catégories de fertilité comme cela a été décrit dans la section Matériels et Méthodes.

Les banques ont été séquencées par la société de service Integragen comme précédemment. Néanmoins, nous avons demandé qu'un minimum de 80 millions séquences (40 millions de séquences en paires) par échantillon soit produit. L'ensemble du processus d'analyse décrit précédemment peut être utilisé sans modification. Les résultats des contrôles qualité du séquençage et de l'alignement des séquences sont donnés Figure 22.A. En utilisant la double digestion enzymatique HaeIII+MspI, le taux d'alignement unique est plus important que celui observé après simple digestion MspI (50,4% et 40,7%, respectivement). Le nombre de CpG analysés est aussi augmenté (6 499 180 et 3 287 154, respectivement). En revanche, la couverture par CpG diminue grandement, avec seulement 40,1% de CpG₁₀ après clivage HaeIII+MspI (ce qui représente en moyenne 2 606 171 CpG₁₀) contre 62,2% de CpG₁₀ avec MspI seul (en moyenne 2 044 610 CpG₁₀). Malgré une augmentation du nombre de séquences par échantillon, la couverture des CpG analysés est moins bonne ce qui doit être pris en considération pour les analyses ultérieures.

a. Par analyse descriptive, il apparait que la double digestion permet d'identifier les CpG₁₀ discriminant mieux les catégories Fertiles+ et Déviants.

En prenant en compte l'ensemble des CpG₁₀, pour les 15 individus, il est possible de réaliser une analyse de corrélation aboutissant à une classification hiérarchique et une ACP. Les résultats peuvent être comparés à ceux obtenus en prenant en compte les CpG₁₀ obtenus pour les banques après simple clivage MspI pour les mêmes individus (Figure 22.B). Les deux classifications hiérarchiques donnent des regroupements similaires : les individus sont classés en deux groupes. Pour les banques issues de la double digestion, le premier groupe correspond aux 5 Fertiles+ et 1 Déviant, le second groupe rassemble les autres déviants et les Fertiles-. La même distribution est observée pour les banques issues de la simple digestion à l'exception d'un Fertile+ qui se regroupe avec les Fertiles- et les déviants. En ACP, pour les banques issues de la double digestion la dimension 1, expliquant 9,92% de la variance, sépare nettement les Fertiles+ et les Déviants. Pour les banques issues de la simple digestion, une telle séparation n'est observée que pour la dimension 2. Les CpG₁₀ identifiés après la double digestion HaeIII+MspI permettent de mieux différencier les Fertiles+ des Déviants que ceux obtenus après la simple digestion avec MspI.



	HaeIII + MspI	Background (%)	DMC (%)	DMR (%)
Nombre de CpG	1 393 907		21 966	698
Eléments génomiques				
Promoteur-TSS	7,6		2,2	3,0
Intron	29,0		29,7	23,5
Exon	10,6		8,6	21,8
TTS	0,5		0,5	1,3
3'UTR	0,7		0,7	1,1
5'UTR	2,7		1,6	0,7
Intergénique	48,9		56,8	48,6
Densité en CpG				
Island	21,1		11,0	29,7
Shore	16,4		14,3	15,6
Shelve	7,6		7,2	4,7
Open Sea	54,9		67,6	50,0
Eléments répétés				
LINE	5,7		13,4	15,3
SINE	9,6		12,8	8,0
LTR	3,2		4,8	3,2
Satellite	6,8		0,4	0,6
Autre	2,0		1,7	3,6
Sans élément répété	72,7		66,8	69,3

Figure 21. Identification et caractérisation des DMC et DMR pour la double digestion HaeIII + MspI

A. DMC identifiées entre les catégories de fertilité avec la double digestion enzymatique. Toutes catégories confondues, on obtient un total de 21 966 DMC. Le nombre de DMC obtenu est plus important qu'en utilisant MspI seul (15 948 DMC). A noter aussi que les DMC identifiées sont à 89% différentes de celles que l'on obtient avec MspI seul. **B.** Identification des gènes associés aux DMC et DMR pour chaque comparaison entre les catégories de fertilité pour la double digestion. Dans le tableau de droite sont comparés les résultats d'annotation entre le *background* et ceux des DMC et DMR.

b. La double digestion permet d'identifier un nombre plus important de DMC entre catégories de fertilité.

Une analyse différentielle avec le programme methylkit a permis d'identifier un total de 21 966 DMC en comparant les catégories de fertilité (Figure 23.A) avec 35,9% de DMC (7888 DMC) identifiées pour la comparaison entre Fertiles+ et Déviants, 20,5% de DMC (4504 DMC) pour la comparaison Fertiles+ et Fertiles-, et seulement 12,6% de DMC (2770 DMC) pour la comparaison entre Fertiles- et Déviants. On peut alors en conclure que globalement la majorité des DMC permettent de distinguer les individus Fertiles+ des individus appartenant aux 2 autres catégories de fertilité.

Comme pour les analyses descriptives, nous avons comparé les résultats d'analyse différentielle pour les banques issues de la double digestion à ceux obtenus pour les banques issues de la simple digestion en ne prenant en compte que les 15 individus précédemment décrits. Le nombre de DMC identifiées dans le cas de la double digestion HaeIII+MspI est 27,6% plus important qu'en utilisant la simple digestion MspI (respectivement, 21 966 DMC contre 15 948 DMC au total). Seules 2402 DMC sont communes entre les deux types de banques, ce qui représente respectivement 10,9 et 15,1% du total de DMC identifiées en utilisant HaeIII+MspI ou MspI, respectivement. Les DMC_[MspI] identifiées se répartissent entre les catégories de fertilité de manière équivalente à celle observée pour les DMC_[HaeIII + MspI] : 39,0% de DMC (6229 DMC) pour la comparaison entre Fertiles+ contre Déviants, 15,9% de DMC (2543 DMC) pour la comparaison Fertiles+ contre Fertiles-, et 11,3% de DMC (1795 DMC) pour la comparaison entre Fertiles- contre Déviants. La double digestion permet donc d'identifier plus de DMC que la simple digestion. Les DMC_[HaeIII + MspI] identifiées sont en grande majorité différentes des DMC_[MspI] mais la distribution des DMC, entre catégories de fertilité reste équivalente dans les deux cas.

c. Annotation génomique des DMC_[HaeIII + MspI].

En prenant en compte comme *background*, les 1 393 907 CpG₁₀ obtenus après double digestion, nous avons déterminé la distribution génomique pour les 21 966 DMC et les 698 DMR. Les résultats sont répertoriés dans le tableau de la Figure 23.B. On remarque que mis à part pour les promoteur-TSS, les DMC se répartissent de façon assez similaire aux CpG du *background* pour ce qui est des éléments génomiques. Ainsi, 29,7% des DMC sont localisées dans les introns de gènes, et 8,6% dans les exons (contre respectivement 29,0 et 10,6% pour les CpG du *background*). En revanche, les DMC_[HaeIII + MspI] couvrent peu de CGI (11,0% contre 21,1% avec le *background*). On les retrouve principalement en *open sea* (67.6%) et

Fertiles+ contre Fertiles-

Note	Nombre de gènes	Significativité	Fonction biologique (Gene Ontology term)
4,65	81	4,20E-02	Small GTPase mediated signal transduction (GO:0007264)
3,94	450	1,90E-02	Ion binding (GO:0043167)
3,14	156	4,10E-02	Cell migration (GO:0016477)
3,08	105	3,20E-02	Regulation of nervous system development (GO:0051960)
3,04	312	2,40E-02	Regulation of signal transduction (GO:0009966)
2,94	41	4,60E-02	GTPase activator activity (GO:0005096)
2,84	10	4,20E-02	Intracellular ligand-gated ion channel activity (GO:0005217)
2,81	74	1,70E-02	Substrate-specific channel activity (GO:0022838)
2,8	87	3,70E-02	Synapse (GO:0045202)
2,8	54	4,10E-02	Phospholipid binding (GO:0005543)
2,26	47	8,10E-02	Kidney development (GO:0001822)
2,24	57	5,70E-02	Ameboidal-type cell migration (GO:0001667)
2,16	151	8,10E-02	Cell junction (GO:0030054)
2,06	142	2,50E-02	Neuron part (GO:0097458)
1,98	277	6,30E-02	Carbohydrate derivative binding (GO:0097367)
1,89	51	2,40E-01	Proteinaceous extracellular matrix (GO:0005578)
1,79	33	9,90E-02	Plasma membrane receptor complex (GO:0098802)
1,79	149	3,60E-01	Anatomical structure formation involved in morphogenesis (GO:0048646)
1,77	75	1,70E-01	Regulation of neuron differentiation (GO:0045664)
1,74	25	4,40E-02	Calcium channel activity (GO:0005262)
1,67	14	6,30E-02	Semaphorin-plexin signaling pathway (GO:0071526)
1,65	75	3,80E-01	Sensory organ development (GO:0007423)
1,62	30	5,40E-01	Regulation of synapse structure or activity (GO:0050803)
1,61	21	4,50E-01	Endothelium development (GO:0003158)
1,61	24	3,20E-01	Protein tyrosine kinase activity (GO:0004713)
1,56	29	2,70E-01	Cytoplasmic side of plasma membrane (GO:0009898)
1,55	6	3,30E-01	Asymmetric protein localization (GO:0008105)
1,54	15	4,70E-01	Caveola (GO:0005901)
1,48	139	1,60E-01	Epithelium development (GO:0060429)
1,44	10	3,90E-01	Integrin complex (GO:0008305)
1,44	29	3,60E-01	Neurotransmitter transport (GO:0006836)
1,44	13	2,20E-01	Regulation of muscle adaptation (GO:0043502)
1,39	74	5,20E-01	Negative regulation of cell proliferation (GO:0008285)
1,37	6	4,40E-01	Macrophage activation involved in immune response (GO:0002281)
1,32	25	4,40E-01	Sensory perception of mechanical stimulus (GO:0050954)
1,31	16	4,50E-01	Endocrine process (GO:0050886)
1,30	14	6,20E-01	Response to alkaloid (GO:0043279)
1,30	60	2,40E-01	Response to wounding (GO:0009611)

Figure 22. Processus biologiques enrichis en DMC – enrichissements pour les comparaisons Fertiles+ contre Fertiles-, pour les taureaux Holstein analysés par double digestion enzymatique

L'analyse a été réalisée avec l'outil DAVID *Functional Annotation Clustering*, en utilisant les termes *Gene Ontology Biological Processes*, *Cellular Component* et *Molecular Function*. Les enrichissements sont faits par rapport aux gènes annotés sur le *background*, soit ici 19 945 gènes. La comparaison Fertiles+ contre Fertiles- sert de témoin de comparaison pour étudier les processus biologiques spécifiquement enrichis entre Déviants/Fertiles+, ainsi que Déviants/Fertiles-. Pour la comparaison entre Fertiles+ contre Fertiles-, l'analyse est réalisée sur 2817 gènes différents comportant au moins une DMC entre ces catégories. Les groupements de termes GO surlignés en vert correspondent aux processus biologiques qui interviennent pendant le développement embryonnaire ; en bleu ceux qui sont relatifs au système nerveux ; en rouge ceux relatifs à la motilité cellulaire.

dans les CpG *shores* (14,3%). Les DMC_[HaeIII + MspI] sont aussi fortement associées aux éléments répétés tels que les LINE, SINE et LTR (respectivement 13,4, 12,8 et 4,8%) comparé aux CpG du *background* (respectivement 5,7%, 9,6% et 3,2%). Les séquences satellites sont au contraire appauvries en DMC (0,4% contre 6,8% pour le *background*). Nous pouvons en conclure que les DMC_[HaeIII + MspI] sont principalement retrouvées dans les introns et les exons des gènes, les CpG *shores*, mais surtout en intergénique et associées à des éléments répétés tels que les LINE et les SINE. Les DMR_[HaeIII + MspI] ne sont pas associées aux mêmes éléments génomiques que les DMC_[HaeIII + MspI]. On les trouve de manière quasi équivalente dans les exons et les introns des gènes (respectivement 21,8 et 23,5%). De plus, les DMR sont retrouvées à 29,7% dans les CGI (comparé aux 11,0% de DMC retrouvées dans les CGI). Enfin, les DMR semblent majoritairement associées aux éléments répétés de type LINE, dont la colocalisation atteint les 15,3%. Au contraire des DMC, il semble que les DMR sont retrouvées de manière équivalente dans les introns et les exons de gènes et sont associées majoritairement aux CGI et aux éléments répétés de type LINE.

▪ **Identification des fonctions biologiques associées aux DMC_[HaeIII + MspI].**

Les gènes identifiés comme comportant au moins une DMC ont été utilisés pour effectuer des analyses d'enrichissement. Cette analyse a été effectuée avec DAVID comme précédemment. Les résultats relatifs aux comparaisons entre Fertiles+ et Fertiles-, Déviants contre Fertiles+ et Déviants contre Fertiles- sont représentés respectivement Figure 24, 25 et 26. Les données d'enrichissements obtenues avec les banques construites après une simple digestion, observait un maximum de groupements biologiques significativement enrichis pour la comparaison entre Déviants et Fertiles-, avec la double digestion, le maximum est observé pour la comparaison Déviants et Fertiles+. En effet, on observe respectivement **38, 43 et 32 groupements de termes GO** significatifs pour les comparaisons **Fertiles+ et Fertiles-** (EASE score entre 1,3 et 4,65), **Déviants et Fertiles+** (EASE score entre 1,3 et 5,32), **Déviants et Fertiles-** (EASE score entre 1,3 et 4,15), respectivement. Pour la comparaison **Fertiles+ et Fertiles-**, 6 groupements GO concernent le système nerveux (« *Regulation of nervous system development* », « *Synapse* », « *Neuron part* », « *Regulation of neuron differentiation* », « *Regulation of synapse structure or activity* », « *Neurotransmitter transport* ») ; 5 ont trait au développement (« *Kidney development* », « *Anatomical structure formation involved in morphogenesis* », « *Sensory organ development* », « *Endothelium development* », « *Epithelium development* »). Enfin, 7 regroupements concernent les voies de transduction impliquant les petites protéines G (« *Small GTPase mediated signal transduction* », « *GTPase activator activity* », « *Regulation of signal transduction* ») et les

Déviants contre Fertiles+

Note	Nombre de gènes	Significativité	Fonction biologique (Gene Ontology term)
5,32	182	1,30E-02	Cell junction (GO:0030054)
4,61	154	1,30E-02	Kinase activity (GO:0016301)
4,51	544	1,30E-03	Metal ion binding (GO:0046872)
4,37	116	5,00E-03	Neuron projection (GO:0043005)
3,64	69	9,20E-03	Cell-substrate adhesion (GO:0031589)
2,80	24	2,30E-02	Excitatory synapse (GO:0060076)
2,78	263	4,10E-02	Regulation of multicellular organismal development (GO:2000026)
2,75	32	4,40E-02	Protein tyrosine kinase activity (GO:0004713)
2,68	52	3,20E-02	Regulation of small GTPase mediated signal transduction (GO:0051056)
2,43	73	9,00E-02	Synapse part (GO:0044456)
2,38	97	3,80E-02	Inorganic cation transmembrane transporter activity (GO:0022890)
2,34	419	2,30E-02	Regulation of cell communication (GO:0010646)
2,30	30	3,50E-02	Cell cortex part (GO:0044448)
2,27	262	3,30E-02	Movement of cell or subcellular component (GO:0006928)
2,13	22	5,00E-02	Memory (GO:0007613)
2,01	70	5,40E-02	Somatodendritic compartment (GO:0036477)
1,87	54	1,50E-01	Membrane region (GO:0098589)
1,87	14	4,90E-01	Regulation of cell adhesion mediated by integrin (GO:0033628)
1,78	50	1,20E-01	Transport vesicle (GO:0030133)
1,75	50	7,10E-02	GTPase regulator activity (GO:0030695)
1,71	33	2,70E-01	Dendrite development (GO:0016358)
1,70	23	2,40E-01	Transport vesicle membrane (GO:0030658)
1,69	21	1,90E-01	Synapse assembly (GO:0007416)
1,69	11	3,50E-01	Calcium activated cation channel activity (GO:0005227)
1,68	53	8,90E-02	Extrinsic component of membrane (GO:0019898)
1,66	36	5,60E-01	Regulation of synapse structure or activity (GO:0050803)
1,64	43	5,30E-01	Skeletal system morphogenesis (GO:0048705)
1,61	7	6,60E-01	Regulation of anoikis (GO:2000209)
1,58	55	3,80E-01	Plasma membrane organization (GO:0007009)
1,53	502	1,30E-01	Membrane-bounded vesicle (GO:0031982)
1,52	11	6,30E-01	Vesicle-mediated transport between endosomal compartments (GO:0098927)
1,49	133	4,30E-02	Single organism cell adhesion (GO:0007155)
1,47	37	2,70E-01	Central nervous system neuron differentiation (GO:0021953)
1,46	20	2,30E-01	Rab GTPase binding (GO:0017137)
1,43	43	3,90E-01	Stress-activated MAPK cascade (GO:0051403)
1,42	39	4,40E-01	Phosphatidylinositol binding (GO:0035091)
1,38	90	4,80E-01	Sensory organ development (GO:0007423)
1,37	8	3,80E-01	Phosphatidylinositol phospholipase C activity (GO:0004435)
1,36	423	2,20E-02	Regulation of signaling (GO:0023051)
1,35	14	6,70E-01	Regulation of postsynaptic membrane potential (GO:0060078)
1,34	81	4,90E-02	Cellular response to nitrogen compound (GO:1901699)
1,32	125	2,40E-01	Central nervous system development (GO:0007417)
1,31	45	7,00E-01	Lipid transport (GO:0006869)

Figure 23. Processus biologiques enrichis en DMC – enrichissements pour les comparaisons Déviants/Fertiles+, pour les taureaux Holstein analysés par double digestion enzymatique

L'analyse a été réalisée avec l'outil DAVID *Functional Annotation Clustering*, en utilisant les termes *Gene Ontology Biological Processes*, *Cellular Component* et *Molecular Function*. Les enrichissements sont faits par rapport aux gènes annotés sur le *background*, soit ici 19 945 gènes. Pour la comparaison entre Déviants/Fertiles+ l'analyse est réalisée sur 3489 gènes différents comportant au moins une DMC entre ces catégories. Les groupements de termes GO surlignés en vert correspondent aux processus biologiques qui interviennent pendant le développement embryonnaire ; en bleu ceux qui sont relatifs au système nerveux ; en rouge ceux relatifs à la motilité cellulaire.

Déviants contre Fertiles-

Note	Nombre de gènes	Significativité	Fonction biologique (Gene Ontology term)
4,15	18	2,60E-02	Rho guanyl-nucleotide exchange factor activity (GO:0005089)
2,94	256	2,30E-02	Regulation of signal transduction (GO:0009966)
2,75	354	1,10E-01	Ion binding (GO:0043167)
2,70	37	4,50E-02	GTPase regulator activity (GO:0030695)
2,50	30	3,70E-02	Divalent inorganic cation transmembrane transporter activity (GO:0072509)
2,26	158	4,50E-02	Locomotion (GO:0040011)
2,25	113	3,30E-02	Neuron part (GO:0097458)
2,10	51	2,10E-01	Synapse part (GO:0044456)
2,05	164	2,10E-01	Regulation of phosphorus metabolic process (GO:0051174)
1,92	28	1,80E-01	Regulation of JNK cascade (GO:0046328)
1,85	194	7,80E-02	Nervous system development (GO:0007399)
1,82	70	2,90E-02	Synapse (GO:0045202)
1,75	15	3,80E-01	Lipid transporter activity (GO:0005319)
1,72	58	1,30E-01	ATPase activity (GO:0016887)
1,67	228	2,20E-01	Cellular localization (GO:0051641)
1,66	10	6,20E-01	Multi-organism behavior (GO:0051705)
1,61	26	6,50E-03	Calcium ion transmembrane transporter activity (GO:0015085)
1,60	11	4,30E-01	Transmembrane receptor protein tyrosine kinase activity (GO:0004714)
1,58	224	9,50E-02	Carbohydrate derivative binding (GO:0097367)
1,54	24	6,30E-01	Learning or memory (GO:0007611)
1,53	88	4,50E-01	Membrane organization (GO:0061024)
1,49	17	7,40E-01	Stem cell population maintenance (GO:0019827)
1,41	4	3,90E-01	Protein-arginine deiminase activity (GO:0004668)
1,41	103	3,70E-01	Negative regulation of multicellular organismal process (GO:0051241)
1,41	117	6,00E-01	Anatomical structure formation involved in morphogenesis (GO:0048646)
1,39	4	4,60E-01	Auditory behavior (GO:0031223)
1,39	9	7,60E-01	Phospholipid transport (GO:0015914)
1,37	8	6,50E-01	Integrin complex (GO:0008305)
1,35	13	7,70E-01	Cellular response to fibroblast growth factor stimulus (GO:0044344)
1,33	25	4,80E-01	Synapse organization (GO:0050808)
1,30	43	2,60E-01	Cell-substrate adhesion (GO:0031589)
1,30	39	4,30E-01	Phospholipid binding (GO:0005543)

Figure 24. Processus biologiques enrichis en DMC – enrichissements pour les comparaisons Déviants/Fertiles-, pour les taureaux Holstein analysés par double digestion enzymatique

L'analyse a été réalisée avec l'outil DAVID *Functional Annotation Clustering*, en utilisant les termes *Gene Ontology Biological Processes*, *Cellular Component* et *Molecular Function*. Les enrichissements sont faits par rapport aux gènes annotés sur le *background*, soit ici 19 945 gènes. Pour la comparaison entre Déviants/Fertiles- l'analyse est réalisée sur 2161 gènes différents comportant au moins une DMC entre ces catégories. Les groupements de termes GO surlignés en vert correspondent aux processus biologiques qui interviennent pendant le développement embryonnaire ; en bleu ceux qui sont relatifs au système nerveux ; en rouge ceux relatifs à la motilité cellulaire.

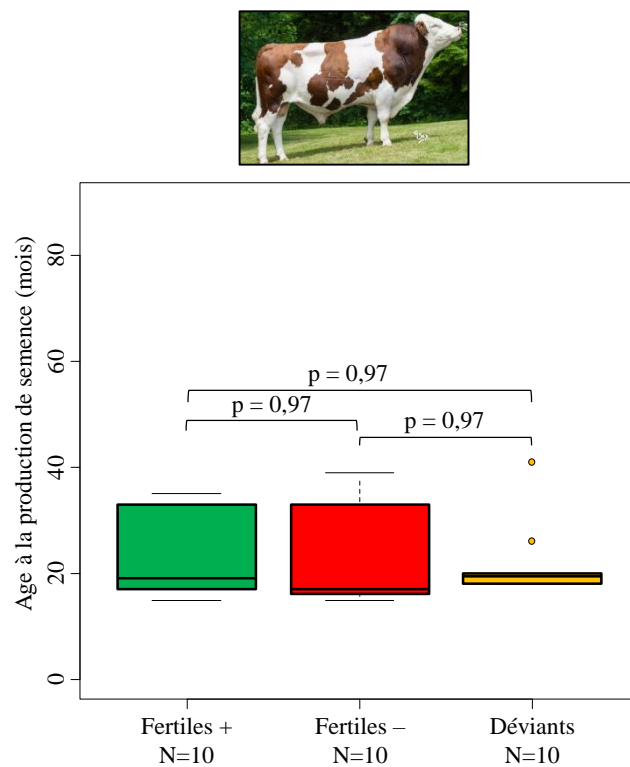
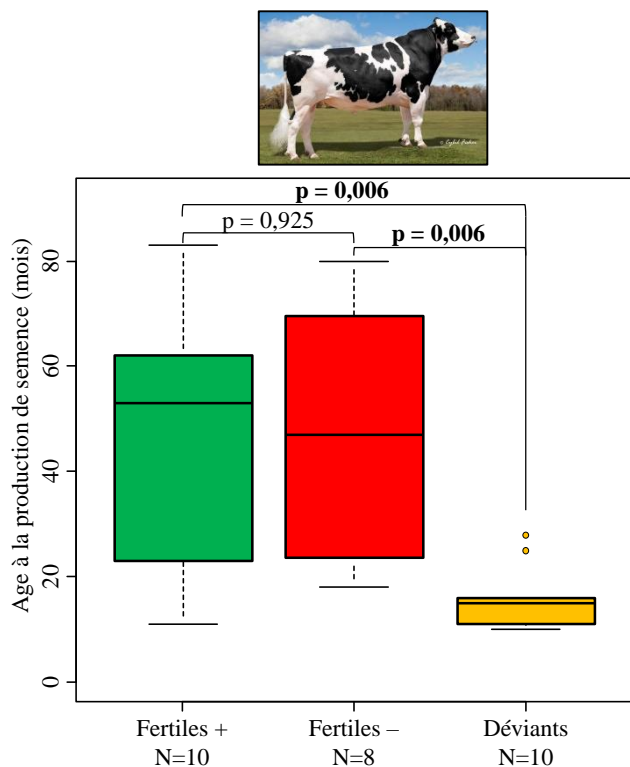


Figure 25. Différentiel d'âge entre les catégories de fertilité

Comparaison des âges à la production de semence pour les 3 catégories de fertilité et les 2 races. Pour les taureaux de race Holstein, l'âge à la production médian est respectivement de 54 ± 24 mois, 47 ± 26 mois et 15 ± 6 mois pour les individus Fertiles+, Fertiles- et Déviants. L'âge à la production des Déviants Holstein est très significativement différent de celui des 2 autres catégories. Pour les taureaux de race Montbéliard, l'âge à la production médian est respectivement de 19 ± 8 mois, 17 ± 9 mois et 20 ± 7 mois pour les individus Fertiles+, Fertiles- et Déviants. Il n'y a aucune différence significative d'âge entre les 3 catégories de fertilité pour cette race. L'ensemble des analyses statistiques ont été effectuées avec des tests de permutation pour séries indépendantes avec échantillonnage Monte Carlo (100 000 tirages), suivis de comparaisons 2 à 2 (1000 tirages) avec correction de Benjamini-Hochberg pour tests multiples.

flux ioniques (« *Ion binding* », « *Intracellular ligand-gated ion channel activity* », « *Substrate-specific channel activity* »). Pour la comparaison **Déviants et Fertiles+**, 10 groupements de termes GO concernent le système nerveux (« *Neuron projection* », « *Excitatory synapse* », « *Synapse part* », « *Memory* », « *Somatodendritic compartment* », « *Dendrite development* », « *Synapse assembly* », « *Regulation of synapse structure or activity* », « *Central nervous system neuron differentiation* », « *Regulation of postsynaptic membrane potential* ») et seulement 4 ont trait au développement (« *Regulation of multicellular organismal development* », « *Skeletal system morphogenesis* », « *Sensory organ development* », « *Central nervous system development* »). Les fonctions biologiques les plus significativement ciblées concernent les interactions cellule-cellule (« *Cell junction* », avec un EASE score de 5,32) et les interactions cellule-matrice extracellulaire (« *Cell-substrate adhesion* » (EASE score 3,64), « *Regulation of cell communication* », « *Regulation of cell adhesion mediated by integrin* », « *Plasma membrane organization* », « *Single organism cell adhesion* »). Les groupements de termes GO concernent les voies de transduction impliquant les petites protéines G et les flux ioniques sont observés dans cette comparaison (« *Regulation of small GTPase mediated signal transduction* », « *GTPase regulator activity* », « *Rab GTPase binding* » ; « *Inorganic cation membrane transporter activity* », « *Calcium activated cation channel activity* »). Concernant la comparaison **Déviants et Fertiles-**, les groupements les plus significatifs concernent aussi les voies de transduction impliquant les petites protéines G et les flux ioniques (« *Rho guanyl-nucleotide exchange factor activity* », « *Regulation of signal transduction* », « *GTPase regulator activity* » ; « *Ion binding* », « *Divalent inorganic cation membrane transporter activity* », « *Calcium ion transmembrane transporter activity* », « *Membrane organization* »).

Ainsi, pour les deux types de banques (simple et double digestion enzymatique), il est observé un enrichissement des fonctions biologiques impliquées dans le développement embryonnaire. Néanmoins, pour les comparaisons **Déviants et Fertilité+** ou **Fertilité-**, l'utilisation des banques issues de la double digestion, permet de mettre en évidence des fonctions biologiques impliquées dans les interactions cellule-cellule, dans la signalisation liée aux petites protéines G et dans la régulation des flux ioniques, suggérant des modifications potentielles au niveau du spermatozoïde pouvant altérer ses fonctions.

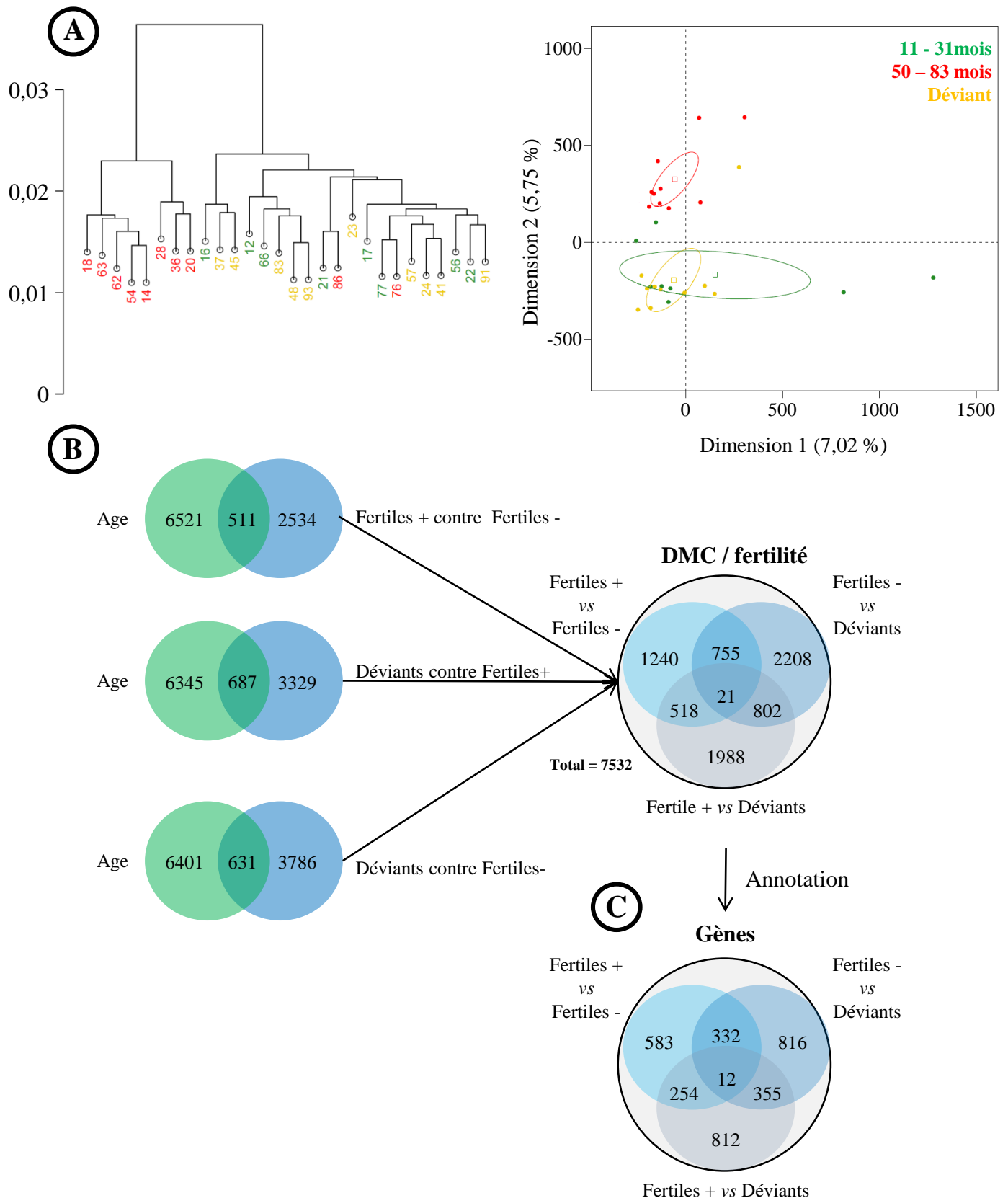


Figure 26. Prise en compte de l'âge à la production

A. Classification hiérarchique de corrélation de Pearson et ACP pour la race Holstein en catégorisant les Fertiles+ et les Fertiles- en fonction de leur âge à la production. Le groupe en vert est composé de 4 individus Fertiles+ et de 4 individus Fertiles-, et leur âge à la production est compris entre 11 et 31 mois. Le groupe en rouge est composé de 6 individus Fertiles+ et de 4 individus Fertiles-, et leur âge à la production est compris entre 50 et 83 mois. L'analyse porte alors sur un total de 1 553 369 CpG. Sur l'ACP une franche séparation entre les individus les plus âgés par rapport aux plus jeunes et aux Déviants est visible sur la dimension 2. Le groupe des individus les plus jeunes est superposé à celui des Déviants (dont la moyenne d'âge à la production est de 16 mois). **B.** Des DMC ont été identifiées entre le groupe des 11-31 mois et les groupes des 50-83 mois et des Déviants. Une partie des DMC ainsi identifiées et liées à l'âge sont en partie communes avec celles identifiées par rapport aux 3 catégories de fertilité. En retranchant les DMC communes des DMC liées à la fertilité, on peut corriger notre jeu de DMC par rapport à l'âge de production. **C.** Nombre de gènes correspondant aux DMC après annotation.

K. Mise en évidence d'un effet « âge » sur le méthylome spermatique - Correction des biomarqueurs de fertilité

Quel effet de l'âge sur le méthylome ?

Comme indiqué sur la Figure 3, il existe une différence significative d'âge à la production de semence des taureaux Holstein Déviants. Nous avons donc décidé d'évaluer les effets de l'âge sur le méthylome spermatique : pour ce faire, nous avons conduit de nouvelles analyses en définissant deux groupes de taureaux différenciés en fonction de leur âge à la production de semence, à partir de la cohorte d'individus Fertiles+ et Fertiles- de la race Holstein ; les déviants n'ont pas été pris en considération pour la constitution de ces deux groupes, puisque les effets « âge » et « catégorie de fertilité » pouvaient être confondus. Le premier groupe est composé d'individus dont la semence analysée a été produite entre 11 et 31 mois (composé de 8 individus, dont 4 Fertiles+ et 4 Fertiles-), le second, d'individus dont la semence analysée a été produite entre 50 et 83 mois (composé de 10 individus, dont 6 Fertiles+ et 4 Fertiles-).

Nous avons repris l'analyse par classification hiérarchique et ACP utilisant les CpG₁₀ pour l'ensemble des taureaux (Figure 27.A, page suivante). La classification hiérarchique sépare de manière claire deux groupes : le premier comprend 6 des 8 individus âgés. Le second groupe est composé des individus les plus jeunes et des 2 individus âgés restants. Comme attendu, les Déviants sont dans ce deuxième groupe. Sur l'ACP, la dimension 2 sépare très nettement le groupe âgé des jeunes (et des déviants). On voit donc que la séparation est meilleure que celle obtenue en groupant les individus en fonction de leur fertilité. En conséquence, nous pouvons considérer que l'âge à la production est un déterminant majeur du méthylome spermatique bovin.

▪ Prise en compte de l'effet âge dans la comparaison des catégories de fertilité

Les analyses différentielles du méthylome par le logiciel methylKit sont basées sur des modèles statistiques de comparaison de groupes 2 à 2, sans possibilité de prise en compte de la variation inter-individuelle ni d'effets tels que l'âge. Nous avons cependant souhaité tenter d'affiner la liste des biomarqueurs en lien avec la Fertilité en tenant compte de l'effet de l'âge à la production de la semence analysée. Pour cela, nous avons simplement soustrait les DMC liées à l'âge de production du jeu de DMC issu des comparaisons entre groupes de fertilité.

Nous avons pu identifier 7032 DMC entre les individus jeunes et âgés : la comparaison des groupes d'âge produit ainsi un plus grand nombre de DMC (+ 131%). Parmi les 7032 DMC_[âge], 511 (16,8%), 687 (17,1%) et 631 (14,3%) des DMC sont communes aux

A**Fertiles+ contre Fertiles-**

Note	Nombre de gènes	Significativité	Fonction biologique (Gene Ontology term)
2,88	39	3,60E-02	Protein serine/threonine kinase activity (GO:0004674)
2,48	53	1,30E-01	Regulation of cellular component movement (GO:0051270)
2,34	85	1,20E-01	Neurogenesis (GO:0022008)
2,30	40	3,90E-01	Anchoring junction (GO:0070161)
2,11	27	4,50E-01	Somatodendritic compartment (GO:0036477)
2,00	4	1,70E-01	Protein-arginine deiminase activity (GO:0004668)
1,86	34	3,10E-01	Skeletal system development (GO:0001501)
1,86	61	1,40E-01	Cardiovascular system development (GO:0072358)
1,75	13	3,50E-01	Regulation of Rho protein signal transduction (GO:0035023)
1,73	41	3,40E-01	Tube development (GO:0035295)
1,69	13	3,00E-01	Regulation of phosphatase activity (GO:0010921)
1,61	37	3,50E-01	Negative regulation of cell proliferation (GO:0008285)
1,49	129	2,70E-01	Intracellular signal transduction (GO:0035556)
1,46	29	1,50E-01	Signal release (GO:0023061)
1,39	11	5,90E-01	Heparin binding (GO:0008201)
1,39	10	4,80E-01	Excitatory synapse (GO:0060076)
1,33	41	3,00E-01	Tissue morphogenesis (GO:0048729)
1,32	25	3,70E-01	Ameboidal-type cell migration (GO:0001667)
1,31	21	5,00E-01	Cytoplasmic vesicle membrane (GO:0030659)

B**Déviants contre Fertiles+**

Note	Nombre de gènes	Significativité	Fonction biologique (Gene Ontology term)
3,87	88	3,30E-02	Cell projection organization (GO:0030030)
3,38	53	3,10E-02	Neuron projection (GO:0043005)
3,26	25	3,20E-02	GTPase activator activity (GO:0005096)
3,20	18	2,30E-02	Ras guanyl-nucleotide exchange factor activity (GO:0005088)
3,11	147	2,10E-02	Carbohydrate derivative binding (GO:0097367)
2,93	37	9,20E-03	Regulation of neuron projection development (GO:0010975)
2,28	42	4,70E-02	Regulation of anatomical structure size (GO:0090066)
2,26	172	2,00E-02	Regulation of signaling (GO:0023051)
2,18	143	3,80E-02	Regulation of cellular component organization (GO:0051128)
2,08	58	1,40E-03	Regulation of neurogenesis (GO:0050767)
1,82	22	1,40E-01	Modulation of synaptic transmission (GO:0050804)
1,79	6	7,70E-02	Semaphorin-plexin signaling pathway involved in neuron projection guidance (GO:1902285)
1,75	109	3,90E-02	Movement of cell or subcellular component (GO:0006928)
1,74	63	3,40E-01	Circulatory system development (GO:0072359)
1,64	7	1,10E-01	Intracellular ligand-gated ion channel activity (GO:0005217)
1,59	31	5,70E-01	Somatodendritic compartment (GO:0036477)
1,53	16	4,60E-01	Dendrite development (GO:0016358)
1,50	21	2,50E-01	Cognition (GO:0050890)
1,41	42	4,80E-01	Synapse (GO:0045202)
1,41	24	6,80E-01	Anatomical structure homeostasis (GO:0060249)
1,39	25	3,90E-01	Camera-type eye development (GO:0043010)
1,36	14	2,00E-01	Cerebral cortex development (GO:0021987)
1,36	148	3,70E-02	Cellular localization (GO:0051641)
1,35	13	1,20E-01	Regulation of axon extension (GO:0030516)
1,34	30	1,00E-01	Cation channel activity (GO:0005261)
1,32	19	6,80E-01	Endosomal transport (GO:0016197)
1,32	17	5,60E-01	Regulation of peptidyl-tyrosine phosphorylation (GO:0050730)

comparaisons Fertiles+ et Fertiles-, Déviants et Fertiles+ et Déviants et Fertiles-, respectivement (Figure 27.B). Il est alors possible d'obtenir un nouveau jeu de DMC, excluant ceux liés à l'âge : 2534 pour la comparaison Fertiles+ vs Fertiles-, 3329 pour la comparaison Déviants vs Fertiles+, 3786 pour la comparaison Déviants vs Fertiles-. Ce jeu de DMC est utilisé pour mettre en évidence les DMC caractéristiques des déviants. Sur le nouveau diagramme de Venn obtenu, un nombre plus important de DMC est observé dans les territoires correspondant aux comparaisons impliquant les taureaux Déviants, que dans celui correspondant à la comparaison entre Fertiles+ et Fertiles-. Les 802 DMC peuvent être considérées comme caractéristiques des Déviants, c'est-à-dire d'une catégorie particulière de fertilité, indépendamment de leur âge à la production de semence.

Comme précédemment les DMC ont été annotées, et des analyses d'enrichissement avec *DAVID Functional Annotation Clustering* ont été effectuées (Figure 28, page suivante). Le background utilisé est le même que celui utilisé pour les enrichissements entre catégories de fertilité. Les nombres de fonctions enrichies sont de **18** (avec un EASE score de 1,3 à 2,88) pour la comparaison entre **Fertiles+ et Fertiles-**, **27** (avec un EASE score de 1,3 à 3,87) pour la comparaison **Déviants et Fertiles+**, et de **39** (avec un EASE score de 1,3 à 4,36) pour la comparaison **Déviants et Fertiles-**. Pour la comparaison **Fertiles+ et Fertiles-**, 5/18 groupements de processus biologiques significativement enrichis sont en lien avec le développement, comme la neurogenèse, le développement du système squelettique, du système cardiovasculaire, des structures en formes de tube (*tube development*), ou la morphogenèse tissulaire. Le groupement le plus significatif est cependant l'activité des kinases (« *Protein serine/threonine kinase activity* », EASE score de 2,88). Pour la comparaison entre **Déviants et Fertiles+**, 11/27 groupements sont en lien avec des fonctions neuronales (contre 3/18 pour la comparaison Fertiles+ vs Fertiles-). Des groupements enrichis ont trait à l'activité enzymatique GTPases (« *GTPase activator activity* » EASE score de 3,26, « *Ras guanyl-nucleotide exchange factor activity* » EASE score de 3,2), la régulation de la taille des structures anatomiques (2,28) ou la régulation de la signalisation (2,26). Pour la comparaison entre **Déviants et Fertiles-**, le groupement le plus enrichi en DMC, et de manière très significative est celui de la régulation de la migration cellulaire (4,36). D'autres groupements particulièrement enrichis sont en rapport avec le développement, comme la régulation de la prolifération cellulaire, l'angiogenèse, le développement embryonnaire et le développement des structures en forme de tube. En recoupant toutes ces informations, il apparaît possible que les Déviants soient impactés par des perturbations de méthylation en lien avec la fonction neuronale et enzymatique, notamment des GTPases. En revanche, les

Déviant contre Fertiles-

Note	Nombre de gènes	Significativité	Fonction biologique (Gene Ontology term)
4,36	59	7,90E-03	Regulation of cell migration (GO:0030334)
4,31	166	2,80E-02	Nucleoside phosphate binding (GO:1901265)
4,13	42	1,40E-02	Cell-substrate junction (GO:0030055)
3,85	48	4,60E-02	Negative regulation of cell proliferation (GO:0008285)
3,84	38	2,70E-02	Angiogenesis (GO:0001525)
3,78	26	4,80E-02	Positive regulation of cell projection organization (GO:0031346)
3,22	156	7,70E-03	Regulation of cellular component organization (GO:0051128)
2,84	12	4,70E-02	Caveola (GO:0005901)
2,69	58	5,20E-03	Tube development (GO:0035295)
2,48	37	5,00E-02	Tube morphogenesis (GO:0035239)
2,47	45	2,30E-03	Protein serine/threonine kinase activity (GO:0004674)
2,32	24	7,30E-02	Protein localization to plasma membrane (GO:0072659)
2,16	61	4,60E-02	Hemopoiesis (GO:0030097)
2,15	44	4,40E-02	Regulation of cell projection organization (GO:0031344)
2,01	19	1,80E-01	Regulation of protein binding (GO:0043393)
1,96	21	2,40E-01	Modulation of synaptic transmission (GO:0050804)
1,85	23	2,70E-01	Negative regulation of cellular component movement (GO:0051271)
1,74	12	1,80E-01	Basement membrane (GO:0005604)
1,72	32	8,00E-01	Ameboidal-type cell migration (GO:0001667)
1,69	15	2,10E-01	Columnar/cuboidal epithelial cell differentiation (GO:0002065)
1,67	21	2,40E-01	Modulation of synaptic transmission (GO:0050804)
1,66	30	8,80E-02	Cell-substrate adhesion (GO:0031589)
1,66	16	1,70E-01	Protein tyrosine kinase activity (GO:0004713)
1,60	21	5,40E-02	Central nervous system neuron differentiation (GO:0021953)
1,55	10	1,40E-01	Regulation of peptidyl-lysine acetylation (GO:2000756)
1,53	4	1,60E-01	Protein-arginine deiminase activity (GO:0004668)
1,50	15	3,30E-01	Ras guanyl-nucleotide exchange factor activity (GO:0005088)
1,50	179	2,10E-02	Regulation of cell communication (GO:0010646)
1,45	8	1,80E-01	T cell selection (GO:0045058)
1,39	47	3,40E-01	Actin filament-based process (GO:0030029)
1,36	11	3,40E-01	Synapse assembly (GO:0007416)
1,32	118	1,50E-02	Regulation of multicellular organismal development (GO:2000026)
1,32	6	5,60E-01	Positive regulation of Notch signaling pathway (GO:0045747)
1,30	8	1,80E-01	Hindbrain morphogenesis (GO:0021575)

Figure 27. Processus biologiques enrichis en DMC – enrichissements pour les comparaisons Fertiles+ contre Fertiles-, Déviant/Fertiles+, Déviant/Fertiles- – taureaux Holstein analysés après retrait des DMC liés à l'âge

L'analyse a été réalisée avec l'outil DAVID *Functional Annotation Clustering*, en utilisant les termes Gene Ontology *Biological Processes*, *Cellular Component* et *Molecular Function*. Les enrichissements sont faits par rapport aux gènes annotés sur le *background* de la race Holstein, soit 18 931 gènes. La comparaison Fertiles+ contre Fertiles- sert de témoin de comparaison pour étudier les processus biologiques spécifiquement enrichis entre Déviant/Fertiles+, et Déviant/Fertiles-. Pour la comparaison entre Fertiles+ contre Fertiles-, l'analyse est réalisée sur 1035 gènes différents comportant au moins une DMC entre ces catégories. Pour la comparaison entre Déviant/Fertiles+, l'analyse est réalisée sur 1215 gènes différents comportant au moins une DMC entre ces catégories. Pour la comparaison entre Déviant/Fertiles-, l'analyse est réalisée sur 1281 gènes différents comportant au moins une DMC entre ces catégories. Les groupements de termes GO surlignés en vert correspondent aux processus biologiques qui interviennent pendant le développement embryonnaire ; en bleu ceux qui sont relatifs au système nerveux ; en rouge ceux relatifs à la motilité cellulaire, et en jaune ceux relatifs aux activités enzymatiques de la machinerie cellulaire.

individus Fertiles- semblent plutôt être impactés pour le développement embryonnaire et la motilité.

En recherchant manuellement des fonctions biologiques impactées par des perturbations de la méthylation, il est apparu que la voie de signalisation Hippo l'est tout particulièrement (Tableau 2). Parmi les 154 gènes impliqués dans cette voie, 23 présentent des DMC (pour un total de 37 DMC) pour la race Holstein après correction de l'âge, et 19 pour la race Montbéliarde (pour un total de 27 DMC et une DMR, sans correction liée à l'âge). Pour tous ces gènes, la plupart des DMC sont retrouvées dans les introns et en contexte de CpG *island*, *shore* ou *shelve*. Il apparaît que la majorité des DMC varient dans le même sens (hypo ou hyperméthylées) chez les Fertiles- et les Déviants par rapport aux Fertiles+.

Tableau 2. Liste des gènes de la voie Hippo impactés par un différentiel de méthylation

Méthylation donnée en moyenne. Les flèches montrent le sens du différentiel de méthylation par rapport aux Fertiles+. Flèche vide: différentiel de plus de 10%, Flèche pleine: différentiel de plus de 25%; Rouge: augmentation; vert: diminution. Pour le gène PRKCZ, les DMC étaient proches et ont alors été réunies en une DMR

Gene	Nom complet	Identifiant	Chromosome	Nombre de DMC	Contexte génomique	Méthylation Fertiles+ (%)	Méthylation Fertiles- (%)	Méthylation Déviants (%)
ATM	Axin1 Telenguectasia Mutated Serine/Threonine Kinase	ENSBTAG00000003111	15	1	CpG shore	56,1 ± 20,3	60,0 ± 17,1	35,1 ± 20,1
APC2	Adenomatous Polyposis Coli 2, WNT Signaling Pathway Regulator	ENSBTAG000000015433	7	1	CpG island	82,7 ± 13,6	89,7 ± 3,6	65,6 ± 11,4
AXIN1	Axin 1	ENSBTAG000000016577	25	3	CpG shore	55,2 ± 35,6	79,7 ± 35,5	76,5 ± 29,0
					CpG shore	55,9 ± 36,2	92,0 ± 13,2	78,4 ± 26,6
					CpG island	18,8 ± 25,7	44,0 ± 42,3	59,9 ± 37,0
					CpG sheave	60,9 ± 35,9	90,0 ± 9,0	56,5 ± 36,6
BMP6	Bone Morphogenetic Protein 6	ENSBTAG000000019234	23	3	CpG shore	33,3 ± 47,1	95,7 ± 3,3	71,4 ± 46,3
					-	34,4 ± 47,0	97,7 ± 2,8	69,6 ± 45,1
FBXW11	F-Box And WD Repeat Domain Containing 11	ENSBTAG000000015376	20	1	CpG shore	30,1 ± 8,6	38,1 ± 18,7	54,5 ± 21,6
GLI2	GLI Family Zinc Finger 2	ENSBTAG000000011682	2	1	-	50,2 ± 10,5	63,4 ± 25,5	35,5 ± 24,2
ITGB2	Integrin Subunit Beta 2	ENSBTAG000000017060	1	1	CpG sheave	73,2 ± 37,6	60,1 ± 45,7	57,3 ± 43,4
LLGL1	Lethal Giant Larvae Homolog 1, Scribble Cell Polarity Complex Component	ENSBTAG000000046535	19	1	CpG sheave	32,9 ± 37,2	43,6 ± 35,2	15,3 ± 18,3
					CpG shore	100 ± 0,0	41,8 ± 35,6	88,5 ± 23,1
NF2	Neurofibromin 2	ENSBTAG000000038992	17	2	CpG island	53,1 ± 29,1	79,7 ± 26,3	72,6 ± 37,0
PAR6B	Par-6 Family Cell Polarity Regulator Beta	ENSBTAG000000021601	13	1	CpG sheave	63,3 ± 31,8	70,1 ± 32,1	50,3 ± 38,0
PPP2R1A	Protein Phosphatase 2, Scaffold Subunit Alpha	ENSBTAG000000019851	18	1	-	70,7 ± 25,8	97,6 ± 4,1	71,0 ± 36,5
PRKCZ	Protein Kinase C Zeta	ENSBTAG000000014119	16	6	CpG island	70,3 ± 8,4	40,6 ± 12,8	56,9 ± 16,2
SMAD1	SMAD Family Member 1	ENSBTAG000000028355	17	1	-	40,3 ± 33,6	93,4 ± 13,3	74,0 ± 35,0
SNAI2	Snail Family Transcriptional Repressor 2	ENSBTAG000000013227	14	1	CpG sheave	50,2 ± 35,0	71,2 ± 24,1	49,1 ± 33,2
STK3	Serine/Threonine Kinase 3	ENSBTAG000000026234	14	1	CpG island	70,2 ± 21,0	90,9 ± 7,6	64,5 ± 19,7
STK4	Serine/Threonine Kinase 4	ENSBTAG00000003257	13	1	-	75,7 ± 35,1	72,0 ± 33,8	53,3 ± 30,8
TCF7L2	Transcription Factor 7-Like 2 (T-Cell Specific, HMG-Box)	ENSBTAG000000016462	24	1	-	63,8 ± 4,0	76,3 ± 4,8	49,2 ± 17,3
TEAD1	TEA Domain Transcription Factor 1	ENSBTAG000000032657	15	1	-	85,0 ± 17,0	79,4 ± 27,0	49,9 ± 23,8
TEAD3	TEA Domain Transcription Factor 3	ENSBTAG000000019495	23	1	CpG shore	38,8 ± 36,3	12,2 ± 17,5	32,2 ± 39,1
TEAD4	TEA Domain Transcription Factor 4	ENSBTAG000000019788	5	1	-	46,5 ± 38,9	0,0 ± 0,0	43,5 ± 39,6
VGLL4	Vesigial Like Family Member 4	ENSBTAG000000011538	22	3	CpG shore	74,5 ± 35,8	98,4 ± 2,3	62,2 ± 45,9
					CpG shore	64,8 ± 34,2	91,6 ± 19,9	69,7 ± 21,8
					CpG shore	82,8 ± 17,2	99,0 ± 2,5	65,3 ± 36,7
WNT5B	Wnt Family Member 5B	ENSBTAG000000001341	5	2	CpG shore	80,4 ± 24,2	64,5 ± 39,3	56,9 ± 39,3
					CpG shore	66,8 ± 43,6	31,6 ± 30,4	53,8 ± 36,8
WNT7A	Wnt Family Member 7A	ENSBTAG000000001668	22	2	CpG sheave	50,6 ± 32,4	75,4 ± 27,5	67,2 ± 18,8
					CpG shore	14,4 ± 27,3	40,0 ± 33,9	54,1 ± 27,1
ACTB	Actin Beta	ENSBTAG000000026199	25	1	-	59,6 ± 39,3	57,2 ± 37,7	87,7 ± 27,1
AMOTL2	Angiomotin Like 2	ENSBTAG000000000742	1	1	CpG island	57,1 ± 41,4	88,9 ± 21,9	89,9 ± 16,6
AXIN1	Axin 1	ENSBTAG000000016577	25	1	CpG island	57,4 ± 37,6	85,0 ± 31,5	72,0 ± 37,6
BMP6	Bone Morphogenetic Protein 6	ENSBTAG000000019234	23	1	CpG sheave	61,4 ± 34,4	54,2 ± 25,7	38,5 ± 38,3
LLGL1	Lethal Giant Larvae Homolog 1, Scribble Cell Polarity Complex Component	ENSBTAG000000046535	19	1	CpG sheave	40,8 ± 38,0	67,7 ± 29,9	59,8 ± 30,8
NF2	Neurofibromin 2	ENSBTAG000000013153	17	1	CpG shore	72,9 ± 31,0	49,0 ± 37,5	85,1 ± 25,1
					-	8,7 ± 22,9	14,2 ± 28,4	31,3 ± 40,7
NKDI	Naked Cuticle Homolog 1	ENSBTAG000000012295	18	4	-	81,8 ± 35,7	79,1 ± 31,7	58,9 ± 39,1
					CpG island	25,8 ± 16,4	39,6 ± 15,1	16,1 ± 8,5
					CpG island	21,6 ± 12,6	38,3 ± 17,2	14,6 ± 12,0
					-	46,3 ± 44,3	78,1 ± 20,3	34,0 ± 36,8
NPHP4	Nephrocystin 4	ENSBTAG000000008126	16	1	-	47,3 ± 20,2	75,5 ± 7,8	73,8 ± 11,5
PPP2R2C	Protein Phosphatase 2 Regulatory Subunit Bgamma	ENSBTAG000000020598	6	1	-	77,0 ± 35,9	61,0 ± 34,9	50,9 ± 41,1
RASSF4	Ras Association Domain Family Member 4	ENSBTAG000000022013	28	1	CpG shore	33,1 ± 8,4	16,3 ± 11,1	5,3 ± 5,2
SCRIB	Scribble Planar Cell Polarity Protein	ENSBTAG000000008421	14	2	CpG island	69,3 ± 33,5	63,9 ± 33,6	86,7 ± 25,9
TCF7L2	Transcription Factor 7-Like 2 (T-Cell Specific, HMG-Box)	ENSBTAG000000016462	24	1	-	87,3 ± 21,6	63,3 ± 38,6	81,2 ± 34,3
TEAD1	TEA Domain Transcription Factor 1	ENSBTAG000000032657	15	3	-	91,5 ± 13,2	50,7 ± 33,8	80,3 ± 25,4
					-	82,0 ± 37,4	41,7 ± 33,2	76,6 ± 28,0
TEAD3	TEA Domain Transcription Factor 3	ENSBTAG000000019495	23	1	-	58,2 ± 35,9	87,0 ± 26,4	89,1 ± 24,4
TEAD4	TEA Domain Transcription Factor 4	ENSBTAG000000019788	5	2	CpG island	34,2 ± 41,2	23,5 ± 34,1	10,4 ± 27,5
WNT3	Wnt Family Member 3	ENSBTAG000000016012	19	1	-	38,6 ± 48,3	68,3 ± 45,8	46,7 ± 39,5
WNT4	Wnt Family Member 4	ENSBTAG000000022671	2	2	-	93,3 ± 14,0	69,5 ± 32,1	70,7 ± 30,9
					-	85,1 ± 23,8	61,9 ± 48,1	88,3 ± 23,6
					-	55,8 ± 32,8	22,2 ± 22,3	65,4 ± 44,0
WNT7A	Wnt Family Member 7A	ENSBTAG000000001668	22	3	CpG shore	28,7 ± 32,2	50,8 ± 46,6	31,2 ± 34,9
WNT10A	Wnt Family Member 10A	ENSBTAG000000009217	2	1	CpG sheave	40,5 ± 46,9	61,8 ± 25,5	65,4 ± 38,0
WWC1	WW And C2 Domain Containing 1	ENSBTAG000000013880	7	1 (DMR)	CpG sheave	48,6 ± 29,1	79,5 ± 23,2	44,0 ± 36,7
					-	34,9 ± 12,2	54,4 ± 19,7	60,7 ± 16,9



Discussion

Dans cette étude, nous avons effectué une analyse pangénomique de la méthylation de l'ADN par RRBS dans les spermatozoïdes bovins, dans le but d'identifier des biomarqueurs de la fertilité mâle. Pour les races Holstein et Montbéliard, nous avons ainsi pu identifier plusieurs milliers de biomarqueurs, dont une proportion importante a pu être associée à des processus biologiques liés au développement embryonnaire, au système nerveux et son fonctionnement, à la fonction spermatique et notamment à la voie de signalisation Hippo. Il est cependant apparu que les profils de méthylation peuvent être influencés par plusieurs facteurs, notamment le génotype et l'âge des individus, qui doivent alors être pris en compte dans un objectif d'optimisation de la caractérisation du méthylome spermatique et de ses variations en fonction de la fertilité.

A. Le méthylome spermatique présente des variations importantes entre races

Nos analyses se sont appuyées sur un dispositif expérimental comportant des taureaux de deux races : les races Holstein et Montbéliarde. Il a été mis en évidence par classification hiérarchique de corrélation une distinction très nette des méthylomes spermatiques entre les taureaux de ces deux races. Il est cependant à noter que les taureaux Holstein proviennent tous du même centre d'insémination et que les taureaux Montbéliard proviennent de 2 centres dépendant de la même entreprise de sélection. Les conduites d'élevage appliquées pour les deux races de taureaux peuvent varier (âge à l'entrée en centre, alimentation, âge à la mise en production...). Il ne peut donc être exclu qu'au moins une partie de cette distinction peut être associée à un effet de l'environnement.

Pour la race Holstein, un plus grand nombre de DMC est observé en mettant en œuvre les comparaisons Fertiles+/Fertiles-, Déviants/Fertiles+ et Déviants/Fertiles-, que pour la race Montbéliarde. Les comparaisons mettant en jeu les taureaux Déviants génèrent aussi un nombre de DMC supérieur à celui de la comparaison Fertiles+/Fertiles-. De plus, ces DMC sont majoritairement hypométhylées chez les taureaux Déviants. Pour la race Montbéliarde, les résultats sont différents : le nombre de DMC est moindre et équivalent entre les trois types de comparaisons et les proportions de DMC hyperméthylées/hypométhylées sont similaires. L'annotation des DMC et la recherche de fonctions biologiques enrichies en gènes associés à ces DMC permet de révéler des enrichissements plus significatif et plus nombreux pour la

race Holstein que pour la race Montbéliarde. Le méthylome spermatique présente donc des spécificités pouvant être directement associées à la race, sans exclure l'influence de l'environnement que nous ne pouvions pas mesurer dans notre dispositif expérimental.

Au cours de notre étude, nous avons mis en évidence que le différentiel de méthylation observé par RRBS pour une région du gène *SMARCD3* était en fait dû à un polymorphisme de séquence (C/A), avec une représentation allélique significativement variable en fonction de la catégorie de fertilité analysée. Cette observation nous a conduit à amorcer une étude pilote dans le but de mieux décrypter le contrôle génétique exercé sur le méthylome des individus. En comparant les données de génotype aux positions des DMC, nous avons ainsi pu observer que 42 SNP sont situés directement sur une DMC pour la race Holstein, et 25 pour la race Montbéliard. Bien que ce cas de figure soit très minoritaire au regard des milliers de DMC que nous avons identifiées pour chaque race, ces observations soulignent l'importance de l'évaluation de la diversité génétique en regard des différentiels de méthylation mis en évidence.

B. Des modifications du méthylome spermatique en fonction de la fertilité ciblent des processus biologiques variés

Lors des analyses conduites en comparant les catégories de fertilité pour les taureaux Holstein (pour les deux types de banques, simple digestion ou double digestion) et pour les taureaux Montbéliard, nous avons identifié des DMC, nous les avons annotées et nous avons recherché les fonctions biologiques enrichies en gènes associées à ces DMC. L'enrichissement en DMC de ces fonctions biologiques est particulièrement significatif pour la race Holstein, mais l'est moins pour la race Montbéliard du fait d'un nombre plus faible de DMC. Pour la race Holstein nous avons effectué une analyse complémentaire sur 15 individus en préparant des banques RRBS issues d'une double digestion HaeIII + MspI. Cette préparation nous a permis de cibler en majeure partie des gènes différents de ceux que nous avons dans un premier temps analysé à partir des banques MspI. Les analyses d'enrichissement DAVID ont cependant montré que le même type de processus biologiques est impacté par les différentiels de méthylation, ce qui conforte nos résultats.

De manière générale, les voies biologiques mises en lumière sont impliquées dans le développement du système nerveux et son fonctionnement, dans le développement embryonnaire incluant les processus de différenciation / prolifération et apoptose cellulaire, dans les interactions cellule-cellule, dans les processus d'adhésion cellulaire et dans les

mécanismes de transduction des signaux intracellulaires en incluant des activités telles les kinases et les GTPases et des flux ioniques. L'ensemble de ces processus sont en lien direct avec la fertilité. Certaines peuvent affecter le fonctionnement du spermatozoïde, d'autres pourraient avoir un rôle prépondérant dans les altérations du développement embryonnaire post-fécondation plus ou moins tardif (perte embryonnaire voire fœtale au cours de la gestation à relier avec l'augmentation de TNR au cours de la gestation).

Le catalogue de DMC identifiées dans notre analyse comporte des points communs avec ceux décrits dans la littérature. Dans une analyse de la méthylation pangénomique par MBD-seq, Kropp et collaborateurs ont comparé 2 groupes de taureaux constitués en fonction du taux de conception (bon *versus* mauvais). Ils ont observé de manière générale une méthylation plus faible pour les individus de plus basse fertilité, et ce résultat est à rapprocher de l'observation d'hypométhylation que nous avons faite pour la catégorie Déviants. Ils ont également pu identifier 76 DMR en comparant les taureaux fertiles et sub-fertiles. L'utilisation de notre pipeline d'annotation en utilisant les positions chromosomiques de ces 76 DMR a permis d'associer ces dernières à 26 gènes. Il s'avère que nous avons pu identifier des DMC pour 9 des 26 gènes, avec notre dispositif. Toujours en utilisant notre pipeline d'annotation, nous avons associé en fonction de ses coordonnées génomiques, une DMR au gène *ImmunoGlobulin Lambda-Like polypeptide 1, IGLL1* ; non identifié dans l'étude de Kropp et al. Un second gène identifié également d'après les coordonnées chromosomiques de Kropp est *U6 spliceosomal RNA (U6)*. Ce gène, présent en de multiples copies dans le génome bovin, code pour un ARN non-codant impliqué dans la formation du spliceosome. Nous avons pu identifier des DMC pour 12 copies du gène et réparties sur 10 chromosomes pour la race Holstein, et 11 copies sur 11 chromosomes pour la race Montbéliarde. De manière intéressante, une étude GWAS portant sur l'âge à la puberté et menée chez des génisses de race *Bos indicus* a montré que parmi les 6 meilleurs gènes candidats, 3 correspondaient à des copies du gène *U6* (Nascimento et al. 2016). Bien qu'il ne s'agisse pas des mêmes copies que celles que nous avons étudiées, ces données renforcent nos observations et soulignent l'important de ce gène dans les processus liés à la fertilité. Un contrôle épigénétique de l'expression de certaines de ces copies peut donc aussi contribuer à ces processus.

Une autre étude réalisée par Verma et collaborateurs en utilisant une micropuce ciblant les CpG localisés dans le promoteur de 2967 gènes a permis d'identifier 73 gènes hyperméthylés pour 5 taureaux Fertiles et 78 pour 5 taureaux sub-fertiles. Les analyses d'enrichissement montraient qu'une implication de ces gènes dans la régulation de la transcription et de la prolifération cellulaire. En particulier, 13 de ces gènes sont décrits comme jouant un rôle

important dans la spermatogenèse, la capacitation ou le développement embryonnaire. Dans notre étude nous retrouvons un différentiel de méthylation pour 4 de ces 13 gènes : *Snail Family Transcriptional Repressor 2 (SNAI2)*, *Leucine-Zipper-like Transcription Regulator 1 (LZTR1)*, *DNA Ligase 3 (LIG3)*, et *Leucine-rich Repeats and Guanylate Kinase domain containing (LRGUK)*. Parmi l'ensemble des 151 gènes différentiellement méthylés observés par Verma et al., nous identifions avec notre approche 17 gènes en commun. Il est à souligner que les approches méthodologiques sont différentes (MeDIP-chip *versus* RRBS). Elles ne ciblent pas les mêmes régions génomiques : l'une cible uniquement la région promotrice de tous les gènes et l'autre est pangénomique.

C. Impact fonctionnel des altérations du méthylome spermatique : conséquences pour la descendance

À ce niveau d'analyse, nous ne pouvons que faire des associations entre différentiel de méthylation et gènes associés sans pouvoir en vérifier l'impact fonctionnel. En effet, pour que les différentiels de méthylation aient effectivement une influence sur le développement embryonnaire, ceux-ci ne doivent pas avoir été reprogrammés lors de la phase de reprogrammation épigénétique du génome paternel qui a lieu après la fécondation. Il s'avère que différentes études ont montré que certains sites CpG échappent à la reprogrammation épigénétique et conservent leur état de méthylation spermatique sur certains *loci* (Tang et al. 2015). Dans un second temps, il faut que les sites CpG ayant échappé à la reprogrammation soient impliqués dans la régulation de l'expression des gènes. Pour la race Holstein comme Montbéliard, près de la moitié des DMC identifiées sont situées dans des CGI, CpG *shores* et CpG *shelves*, des régions qui sont impliquées dans la régulation de l'expression génique. Il est cependant difficile de statuer de manière générale sur l'effet qu'auront les différentiels sur l'expression des gènes car cela dépend du contexte génomique de chaque DMC. Les DMR sont plus informatives à ce sujet. Nous avons observé qu'elles sont fréquemment retrouvées dans les CGI (à 38,6% chez les Holstein et 30,0% chez les Montbéliard), ainsi que très majoritairement en intragénique (introns et exons). Il est alors possible que ces DMR soient impliquées dans des processus d'épissages alternatifs ou soient situées dans des promoteurs intragéniques (Maunakea et al. 2010) (Maunakea et al. 2013).

Pour la suite de nos travaux, il peut être proposé de mettre en place une production d'embryons par fécondation *in vitro* en utilisant de la semence des taureaux porteurs des altérations de méthylation décrites ci-dessus. Ces études nous permettraient de préciser la

possibilité d'une transmission d'information non génétique et de mieux définir les effets d'altérations épigénétiques portées par le génome paternel. Non seulement, le développement embryonnaire pourrait être suivi en termes de cinétique de clivage cellulaire et de morphologie, mais il serait aussi nécessaire d'effectuer un suivi de la reprogrammation épigénétique embryonnaire par immunocytochimie afin d'étudier la distribution des marques d'histones en domaine répressif et permissif. Une analyse de la méthylation de l'ADN, après l'activation du génome embryonnaire (c'est-à-dire aux stades 8 à 16 cellules) et au stade blastocyste en séparant le trophoctoderme et les cellules de la masse interne (ICM) pourrait être conduite en adaptant les différentes étapes du protocole RRBS à l'utilisation de faibles quantités de matériels biologiques (Wang et al. 2015). Couplée à ces analyses, des analyses transcriptomiques nous paraissent aussi indispensables (Hosseini et al. 2015).

D. Le catalogue de DMC identifiées peut-il servir de base pour le développement d'un outil d'épigénotypage de la semence bovine ?

Nous avons tenu à procéder à la validation de nos données, pour s'assurer de la robustesse de l'analyse par RRBS. Pour cela, nous avons alors sélectionné dans un premier temps des gènes présentant une DMR spécifique aux Déviants pour la race Holstein. Nous avons également sélectionné des DMR répondant à divers critères (le nombre d'échantillons minimum par catégorie utilisé, le seuil de différentiel de méthylation, la position dans le gène). Nous avons aussi sélectionné des gènes présentant des DMC communes entre les races Holstein et Montbéliard.

Ces résultats ont montré qu'il existe de très bonnes corrélations entre les pourcentages de méthylation calculés par RRBS et ceux calculés par pyroséquençage, CpG par CpG pour l'ensemble des individus analysés et les cinq DMR choisies. Deux des cinq DMR, *WWC2* et *SNIP1*, présentent aussi un différentiel de méthylation moyen statistiquement validées après pyroséquençage. On observe également une nette tendance à l'hypométhylation chez les Déviants pour les gènes *RARβ* et *TLE3*, mais les différentiels ne sont pas statistiquement significatifs. L'absence de significativité après pyroséquençage, malgré la bonne corrélation qui existe entre les résultats de RRBS et de pyroséquençage, est liée aux méthodes statistiques utilisées. En effet, en RRBS, le logiciel methylKit identifie les DMC sans prendre en compte la variabilité inter-individuelle : pour chaque position CpG, le logiciel comptabilise le nombre de C et le nombre de T en totalisant les séquences pour tous les individus par catégorie. C'est ce ratio C/T par catégorie qui est comparé entre catégories, et donc la variabilité inter-

individus n'est pas prise en compte. Une importante variabilité inter-individuelle des taux de méthylation est néanmoins observée par pyroséquençage, ce qui ne permet pas de valider les résultats. En effet, il n'est pas rare qu'au sein d'une même catégorie de fertilité, le pourcentage de méthylation moyen puisse varier de $\pm 30\%$ entre les individus, voire parfois beaucoup plus. Les profils de méthylation obtenus par pyroséquençage sont donc peu discriminants entre catégories de fertilité. La bonne concordance des résultats entre les 2 approches techniques nous donne cependant confiance dans les différentiels de méthylation identifiés par RRBS. Ces données suggèrent qu'une combinaison entre de nombreuses DMC/DMR soit nécessaire pour prédire de manière fiable le statut de fertilité d'un taureau. Un outil d'épigénotypage haut débit devra donc intégrer les biomarqueurs les plus pertinents.

Les DMR à valider ont été sélectionnées sur la base de plusieurs critères :

- Le différentiel de méthylation concerne les Déviants par rapports aux Fertiles+ et Fertiles-. De plus, les DMR sont associées à des gènes impliqués dans le développement embryonnaire (approche d'un point de vue biologique)
- La couverture des DMC qui composent la DMR et le nombre d'individus couverts par catégorie est variable (approche d'un point de vue technique)

Pour les gènes *RAR β* et *TLE3*, la DMR est présente en région intronique ; elle est exonique pour les gènes *WWC2* et *SNIP1*. Celle de *LAMB2* est située dans le promoteur du gène. De plus, alors que pour les gènes *WWC2*, *RAR β* et *LAMB2* les DMR et DMC ne sont pas associés à un des types de régions denses en CpG (CGI, *shore* ou *shelve*), les DMR de *SNIP1* et *TLE3* sont situés dans un CGI. Des études ont montré que dans certains types cellulaires, l'expression de ces gènes est en partie contrôlée par la méthylation. Il a été montré que l'hyperméthylation du CGI du promoteur de *RAR β* est associée à une diminution de l'expression du gène dans des lignées de cellules ovariennes tumorales humaines (Pronina et al. 2016). Dans une autre étude, une hyperméthylation du promoteur de ce gène a été retrouvée dans 14% des échantillons de moelle osseuse de patients atteints de leucémie myéloïde infantile (Sakaguchi et al. 2015). Des aberrations de la méthylation du gène *TLE3* ont été relevées dans des cas de cancers colorectaux chez l'Homme (Farkas et al. 2014). Concernant *WWC2*, il n'existe à ce jour aucune donnée sur la possible régulation de ce gène par la méthylation. Il a cependant été décrit que la diminution de l'expression de ce gène est associée à une augmentation de la croissance tumorale dans des cas de carcinomes hépatocellulaires chez l'Homme (Zhang et al. 2017). Il n'existe pas non plus de données de ce type pour les gènes *SNIP1* et *LAMB2*.

Bien que les différentiels que l'on observe ne soient pas situés dans le promoteur des gènes, il est possible que ces différentiels aient tout de même une influence sur leur expression. Il a notamment été montré qu'il existe une forte corrélation entre hyperméthylation de région génomique localisée dans le corps d'un gène et son expression, dans le cas où la DMR n'est pas retrouvée en contexte de CGI (Schultz et al. 2015), ce qui est le cas pour les gènes *WWC2* et *RARβ*. Dans ce contexte, on peut émettre l'hypothèse que l'hypométhylation retrouvée chez les Déviants pourrait être associée à une diminution de l'expression de ces gènes (dans l'optique où cette hypométhylation serait conservée après la reprogrammation épigénétique de l'embryon précoce). Pour les gènes *SNIP1* et *TLE3*, les DMR sont retrouvées dans le corps des gènes et en contexte CGI. Dans ce cas, l'hypométhylation retrouvée chez les Déviants pourrait conduire à des altérations des mécanismes d'épissages alternatifs de ces gènes (Laurent et al. 2010), et donc impliquer l'expression d'isoformes protéiques. Ces considérations sont cependant uniquement spéculatives, et il serait nécessaire de procéder aux analyses sur embryons tels que décrits dans la section précédente.

E. Effet de l'âge à la production de semence sur le méthylome spermatique

En plus de la fertilité, le facteur le plus susceptible de contribuer aux différentiels de méthylation que nous avons observés est celui de l'âge des individus. En effet, il existe une différence importante de l'âge à la production de semence des Déviants par rapport aux 2 autres catégories de fertilité : les taureaux Déviants Holstein ont produit les paillettes de semence à un âge significativement plus précoce (médiane de 15 ± 6 mois) que les taureaux Fertiles+ et Fertiles- (respectivement 54 ± 24 et 47 ± 26 mois). Cette caractéristique n'est pas retrouvée pour les taureaux Montbéliard. Ainsi, pour les Holstein, alors que les taureaux Fertiles+ et Fertiles- ont été utilisés à un âge adulte, les Déviants ont été employés en production peu de temps après leur puberté.

Il existe un nombre important de données dans la littérature montrant l'impact de l'âge sur le méthylome. Dans les tissus somatiques, alors qu'on observe une augmentation générale du niveau de méthylation de l'ADN des cellules sanguines au cours de la première année après la naissance chez l'Homme (Herbstman et al. 2013), à l'âge adulte, on observe une hypométhylation globale corrélée avec l'âge des individus. L'hypométhylation est retrouvée préférentiellement dans certains types d'éléments répétés comme les séquences *Alu* et *LINE1* chez l'Homme (Bollati et al. 2009) ainsi que dans des régions intergéniques peu denses en

CpG. Au contraire, il a été montré que les régions riches en CpG, comme les CGI et les CpG *shores*, présentent une hyperméthylation avec l'âge (McClay et al. 2014).

Les études relatant l'évolution de l'épigénome des spermatozoïdes avec l'âge vont à l'encontre de ce qu'on observe pour les tissus somatiques. Dans une étude longitudinale réalisée sur 17 Hommes, pour qui 2 prélèvements de semence ont été réalisés avec entre 9 et 19 ans d'écart, il a été observé une hyperméthylation globale mesurée par LINE1-pyroséquençage (Jenkins et al. 2014). De façon similaire, une étude récente a montré une augmentation globale de la méthylation avec l'âge dans les spermatozoïdes de taureaux de races Holstein et *Japanese Black* (Takeda et al. 2017). Il s'agit des mêmes tendances que ce qu'on a observé dans notre étude. Puisque les Déviants sont globalement plus jeunes que tous les autres taureaux, on constate :

- Une hypométhylation des Déviants pour les CpG₅₀₀, dont le pourcentage de méthylation global est de 16,9% contre respectivement 19,3 et 21,9% pour les taureaux Fertiles+ et Fertiles-
- Une hypométhylation des Déviants pour les DMC, dont la proportion est de respectivement 67,3 et 62,4% par rapports aux catégories Fertiles+ et Fertiles-

Il est d'ailleurs possible que les différences liées à l'âge soient sous-évaluées. En effet, alors que les séquences répétées semblent particulièrement impactées par les changements de méthylation liés à l'âge, la majorité d'entre elles sont éliminés lors de l'étape de sélection de taille en RRBS et pendant l'alignement des séquences sur le génome de référence.

Il est intéressant de noter que chez l'Homme, l'augmentation de l'âge à la reproduction est associée à une prévalence augmentée de troubles neuropsychiatriques chez les descendants, comme l'autisme, les troubles bipolaires ou encore schizophrénie (Miller et al. 2011). De plus, chez la souris, l'âge à la reproduction du père implique des différentiels de méthylation dans le cerveau des descendants et peut être associé à des changements d'expression de gènes et des troubles du comportement (Smith et al. 2013) (Milekic et al. 2015). Or, quand on regarde les enrichissements DAVID que nous avons obtenus pour les DMC entre Fertiles+ et Déviants, on observe que beaucoup de processus biologiques sont liés à au fonctionnement neuronal. Etant donné que pour cette analyse les effets de la fertilité et de l'âge ne peuvent pas être dissociés, il est possible que les DMC associées à l'âge soient celles impactant des gènes impliqués dans ces processus.

En plus de l'âge à la production de semence, un second paramètre pouvant expliquer la présence surnuméraire de DMC chez les Déviants concerne leur année de naissance. En effet,

la plupart des taureaux de catégorie Fertiles+ et Fertiles- sont nés avant l'année 2008 (allant de l'année 2000 à l'année 2011, avec l'année 2006 comme médiane). Ce sont donc des taureaux nés avant l'utilisation en routine de la génomique et qui ont été employés en testage sur descendance, c'est pourquoi leur âge à la production est plus élevé que pour les taureaux Déviants. En revanche, les individus Déviants de race Holstein sont tous nés entre 2009 et 2010, c'est-à-dire que ce sont des taureaux génomiques, ce qui peut expliquer leur âge à la production. Il est possible qu'entre les années 2000 et 2010 il y ait eu des changements dans les conduites d'élevage pouvant avoir un impact sur la méthylation de l'ADN spermatique. Ces changements pourraient être :

- Une modification de l'alimentation, notamment de la supplémentation. Certains des composés présents dans les compléments alimentaires peuvent avoir un impact sur l'état de méthylation de l'ADN, notamment des facteurs antioxydants ou le Sélénium. Ce dernier est en effet qui est impliqué dans le fonctionnement d'enzymes à action antioxydante, et a une influence sur la fertilité (Shalini et Bansal 2005)
- Un changement de diluant (ou de la composition du diluant) utilisé pour produire et conserver les paillettes de semence

F. Nouvelle analyse du méthylome spermatique en fonction de la catégorie de fertilité en excluant les effets liés à l'âge à la production de la semence

Lors des premières analyses d'identification des DMC, ces aspects liés à l'âge n'ont pas été pris en considération et sont donc confondus avec les effets de la fertilité. Après avoir identifié les DMC associés à l'âge, nous avons procédé à une simple soustraction pour obtenir un nouveau jeu de DMC. Les analyses d'annotation puis d'enrichissement ont été effectuées à nouveau, en utilisant cette fois les données de DMC corrigées. Cette correction des données exacerbe les différences d'enrichissement entre les catégories de fertilité, à savoir un fort enrichissement des fonctions biologiques concernant les processus développementaux, la régulation de la migration cellulaire chez les Fertiles-, ainsi que la régulation de la fonction neuronale pour les Déviants.

En plus des analyses d'enrichissement par DAVID, en effectuant des recherches manuelles de processus biologiques enrichis en DMC nous avons observé que les gènes de la voie de signalisation de l'Hippo semblent particulièrement impactés par des différentiels de méthylation. La voie de signalisation Hippo a été identifiée il y a 20 ans chez la Drosophile.

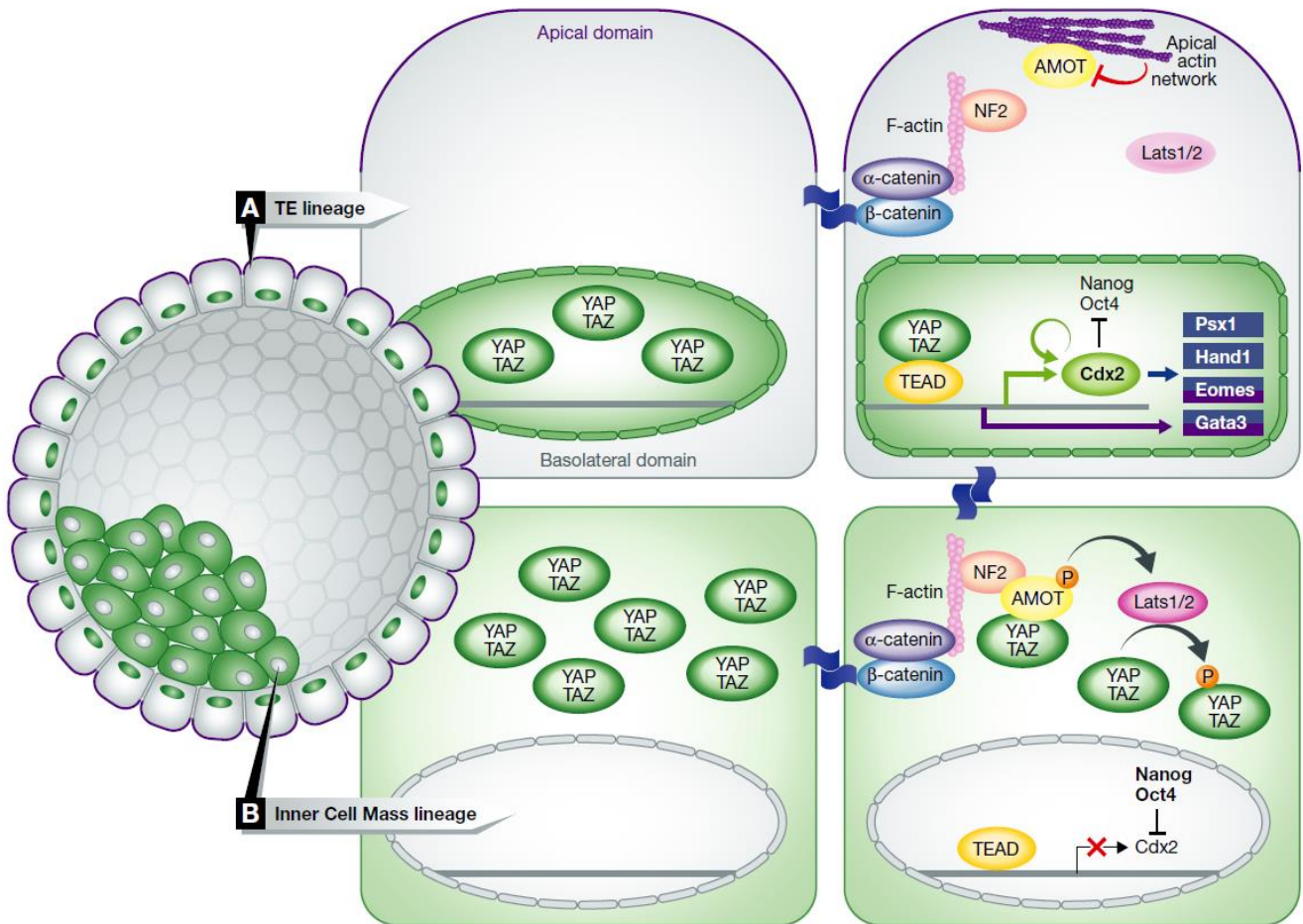


Figure 28. Régulation du lignage Trophectoderme-ICM par la voie Hippo dans l'embryon préimplantatoire

Dans les blastomères formant l'enveloppe externe de l'embryon, la localisation nucléaire de YAP et de TEAD4 permet l'activation de gènes comme Gata3, Cdx2 et Eomes (A). Ces cellules ont vocation à se différencier pour former le trophoctoderme (TE). Dans les blastomères situés à l'intérieur de l'embryon, les adhésions cellulaires permettent l'activation de la voie Hippo. La voie Hippo empêche la localisation nucléaire d'YAP, et en conséquence limite l'activation de la transcription des gènes cibles de TEAD4 (B). Ces cellules formeront la masse cellulaire interne (*Inner Cell Mass*, ICM), dont le caractère pluripotent est maintenu par l'activation de Nanog et Oct4.

(tiré de Mo *et al.*, 2014)

Hautement conservée entre espèces, elle intervient dans la régulation de la prolifération cellulaire, de la croissance des organes et de l'apoptose. Il est ainsi apparu que la voie Hippo est dérégulée dans une grande variété de cancers chez l'Homme (Nishio et al. 2017). Il a également été décrit que la voie Hippo a une implication fondamentale dans la régulation de la différenciation des lignages cellulaires entre trophectoderme et ICM au stade blastocyste (Figure 29). Sur les 159 gènes décrits comme jouant un rôle dans cette voie, 23 contiennent au moins une DMC (toute comparaison confondue) pour la race Holstein (même après avoir corrigé les effets liés à l'âge des individus) et 20 pour la race Montbéliard. Par exemple, pour les 2 races nous avons retrouvé des DMC dans le gène *Neurofibromin 2 (NF2)*, dont il a été décrit que la déplétion résulte en une spécification aberrante des cellules en ICM (Cockburn et al. 2013). Nous avons également identifié des DMC pour le gène *TEA Domain Transcription Factor 1 (TEAD1)*, dont il a été montré que l'inactivation est létale au stade embryonnaire à E11.5 chez la souris (Chen, Friedrich, et Soriano 1994), ou encore le gène *TEAD4*, dont la perte de fonction conduit à un échec de formation du blastocœle (Yagi et al. 2007).

Pour l'ensemble des DMC identifiées sur les gènes de la voie Hippo, pour la race Holstein et Montbéliard, dans respectivement 72% et 83% des cas où l'on peut observer un différentiel de méthylation de plus de 10% chez les Déviants et les Fertiles- par rapport aux Fertiles+, le différentiel va dans le même sens (hypo ou hyperméthylation). Cependant, il ne se dégage pas de tendance générale à l'hyper- ou hypométhylation dans une catégorie de fertilité par rapport à une autre. En revanche, il est à noter que chez les Holstein plus de 70% des DMC sont en contexte de CGI, CpG *shore* ou CpG *shelve*, ainsi qu'en contexte intronique. Le constat est similaire pour la race Montbéliarde. Ainsi, comme pour les autres DMC identifiées, il est possible que ces différentiels de méthylation puissent avoir un impact sur l'expression de ces gènes au stade embryonnaire, s'ils échappent à la reprogrammation. Une hypothèse que nous pouvons formuler est que de multiples petits différentiels d'expression pour certains des gènes de la voie de signalisation de l'Hippo puissent avoir pour conséquence une dérégulation de la mise en place des premiers lignages cellulaires au cours du développement embryonnaire. Ces perturbations pourraient conduire dans certains cas à des pertes embryonnaires, qui se traduiraient par une apparente sub-fertilité des taureaux ayant fourni la semence.

En conclusion, nous avons pu établir un catalogue de DMC liées à la fertilité mâle dans la semence bovine. Ce travail n'est cependant pas achevé et va nécessiter de mener plusieurs analyses complémentaires.

En premier lieu, étant donné que dans le groupe des Déviants les individus sont tous très jeunes, il reste à préciser les DMC influencées par leur âge ou leur sub-fertilité. Pour approfondir ce volet de notre analyse il sera indispensable d'ajouter au dispositif expérimental un groupe d'individus fertiles du même âge. En identifiant les DMC entre le groupe de jeunes fertiles et le groupe des jeunes déviants, on pourra précisément identifier les DMC liées à la fertilité. A partir de ces données et pour raffiner notre analyse, il sera nécessaire de refaire les analyses discriminantes et d'intégration avec les SNP en prenant en compte ces DMC corrigées de l'âge des individus. Idéalement, il faudrait ajouter à l'analyse le plus de paramètres possibles (l'âge de naissance, la nutrition, etc.). Cette approche conduira à réduire le nombre de DMC pour ne se focaliser que sur ceux qui sont uniquement révélateur de la fertilité des individus, sans être influencé par tous ces paramètres extérieurs. Nous obtiendrons alors un catalogue de DMC. Nous pourrions seulement à ce stade extrapoler sur l'impact que peuvent avoir ces différentiels de méthylation sur la fertilité. Il sera donc nécessaire d'obtenir des données d'expression de ces gènes dans des embryons produits à partir de la semence des individus du dispositif. Ces données pourront être complétées de données fonctionnelles des embryons. Par exemple, si effectivement la voie de signalisation Hippo est altérée, on peut s'attendre à une perturbation de la mise en place du premier lignage cellulaire et donc à une altération de la formation du trophoctoderme par rapport à l'ICM. Des analyses de la méthylation des gènes identifiés permettront aussi de voir si les différentiels sont toujours observés au stade blastocyste. Comme la voie de signalisation Hippo est également impliquée dans la polarité de l'axe embryonnaire, on pourra potentiellement observer macroscopiquement des modifications de la morphologie des blastocystes.

Discussion générale

Points technologiques critiques pour une fiabilité des biomarqueurs

L'objectif principal de la thèse consistait en l'identification de biomarqueurs de la fertilité mâle bovine dans le cadre du projet SeQuaMol (les principaux résultats étant exposés dans la seconde partie du chapitre « Résultats »). Le catalogue de biomarqueurs doit servir de base pour la seconde phase du projet, à savoir la phase de validation de ces marqueurs sur une cohorte de taureaux plus importante (d'une centaine d'individus par race). Pour s'assurer de leur fiabilité, un effort important a été fourni pour l'optimisation et la standardisation des protocoles, pour la mise en place de processus de contrôles qualité pour l'ensemble de la chaîne de production des données. Dans cette partie, seront discutés différents points critiques ayant attiré notre attention dans la recherche de biomarqueurs sous forme de différentiels de méthylation de l'ADN en fonction de la fertilité des taureaux.

Homogénéité de la population cellulaire

Avant l'extraction d'ADN génomique, une fraction de chaque éjaculat a été analysée au CASA et en cytométrie en flux, pour effectuer respectivement les analyses fonctionnelles et la mesure de la concentration spermatique. A cette étape, il est possible de déterminer l'état de contamination de l'éjaculat par d'autres types cellulaires. Une contamination, même de faible envergure, peut avoir une influence non-négligeable sur les résultats des analyses de méthylation, car comme nous l'avons vu, il existe d'importantes différences de méthylation globale entre les types cellulaires. Dans un éjaculat, la contamination pourrait être le fait de leucocytes (dans le cas d'une inflammation), de cellules épithéliales ou de bactéries. Les analyses systématiques n'ont révélé aucune contamination pour l'ensemble des éjaculats. La population cellulaire est homogène, et en conséquence les pourcentages de méthylation calculés sont bien révélateurs du méthylome spermatique uniquement. Dans ce cas, il n'est pas paru essentiel de procéder à un tri cellulaire. Il est nécessaire de garder à l'esprit que dans le cas contamination par d'autres types cellulaires, un tri ou une sélection des spermatozoïdes n'est pas sans conséquence : par exemple chez l'Homme, la sélection des spermatozoïdes s'effectue souvent sur gradient et dans ce cas seuls les spermatozoïdes mobiles sont conservés et analysés. Chez le taureau, nous avons écarté ce biais pour nos analyses.

Extraction d'ADNg

Les spermatozoïdes sont des cellules dont l'une des caractéristiques est l'extrême compaction du noyau. Ce sont donc des cellules réputées pour la difficulté d'extraction de leur ADNg. Il alors est indispensable d'ajouter des agents chimiques réducteurs dont le rôle est de faciliter la décompaction du noyau en dissociant les ponts disulfures, notamment du 2-mercaptoéthanol ou du dithiothréitol (DTT). Plusieurs protocoles d'extraction de l'ADNg à partir des paillettes de semence ont été expérimenté dans le but d'optimiser le rendement d'extraction tout en évitant la dégradation de l'ADN :

- Le protocole du laboratoire (tel que décrit dans les 2 articles) précédé d'une incubation dans un tampon de lyse complémenté avec de l'héparine (Jager, Wijchman, et Kremer 1990). L'héparine est un polyanion qui entre en compétition avec l'ADN (également un polyanion) dans sa liaison avec les protamines
- Le protocole standard précédé d'une incubation dans un tampon de lyse complémenté avec de la papaïne, qui est une protéase à cystéine (Bugno-Poniewierska, Jabłońska, et Słota 2010)
- Un protocole utilisant du guanidium et du 2-mercaptoéthanol (Hossain et al. 1997), dont on a fait varier les durées d'incubation (4h ou sur la nuit)
- Nous avons également effectué des extractions selon le protocole standard en faisant varier la concentration en DTT (0 ; 0,1 ; 1 ; 10 et 50 mM)

Le protocole du laboratoire s'est révélé être le plus efficace en termes de rendement, sans introduire de dégradation de l'ADN (contrôler après migration électrophorétique, par LUMA et génotypage). Nous avons observé qu'une concentration de 10 mM de DTT était absolument nécessaire pour obtenir une destruction des têtes de spermatozoïdes (à 1 mM de DTT, il est encore possible d'observer au microscope optique les têtes des spermatozoïdes intègres). A titre de comparaison, dans l'étude réalisée par Kropp, il n'est pas fait mention de l'utilisation de DTT pour l'extraction d'ADN. Dans ce cas, il est possible que seule une fraction de la population de spermatozoïdes (par exemple, celle présentant un début de décompaction du noyau) soit correctement extraite, et c'est alors prendre le risque de n'analyser que cette sous-population cellulaire.

Qualité de l'ADN génomique issu de l'extraction

Il est décrit dans la littérature que la dégradation de l'ADN peut induire une hypométhylation en LUMA (Head, Mittal, et Basu 2014). Dans nos analyses LUMA, nous avons contrôlé la corrélation entre les mesures de sites MspI et de sites EcoRI, révélatrice de l'intégrité des ADN génomiques utilisés. Nous avons alors pensé que ce phénomène pouvait potentiellement être expliqué par une perte de matériel au cours de l'extraction. En effet, une mauvaise remise en suspension de certaines régions hyperméthylées de l'ADN au cours de l'extraction aurait pu résulter en une hypométhylation globale lue LUMA. Dans le but de vérifier que l'ADN génomique préparé selon notre protocole couvre bien l'ensemble du génome, nous avons procédé à un génotypage. A titre de comparaison, nous avons également extrait de l'ADN à partir de sang en utilisant le même protocole que celui utilisé pour les spermatozoïdes, pour les mêmes individus. Nous avons observé, que ce soit pour les spermatozoïdes ou les cellules sanguines, que l'ensemble des SNP étaient représentés dans les résultats de génotypage. Nous en avons donc conclu que l'hypométhylation globale des spermatozoïdes bovins ne provenait ni d'une dégradation de l'ADN, ni d'une perte de matériel biologique. De plus, cette hypométhylation mesurée par LUMA a été également observée en utilisant d'autres méthodes bien moins sensibles à la dégradation de l'ADN (MeDIP-chip et RRBS).

Préparation et standardisation de la préparation des banques RRBS

Une des particularités du protocole de RRBS que nous employons est l'application de 2 conversions successives de l'ADN au bisulfite de sodium. Alors que la première conversion permet d'obtenir un taux de conversion de l'ordre de 95%, la seconde permet de passer à plus de 99%. Cette différence de quelques pourcents de conversion a un impact fondamental sur les calculs de méthylation. En effet, à la suite du séquençage d'une des séries de préparation de 11 banques RRBS, l'analyse des contrôles qualité FastQC a révélé une mauvaise conversion au bisulfite de sodium. Après alignement des séquences, il s'est avéré que le taux de conversion était d'en moyenne $95,4 \pm 0,9$ %. Le taux de méthylation global des CpG₁₀ de ces échantillons était alors de $54,5 \pm 0,7$ %. De nouvelles banques, préparées à partir des mêmes échantillons, et pour lesquelles le taux de conversion était cette fois d'en moyenne $99,8 \pm 0,1$ %, a fourni un taux de méthylation moyen des CpG₁₀ de $51,7 \pm 0,8$ %. Cette anecdote souligne l'importance d'obtenir un taux de conversion au bisulfite de sodium le plus proche possible de 100%.

Mise au point de la procédure de traitement bio-informatique

Cette mise au point a nécessité de tester plusieurs logiciels, afin de sélectionner le plus efficace. Pour l'alignement, nous avons ainsi essayé Bismark (Krueger et Andrews 2011) et BSMAP (Xi et Li 2009), en exécutant avec chacun des logiciels plusieurs analyses identiques sur un même jeu de donnée. Les résultats fournis par les 2 logiciels ont dans un premier temps paru très similaires. Cependant, il est apparu que Bismark fourni des résultats identiques d'une analyse à l'autre, alors que sur un même jeu de données on observait des variations importantes avec BSMAP. La reproductibilité de Bismark est la raison principale pour laquelle nous avons sélectionné ce logiciel.

En ce qui concerne les options du logiciel, nous avons sélectionné un nombre de mésappariement de 1. Il est important de fixer un seuil ni trop élevé (pour éviter les mauvais alignements), ni trop strict (pour permettre une certaine flexibilité de l'alignement, par exemple s'il y a présence de SNP sur une séquence). Un seuil de mésappariement de 1 est assez strict par rapport à ce qui est parfois utilisé dans les autres études. Certaines études (M. Choi et al. 2015) préconisent un seuil de 4 mésappariements, ce qui laisse présager de la présence de nombreux faux positifs.

Pour l'analyse différentielle, nous avons comparé les résultats de methylKit (Akalın et al. 2012) et de MethylSig (Park et al. 2014). Le premier logiciel utilise un modèle de régression logistique et permet une détection assez large des DMC, tandis que le second utilise un modèle bêta-binomial et est plus drastique, avec notamment la prise en compte de la variabilité inter-individuelle. Pour la race Holstein, le logiciel methylKit a identifié 8884 DMC (toutes catégories de fertilité confondues), alors que le logiciel MethylSig (en utilisant un seuil de significativité avec p value $<0,01$) en identifie 6559. De plus, seules 2370 DMC sont en commun pour les analyses avec les 2 logiciels. Nous avons finalement décidé d'utiliser le logiciel methylKit ; permettant une détection de DMC avec une couverture plus faible et ayant une option pour définir le seuil du différentiel de méthylation d'intérêt (nous avons utilisé un seuil de 25%). Dans la littérature, le seuil de différentiel de 25% est couramment utilisé.

Validations par pyroséquençage

Les validations de DMR ont montré une bonne corrélation entre les résultats de RRBS et ceux de pyroséquençage. Néanmoins, elles ont aussi permis de souligner l'importance à accorder au seuil de couverture de séquençage pour l'obtention d'un taux de méthylation le plus fiable possible. Le seuil de couverture doit être suffisant pour estimer correctement le taux de méthylation (un minimum de 10) et limité (inférieur à 500) afin d'éliminer les biais potentiels d'amplification. Dans la littérature, le seuil choisi est souvent de 5, 8 ou 10 (Zhou et al. 2016) (Korkmaz et Kerr 2017). Quand nous avons établi le méthylome spermatique en comparaison avec d'autres types cellulaires somatiques, les différentiels de méthylation attendus étaient très importants. Se placer à un seuil de couverture de 5 était alors suffisant pour révéler des DMC pertinentes. En revanche, quand nous recherchons des variations du méthylome spermatique en fonction de la fertilité ou de l'âge, les différentiels de méthylation sont plus subtils, et un seuil de couverture de 10 est un minimum.

En effet, nous avons pu observer que la marge d'erreur entre la valeur de méthylation théorique et celle calculée en RRBS se stabilise à moins de 20% qu'à partir d'une couverture de 30 (voir Figure 18, Résultats n°2). Idéalement, c'est à ce seuil qu'il faudrait se positionner pour obtenir les résultats les plus fiables. Cependant, la contrepartie serait la perte importante de CpG analysables, et donc de biomarqueurs potentiels. Par exemple, pour la race Holstein, le nombre de CpG couverts plus de 30 fois est d'en moyenne 762 000, ce qui représente une perte de près des 2/3 des CpG₁₀ (1 962 235 en moyenne).

Dans l'optique où le seuil de couverture serait rehaussé, et pour éviter de perdre une quantité trop importante de CpG, il serait nécessaire d'augmenter la profondeur de séquençage. Pour nos analyses les banques ont été séquencées à profondeur de 25 millions de séquences (\pm 15%). Un essai de séquençage à une profondeur de 40 millions de séquences a révélé que cela permettait d'ajouter 300 000 CpG₁₀ supplémentaires. Un bon compromis pourrait être de séquencer les banques à une profondeur de 40 millions, et de se placer sur un seuil de couverture intermédiaire (mais largement suffisant) de 20.

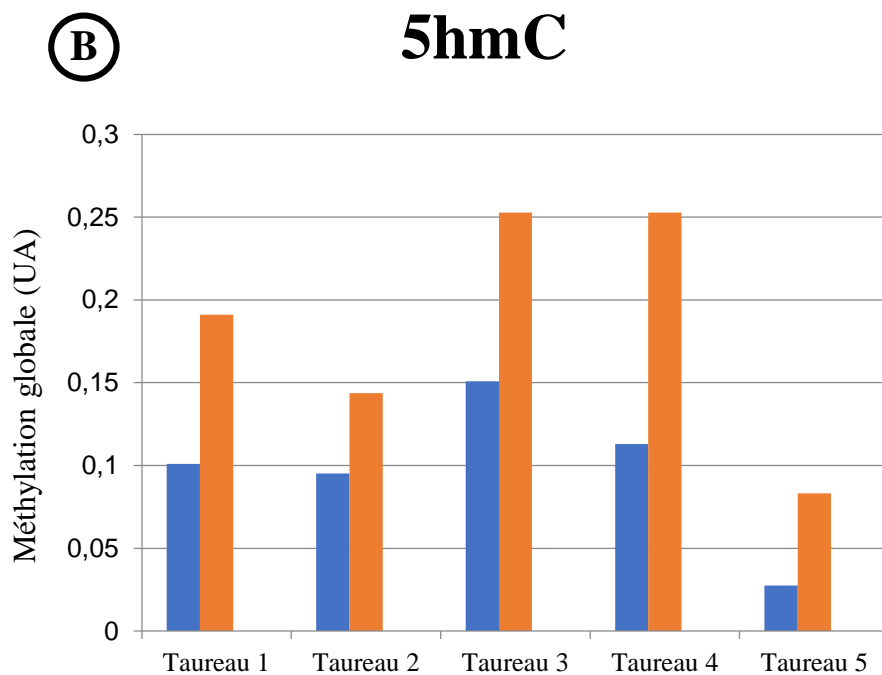
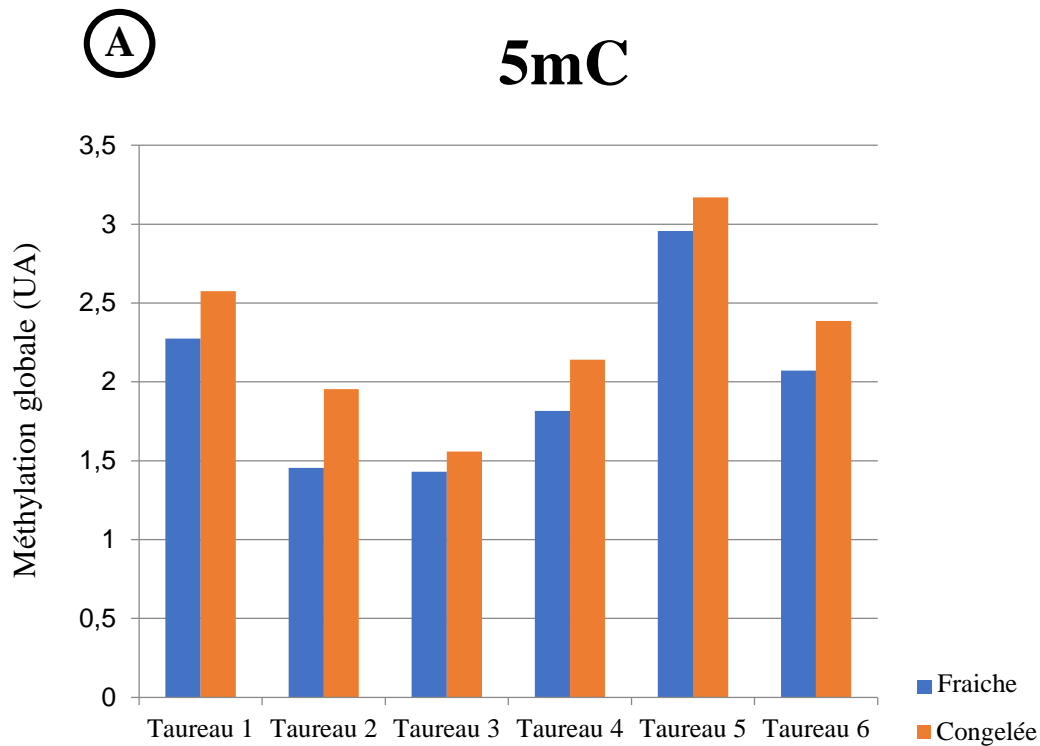


Figure 1. Mesure de la 5mC et 5hmC globale par ELISA dans la semence fraiche et congelée de taureaux

A. Mesure de la 5mC. Aucune différence significative n'a été observée entre la semence fraiche et la semence congelée, bien qu'on observe une légère tendance à l'hyperméthylation en semence congelée. **B.** Mesure de la 5hmC. Comme pour la 5mC, on n'observe pas de différence significative d'hydroxyméthylation entre semence fraiche et congelée. En revanche on observe également une tendance à l'hyper-hydroxyméthylation en semence congelée. La différence relative entre semence fraiche et congelée est aussi plus importante que quand on considère la 5mC. (expérimentations réalisées par Audrey Prézélin)

Une limite de l'approche par RRBS : l'absence de données d'hydroxyméthylation

L'utilisation du RRBS pour l'identification des biomarqueurs est une bonne approche car c'est une technique à bonne résolution (au CpG près), qui permet de cibler des régions génomiques d'intérêt (près de la moitié des CpG couverts sont en contexte CGI, shore ou shelve, et associés à des gènes), tout en limitant les coûts de séquençage par rapport au WGBS. Cette méthode n'est cependant pas sans présenter des limites techniques. En effet, il a été montré que lorsque les sites CpG sont hydroxyméthylés (5hmC), le clivage par MspI est 170 fois plus lent que lorsque le site est non-méthylé, et 70 fois plus lents que lorsque le site est méthylé (Ichiyanagi 2012). Néanmoins, dans nos conditions de clivage enzymatique, nous n'avons pas observé de différences de quantification par LUMA (qui utilise la même enzyme que le RRBS, MspI) des sites CCGG quel que soit le temps d'incubation (entre 4h et 24H). Par ailleurs, il a été décrit que la conversion de l'ADN au bisulfite de sodium ne peut distinguer 5hmC de 5mC ; ainsi les deux marques se traduisent par le maintien des C au niveau des CpG, que ceux-ci aient été 5mC ou 5hmC (Huang et al. 2010).

Quelques données concernant la possibilité d'observer des 5hmC existent chez l'Homme. Il a notamment été décrit que le taux de 5hmC dans les spermatozoïdes est 3 fois moins important que dans les leucocytes (Jenkins et al. 2013), alors que ces derniers sont décrits comme ne présentant que 0,1% d'hydroxyméthylation globale (Nestor et al. 2012). On peut penser que de la même manière, le taux de 5hmC dans les spermatozoïdes bovins doit être très faible. Pour avoir un élément de réponse à ce sujet, nous avons mesuré le taux spermatique global de 5hmC par une approche *Enzyme-Linked ImmunoSorbent Assay* (ELISA). Il s'agit d'une méthode utilisant un anticorps ciblant les sites hydroxyméthylés, couplée à une mesure colorimétrique. Cette méthode a été appliquée sur la semence de 6 taureaux, pour lesquels nous avons obtenu de la semence fraîche. Une partie de cette semence a été cryoconservée, et l'analyse été effectuée sur la semence fraîche et congelée (Figure 1). Nous avons également mesuré le taux de 5mC avec la même méthode. Nous n'avons pas observé de différence significative du taux de 5mC et de 5hmC entre les taureaux et les conditions, mais cependant une tendance à la hausse des taux après cryoconservation. La proportion de 5hmC reste extrêmement faible par rapport à celle de 5mC.

Il serait pourtant particulièrement intéressant de connaître la distribution génomique de cette marque épigénétique, car des liens entre 5hmC et fertilité mâle ont été décrits chez l'Homme.

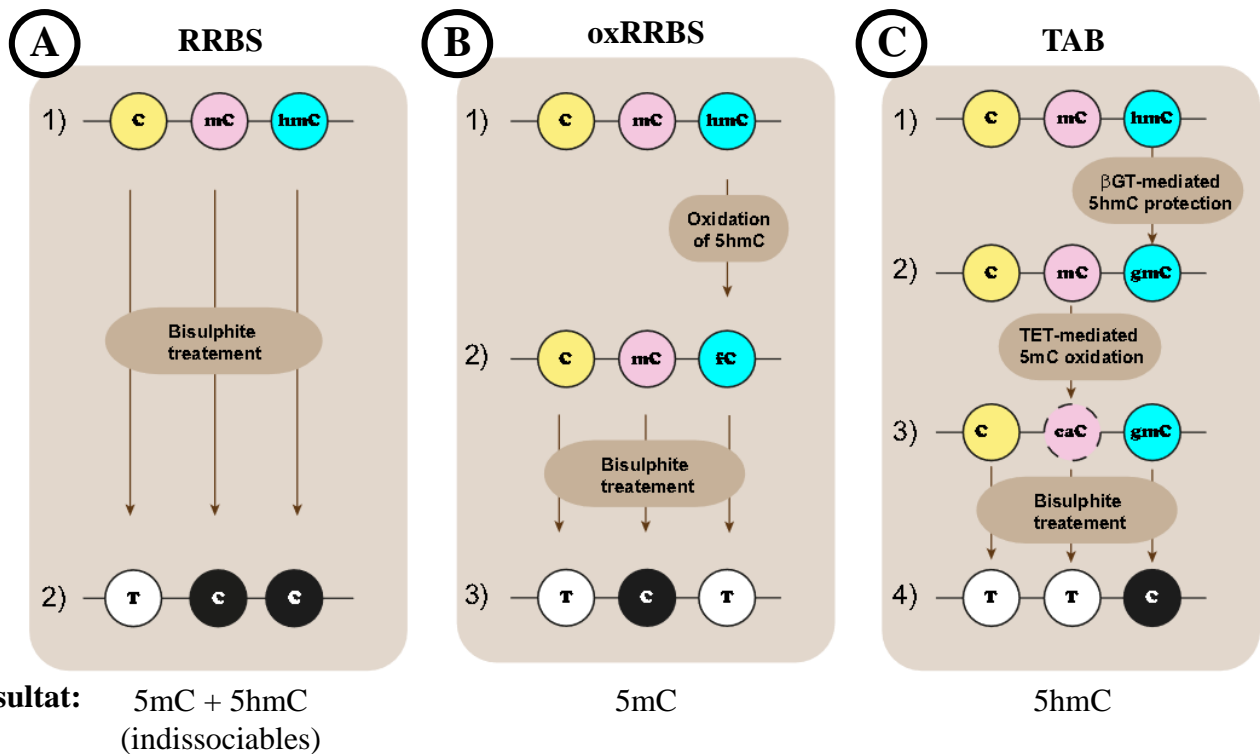


Figure 2. Représentation schématique des méthodes de RRBS, de oxRRBS et de TAB-seq

A. Avec le RRBS, la mesure du taux de 5mC est indissociable de la mesure de 5hmC, car la méthode repose sur la simple conversion de l'ADN au bisulfite de sodium qui ne fait pas la distinction entre les 2 marques. **B.** Avec le oxRRBS, le traitement préalable de l'ADN par le K₂Cr₂O₇ permet de convertir sélectivement les 5hmC en 5fC. La 5fC sera ensuite converti par le bisulfite de sodium, comme les cytosines non-méthylées. Au final, l'information analysable est bien représentative du taux de 5mC et non du taux de 5mC + 5hmC comme en RRBS. **C.** Avec le TAB-seq, le premier traitement consiste en la protection des 5hmC par la β -glucosyltransférase. L'utilisation de l'enzyme TET permet ensuite de d'oxyder les 5mC en 5caC. Avec le traitement au bisulfite de sodium, les cytosines comme les 5caC sont converti, et seuls les sites 5hmC protégés ne le sont pas. Cette méthode permet alors de ne récupérer que l'information de 5hmC. (adapté de Skvortsova et al. 2017)

En effet, il a été montré que des altérations des taux globaux de 5hmC sont retrouvés chez des patients globozoospermiques, c'est-à-dire dont les spermatozoïdes ont une morphologie de la tête anormale, et dont l'acrosome est absent (Wang et al. 2015). Cette étude montre également que la 5hmC a été observée très majoritairement en région intronique. Or, dans notre étude, une partie des DMC a aussi été identifiée en régions introniques (32,4% des DMC pour les Holstein et 36,1% pour les Montbéliard). Il a également été montré que DNMT1 ne reconnaît pas la 5hmC pendant la réplication de l'ADN (Valinluck et Sowers 2007). En conséquence, une augmentation du taux de 5hmC dans les spermatozoïdes pourrait conduire à une hypométhylation anormale après les premières divisions cellulaires au cours du développement embryonnaire précoce, et donc avoir un impact sur la fertilité.

Il est possible de proposer d'améliorer le protocole de RRBS dans le but de pouvoir interroger également les taux de 5hmC. Dans ce but, 2 méthodes ont été mises au point le oxRRBS (Booth et al. 2012) et le *TET-Assisted Bisulphite sequencing* (TAB-seq ; (Yu et al. 2012)) (Figure 2). Le oxRRBS repose sur l'oxydation spécifique des 5hmC en 5fC et 5caC, par l'action du perruthenate de potassium (KRuO_4). Ces bases sont converties en U lors du traitement au bisulfite de sodium. Ainsi, seules les 5mC ne sont pas converties ; donc après séquençage, les pourcentages de méthylation calculés sont bien uniquement ceux du fait de la 5mC. Pour un échantillon donné, en produisant une banque par RRBS classique, et une autre par oxRRBS, il est théoriquement possible d'obtenir également l'information pour la 5hmC en mettant en parallèle les résultats. En TAB-seq, l'ADNg est traité par la β -glucosyltransférase (β -GT), ce qui a pour conséquence l'ajout d'un résidu glucose aux 5hmC. Un traitement par la TET permet alors d'oxyder les 5mC en 5caC, mais reste inactif sur les 5hmC qui sont alors protégées par le résidu glucose. Le traitement au bisulfite de sodium va alors pouvoir révéler le taux de 5hmC uniquement. Le séquençage déterminera alors les régions génomiques concernées par cette marque épigénétique.

Hypométhylation de l'ADN spermatique bovin et compaction nucléaire

Une des caractéristiques du génome spermatique est son hypométhylation globale, une particularité observée chez de nombreuses espèces mammifères. Cette hypométhylation est retrouvée dans les promoteurs de gènes impliqués dans le développement embryonnaire, mais aussi de façon très importante dans les régions satellites péri-centromériques (Molaro et al. 2011) (Yamagata et al. 2007). Le travail de thèse a été amorcé par une analyse globale de la méthylation par LUMA en comparant les spermatozoïdes bovins, à d'autres tissus bovins ou des spermatozoïdes de différentes espèces mammifères. Une hypométhylation très marquée des spermatozoïdes bovins par rapport à tous les autres types cellulaires a été observée. Il s'est avéré que nous avons confirmé cette hypométhylation globale par d'autres approches méthodologiques, à un niveau pangénomique par MeDIP-chip et RRBS, et sur certaines régions candidates par bisulfite-pyrosequencing.

Après avoir effectué des recherches bibliographiques à ce sujet, il s'est avéré que cette information est connue depuis longtemps. En effet, plusieurs études réalisées au début des années 1980 avaient mis à jour cette caractéristique du génome bovin en précisant que l'hypométhylation est associée en partie aux séquences satellites (Kaput et Sneider 1979) (Sturm et Taylor 1981) (Pages et Roizes 1982). Ces séquences représentent une part importante du génome bovin : près de 23% de sa taille totale, dont 4,3% uniquement le fait des satellites de type IV (Macaya, Cortadas, et Bernardi 1978). Elles sont impliquées dans l'organisation spatiale et la structure de la chromatine. Chez l'Homme, l'hypométhylation des séquences satellites péri-centromériques, dans les cas de syndrome *d'Immunodeficiency with Centromeric instability and Facial anomalies* (ICF), est associée à une décondensation de ces séquences et une modification du positionnement des chromosomes dans le noyau (Jefferson et al. 2010).

Une autre des caractéristiques du spermatozoïde est l'extrême compaction de son génome. La compaction est nécessaire pour protéger le génome de l'environnement extérieur : le spermatozoïde est en effet le seul type cellulaire dont la fonction est à effectuer à distance de son organisme d'origine. Cette compaction du génome est le fait du remplacement de 85% des histones par les protamines chez le bovin (Samans et al. 2014). Alors que chez l'Homme et la souris, des protamines de types I et II ont été décrites, chez le bovin, le porc ou l'ovin seule la protamine I a été à ce jour détectée. De plus, il a été montré que le noyau du

spermatozoïde bovin présente une aire respectivement 1,8 et 12,6% plus importante que celui du bélier et du verrat (Yániz et al. 2016), alors que la taille de leur génome est de même ordre (2,7 Gb pour le bovin, 2,6 Gb pour le bélier et 2,8 Gb pour le verrat). Cela suggère que la compaction du génome bovin est moins importante que celle des 2 autres espèces, en particulier par rapport au verrat, tout en étant particulièrement résistant à la lyse (la préparation d'ADN génomique nécessitant une forte concentration de DTT).

La compaction moindre du génome bovin est-elle liée à son hypométhylation ? Une analyse par FRET a montré que la méthylation tendrait à rigidifier l'ADN et à augmenter son degré d'interaction avec les histones, ce qui induirait une augmentation de la compaction de l'ADN autour des nucléosomes (Lee et Lee 2012). Il n'existe cependant pas de donnée à ce sujet concernant les protamines, mais il est probable que l'effet est similaire. D'autres publications indiquent que la méthylation de l'ADN n'affecte pas la compaction de la chromatine, mais qu'elle aurait plutôt une influence sur la structure de la chromatine (Gilbert et al. 2007).

A la lumière de ces éléments, il a paru nécessaire d'éclaircir les conditions spatio-temporelles dans lesquelles s'établit la méthylation de l'ADN dans les spermatozoïdes bovins. Un nouveau projet a alors vu le jour, le projet GerMiMale, dont l'objectif sera d'analyser la cinétique de méthylation de l'ADN à plusieurs étapes clés de la différenciation des cellules germinales en phase fœtale et chez l'adulte, en comparant plusieurs espèces de ruminants (bovin, ovin et caprin).

Dans le cadre du projet SeQuaMol, il serait également intéressant d'avoir des éléments concernant le degré de compaction et la structure du génome bovin. Une méthode d'imagerie tridimensionnelle a été développée et permet de visualiser l'organisation de la chromatine en même temps que la répartition chromosomique des marques de méthylation de l'ADN. Cette méthode, appelée *3D quantitative DNA methylation imaging* (Tajbakhsh 2013), n'a encore jamais été utilisée sur des spermatozoïdes. La méthode est basée sur le traitement des cellules par des anticorps anti-5mC en parallèle d'une coloration globale du génome au 4,6-DiAmidino-2-PhénylIndole (DAPI). L'intensité du signal est mesurée par microscopie confocale à fluorescence, et la distribution relative du signal associé à la méthylation de l'ADN est alors comparée à celle du DAPI. Une représentation 3D de la structure génomique peut être reconstruite par l'utilisation d'un logiciel adapté. La mise en œuvre de cette méthode sur différentes espèces nous fournirait de précieuses informations sur le lien qui peut exister entre hypométhylation et conformation chromatinienne. Cette méthode pourrait également être réalisée pour comparer les taureaux des différentes catégories de fertilité.

Des biomarqueurs révélateurs de la fertilité à l'éjaculat

Au cours de la carrière d'un reproducteur, chaque éjaculat récolté, ayant passé l'ensemble des contrôles qualité, contient suffisamment de spermatozoïdes pour être conditionné en milliers de paillettes. Chaque paillette permettant d'effectuer une insémination, le nombre d'IA effectuées par éjaculat est suffisant pour établir un TNR pour l'éjaculat. Que le taureau soit bon ou mauvais reproducteur, il existe une certaine variabilité dans les TNR associés à chaque éjaculat. La cause de cette variabilité est difficile à évaluer car elle est la résultante d'une multitude d'effets liés au taureau, à son environnement (santé, nutrition, climat...), ainsi qu'à des facteurs techniques de manipulation de la semence.

Lors de la recherche de biomarqueurs de la fertilité, les taureaux dits « Déviants » ont été sélectionnés sur la base d'une non-adéquation entre une prédiction génomique de la fertilité et un indicateur de fertilité. La prédiction génomique est basée sur un modèle établi à partir d'une population de référence, en prenant en compte la base des données de génotypage et le TNR₂₈₂ moyen calculé sur l'ensemble des éjaculats produits au cours de la carrière de l'individu (l'indicateur de fertilité). Une fois le modèle de prédiction génomique établi, celui-ci a été appliqué aux taureaux de la cohorte SeQuaMol ne faisant pas partie de la population de référence. L'indicateur de fertilité utilisé est un indicateur au « niveau taureau », qui moyenne la variabilité des TNR qui peut exister entre les éjaculats, pour un même taureau. En revanche, pour chaque individu, les analyses de méthylation ont été réalisées à partir de l'ADNg extrait d'une paillette unique de semence, produite à partir d'un seul éjaculat, et associé à un TNR. Comme ce TNR n'est pas toujours représentatif de la fertilité générale de l'individu, les biomarqueurs que nous avons identifiés sont des biomarqueurs de niveau « éjaculat », et ne peuvent pas être utilisés pour évaluer la fertilité au niveau « taureau ». D'un point de vue pratique, cela signifie que l'outil de prédiction de la fertilité, à développer dans la suite du projet SeQuaMol, doit être pensé pour une utilisation à l'éjaculat.

Pour compléter notre analyse et étendre la possibilité d'utiliser ces biomarqueurs au niveau « taureau », il serait intéressant de savoir comment évoluent les biomarqueurs d'un éjaculat à l'autre, pour un même individu. Pour avoir des données à ce sujet, il est nécessaire de procéder à une étude longitudinale, réalisée sur plusieurs éjaculats récoltés tout au long de la carrière d'un reproducteur. Comme d'un éjaculat à l'autre le génome du reproducteur est fixe, ce type d'analyse permettrait en outre de ne pouvoir se focaliser que sur les variations épigénétiques indépendantes de la génétique du taureau.

Deux types de dispositifs expérimentaux peuvent être mis en place pour effectuer l'étude longitudinale. Le premier serait composé d'individus pour lesquels les conditions environnementales seraient standardisées. Le but de la standardisation est de limiter la variabilité épigénétique liés aux facteurs environnementaux, pouvant avoir une influence sur le méthylome. Pour ce faire, il est nécessaire que les individus proviennent du même centre d'élevage, qu'ils soient nourris de la même façon, soient en contact avec le même personnel (taureliers, vétérinaires...), etc. La standardisation des conduites d'élevage devrait être d'autant plus drastique pendant la période juvénile et prépubère des taureaux, période pendant laquelle on assiste au développement des organes reproducteurs et la mise en place de l'axe hypothalamo-hypophyso-testiculaire conduisant à la puberté. On pourrait même imaginer appliquer cette standardisation aux mères pendant la gestation des futurs reproducteurs, car, selon les concepts de la DOHaD (*Developmental Origins of Health and Diseases*), l'environnement maternel en période conceptionnelle peut avoir un impact sur l'expression et l'activité des enzymes de la machinerie épigénétique, avec des répercussions sur le développement embryonnaire (Panchenko et al. 2016), et potentiellement la gamétogenèse.

Un second dispositif expérimental serait d'effectuer les analyses sur des individus pour lesquels on tolérerait des variations plus importantes des conduites d'élevages (taureaux provenant de divers centres, par exemple). En revanche, pour pouvoir prendre en considération les effets de la variabilité épigénétique induite par les conditions d'élevages, ce protocole nécessiterait d'impliquer un nombre plus important d'individus. On serait à même de pouvoir évaluer l'influence qu'ont les paramètres environnementaux sur le méthylome.

Pour revenir au contexte de notre étude, nous avons vu que les analyses de méthylation ont été réalisées sur de l'ADNg extrait à partir d'une seule paillette de semence. Au sein d'un éjaculat, il existe une variété de sous-populations (motiles ou non, dégradés ou non, présentant des anomalies morphologiques...). Lors de la mise en paillette, les analyses sont effectuées sur un aliquote, qui n'est pas nécessairement représentatif de la population cellulaire (Valverde et al. 2016). Afin d'estimer la variabilité inter-paillettes, Utt a suggéré de pratiquer les mesures sur aux moins 2 paillettes pour chaque analyse phénotypique de la semence ; cette suggestion peut également s'appliquer aux analyses de méthylation (Utt 2016). Pour les analyses phénotypiques, nous avons réalisé un réplica technique et non biologique comme suggéré par Utt. Dans le cas de l'analyse longitudinale ou pour la phase de validation du projet SeQuaMol, la situation optimale serait d'avoir à disposition 2 (ou plus) paillettes de semence par éjaculat. Mélanger le contenu des paillettes, avant d'effectuer tous types d'analyses, permettrait ainsi de réduire la part de variabilité inter-paillette.

Quelles sont les causes et les conséquences des différentiels de méthylation ?

Les causes

Après avoir pris en considération ces aspects techniques, nous pouvons considérer que notre travail a contribué à établir un catalogue de biomarqueur basés sur des différentiels de méthylation. Il est important de pouvoir interpréter les différentiels de méthylation observés. Dans les spermatozoïdes matures, le noyau spermatique est compacté et la chromatine n'est plus accessible à la machinerie épigénétique. En conséquence, mise à part des processus chimique de dégradation de la méthylation, le méthylome spermatique n'est pas modifiable. Par contre, les fenêtres de reprogrammation épigénétique, au cours de la spermatogenèse ou au cours des premières phases de développement embryonnaire, sont potentiellement des fenêtres de sensibilité du méthylome. Quelle est donc l'origine des différentiels de méthylation entre les catégories de fertilité ? Plusieurs hypothèses à ce sujet peuvent être émises.

Les différentiels de méthylation sont-ils liés à la manipulation de la semence dans le cadre de l'IA ?

Les données concernant l'impact qu'aurait la cryoconservation sur le méthylome spermatique varient suivant l'espèce considérée. Certaines études montrent que les agents cryoprotectants induisent une hypométhylation globale du génome chez le tabaqui (de Mello et al. 2017), ou au contraire que le processus de cryoconservation induit une hyperméthylation globale chez le cheval (Aurich et al. 2016). Chez l'Homme, il a été montré que la cryoconservation n'induit pas de modification significative du niveau de méthylation de 5 gènes soumis à l'empreinte (*MEG3*, *H19*, *LIT1*, *SNRPN* et *MEST*), d'un gène impliqué dans la spermatogenèse (*VASA*), ou dans le cycle du folate (*MTHFR*), ainsi que les séquences répétées *LINE1* et *Alu* (Kläver et al. 2012). En revanche, dans cette dernière étude, les spermatozoïdes ont été triés par la méthode du « swim-up », qui permet d'éliminer une grande partie des spermatozoïdes immobiles. Il y a donc eu sélection d'une sous-population cellulaire, et vraisemblablement celle qui a été la moins impactée par la congélation, donc ces données sont difficilement comparables avec les nôtres.

Dans le contexte de notre étude, la semence de tous les taureaux provient de paillettes qui ont subi une cryoconservation. Si ce traitement est effectivement la cause du différentiel de

méthylation observé entre les catégories de fertilité, alors cela suggère que la semence de certains individus résiste plus ou moins efficacement aux dommages causés par le processus de cryoconservation. Cette caractéristique, décrite comme étant probablement d'origine génétique (Thurston, Watson, et Holt 2002), induirait aussi une sélection de la population de spermatozoïdes analysée. Dans ce cas, une cause de la différence de fertilité entre les groupes pourrait provenir d'une différence de sensibilité qui est génétiquement déterminée. On ne peut également pas écarter qu'une moins bonne résistance de la semence aux processus de congélation ait une conséquence sur la conservation des marques épigénétiques, et qui pourrait être vue comme de l'hypométhylation.

Certaines données générées par nos analyses laissent à penser que la résistance à la congélation n'est pas (ou en tout cas, principalement pas) la raison du différentiel de méthylation observé entre les catégories de fertilité. En effet :

- Les analyses fonctionnelles et morphologiques des spermatozoïdes (Résultats partie n°2) ne révèlent pas de différences significatives pour les paramètres qui sont susceptibles d'être impactés par la congélation. Par exemple, nous n'avons pas observé de différence significative dans le ratio spermatozoïdes avec membrane intacte / spermatozoïdes avec membrane dégradée pour les 2 races, ni de différentiel de motilité entre catégories pour la race Holstein. En revanche, nous avons bien observé un différentiel significatif pour la motilité entre les catégories Fertiles+ et Déviants pour la race Montbéliard.
- Par LUMA, nous avons évalué l'impact de la cryoconservation pour la semence de 6 taureaux de race Holstein et nous n'avons pas observé de différentiel significatif en méthylation comme en hydroxyméthylation globale entre semence fraîche et semence congelée d'un même taureau.

Les différentiels de méthylation sont-elles le reflet de perturbations antérieures ?

Il est possible que les perturbations de la méthylation observées dans les spermatozoïdes soient la conséquence d'altérations ayant eu lieu au cours de la différenciation des PGC, ou lors de la spermatogenèse. En effet, au moment de la migration des PGC et de la colonisation des crêtes génitales, a lieu la première phase de reprogrammation épigénétique. Elle consiste notamment en un effacement quasi-total des marques de méthylation de l'ADN. La phase de méthylation *de novo* se déroule plus tardivement lors du développement fœtal (à partir de E14,5 chez la souris). C'est également à ce moment que débute l'apposition de la méthylation sur les gènes soumis à l'empreinte, sur les séquences répétées et sur les séquences satellites

par l'action de DNMT3b en interaction avec DNMT3L (Kato et al. 2007). Des perturbations au moment de cette vague de méthylation *de novo* pourraient consister soit en une apposition défectueuse des marques de méthylation sur les séquences cibles soit en une apposition de méthylation sur d'autres régions génomiques. Ces perturbations pourraient être induites par des altérations de l'environnement maternel ayant des répercussions sur le développement du fœtus et son épigénome. Etant donné que nous avons observé une majorité de DMC hypométhylées chez les Déviants, l'hypothèse d'une perturbation de la méthylation *de novo* en période fœtale pour ces individus est crédible.

Il est également possible que ces altérations de la méthylation aient lieu à l'âge adulte, au cours de la spermatogenèse. En effet, chez l'Homme les enzymes DNMT3a, DNMT3b et DNMT1 sont exprimées tout au long de la spermatogenèse (Chen et Chan 2014). Le méthylome spermatique est alors sujet à des modifications subtiles, avec des remaniements du méthylome de l'ordre de 10% au total (Oakes et al. 2007). Il est alors possible que certaines conditions environnementales puissent, à un moment donné et pour certains individus, avoir une influence sur l'épigénome spermatique. Par exemple, les conditions d'élevage comme la température, la densité en animaux dans l'élevage, le bruit et bien d'autres facteurs environnementaux, peuvent induire du stress chez les animaux. Le niveau de stress est lui-même associé à des modifications des sécrétions hormonales, pouvant avoir une influence négative sur les performances en reproduction (Córdova-Izquierdo et al. 2014) (Bova et al. 2014). Il est possible que les taureaux soient plus ou moins sensibles aux facteurs de stress : ainsi, bien que soumis aux mêmes facteurs, les taureaux en subissent les conséquences différemment ; cela pourrait s'exprimer sous forme de pertes de performances reproductives liées à une modification de l'épigénome spermatique. L'hypothèse de perturbations ponctuelles des processus de méthylation au cours de la spermatogenèse est crédible pour les taureaux Fertiles-. En effet, ces taureaux ont été sélectionnés sur la base d'un TNR faible à l'éjaculat mais peuvent présenter un TNR moyen sur l'ensemble de la carrière correct. Certains cycles de la spermatogenèse pourraient donc avoir été perturbés par des conditions environnementales délétères ponctuelles et l'épigénome perturbé en serait le reflet.

Les conséquences

Après avoir étudié les causes potentielles des différentiels de méthylation que nous avons observé entre les catégories de fertilité, nous pouvons nous intéresser aux conséquences qu'ils peuvent avoir sur la fertilité. Pour qu'une altération de la méthylation puisse avoir un impact sur la fertilité, il faut que le spermatozoïde qui en est porteur soit le spermatozoïde fécondant. En post fécondation, le premier évènement majeur est la reprogrammation active du génome paternel, qui consiste essentiellement en un effacement quasi-général des marques de méthylation. Ainsi, pour que l'altération épigénétique puisse avoir un impact, il faut que celle-ci soit échappée à cette reprogrammation, soit perturbe la reprogrammation. Aujourd'hui, on sait que certains sites échappent effectivement à la reprogrammation épigénétique, des sites appelés « *escapees* ». Ces sites sont, dans une très large majorité des cas (plus de 95% chez la souris) associés à des éléments répétés (Hackett et al. 2013). Certains de ces *escapees* sont également associés à des gènes, comme *Vmn2r29*, *Sfi1* ou *Srrm2* chez la souris (Hackett et al. 2013), ou encore *TACC2* chez l'Homme (Tang et al. 2015). Or, il se trouve justement que nous avons obtenu une DMC dans un intron du gène *Sfi1*, en contexte de CpG *shore*, et que le différentiel est spécifique aux Déviants pour la race Holstein comme pour la race Montbéliard. Nous avons également obtenu plusieurs DMC dans un intron du gène *TACC2*, dont l'un est spécifique aux Déviants pour la race Montbéliarde et spécifique aux Fertiles pour la race Holstein. Ainsi, il s'avère que nous retrouvons des DMC sur des gènes décrits comme possédant des sites CpG échappant à la reprogrammation. Des analyses plus approfondies à ce sujet sont actuellement en cours sur notre jeu de données, dont les résultats seront particulièrement importants pour la poursuite du projet.

Pour que les altérations de la méthylation, transmises à l'embryon et non reprogrammées, aient un impact sur la fertilité, il faut que celles-ci aient une influence sur la régulation des gènes auxquels elles sont associées. Cela dépend, pour chacun des gènes impactés, du contexte génomique dans lequel la perturbation est retrouvée. L'association entre différentiel de méthylation, contexte génomique et expression du gène doit être considérée au cas par cas. De plus, que le différentiel de méthylation soit retrouvé sous forme de DMR, ou que celui-ci soit retrouvé sous forme de DMC, doit être interprété de manière différente.

La plupart des études cherchant à démontrer une corrélation entre expression génique et différentiel de méthylation se focalisent sur l'analyse de DMR. Un des domaines de recherche majeur à ce sujet est celui du cancer. De manière très générale, les cellules tumorales présentent une hypométhylation globale du génome, associée à des hyperméthylations

localisées, et en particulier sur les promoteurs de gènes suppresseurs de tumeur. C'est par exemple le cas du gène *RARβ* dans certains types de cancer du poumon (Li et al. 2017). L'hyperméthylation du promoteur d'un gène suppresseur de tumeur a pour conséquence son inactivation transcriptionnelle, ce qui contribue à perturber le processus biologique auquel il participe, comme la réparation de l'ADN, le contrôle du cycle cellulaire, l'apoptose, ou l'adhérence cellulaire. Ces processus biologiques sont également fondamentaux pour cours du développement embryonnaire. En conséquence, des DMR localisées en régions régulatrices de l'expression génique et transmises à l'embryon pourraient induire des perturbations des processus de prolifération ou d'adhérence cellulaire, conduisant à un arrêt développemental, et qui se traduirait par une sub-fertilité chez le mâle dont est issue la semence.

Dans le cadre des recherches sur le lien entre perturbations de la méthylation de l'ADN et sub-fertilité, plusieurs études se sont focalisées sur les gènes soumis à l'empreinte. Sur les 23 gènes soumis à l'empreinte connus chez le bovin, nous avons identifié pour les 2 races une DMR pour le gène *IGF2*, située en amont du gène et hypométhylée chez les Déviants. Concernant *IGF2*, une étude a montré que dans le sang de cordon humain, un changement de méthylation de seulement 1% sur la DMR associée au gène a pour conséquence une diminution d'un facteur 2 (dans le cas d'une hyperméthylation) ou un doublement (dans le cas d'une hypométhylation) de l'expression du gène (Murphy et al. 2012). Pour notre analyse, il est donc également possible qu'une perturbation de la méthylation du gène soumis à l'empreinte *IGF2* puisse être impliquée dans la sub-fertilité des individus.

Au total, grâce à nos analyses, nous avons identifié 267 DMR pour la race Holstein et 80 pour la race Montbéliarde. Bien que celles-ci aient été retrouvées en majorité en régions introniques, il est possible qu'elles aient une influence sur l'expression des gènes. En plus de ces DMR, nous avons identifié plusieurs milliers de DMC. L'influence que ces DMC isolées peuvent avoir sur l'expression des gènes est plus difficile à statuer que pour les DMR. En effet, les équipes de recherche ne se focalisent la plupart du temps que sur l'étude des DMR, car il est toujours possible que des DMC isolées soient en réalité des faux-positifs. Il est également difficile de statuer sur l'influence que peut avoir une seule DMC sur l'expression. Katari a montré que des différentiels de méthylation sur seulement 2 CpG associés à un gène peut avoir des répercussions (bien que modestes) sur l'expression génique (Katari et al. 2009). Si 2 DMC proches peuvent impacter l'expression d'un gène, qu'en est-il d'une DMC ?

De plus, il faut garder à l'esprit qu'il est possible que la méthode d'identification des DMC puisse donner l'impression qu'une DMC est isolée, alors qu'elle participe en réalité à une DMR. Dans le cas de notre analyse RRBS, il est en effet possible que :

- Les sites CpG situés autour de la DMC n'aient pas passé les seuils de significativité, mais présentent bien une tendance à un différentiel de méthylation. Par exemple, si entre 2 catégories de fertilité on observe pour un CpG un différentiel de méthylation de 24%, cette position ne sera pas considérée comme DMC, alors que la position voisine présentant un différentiel de 25% sera considérée comme telle
- Les sites CpG ne présentent pas, après séquençage, une couverture suffisante pour être intégrés aux analyses. Le CpG n'a par exemple été couvert que par 9 séquences, il est alors éliminé, même s'il présentait théoriquement un différentiel de méthylation entre 2 groupes bien plus important que 25%
- Les sites CpG n'ont pas été séquencés car ils sont situés sur un autre fragment d'ADN que celui de la DMC, et ce fragment a été éliminé lors de l'étape de sélection de taille

Ainsi, dans l'idéal toutes les DMC devraient être considérées au cas par cas. Si l'on considère que celles-ci peuvent effectivement avoir une influence, même modeste, sur l'expression, alors par effet de nombre il est envisageable qu'une multitude de DMC impactant un grand nombre de gènes puisse perturber des voies biologiques à grande échelle. C'est une des raisons qui nous a amené à nous intéresser à la voie de signalisation Hippo, dont une proportion importante de gènes présente des DMC. Cette voie est impliquée très précocement au cours du développement, dans la formation des premiers lignages.

Jusqu'à présent, nous n'avons aucune information sur l'influence que peuvent avoir les différentiels de méthylation sur l'expression dans l'embryon. Pour la suite du projet, une analyse transcriptomique sur des blastocystes (en séparant le trophoctoderme de l'ICM) s'avère essentielle. Les embryons devront avoir été produits à partir des mêmes éjaculats que ceux que nous avons analysés en RRBS. L'analyse des marques de méthylation et d'hydroxyméthylation, par marquage fluorescent suivie d'une observation en microscopie confocale, permettra d'observer le déroulement de la reprogrammation. Il sera également nécessaire de procéder à des analyses gènes-candidats par bisulfite-pyroséquençage pour savoir si les différentiels de méthylation spermatiques sont bien transmis à l'embryon.

Un point important à prendre en considération est qu'il faudra être en mesure de distinguer les différentiels de méthylation transmis par le spermatozoïde à l'embryon des différentiels induits par la FIV elle-même. En effet, des données montrent que la culture *in vitro* d'embryons, et notamment dans le cadre de la FIV, peuvent introduire des perturbations de la méthylation (Sutcliffe et al. 2006) (Sanchez-Albisua et al. 2007).

Prise en compte de l'âge à la production de semence pour l'identification de biomarqueurs pertinents

Les paillettes utilisées pour les analyses proviennent des ES adhérents d'Allice qui souhaitent prendre part au projet SeQuaMol. Les réflexions menées en amont du projet et concernant la mise en place du dispositif expérimental (répartitions des individus en fonction de catégories de fertilité, analyses au niveau éjaculat, utilisation de plusieurs races, etc.) nous ont conduit à établir une liste de taureaux répondant à des critères d'âge, de TNR et de disponibilité des paillettes (10 paillettes étaient requises pour établir en parallèle le catalogue de DMC et un catalogue d'expression de miRNA (étude menée par Eli Sellem), déterminer le ratio protamines/histones et identifier certaines MPT (étude menée par C. Le Danvic). Les ES n'ont pu fournir que les paillettes des individus qu'ils avaient à disposition, et en nombre suffisant. Ainsi, bien que les individus fussent être homogènes en termes d'âge à la production de semence, il s'est avéré que l'âge des taureaux variait de façon très significative entre les catégories de fertilité pour la race Holstein. En effet, les taureaux de la catégorie des Déviants ont été employés en production bien plus jeunes en moyenne que ceux des autres catégories. Il est ainsi apparu que ce paramètre a une influence non-négligeable sur nos résultats, et qu'il est essentiel de la prendre en considération.

Dans les autres études sur la méthylation pangénomique de l'ADN spermatique chez le bovin, les individus analysés ont des âges à la production compris entre 3 et 5 ans (Verma et al. 2014) ou 3 et 4 ans (Kropp et al. 2017) ; pour ces deux études ils sont à l'âge adulte. Les individus sont homogènes en termes d'âge et leur catalogue de biomarqueurs ne peut donc documenter les effets de ce facteur. Or, dans le contexte de la sélection génomique, les individus sont utilisés à un âge de plus en plus précoce, parfois avant un an. Comme nous avons observé un fort différentiel de méthylation entre les individus très jeunes et les adultes, il est important d'utiliser ces biomarqueurs dans le contexte d'une utilisation de taureaux génomiques.

Ainsi, pour rester dans le contexte de la sélection génomique, il apparaît important de travailler sur la semence d'individus jeunes plutôt qu'à un âge adulte avancé. C'est pourquoi, l'analyse va être complétée pour la production de nouvelles banques RRBS à partir de la semence de 10 taureaux Fertiles+ de même âge que les Déviants. Egalement, la mise en œuvre d'une analyse longitudinale, telle que décrite dans la partie précédente, pourrait permettre d'apporter un éclairage sur l'évolution du méthylome spermatique avec l'âge.

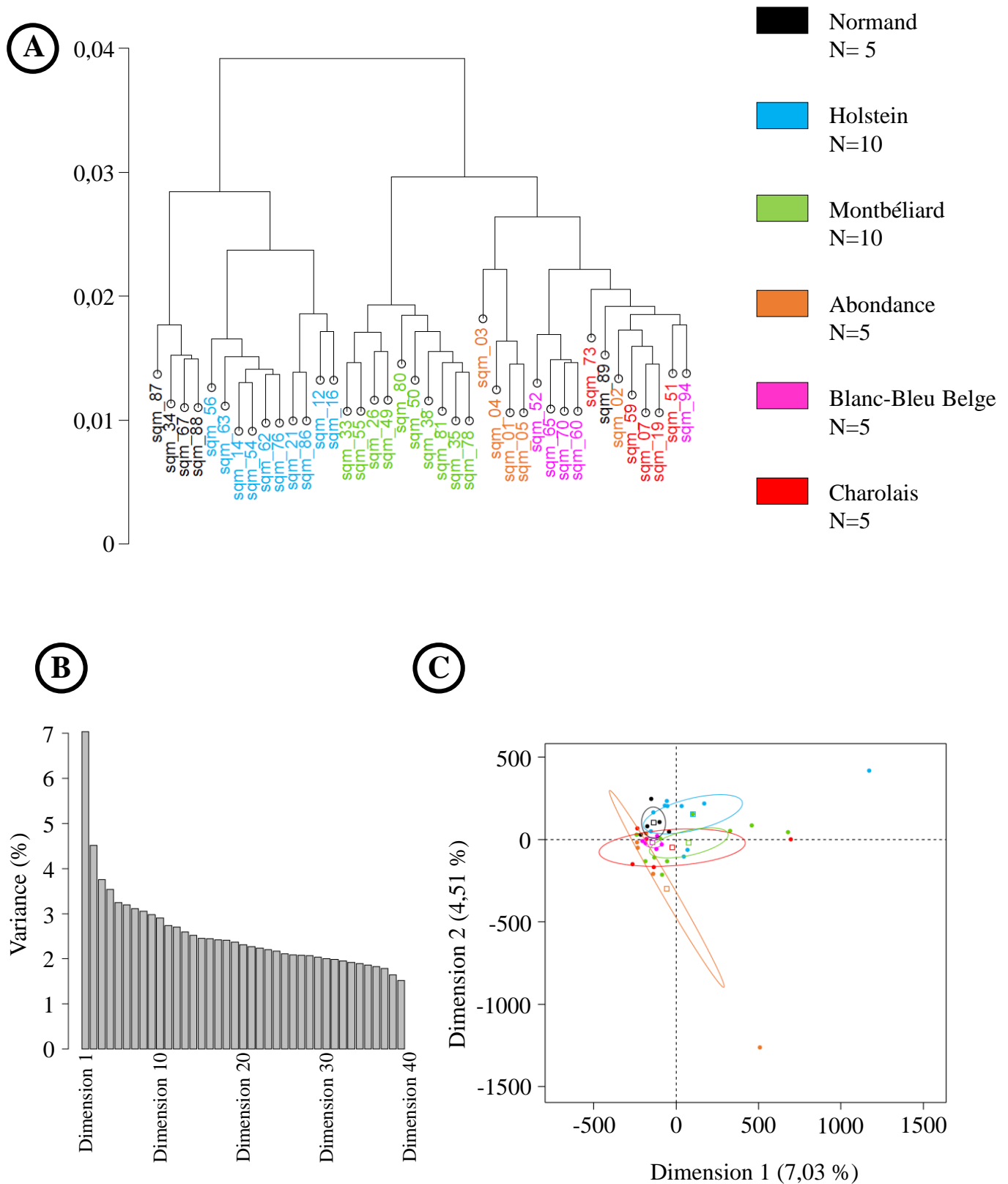


Figure 3. Analyse descriptive prenant en compte toutes les races du dispositif

Les CpG utilisés pour cette analyse descriptive sont couverts par un minimum de 10 séquences pour au moins 3 individus par race. Ainsi, l'analyse porte sur un total de 1 310 964 CpG. Pour les races Holstein et Montbéliarde, seuls les individus des groupes Fertiles+ ont été utilisés. Pour la race Blanc-Bleu-Belge, ce sont les 5 individus avec le meilleur TNR56. **A.** Classification hiérarchique de corrélation de Pearson. On observe une nette séparation des individus en fonction de leur race. Les races Holstein et Normandes forment un premier embranchement. Le second embranchement contient les 4 autres races mais il semble que les Blanc-Bleu-Belge et les Charolais soient les mieux regroupés. A noter aussi que plusieurs individus des autres races se regroupent avec les Charolais. **B.** Digramme en éboulis pour l'Analyse en Composantes Principales (ACP). Cette représentation permet de montrer la contribution de chaque axe à la variabilité totale. **C.** L'ACP permet de visualiser la séparation entre les 2 races selon la dimension 1. On observe que la séparation est très mauvaise entre toutes les races, avec seulement 7,03% de la variance expliquée sur la dimension 1 et 4,51% sur la dimension 2. Il faut aussi noter la présence d'individus extrêmes pour les races Holstein, Abondance, Normande et Montbéliarde.

Intégration des données épigénétiques et génétiques pour la compréhension du phénotype spermatique.

L'intégration des données de génotype aux données de méthylation a également montré que la génétique des taureaux avait une influence non négligeable sur la méthylation de certains CpG. L'effet du génotype n'est étonnamment pas si influent par rapport aux effets environnementaux. Par exemple, en effectuant une analyse différentielle comparant les 10 individus Fertiles+ Holstein et Montbéliard, nous avons pu mettre en évidence 10 898 DMC. En parallèle, nous avons vu que 7032 DMC avaient été identifiées en catégorisant les Fertiles+ et Fertiles- Holstein en fonction de leur âge, soit un nombre de DMC inférieur mais restant du même ordre de grandeur.

Une autre analyse permet d'apporter des informations à ce sujet. Le dispositif expérimental de SeQuaMol comprend 94 taureaux, de 6 races différentes. Dans ce document n'a pu être décrit que les analyses de fertilité pour les races Holstein et Montbéliard, mais quelques analyses complémentaires ont pu être effectuées pour l'ensemble du dispositif incluant les 6 races. Une analyse descriptive (classification hiérarchique et ACP) est présentée Figure 3. Celle-ci a été effectuée en utilisant les taureaux Fertiles+ Holstein et Montbéliard, et les taureaux des autres races (pour la race Blanc-Bleu Belge, seuls les 5 individus les plus fertiles ont été utilisés). La classification hiérarchique met en évidence une nette séparation des individus en fonction de la race. Cependant, certaines races semblent se regrouper entre elles, notamment les Holstein et les Normandes qui clustérisent ensemble avant de se séparer suivant leur race respective. Alors que les races Holstein et Normande ne présentent pas de parenté, et que contrairement à la race Holstein, la race Normande est totalement fermée aux apports génétiques des populations extérieures (Mattalia et al. 2006), il s'avère que les taureaux Holstein et Normand proviennent du même centre de production de semence, et ont donc été élevés de façon identique. Ces données suggèrent qu'il est possible de séparer les individus d'abord en fonction de leurs conditions d'élevage puis en fonction de leur race.

Les 2 points décrits précédemment, âge et génotype, montrent à quel point il est important de prendre en considération tous les paramètres ayant trait au taureau et soulignent que les effets environnements sont particulièrement importants dans l'établissement de l'épigénome, même s'il ne peut être écarté un effet génétique.

La plus-value du projet SeQuaMol est d'incorporer des résultats provenant de l'analyse des autres marques épigénétiques, que sont les miRNA, les MPT (en particulier l'acétylation et la

méthylation des histones) et le ratio histones-protamines sur la même cohorte de taureaux. L'ensemble des données générées pour les différentes marques épigénétiques vont être intégrées et permettront de déterminer les interactions possibles entre elles pour le contrôle de la fertilité. Cette démarche nécessite la mise en place de modèles statistiques élaborés qui feront l'objet d'une nouvelle thèse (thèse SeQuaStats).

Considérations pour la mise au point d'un outil d'épigénotypage

Une fois que les biomarqueurs de méthylation auront été validés au cours de la seconde phase du projet SeQuaMol, la troisième et dernière phase du projet consistera à mettre au point une équation de prédiction de la fertilité. Cette équation a pour vocation l'intégration des données elles-mêmes issues de l'intégration des données de méthylation de l'ADN, de microRNA et de MPT d'histones. Il est aussi préconisé le développement technologique d'outils moléculaires permettant d'analyser en routine les trois sets de marques épigénétiques. Les objectifs à long terme sont donc de fournir des outils moléculaires produisant les données et une équation de prédiction permettant l'indexation des taureaux pour leur fertilité. Une voie d'exploitation serait une miniaturisation du nombre de biomarqueurs pertinents afin de permettre une exploitation en routine par les entreprises de sélection soit au niveau du taureau (avant la mise à la production officielle) ou au niveau de l'éjaculat pour améliorer la caractérisation du potentiel fécondant pendant les contrôles qualités post-collecte.

Il est possible qu'une partie des biomarqueurs identifiés soit éliminée lors de la phase de validation, mais nous pensons que plusieurs centaines, voire plusieurs milliers soient toutefois validées. Il sera alors nécessaire de procéder à une sélection des biomarqueurs à incorporer sur l'outil. Les critères à considérer pour la sélection seront :

- La variabilité du taux de méthylation inter-individuelle. Nous avons vu qu'elle pouvait être de plus de 30% sur certaines positions. Le choix doit se porter sur celles dont la variabilité est minimale (5 à 10% maximum)
- Dans le cas des DMR, il peut être suffisant de n'interroger qu'une seule DMC. En effet, les DMC d'une DMR suivent généralement les mêmes tendances de méthylation. On peut cependant sélectionner la meilleure DMC, en suivant les autres critères décrits ici

- Les conditions techniques de validation. Tous autres critères égaux, il est préférable de favoriser les biomarqueurs identifiés selon les conditions techniques les plus strictes (par exemple, celui avec la meilleure couverture de séquençage en RRBS)
- Le seuil de différentiel. Le différentiel de méthylation minimal utilisé dans nos analyses, pour considérer un CpG comme DMC, est de 25%. Ce différentiel est dans certains cas bien plus important. Il est préférable de choisir les DMC présentant le plus fort différentiel de méthylation entre catégories de fertilité afin d'assurer la robustesse du système

Ce dernier point est fondamental. Quand la méthylation de l'ADN est calculée après RRBS, le résultat est une valeur de méthylation continue (de 0 à 100% de méthylation). Ce résultat n'a pas de sens biologique au niveau cellulaire, étant donné que la méthylation est une donnée qualitative : le groupement méthyle est présent ou absent d'un CpG. Dans une cellule somatique, pour un CpG donné, l'état de méthylation ne peut donc qu'être 100% (lorsque les 2 allèles sont méthylés), 50% (lorsque seul un des 2 allèles est méthylé) ou 0% (lorsque aucun allèle n'est méthylé). Analyser le méthylome spermatique est un cas particulier puisque le spermatozoïde est une cellule haploïde. Dans cette configuration, l'état de méthylation ne peut être que de 0 ou 100% à chaque CpG. En analysant une paillette de spermatozoïdes, la valeur de méthylation continue calculée est en réalité le reflet de l'état de méthylation de niveau de la population cellulaire : un pourcentage de méthylation calculé à 75% signifie que 75% des spermatozoïdes analysés sont méthylés sur la position et que 25% ne le sont pas. Les biomarqueurs que nous avons identifiés, sont donc aussi révélateurs de la proportion de spermatozoïdes porteurs ou non d'un groupement méthyle sur les CpG considérés.

C'est probablement une raison pour laquelle une forte variabilité inter-individuelle de méthylation peut être observée pour chaque DMC. Cela signifierait que dans chaque éjaculat, est présent une quantité variable de spermatozoïdes porteurs ou non d'une aberration de méthylation en un *locus* donné. Comme au cours d'une IA, il est impossible de savoir quel spermatozoïde est fécondant, il est impossible de savoir si c'est un spermatozoïde porteur d'une altération de la méthylation ou non. En revanche, on peut imaginer que plus on observe la présence d'un biomarqueur dans un éjaculat, plus la probabilité que celui-ci soit transmis à l'embryon est élevée. Dans le contexte d'un contrôle épigénétique de la qualité de la semence, il serait possible de fixer des seuils de tolérance, à la manière de ce qui est couramment utilisé concernant les anomalies morphologiques de la semence. Par exemple, un éjaculat serait considéré comme mauvais, et éliminé, si on observe la présence d'un biomarqueur dans plus de 20% des spermatozoïdes (i.e., on aurait en réalité observé un taux de méthylation de 20%).

En conclusion, ce travail de thèse a permis dans un premier temps de caractériser de manière approfondie le méthylome du spermatozoïde bovin. A notamment été pointée son hypométhylation très marquée, qui n'est pas liée à la cryocongélation mais pourrait jouer un rôle dans la structuration du génome. Dans un second temps, ce travail a mis en évidence plusieurs milliers de biomarqueurs de méthylation liés à la fertilité et à l'âge à la production de semence des taureaux. Il est cependant nécessaire d'approfondir les analyses sur le lien qui existe entre épigénome et génotype, et notamment en exploitant les différentiels entre races. Pour la suite du projet SeQuaMol, les biomarqueurs nécessiteront d'être validés sur une cohorte plus importante, où sera prise en compte une plus grande variété de paramètres environnementaux pouvant affecter le méthylome de l'ADN spermatique.

Bibliographie

- Adelson, DL, Raison, JM, et Edgar, RC. 2009. « Characterization and distribution of retrotransposons and simple sequence repeats in the bovine genome ». *Proceedings of the National Academy of Sciences* 106 (31):12855–12860.
- Adham, IM. 2000. « The Role of the Testicular Factor INSL3 in Establishing the Gonadal Position ». *Mol Cell Endocrinol* 160 (1-2):11-16.
- Akalin, A, Kormaksson, M, Li, S, Francine, E, Garrett-Bakelman, FE, Figueroa, ME, Melnick, A, et Mason, CE. 2012. « MethylKit: A Comprehensive R Package for the Analysis of Genome-Wide DNA Methylation Profiles ». *Genome Biology* 13 (10):R87. <https://doi.org/10.1186/gb-2012-13-10-r87>.
- Alexander, JH. 2008. « Bull Breeding Soundness Evaluation: A Practitioner's Perspective ». *Theriogenology* 70 (3):469-72. <https://doi.org/10.1016/j.theriogenology.2008.05.030>.
- Amann, RP, et DeJarnette, JM. 2012. « Impact of Genomic Selection of AI Dairy Sires on Their Likely Utilization and Methods to Estimate Fertility: A Paradigm Shift ». *Theriogenology* 77 (5):795-817. <https://doi.org/10.1016/j.theriogenology.2011.09.002>.
- Amann, RP, et Katz, DF. 2004. « Andrology lab corner: Reflections on CASA after 25 years ». *Journal of andrology* 25 (3):317–325.
- Amann, RP, et Waberski, D. 2014. « Computer-Assisted Sperm Analysis (CASA): Capabilities and Potential Developments ». *Theriogenology* 81 (1):5-17.e3. <https://doi.org/10.1016/j.theriogenology.2013.09.004>.
- Andersson, M, Aalto, J, et Gustavsson, I. 1992. « Embryo quality and andrological study of two subfertile bulls versus five control bulls with normal fertility ». *Theriogenology* 38 (4):623–631.
- Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Anway, MD, Cupp, AS, Uzumcu, M, et Skinner, MK. 2005. « Epigenetic transgenerational actions of endocrine disruptors and male fertility ». *Science* 308 (5727):1466–1469.
- Aran, D, Toperoff, G, Rosenberg, M, et Hellman, A. 2011. « Replication Timing-Related and Gene Body-Specific Methylation of Active Human Genes ». *Human Molecular Genetics* 20 (4):670-80. <https://doi.org/10.1093/hmg/ddq513>.
- Aravin, AA, Hannon, GJ, et Brennecke, J. 2007. « The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race ». *Science* 318 (5851):761–764.
- Aravin, AA, Sachidanandam, R, Bourc'his, D, Schaefer, C, Pezic, D, Fejes Toth, K, Bestor, T, et Hannon, GJ. 2008. « A PiRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice ». *Molecular Cell* 31 (6):785-99. <https://doi.org/10.1016/j.molcel.2008.09.003>.
- Aravindan, GR, Bjordahl, J, Jost, LK, et Evenson, DP. 1997. « Susceptibility of Human Sperm to in Situ DNA Denaturation Is Strongly Correlated with DNA Strand Breaks Identified by Single-Cell Electrophoresis ». *Experimental Cell Research* 236 (1):231-37. <https://doi.org/10.1006/excr.1997.3719>.
- Auclair, G, Guibert, S, Bender, A, et Weber, M. 2014. « Ontogeny of CpG Island Methylation and Specificity of DNMT3 Methyltransferases during Embryonic Development in the Mouse ». *Genome Biology* 15 (12). <https://doi.org/10.1186/s13059-014-0545-5>.
- Baccarelli, A, Wright, RO, Bollati, V, Tarantini, L, Litonjua, AA, Suh, HH, Zanobetti, A, Sparrow, D, Vokonas, PS, et Schwartz, J. 2009. « Rapid DNA Methylation Changes after Exposure to Traffic Particles ». *American Journal of Respiratory and Critical Care Medicine* 179 (7):572-78. <https://doi.org/10.1164/rccm.200807-1097OC>.
- Bailey, J, Morrier, A, et Cormier, N. 2003. « Semen cryopreservation: successes and persistent problems in farm species ». *Canadian journal of animal science* 83 (3):393–401.
- Ballachey, BE, Miller, HL, Jost, LK, et Evenson, DP. 1986. « Flow Cytometry Evaluation of Testicular and Sperm Cells Obtained from Bulls Implanted with Zeranol ». *Journal of Animal Science* 63 (3):995-1004.
- Barbat, A, Le Mézec, P, Ducrocq, V, Mattalia, S, Fritz, S, Boichard, D, Ponsart, C, et Humblot, P. 2010. « Female Fertility in French Dairy Breeds: Current Situation and Strategies for Improvement ». *The Journal of Reproduction and Development* 56 Suppl (janvier):S15-21.
- Barth, AD. 2007. « Evaluation of potential breeding soundness of the bull ». In: Youngquist RS, Threlfall WR (eds), *Current Therapy in Large Animal Theriogenology* 2. Saunders Elsevier, Philadelphia, pp. 228–240
- Barth, AD, et Oko, RJ. 1989. « Abnormal morphology of bovine spermatozoa ». Iowa State University Press
- Baubec, T, et Schübeler, D. 2014. « Genomic patterns and context specific interpretation of DNA methylation ». *Current Opinion in Genetics & Development* 2014, 25:85–92.
- Baubec, T, Colombo, DF, Wirbelauer, C, Schmidt, J, Burger, L, Krebs, AR, Akalin, A, et Schübeler, D. 2015. « Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation ». *Nature* 520 (7546):243-47. <https://doi.org/10.1038/nature14176>.
- Baurau, J, Teissandier, A, Zamudio, N, Roy, S, Nalesso, V, Héroult, Y, Guillou, F, et Bourc'his, D. 2016. « The DNA methyltransferase DNMT3C protects male germ cells from transposon activity ». *Science* 354 (6314), 909-912.
- Begum, G, Stevens, A, Smith, EB, Connor, K, Challis, JRG, Bloomfield, F, et White, A. 2012. « Epigenetic changes in fetal hypothalamic energy regulating pathways are associated with maternal undernutrition and twinning ». *The FASEB Journal*, 26(4), 1694–1703. <http://doi.org/10.1096/fj.11-198762>.
- Benchaib, M. 2005. « Influence of Global Sperm DNA Methylation on IVF Results ». *Human Reproduction* 20 (3):768-73. <https://doi.org/10.1093/humrep/deh684>.
- Beraldi, R, Pittoggi, C, Sciamanna, I, Mattei, E, et Spadafora, C. 2006. « Expression of LINE-1 Retroposons Is Essential for Murine Preimplantation Development ». *Molecular Reproduction and Development* 73 (3):279-87. <https://doi.org/10.1002/mrd.20423>.
- Berg, DK, Smith, CS, Pearton, DJ, Wells, DN, Broadhurst, R, Donnison, M, et Pfeffer, PL. 2011. « Trophectoderm Lineage Determination in Cattle ». *Developmental Cell* 20 (2):244-55. <https://doi.org/10.1016/j.devcel.2011.01.003>.
- Berndson, WE, et Desjardins, C. 1974. « The Cycle of the Seminiferous Epithelium and Spermatogenesis in the Bovine Testis ». *The American Journal of Anatomy* 140 (2):167-79. <https://doi.org/10.1002/aja.1001400204>.

- Berry, DP, Buckley, F, Dillon, P, Evans, RD, Rath, M, et Veerkamp, RF. 2003. « Genetic relationships among body condition score, body weight, milk yield, and fertility in dairy cows ». *Journal of Dairy Science* 86 (6):2193–2204.
- Berry, DP, Wall, E, et Pryce, JE. 2014. « Genetics and Genomics of Reproductive Performance in Dairy and Beef Cattle ». *Animal* 8 (s1):105-21. <https://doi.org/10.1017/S1751731114000743>.
- Bezerra, LR, Sarmiento, JL, Neto, SG, de Paula, NR, Oliveira, RL, et do Rêgo, WM. 2013. « Residual Feed Intake: A Nutritional Tool for Genetic Improvement ». *Tropical Animal Health and Production* 45 (8):1649-61. <https://doi.org/10.1007/s11250-013-0435-y>.
- Bird, A. 2002. « DNA methylation patterns and epigenetic memory ». *Genes & development* 16 (1):6–21.
- Bird, A. 2007. « Perceptions of epigenetics ». *Nature* 447 (7143):396-98. <https://doi.org/10.1038/nature05913>.
- Bird, A, Taggart, M, Frommer, M, Miller, OJ, et Macleod, D. 1985. « A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA ». *Cell* 40 (1):91–99.
- Birney, E, Andrews, TD, Bevan, P, Caccamo, M, Chen, Y, Clarke, L, Coates, G, Cuff, J, Curwen, V, Cutts, T, Down, T, Eyra, E, Fernandez-Suarez, XM, Gane, P, Gibbins, B, Gilbert, J, Hammond, M, Hotz, RH, Iyer, V, Jekosch, K, Kahari, A, Kasprzyk, A, Keefe, D, Keenan, S, Lehvaslaiho, S, McVicker, G, Melsopp, C, Meidl, P, Mongin, E, Pettett, R, Potter, S, Proctor, G, Rae, M, Searle, S, Slater, G, Smedley, D, Smith, J, Spooner, W, Stabenau, A, Stalker, J, Storey, R, Ureta-Vidal, A, Woodwark, KC, Cameron, G, Durbin, R, Cox, A, Hubbard, T, et Clamp, M. 2004. « An Overview of Ensembl ». *Cold Spring Harbor Laboratory Press* 14:925–928.
- Bock, C, Halbritter, F, Carmona, FJ, Tierling, S, Datlinger, P, Assenov, Y, Berdasco, M, et al. 2016. « Quantitative comparison of DNA methylation assays for biomarker development and clinical applications ». *Nature Biotechnology* 34 (7):726-37. <https://doi.org/10.1038/nbt.3605>.
- Bock, C, Tomazou, EM, Brinkman, AB, Müller, F, Simmer, F, Gu, H, Jäger, N, Gnirke, A, Stunnenberg, HG, et Meissner, A. 2010. « Quantitative comparison of genome-wide DNA methylation mapping technologies ». *Nature Biotechnology* 28 (10):1106-14. <https://doi.org/10.1038/nbt.1681>.
- Bogdanović, O., Lister R. 2017. DNA methylation and the preservation of cell identity. *Current Opinion in Genetics & Development* 2017, 46:9–14
- Boissonnas, CC, El Abdalaoui, H, Haelewyn, V, Fauque, P, Dupont, JM, Gut, I, Vaiman, D, Jouannet, P, Tost, J, et Jammes, H. 2010. « Specific Epigenetic Alterations of IGF2-H19 Locus in Spermatozoa from Infertile Men ». *European Journal of Human Genetics: EJHG* 18 (1):73-80. <https://doi.org/10.1038/ejhg.2009.117>.
- Boissonnas, CC, Jouannet, P, et Jammes, H. 2013. « Epigenetic disorders and male subfertility ». *Fertil Steril.* 2013 Mar 1;99(3):624-31. doi: 10.1016/j.fertnstert.2013.01.124.
- Bollati, V, Baccarelli, A, Hou, L, Bonzini, M, Fustinoni, S, Cavallo, D, Byun, HM, Jiang, J, Marinelli, B, et Pesatori, AC. 2007. « Changes in DNA methylation patterns in subjects exposed to low-dose benzene ». *Cancer research* 67 (3):876–880.
- Booth, MJ, Branco, MR, Ficz, G, Oxley, D, Krueger, F, Reik, W, et Balasubramanian, S. 2012. « Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution ». *Science (New York, N.Y.)* 336 (6083):934-37. <https://doi.org/10.1126/science.1220671>.
- Bourc'his, D. 2001. « Dnmt3L and the Establishment of Maternal Genomic Imprints ». *Science* 294 (5551):2536-39. <https://doi.org/10.1126/science.1065848>.
- Bourc'his, D, et Bestor, TH. 2004. « Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L ». *Nature* 431 (7004):92–96.
- Bostick, M, Kim, JK, Estève, P, Clark, A, Pradhan, S, et Jacobsen, SE. 2007. « UHRF1 plays a role in maintaining DNA methylation in mammalian cells ». *Science* 317, 1760-1764.
- Bova, TL, Chiavaccini, L, Cline, GF, Hart, CG, Matheny, K, Muth, AM, Voelz, BE, Kesler, D, et Memili, E. 2014. « Environmental stressors influencing hormones and systems physiology in cattle ». *Reproductive Biology and Endocrinology* 2014, 12:58.
- Bowdin, S, Allen, C, Kirby, G, Brueton, L, Afnan, M, Barratt, C, Kirkman-Brown, J, Harrison, R, Maher, ER, et Reardon, W. 2007. « A Survey of Assisted Reproductive Technology Births and Imprinting Disorders ». *Human Reproduction* 22 (12):3237-40. <https://doi.org/10.1093/humrep/dem268>.
- Brandriff, B., et R. A. Pedersen. 1981. « Repair of the Ultraviolet-Irradiated Male Genome in Fertilized Mouse Eggs ». *Science (New York, N.Y.)* 211 (4489):1431-33.
- Braunschweig, M, Jagannathan, V, Gutzwiller, A, et Bee, G .2012. « Investigations on Transgenerational Epigenetic Response Down the Male Line in F2 Pigs ». *PLoS ONE* 7(2): e30583. doi:10.1371/journal.pone.0030583.
- Brenet, F, Moh, M, Funk, P, Feierstein, E, Viale, AJ, Succi, ND, et Scandura, JM. 2011. « DNA Methylation of the First Exon Is Tightly Linked to Transcriptional Silencing ». Édité par Nina Papavasiliou. *PLoS ONE* 6 (1):e14524. <https://doi.org/10.1371/journal.pone.0014524>.
- Brito, LFC, Barth, AD, Rawlings, NC, Wilde, RE, Crews, DH, Boisclair, YR, Ehrhardt, RA, et Kastelic, JP. 2007. « Effect of Feed Restriction during Calfhood on Serum Concentrations of Metabolic Hormones, Gonadotropins, Testosterone, and on Sexual Development in Bulls ». *Reproduction* 134 (1):171-81. <https://doi.org/10.1530/REP-06-0353>.
- Brito, LFC, Barth, AD, Rawlings, NC, Wilde, RE, Crews, DH, Mir, PS, et Kastelic, JP. 2007a. « Effect of Nutrition during Calfhood and Peripubertal Period on Serum Metabolic Hormones, Gonadotropins and Testosterone Concentrations, and on Sexual Development in Bulls ». *Domestic Animal Endocrinology* 33 (1):1-18. <https://doi.org/10.1016/j.domaniend.2006.04.001>.
- Brito, LFC. 2007b. « Effect of Improved Nutrition during Calfhood on Serum Metabolic Hormones, Gonadotropins, and Testosterone Concentrations, and on Testicular Development in Bulls ». *Domestic Animal Endocrinology* 33 (4):460-69. <https://doi.org/10.1016/j.domaniend.2006.09.004>.
- Brockdorff, N, et Turner, BM. 2015. « Dosage Compensation in Mammals ». *Cold Spring Harb Perspect Biol* 2015;7:a019406.

- Bromfield, JJ, Schjenken, JE, Chin, PY, Care, AS, Jasper, MJ, et Robertson, SA. 2014. « Maternal Tract Factors Contribute to Paternal Seminal Fluid Impact on Metabolic Phenotype in Offspring ». *Proceedings of the National Academy of Sciences* 111 (6):2200-2205. <https://doi.org/10.1073/pnas.1305609111>.
- Brøndum, RF, Su, G, Janss, L, Sahana, G, Guldbandsen, B, Boichard, D, et Lund, MS. 2015. « Quantitative Trait Loci Markers Derived from Whole Genome Sequence Data Increases the Reliability of Genomic Prediction ». *Journal of Dairy Science* 98 (6):4107-16. <https://doi.org/10.3168/jds.2014-9005>.
- Brykczynska, U, Hisano, M, Erkek, S, Ramos, L, Oakeley, EJ, Roloff, TC, Beisel, C, Schübeler, D, Stadler, MB, et Peters, AH. 2010. « Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa ». *Nature Structural & Molecular Biology* 17 (6):679-87. <https://doi.org/10.1038/nsmb.1821>.
- Buck-Koehntop, BA, et Defossez, PA. 2013 « On how mammalian transcription factors recognize methylated DNA ». *Epigenetics*, 8:2, 131-137, DOI:10.4161/epi.23632.
- Bui, LC, Evsikov, AV, Khan, DR, Archilla, C, Peynot, N, Henaut, A, Le Bourhis, D, Vignon, X, Renard, JP, et Duranthon, V. 2009. « Retrotransposon Expression as a Defining Event of Genome Reprogramming in Fertilized and Cloned Bovine Embryos ». *Reproduction* 138 (2):289-99. <https://doi.org/10.1530/REP-09-0042>.
- Burkman, LJ, Coddington, CC, Franken, DR, Krugen, TF, Rosenwaks, F, et Hogen, GD. 1988. « The Hemizona Assay (HZA): Development of a Diagnostic Test for the Binding of Human Spermatozoa to the Human Hemizona Pellucida to Predict Fertilization Potential ». *Fertility and Sterility* 49 (4):688-97.
- Burns, BM, Gazzola, C, Holroyd, RG, Crisp, J, et McGowan, MR. 2011. « Male Reproductive Traits and Their Relationship to Reproductive Traits in Their Female Progeny: A Systematic Review: Male and Female Reproductive Traits ». *Reproduction in Domestic Animals* 46 (3):534-53. <https://doi.org/10.1111/j.1439-0531.2011.01748.x>.
- Burnside, EB, Kowalchuk, SB, Lambroughton, DB, et MacLeod, NM. 1971. « Canadian dairy cow disposals: I. Differences between breeds, lactation numbers and seasons ». *Canadian Journal of Animal Science* 51 (1):75-83.
- Burton, A, et Torres-Padilla, ME. 2010. « Epigenetic Reprogramming and Development : A Unique Heterochromatin Organization in the Preimplantation Mouse Embryo ». *Briefings in Functional Genomics* 9 (5-6):444-54. <https://doi.org/10.1093/bfpg/elq027>.
- Butler, ST, Hutchinson, IA, Cromie, AR, et Shalloo, L. 2014. « Applications and Cost Benefits of Sexed Semen in Pasture-Based Dairy Production Systems ». *Animal* 8 (s1):165-72. <https://doi.org/10.1017/S1751731114000664>.
- Butler, WR. 1998. « Effect of protein nutrition on ovarian and uterine physiology in dairy cattle ». *Journal of dairy science* 81 (9):2533-2539.
- Cai, D, Jia Y, Song H, Sui S, Lu J, et al. (2014) Betaine Supplementation in Maternal Diet Modulates the Epigenetic Regulation of Hepatic Gluconeogenic Genes in Neonatal Piglets. *PLoS ONE* 9(8): e105504. doi:10.1371/journal.pone.0105504
- Cammack, K.M., Thomas, M.G., Enns, R.M., 2009. Reproductive traits and their heritabilities in beef cattle. *Prof. Anim. Sci.* 25, 517-528.
- Capel, B. 2000. « The battle of the sexes ». *Mechanisms of development* 92 (1):89-103.
- Cardoso, MC, Sporbert, A, et Leonhardt, H. 1999. « Structure and function in the nucleus: subnuclear trafficking of DNA replication factors ». *Journal of cellular biochemistry* 75 (s 32):15-23.
- Carrel, L, et Willard, HF. 2005. « X-inactivation profile reveals extensive variability in X-linked gene expression in females ». *Nature* 434 (7031):400-404.
- Carrell DT, 2012. « Epigenetics of the male gamete ». *Fertility and Sterility* Vol. 97, No. 2, February 2012 0015-0282. doi:10.1016/j.fertnstert.2011.12.036.
- Cassell, BG, Adamec, V, et Pearson, RE. 2003. « Maternal and fetal inbreeding depression for 70-day nonreturn and calving rate in Holsteins and Jerseys ». *Journal of dairy science* 86 (9):2977-2983.
- Cedar, H, Solage, A, Glaser, G, et Razin, A. 1979. « Direct detection of methylated cytosine in DNA by use of the restriction enzyme MspI ». *Nucleic acids research* 6 (6):2125-2132.
- Charlier, C, Agerholm, JS, Coppieters, W, Karlskov-Mortensen, P, Li, W, de Jong, G, Fasquelle, C. 2012. « A Deletion in the Bovine FANCI Gene Compromises Fertility by Causing Fetal Death and Brachyspina ». Édité par Reiner Albert Veitia. *PLoS ONE* 7 (8):e43085. <https://doi.org/10.1371/journal.pone.0043085>.
- Chatterjee, A, Ozaki, Y, Stockwell, PA, Horsfield, JA, Morison, IM, et Nakagawa, S. 2013. « Mapping the Zebrafish Brain Methylome Using Reduced Representation Bisulfite Sequencing ». *Epigenetics* 8 (9):979-89. <https://doi.org/10.4161/epi.25797>.
- Chavatte-Palmer, P, et Tarrade, A. 2016. « Placentation in Different Mammalian Species ». *Annales D'endocrinologie* 77 (2):67-74. <https://doi.org/10.1016/j.ando.2016.04.006>.
- Chaves, R, Atega, F, Heslop-Harrison, JS, Guedes-Pinto, H, et Wienberg, J. 2003. « Complex satellite DNA reshuffling in the polymorphic t (1; 29) Robertsonian translocation and evolutionarily derived chromosomes in cattle ». *Chromosome Research* 11 (7):641-648.
- Chen, BF, et Chan, WY. 2014. « The de Novo DNA Methyltransferase DNMT3A in Development and Cancer ». *Epigenetics* 9 (5):669-77. <https://doi.org/10.4161/epi.28324>.
- Chen, L, Chen, K, Lavery, LA, Baker, SA, Shaw, CA, Li, W, et Zoghbi, HY. 2015. « MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome ». *Proceedings of the National Academy of Sciences* 112 (17):5509-5514.
- Chen, SL, Shi, XY, Zheng, HY, Wu, FR, et Luo, C. 2010. « Aberrant DNA Methylation of Imprinted H19 Gene in Human Preimplantation Embryos ». *Fertility and Sterility* 94 (6):2356-2358.e1. <https://doi.org/10.1016/j.fertnstert.2010.01.120>.
- Choi, E, Han, C, Park, I, Lee, B, Jin, S, Choi, H, Kim, DH, Park, ZY, Eddy, EM, et Cho, C. 2008. « A Novel Germ Cell-Specific Protein, SHIP1, Forms a Complex with Chromatin Remodeling Activity during Spermatogenesis ». *Journal of Biological Chemistry* 283 (50):35283-94. <https://doi.org/10.1074/jbc.M805590200>.
- Clark, SJ, Harrison, J, Paul, CL, et Frommer, M. 1994. « High sensitivity mapping of methylated cytosines ». *Nucleic acids research* 22 (15):2990-2997.

- Clift, D, et Schuh, M. 2013. « Restarting life: fertilization and the transition from meiosis to mitosis ». *Nature* Vol. 14; 549-562; doi:10.1038/nrm3643.
- Colleau, JJ, Fritz, S, et Guillaume, F. 2009. « Simulating the potential of genomic selection in dairy cattle breeding ». *Rencontres Recherche Ruminants*, vol. 16, p. 419.
- Cornwall, GA. 2008. « New Insights into Epididymal Biology and Function ». *Human Reproduction Update* 15 (2):213-27. <https://doi.org/10.1093/humupd/dmn055>.
- Cortellino, S, Xu, J, Sannai, M, Moore, R, Caretti, E, Cigliano, A, Le Coz, M, et al. 2011. « Thymine DNA Glycosylase Is Essential for Active DNA Demethylation by Linked Deamination-Base Excision Repair ». *Cell* 146 (1):67-79. <https://doi.org/10.1016/j.cell.2011.06.020>.
- Couldrey, C, Brauning, R, Bracegirdle, J, Maclean, P, Henderson, HV, et McEwan, JC. 2014. « Genome-Wide DNA Methylation Patterns and Transcription Analysis in Sheep Muscle ». Édité par Heiner Niemann. *PLoS ONE* 9 (7):e101853. <https://doi.org/10.1371/journal.pone.0101853>.
- Couldrey, C, et Lee, RSF. 2010. « DNA methylation patterns in tissues from mid-gestation bovine fetuses produced by somatic cell nuclear transfer show subtle abnormalities in nuclear reprogramming ». *BMC developmental biology* 10 (1):27.
- Couldrey, Christine, et David N. Wells. 2013. « DNA Methylation at a Bovine Alpha Satellite I Repeat CpG Site during Development Following Fertilization and Somatic Cell Nuclear Transfer ». Édité par Xiuhcun (Cindy) Tian. *PLoS ONE* 8 (2):e55153. <https://doi.org/10.1371/journal.pone.0055153>.
- Couldrey, C, Wells, DN, et Lee, RSF. 2011. « DNA Methylation Patterns Are Appropriately Established in the Sperm of Bulls Generated by Somatic Cell Nuclear Transfer ». *Cellular Reprogramming (Formerly « Cloning and Stem Cells »)* 13 (2):171-77. <https://doi.org/10.1089/cell.2010.0065>.
- Coulter, GH, et Bailey, DR. 1988. « Effects of Ultrasonography on the Bovine Testis and Semen Quality ». *Theriogenology* 30 (4):743-49.
- Dada, R, Kumar, M, Jesudasan, R, Fernández, JL, Gosálvez, J, et Agarwal, A. 2012. « Epigenetics and Its Role in Male Infertility ». *Journal of Assisted Reproduction and Genetics* 29 (3):213-23. <https://doi.org/10.1007/s10815-012-9715-0>.
- Dadoune, JP. 2009. « Spermatozoal RNAs: What about Their Functions? » *Microscopy Research and Technique* 72 (8):536-51. <https://doi.org/10.1002/jemt.20697>.
- Daetwyler, Hans D, Aurélien Capitan, Hubert Pausch, Paul Stothard, Rianne van Binsbergen, Rasmus F Brøndum, Xiaoping Liao, et al. 2014. « Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle ». *Nature Genetics* 46 (8):858-65. <https://doi.org/10.1038/ng.3034>.
- Dai, J, Zhan, C, Xu, W, Wang, Z, Nie, D, Zhao, X, Zhang, D, et al. 2015. « Nicotine Elevates Sperm Motility and Induces *Pfn1* Promoter Hypomethylation in Mouse Testis ». *Andrology* 3 (5):967-78. <https://doi.org/10.1111/andr.12072>.
- Dance, A, Thundathil, J, Wilde, R, Blondin, P, et Kastelic, J. 2015. « Enhanced Early-Life Nutrition Promotes Hormone Production and Reproductive Development in Holstein Bulls ». *Journal of Dairy Science* 98 (2):987-98. <https://doi.org/10.3168/jds.2014-8564>.
- Danchin-Burge, C., Leroy, G., Moureaux, S., Verrier, E. 2012. « Evolution of the genetic variability of 8 French dairy cattle breeds assessed by pedigree analysis ». *Journal of Animal Breeding and Genetics* 129, 206-217.
- Daniel, JM, Spring, CM, Crawford, HC, Reynolds, AB, et Baig, A. 2002. « The P120(Ctn)-Binding Partner Kaiso Is a Bi-Modal DNA-Binding Protein That Recognizes Both a Sequence-Specific Consensus and Methylated CpG Dinucleotides ». *Nucleic Acids Research* 30 (13):2911-19.
- Daujat, S, Weiss, T, Mohn, F, Lange, UC, Ziegler-Birling, C, Zeissler, U, Lappe, M, Schübeler, D, Torres-Padilla, ME, et Schneider, R. 2009. « H3K64 trimethylation marks heterochromatin and is dynamically remodeled during developmental reprogramming ». *Nature Structural & Molecular Biology* 16 (7):777-81. <https://doi.org/10.1038/nsmb.1629>.
- De Felici, MASSIMO. 2000. « Regulation of primordial germ cell development in the mouse. » *International journal of developmental biology* 44 (6):575-580.
- De Meyer, T, Mampaey, E, Vlemmix, M, Denil, S, Trooskens, G, Renard, JP, De Keulenaer, S, Dehan, P, Menschaert, G, et Van Criekinge, W. 2013. « Quality Evaluation of Methyl Binding Domain Based Kits for Enrichment DNA-Methylation Sequencing ». Édité par Osman El-Maarri. *PLoS ONE* 8 (3):e59068. <https://doi.org/10.1371/journal.pone.0059068>.
- De Montera, B, El Zeihery, D, Müller, S, Jammes, H, Brem, G, Reichenbach, HD, Scheipl, F, Chavatte-Palmer, P, Zakhartchenko, V, et Schmitz, OJ. 2010. « Quantification of leukocyte genomic 5-methylcytosine levels reveals epigenetic plasticity in healthy adult cloned cattle ». *Cellular Reprogramming (Formerly "Cloning and Stem Cells")* 12 (2):175-181.
- Degrelle, SA, Champion, E, Cabau, C, Piumi, F, Reinaud, P, Richard, C, Renard, JP, et Hue, I. 2005. « Molecular Evidence for a Critical Period in Mural Trophoblast Development in Bovine Blastocysts ». *Developmental Biology* 288 (2):448-60. <https://doi.org/10.1016/j.ydbio.2005.09.043>.
- DeJarnette, JM, Marshall, CE, Lenz, RW, Monke, DR, Ayars, WH, et Sattler, CG. 2004. « Sustaining the fertility of artificially inseminated dairy cattle: the role of the artificial insemination industry ». *Journal of dairy Science* 87:E93-E104.
- DeJarnette, JM, McCleary, CR, Leach, MA, Moreno, JF, Nebel, RL, et Marshall, CE. 2010. « Effects of 2.1 and 3.5×10⁶ Sex-Sorted Sperm Dosages on Conception Rates of Holstein Cows and Heifers ». *Journal of Dairy Science* 93 (9):4079-85. <https://doi.org/10.3168/jds.2010-3181>.
- Den Daas, JH, De Jong, G, Lansbergen, LM, et Van Wagendonk-De Leeuw, AM. 1998. « The relationship between the number of spermatozoa inseminated and the reproductive efficiency of individual dairy bulls ». *Journal of Dairy Science* 81 (6):1714-1723.

- Dhayalan, A, Rajavelu, A, Rathert, P, Tamas, R, Jurkowska, RZ, Ragozin, S, et Jeltsch, A. 2010. « The Dnmt3a PWWP Domain Reads Histone 3 Lysine 36 Trimethylation and Guides DNA Methylation ». *Journal of Biological Chemistry* 285 (34):26114-20. <https://doi.org/10.1074/jbc.M109.089433>.
- Dias, BG, et Ressler, KJ. 2014. « Parental olfactory experience influences behavior and neural structure in subsequent generations ». *Nature neuroscience* 17 (1):89-96.
- Diskin, MG, et Morris, DG. 2008. « Embryonic and Early Foetal Losses in Cattle and Other Ruminants ». *Reproduction in Domestic Animals* 43 (juillet):260-67. <https://doi.org/10.1111/j.1439-0531.2008.01171.x>.
- Diskin, MG, Murphy, JJ, et Sreenan, JM. 2006. « Embryo Survival in Dairy Cows Managed under Pastoral Conditions ». *Animal Reproduction Science* 96 (3-4):297-311. <https://doi.org/10.1016/j.anireprosci.2006.08.008>.
- Doi, A, Park, IH, Wen, B, Murakami, P, Aryee, MJ, Irizarry, R, Herb, B, et al. 2009. « Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts ». *Nature Genetics* 41 (12):1350-53. <https://doi.org/10.1038/ng.471>.
- Ducos, A, Dumont, P, Séguéla, A, Pinton, A, Berland, H, Brun-Baronnat, C, Darré, A, et al. 2000. « A New Reciprocal Translocation in a Subfertile Bull ». *Genetics, Selection, Evolution: GSE* 32 (6):589-98. <https://doi.org/10.1051/gse:2000139>.
- Duncan, BK, et Miller, JH. 1980. « Mutagenic Deamination of Cytosine Residues in DNA ». *Nature* 287 (5782):560-61.
- Dunne, LD, Diskin, MG, et Sreenan, JM. 2000. « Embryo and Foetal Loss in Beef Heifers between Day 14 of Gestation and Full Term ». *Animal Reproduction Science* 58 (1-2):39-44.
- Dupont, JM, Tost, J, Jammes, H, et Gut, IG. 2004. « De Novo Quantitative Bisulfite Sequencing Using the Pyrosequencing Technology ». *Analytical Biochemistry* 333 (1):119-27. <https://doi.org/10.1016/j.ab.2004.05.007>.
- Egger, G, Jeong, S, Escobar, SG, Cortez, CC, Li, TW, Saito, Y, Yoo, CB, Jones, PA, et Liang, G. 2006. « Identification of DNMT1 (DNA Methyltransferase 1) Hypomorphs in Somatic Knockouts Suggests an Essential Role for DNMT1 in Cell Survival ». *Proceedings of the National Academy of Sciences of the United States of America* 103 (38):14080-85. <https://doi.org/10.1073/pnas.0604602103>.
- El Hajj, N, Trapphoff, T, Linke, M, May, A, Hansmann, T, Kuhlitz, J, Reifenberg, K, et al. 2011. « Limiting Dilution Bisulfite (Pyro)Sequencing Reveals Parent-Specific Methylation Patterns in Single Early Mouse Embryos and Bovine Oocytes ». *Epigenetics* 6 (10):1176-88. <https://doi.org/10.4161/epi.6.10.17202>.
- Engelken, TJ. 2008a. « Developing Replacement Beef Heifers ». *Theriogenology* 70 (3):569-72. <https://doi.org/10.1016/j.theriogenology.2008.05.032>.
- Engelken, TJ. 2008b. « The Development of Beef Breeding Bulls ». *Theriogenology* 70 (3):573-75. <https://doi.org/10.1016/j.theriogenology.2008.05.038>.
- Erkek, S, Hisano, M, Liang, CY, Gill, M, Murr, R, Dieker, J, Schübeler, D, Van der Vlag, J, Stadler, MB, et Peters, AH. 2013. « Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa ». *Nature Structural & Molecular Biology* 20 (7):868-75. <https://doi.org/10.1038/nsmb.2599>.
- Evans, JP. 2012. « Sperm-Egg Interaction ». *Annual Review of Physiology* 74 (1):477-502. <https://doi.org/10.1146/annurev-physiol-020911-153339>.
- Evenson, DP. 2016. « The Sperm Chromatin Structure Assay (SCSA®) and Other Sperm DNA Fragmentation Tests for Evaluation of Sperm Nuclear DNA Integrity as Related to Fertility ». *Animal Reproduction Science* 169 (juin):56-75. <https://doi.org/10.1016/j.anireprosci.2016.01.017>.
- Evenson, DP, Darzynkiewicz, Z, et Melamed, MR. 1980. « Comparison of Human and Mouse Sperm Chromatin Structure by Flow Cytometry ». *Chromosoma* 78 (2):225-38.
- Eymery, A, Horard, B, Atifi-Borel, ME, Fourel, G, Berger, F, Vitte, AL, Van den Broeck, A, et al. 2009. « A Transcriptomic Analysis of Human Centromeric and Pericentric Sequences in Normal and Tumor Cells ». *Nucleic Acids Research* 37 (19):6340-54. <https://doi.org/10.1093/nar/gkp639>.
- Farrell, PB, Presicce, GA, Brockett, CC, et Foote, RH. 1998. « Quantification of Bull Sperm Characteristics Measured by Computer-Assisted Sperm Analysis (CASA) and the Relationship to Fertility ». *Theriogenology* 49 (4):871-79. [https://doi.org/10.1016/S0093-691X\(98\)00036-3](https://doi.org/10.1016/S0093-691X(98)00036-3).
- Farthing, CR, Ficiz, G, Ng, RK, Chan, CF, Andrews, S, Dean, W, Hemberger, M, et Reik, W. 2008. « Global Mapping of DNA Methylation in Mouse Promoters Reveals Epigenetic Reprogramming of Pluripotency Genes ». Édité par Wayne N. Frankel. *PLoS Genetics* 4 (6):e1000116. <https://doi.org/10.1371/journal.pgen.1000116>.
- Fatehi, AN. 2006. « DNA Damage in Bovine Sperm Does Not Block Fertilization and Early Embryonic Development But Induces Apoptosis After the First Cleavages ». *Journal of Andrology* 27 (2):176-88. <https://doi.org/10.2164/jandrol.04152>.
- Fazzari, MJ, et Grealley, JM. 2004. « Epigenomics: beyond CpG islands ». *Nature Reviews Genetics* 5 (6):446-55. <https://doi.org/10.1038/nrg1349>.
- Ficz, G, Branco, MR, Seisenberger, S, Santos, F, Krueger, F, Hore, TA, Marques, CJ, Andrews, S, et Reik, W. 2011. « Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation ». *Nature* 473 (7347):398-402. <https://doi.org/10.1038/nature10008>.
- Flowers, WL. 2013. « Triennial Reproduction Symposium: Sperm Characteristics That Limit Success of Fertilization ». *Journal of Animal Science* 91 (7):3022-29. <https://doi.org/10.2527/jas.2012-5945>.
- Fortes, MR, Deatley, KL, Lehnert, SA, Burns, BM, Reverter, A, Hawken, RJ, Boe-Hansen, G, Moore, SS, et Thomas, MG. 2013. « Genomic Regions Associated with Fertility Traits in Male and Female Cattle: Advances from Microsatellites to High-Density Chips and Beyond ». *Animal Reproduction Science* 141 (1-2):1-19. <https://doi.org/10.1016/j.anireprosci.2013.07.002>.
- Frauer, C, Hoffmann, T, Bultmann, S, Casa, V, Cardoso, MC, Antes, I, et Leonhardt, H. 2011. « Recognition of 5-Hydroxymethylcytosine by the Uhrf1 SRA Domain ». Édité par Shuang-yong Xu. *PLoS ONE* 6 (6):e21306. <https://doi.org/10.1371/journal.pone.0021306>.
- Frei, RE, Schultz, GA, et Church, RB. 1989. « Qualitative and Quantitative Changes in Protein Synthesis Occur at the 8-16-Cell Stage of Embryogenesis in the Cow ». *Journal of Reproduction and Fertility* 86 (2):637-41.

- Freret, S, Ponsart, C, Rai, DB, Jeanguyot, N, Paccard, P, et Humblot, P. 2006. « Facteurs de variation de la fertilité en première insémination et des taux de mortalités embryonnaires en élevages laitiers Prim'Holstein ». *Rencontres Recherches Ruminants* 13 (janvier):281-84.
- Fritz, S, Capitan, A, Djari, A, Rodriguez, SC, Barbat, A, Baur, A, Grohs, C, et al. 2013. « Detection of Haplotypes Associated with Prenatal Death in Dairy Cattle and Identification of Deleterious Mutations in GART, SHBG and SLC37A2 ». Édité par Reiner Albert Veitia. *PLoS ONE* 8 (6):e65550. <https://doi.org/10.1371/journal.pone.0065550>.
- Frommer, M, McDonald, LE, Millar, DS, Collis, CM, Watt, F, Grigg, GW, Molloy, PL, et Paul, CL. 1992. « A Genomic Sequencing Protocol That Yields a Positive Display of 5-Methylcytosine Residues in Individual DNA Strands ». *Proceedings of the National Academy of Sciences of the United States of America* 89 (5):1827-31.
- Fujita, N, Watanabe, S, Ichimura, T, Tsuruzoe, S, Shinkai, Y, Tachibana, M, Chiba, T, et Nakao, M. 2003. « Methyl-CpG Binding Domain 1 (MBD1) Interacts with the Suv39h1-HP1 Heterochromatin Complex for DNA Methylation-Based Transcriptional Repression ». *Journal of Biological Chemistry* 278 (26):24132-38. <https://doi.org/10.1074/jbc.M302283200>.
- Fuks, F. 2003. « The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase ». *Nucleic Acids Research* 31 (9):2305-12. <https://doi.org/10.1093/nar/gkg332>.
- Fuks, F, Burgers, WA, Brehm, A, Hughes-Davies, L, et Kouzarides, T. 2000. « DNA Methyltransferase Dnmt1 Associates with Histone Deacetylase Activity ». *Nature Genetics* 24 (1):88-91. <https://doi.org/10.1038/17150>.
- Gabory, A, Ripoche, MA, Le Digarcher, A, Watrin, F, Ziyat, A, Forne, T, Jammes, H, et al. 2009. « H19 Acts as a Trans Regulator of the Imprinted Gene Network Controlling Growth in Mice ». *Development* 136 (20):3413-21. <https://doi.org/10.1242/dev.036061>.
- Galli, A, Basetti, M, Balduzzi, D, Martignoni, M, Bornaghi, V, et Maffii, M. 1991. « Frozen Bovine Semen Quality and Bovine Cervical Mucus Penetration Test ». *Theriogenology* 35 (4):837-44.
- Gal-Yam, EN, Egger, G, Iniguez, L, Holster, H, Einarsson, S, Zhang, X, Lin, JC, Liang, G, Jones, PA, et Tanay, A. 2008. « Frequent Switching of Polycomb Repressive Marks and DNA Hypermethylation in the PC3 Prostate Cancer Cell Line ». *Proceedings of the National Academy of Sciences of the United States of America* 105 (35):12979-84. <https://doi.org/10.1073/pnas.0806437105>.
- García-Macías, Vanesa, Paulino de Paz, Felipe Martínez-Pastor, Mercedes Álvarez, Susana Gomes-Alves, Jana Bernardo, Enrique Anel, et Luis Anel. 2007. « DNA Fragmentation Assessment by Flow Cytometry and Sperm?Bos?Halomax (Bright-Field Microscopy and Fluorescence Microscopy) in Bull Sperm ». *International Journal of Andrology* 30 (2):88-98. <https://doi.org/10.1111/j.1365-2605.2006.00723.x>.
- Gardiner-Garden, M, et Frommer, M. 1987. « CpG Islands in Vertebrate Genomes ». *Journal of Molecular Biology* 196 (2):261-82.
- Garner, DL. 2009. « Hoechst 33342: The Dye That Enabled Differentiation of Living X-and Y-Chromosome Bearing Mammalian Sperm ». *Theriogenology* 71 (1):11-21. <https://doi.org/10.1016/j.theriogenology.2008.09.023>.
- Garrett, C. 2003. « Automated Semen Analysis: "Zona Pellucida Preferred" Sperm Morphometry and Straight-Line Velocity Are Related to Pregnancy Rate in Subfertile Couples ». *Human Reproduction* 18 (8):1643-49. <https://doi.org/10.1093/humrep/deg306>.
- Gerard, O, Ponsart, C, Petit, M, et Humblot, P. 2008. « Evolution des techniques de préparation de la semence et d'insémination artificielle chez les bovins ». In *Rencontres autour des recherches sur les ruminants*, 351-54. Institut de l'élevage. <http://cat.inist.fr/?aModele=afficheN&cpsid=21823726>.
- Gérard-Monnier, D, Erdelmeier, I, Régnard, K, Moze-Henry, N, Yadan, JC, et Chaudière, J. 1998. « Reactions of 1-Methyl-2-Phenylindole with Malondialdehyde and 4-Hydroxyalkenals. Analytical Applications to a Colorimetric Assay of Lipid Peroxidation ». *Chemical Research in Toxicology* 11 (10):1176-83. <https://doi.org/10.1021/tx9701790>.
- Ghirlando, R, et Felsenfeld, G. 2016. « CTCF: making the right connections ». *Genes & Development* 30:881-891. Article is online at <http://www.genesdev.org/cgi/doi/10.1101/gad.277863.116>.
- Giorgetti, L, Lajoie, BR, Carter, AC, Attia, M, Zhan, Y, Xu, J, Chen, CJ, et al. 2016. « Structural organization of the inactive X chromosome in the mouse ». *Nature* 535 (7613):575-79. <https://doi.org/10.1038/nature18589>.
- Glass, JL, Thompson, RF, Khulan, B, Figueroa, ME, Olivier, EN, Oakley, EJ, Van Zant, G, et al. 2007. « CG Dinucleotide Clustering Is a Species-Specific Property of the Genome ». *Nucleic Acids Research* 35 (20):6798-6807. <https://doi.org/10.1093/nar/gkm489>.
- Gorczyca, W, Traganos, F, Jesionowska, H, et Darzynkiewicz, Z. 1993. « Presence of DNA strand breaks and increased sensitivity of DNA in situ to denaturation in abnormal human sperm cells: analogy to apoptosis of somatic cells ». *Exp Cell Res*. 1993 Jul;207(1):202-5.
- Goudet, G, Mugnier, S, Le Foll, N, Chastant-Maillard, S, et Wolf, JP. 2014. « La fécondation ». In: Marie Saint-Dizier, Sylvie Chastant-Maillard, dir., *La reproduction animale et humaine* (p. 315-330). Synthèses (Quae). Versailles, FRA : Editions Quae. <http://prodinra.inra.fr/record/272871>.
- Grandi, FC, Rosser, JM, Newkirk, SJ, Yin, J, Jiang, X, Xing, Z, Whitmore, L, et al. 2015. « Retrotransposition Creates Sloping Shores: A Graded Influence of Hypomethylated CpG Islands on Flanking CpG Sites ». *Genome Research* 25 (8):1135-46. <https://doi.org/10.1101/gr.185132.114>.
- Gravance, CG, Garner, DL, Pitt, C, Vishwanath, R, Sax-Gravance, SK, et Casey, PJ. 1999. « Replicate and Technician Variation Associated with Computer Aided Bull Sperm Head Morphometry Analysis (ASMA) ». *International Journal of Andrology* 22 (2):77-82.
- Greathouse, JR, Hunt, MC, Dikeman, ME, Corah, LR, Kastner, CL, et Kropf, DH. 1983. « Ralgro-implanted bulls: performance, carcass characteristics, longissimus palatability and carcass electrical stimulation ». *Journal of Animal Science*, Vol. 57, No. 2.
- Grunau, C, Clark, SJ, et Rosenthal, A. 2001. « Bisulfite Genomic Sequencing: Systematic Investigation of Critical Experimental Parameters ». *Nucleic Acids Research* 29 (13):E65-65.

- Grunewald, S, Paasch, U, Glander, HJ, et Anderegg, U. 2005. « Mature Human Spermatozoa Do Not Transcribe Novel RNA ». *Andrologia* 37 (2-3):69-71. <https://doi.org/10.1111/j.1439-0272.2005.00656.x>.
- Gu, H, Bock, C, Mikkelsen, TS, Jäger, N, Smith, ZD, Tomazou, E, Gnirke, A, Lander, ES, et Meissner, A. 2010. « Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution ». *Nature Methods* 7 (2):133-36. <https://doi.org/10.1038/nmeth.1414>.
- Gu, Y, Li, Q, Pan, Z, Li, M, Luo, H, et Xie, Z. 2013. « Molecular Cloning, Gene Expression and Methylation Status Analysis of PIWIL1 in Cattle-Yaks and the Parental Generation ». *Animal Reproduction Science* 140 (3-4):131-37. <https://doi.org/10.1016/j.anireprosci.2013.05.010>.
- Guerrero-Bosagna, C. 2016. « High Type II Error and Interpretation Inconsistencies When Attempting to Refute Transgenerational Epigenetic Inheritance ». *Genome Biology* 17 (1). <https://doi.org/10.1186/s13059-016-0982-4>.
- Guibert, S, Fome, T, et Weber, M. 2012. « Global Profiling of DNA Methylation Erasure in Mouse Primordial Germ Cells ». *Genome Research* 22 (4):633-41. <https://doi.org/10.1101/gr.130997.111>.
- Guo, F, Li, X, Liang, D, Li, T, Zhu, P, Guo, H, Wu, X, et al. 2014. « Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote ». *Cell Stem Cell* 15 (4):447-59. <https://doi.org/10.1016/j.stem.2014.08.003>.
- Guo, F, Yang, B, Ju, ZH, Wang, XG, Qi, C, Zhang, Y, Wang, CF, et al. 2013. « Alternative Splicing, Promoter Methylation, and Functional SNPs of Sperm Flagella 2 Gene in Testis and Mature Spermatozoa of Holstein Bulls ». *Reproduction* 147 (2):241-52. <https://doi.org/10.1530/REP-13-0343>.
- Gutiérrez-Adán, A, Behboodi, E, Murray, JD, et Anderson, GB. 1997. « Early Transcription of the SRY Gene by Bovine Preimplantation Embryos ». *Molecular Reproduction and Development* 48 (2):246-50. [https://doi.org/10.1002/\(SICI\)1098-2795\(199710\)48:2<246::AID-MRD12>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1098-2795(199710)48:2<246::AID-MRD12>3.0.CO;2-W).
- Hackenberg, M, Previti, P, Luque-Escamilla, PL, Carpena, P, Martínez-Aroza, J, et Oliver, JL. 2006. « CpGcluster: a distance-based algorithm for CpG-island detection ». *BMC Bioinformatics* 2006, 7:446 doi:10.1186/1471-2105-7-446.
- Hackett, JA, Sengupta, R, Zyllicz, JJ, Murakami, K, Lee, C, Down, TA, et Surani, MA. 2013. « Germline DNA Demethylation Dynamics and Imprint Erasure Through 5-Hydroxymethylcytosine ». *Science* 339 (6118):448-52. <https://doi.org/10.1126/science.1229277>.
- Hagen, N, et Gayrard, V. 2005. « Mémento des critères numériques de reproduction des mammifères domestiques ». Ecole Vétérinaire de Toulouse.
- Hajkova, P, Erhardt, S, Lane, N, Haaf, T, El-Maarri, O, Reik, W, Walter, J, et Surani, MA. 2002. « Epigenetic Reprogramming in Mouse Primordial Germ Cells ». *Mechanisms of Development* 117 (1-2):15-23.
- Hajkova, P, Jeffries, SJ, Lee, C, Miller, N, Jackson, SP, et Surani, MA. 2010. « Genome-Wide Reprogramming in the Mouse Germ Line Entails the Base Excision Repair Pathway ». *Science* 329 (5987):78-82. <https://doi.org/10.1126/science.1187945>.
- Hammoud, SS, Nix, DA, Zhang, H, Purwar, J, Carrell, DT, et Cairns, BR. 2009. « Distinctive chromatin in human sperm packages genes for embryo development ». *Nature*, juin. <https://doi.org/10.1038/nature08162>.
- Hammoud, SS, Purwar, J, Pflueger, C, Cairns, BR, et Carrell, DT. 2010. « Alterations in Sperm DNA Methylation Patterns at Imprinted Loci in Two Classes of Infertility ». *Fertility and Sterility* 94 (5):1728-33. <https://doi.org/10.1016/j.fertnstert.2009.09.010>.
- Hammoud, SS, Nix, DA, Hammoud, AO, Gibson, M, Cairns, BR, et Carrell, DT. 2011. « Genome-Wide Analysis Identifies Changes in Histone Retention and Epigenetic Modifications at Developmental and Imprinted Gene Loci in the Sperm of Infertile Men ». *Human Reproduction* 26 (9):2558-69. <https://doi.org/10.1093/humrep/der192>.
- Hammoud, SS, Low, DH, Yi, C, Carrell, DT, Guccione, E, et Cairns, BR. 2014. « Chromatin and Transcription Transitions of Mammalian Adult Germline Stem Cells and Spermatogenesis ». *Cell Stem Cell* 15 (2):239-53. <https://doi.org/10.1016/j.stem.2014.04.006>.
- Hayashi, K, Chuva de Sousa Lopes, SM, Kaneda, M, Tang, F, Hajkova, P, Lao, K, O'Carroll, D, et al. 2008. « MicroRNA Biogenesis Is Required for Mouse Primordial Germ Cell Development and Spermatogenesis ». Édité par Edith Heard. *PLoS ONE* 3 (3):e1738. <https://doi.org/10.1371/journal.pone.0001738>.
- Hayashi, S. 2003. « Mouse Preimplantation Embryos Developed from Oocytes Injected with Round Spermatids or Spermatozoa Have Similar but Distinct Patterns of Early Messenger RNA Expression ». *Biology of Reproduction* 69 (4):1170-76. <https://doi.org/10.1095/biolreprod.103.016832>.
- Hayashi, T, Miyata, A, et Yamada, T. 2008. « The Impact of Commonly Prescribed Drugs on Male Fertility ». *Human Fertility (Cambridge, England)* 11 (3):191-96. <https://doi.org/10.1080/14647270701739566>.
- Hayes, BJ, Bowman, PJ, Chamberlain, AJ, et Goddard, ME. 2009. « Invited Review: Genomic Selection in Dairy Cattle: Progress and Challenges ». *Journal of Dairy Science* 92 (2):433-43. <https://doi.org/10.3168/jds.2008-1646>.
- Hayes, B, MacLeod, I, Daetwyler, H, Bowman, P, Chamberlain, A, et al. 2014. « Genomic prediction from whole genome sequence in livestock: the 1000 bull genomes project ». In Proceedings of the 10th World Congress of Genetics Applied to Livestock Production.
- Hazzouri, M, Pivot-Pajot, C, Faure, AK, Usson, Y, Pelletier, R, Sèle, B, Khochbin, S, et Rousseaux, S. 2000. « Regulated Hyperacetylation of Core Histones during Mouse Spermatogenesis: Involvement of Histone Deacetylases ». *European Journal of Cell Biology* 79 (12):950-60. <https://doi.org/10.1078/0171-9335-00123>.
- Heijmans, BT, Tobi, EW, Stein, AD, Putter, H, Blauw, GJ, Susser, ES, Slagboom, PE, et Lumey, LH. 2008. « Persistent Epigenetic Differences Associated with Prenatal Exposure to Famine in Humans ». *Proceedings of the National Academy of Sciences of the United States of America* 105 (44):17046-49. <https://doi.org/10.1073/pnas.0806560105>.
- Henckel, A, Chebli, K, Kota, SK, Arnaud, P, et Feil, R. 2012. « Transcription and Histone Methylation Changes Correlate with Imprint Acquisition in Male Germ Cells ». *The EMBO Journal* 31 (3):606-15. <https://doi.org/10.1038/emboj.2011.425>.

- Hereng, TH, Elgstøen, KBP, Eide, L, Rosendal, KR, et Skålhegg, BS. 2014. « Serum Albumin and HCO₃⁻ Regulate Separate Pools of ATP in Human Spermatozoa ». *Human Reproduction* 29 (5):918-30. <https://doi.org/10.1093/humrep/deu028>.
- Hering, DM, Olenksi, K, et Kaminski, S. 2014. « Genome-Wide Association Study for Poor Sperm Motility in Holstein-Friesian Bulls ». *Animal Reproduction Science* 146 (3-4):89-97. <https://doi.org/10.1016/j.anireprosci.2014.01.012>.
- Hering, DM, Oleński, K, Ruś, A, et Kaminski, S. 2014. « Genome-Wide Association Study for Semen Volume and Total Number of Sperm in Holstein-Friesian Bulls ». *Animal Reproduction Science* 151 (3-4):126-30. <https://doi.org/10.1016/j.anireprosci.2014.10.022>.
- Herman, JG, Graff, JR, Myöhänen, S, Nelkin, BD, et Baylin, SB. 1996. « Methylation-Specific PCR: A Novel PCR Assay for Methylation Status of CpG Islands ». *Proceedings of the National Academy of Sciences of the United States of America* 93 (18):9821-26.
- Heyn, H, Ferreira, HJ, Bassas, L, Bonache, S, Sayols, S, Sandoval, J, Esteller, M, et Larriba, S. 2012. « Epigenetic Disruption of the PIWI Pathway in Human Spermatogenic Disorders ». Édité par Carmen J. Marsit. *PLoS ONE* 7 (10):e47892. <https://doi.org/10.1371/journal.pone.0047892>.
- Hinrichs, D, et Thaller, G. 2011. « Pedigree Analysis and Inbreeding Effects on Calving Traits in Large Dairy Herds in Germany ». *Journal of Dairy Science* 94 (9):4726-33. <https://doi.org/10.3168/jds.2010-4100>.
- Hinsch, E, Hinsch, KD, Boehm, J, Schill, WB, et Mueller-Schloesser, F. 1997. « Functional Parameters and Fertilization Success of Bovine Semen Cryopreserved in Egg-Yolk Free and Egg-Yolk Containing Extenders ». *Reproduction in Domestic Animals* 32 (3):143-49. <https://doi.org/10.1111/j.1439-0531.1997.tb01272.x>.
- Hirasawa, R, Chiba, H, Kaneda, M, Tajima, S, Li, E, Jaenisch, R, et Sasaki, H. 2008. « Maternal and Zygotic Dnmt1 Are Necessary and Sufficient for the Maintenance of DNA Methylation Imprints during Preimplantation Development ». *Genes & Development* 22 (12):1607-16. <https://doi.org/10.1101/gad.1667008>.
- Hocdé, A, et Joly, PB, 2013. « Analyse Socio-économique des Impacts de la Recherche Publique Agricole : La sélection génomique bovine ». *Executive Summary ASIRPA*.
- Hochereau-de Reviers, MT, Courot, M, et Ortavant, R. 1964. « Durée de la spermatogenèse chez le taureau : étude par autoradiographie testiculaire ». In *5th International Congress on Animal Reproduction and Artificial Insemination*. Volume III 541-546.
- Holliday, R, et Pugh, JE. 1975. « DNA Modification Mechanisms and Gene Activity during Development ». *Science (New York, N.Y.)* 187 (4173):226-32.
- Holm, P, Shukri, NN, Vajta, G, Booth, P, Bendixen, C, et Callesen, H. 1998. « Developmental Kinetics of the First Cell Cycles of Bovine In Vitro Produced Embryos in Relation to Their In Vitro Viability and Sex ». *Theriogenology* 50 (8):1285-99.
- Hosack, DA, Dennis, G, Sherman, BT, Lane, HC, et Lempicki, RA. 2003. « Identifying Biological Themes within Lists of Genes with EASE ». *Genome Biology* 4 (10):R70. <https://doi.org/10.1186/gb-2003-4-10-r70>.
- Hosseini, SM, Dufort, I, Caballero, J, Moulavi, F, Ghanaei, HR, et Sirard, MA. 2015. « Transcriptome Profiling of Bovine Inner Cell Mass and Trophectoderm Derived from in Vivo Generated Blastocysts ». *BMC Developmental Biology* 15 (1). <https://doi.org/10.1186/s12861-015-0096-3>.
- Houshdaran, S, Cortessis, VK, Siegmund, K, Yang, A, Laird, PW, et Sokol, RZ. 2007. « Widespread Epigenetic Abnormalities Suggest a Broad DNA Methylation Erasure Defect in Abnormal Human Sperm ». Édité par Rodolfo Aramayo. *PLoS ONE* 2 (12):e1289. <https://doi.org/10.1371/journal.pone.0001289>.
- Huang, DW, Sherman, BT, et Lempicki, RA. 2008. « Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources ». *Nature Protocols* 4 (1):44-57. <https://doi.org/10.1038/nprot.2008.211>.
- Huang da, W, Sherman, BT, et Lempicki, RA. 2009. « Bioinformatics Enrichment Tools: Paths toward the Comprehensive Functional Analysis of Large Gene Lists ». *Nucleic Acids Research* 37 (1):1-13. <https://doi.org/10.1093/nar/gkn923>.
- Huang, Y, Pastor, WA, Shen, Y, Tahiliani, M, Liu, DR, et Rao, A. 2010. « The Behaviour of 5-Hydroxymethylcytosine in Bisulfite Sequencing ». Édité par Jun Liu. *PLoS ONE* 5 (1):e8888. <https://doi.org/10.1371/journal.pone.0008888>.
- Hu, Y, Sun, Q, Li, X, Wang, M, Cai, D, Li, X, et Zhao, R. 2015. « In Ovo Injection of Betaine Affects Hepatic Cholesterol Metabolism through Epigenetic Gene Regulation in Newly Hatched Chicks ». *PLoS ONE* 10(4): e0122643. doi:10.1371/journal.pone.0122643.
- Hue, I, Degrelle, SA, et Turenne, N. 2012. « Conceptus Elongation in Cattle: Genes, Models and Questions ». *Animal Reproduction Science* 134 (1-2):19-28. <https://doi.org/10.1016/j.anireprosci.2012.08.007>.
- Humblot, P. 2001. « Use of Pregnancy Specific Proteins and Progesterone Assays to Monitor Pregnancy and Determine the Timing, Frequencies and Sources of Embryonic Mortality in Ruminants ». *Theriogenology* 56 (9):1417-33.
- Illingworth, RS, Gruenewald-Schneider, U, Webb, S, Kerr, AR, James, KD, Turner, DJ, Smith, C, Harrison, DJ, Andrews, R, et Bird, A. 2010. « Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome ». Édité par Wolf Reik. *PLoS Genetics* 6 (9):e1001134. <https://doi.org/10.1371/journal.pgen.1001134>.
- Intarasunanont, P, Navasumrit, P, Waraprasit, S, Chaisatra, K, Suk, WA, Mahidol, C, et Ruchirawat, M. 2012. « Effects of Arsenic Exposure on DNA Methylation in Cord Blood Samples from Newborn Babies and in a Human Lymphoblast Cell Line ». *Environmental Health: A Global Access Science Source* 11 (mai):31. <https://doi.org/10.1186/1476-069X-11-31>.
- Iqbal, K, Tran, DA, Li, AX, Warden, C, Bai, AY, Singh, P, Wu, X, Pfeifer, GP, et Szabó, PE. 2015. « Deleterious Effects of Endocrine Disruptors Are Corrected in the Mammalian Germline by Epigenome Reprogramming ». *Genome Biology* 16 (1). <https://doi.org/10.1186/s13059-015-0619-z>.
- Irizarry, RA, Ladd-Acosta, C, Wen, B, Wu, Z, Montano, C, Onyango, P, Cui, H, et al. 2009. « The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores ». *Nature Genetics* 41 (2):178-86. <https://doi.org/10.1038/ng.298>.
- Ito, J, et Kashiwazaki, N. 2012. « Molecular Mechanism of Fertilization in the Pig: Molecular mechanism of pig fertilization ». *Animal Science Journal* 83 (10):669-82. <https://doi.org/10.1111/j.1740-0929.2012.01044.x>.

- Iuso, D, Czernik, M, Toschi, P, Fidanza, A, Zacchini, F, Feil, R, Curtet, S, et al. 2015. « Exogenous Expression of Human Protamine 1 (hPrm1) Remodels Fibroblast Nuclei into Spermatid-like Structures ». *Cell Reports* 13 (9):1765-71. <https://doi.org/10.1016/j.celrep.2015.10.066>.
- Januskaukas, A, Johannisson, A, et Rodriguez-Martinez, H. 2003. « Subtle Membrane Changes in Cryopreserved Bull Semen in Relation with Sperm Viability, Chromatin Structure, and Field Fertility ». *Theriogenology* 60 (4):743-58. [https://doi.org/10.1016/S0093-691X\(03\)00050-5](https://doi.org/10.1016/S0093-691X(03)00050-5).
- Jefferson, A, Colella, S, Moralli, D, Wilson, N, Yusuf, M, Gimelli, G, Ragoussis, J, et Volpi, EV. 2010. « Altered Intra-Nuclear Organisation of Heterochromatin and Genes in ICF Syndrome ». Édité par Beth A. Sullivan. *PLoS ONE* 5 (6):e11364. <https://doi.org/10.1371/journal.pone.0011364>.
- Jégou, B, Rolland, AD, Albert, O. 2014. Le testicule. In : Marie Saint-Dizier, Sylvie Chastant-Maillard, dir., « La reproduction animale et humaine » (p. 315-330). Synthèses (Quae). Versailles, FRA : Editions Quae.
- Jena, SC, Kumar, S, Rajput, S, Roy, B, Verma, A, Kumaresan, A, Mohanty, TK, De, S, Kumar, R, et Datta, TK. 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 : *IGF2-H19* methylation and bull fertility ». *Molecular Reproduction and Development* 81 (4):350-62. <https://doi.org/10.1002/mrd.22303>.
- Jenkins, TG, et Carrell, DT. 2012. « The Sperm Epigenome and Potential Implications for the Developing Embryo ». *Reproduction* 143 (6):727-34. <https://doi.org/10.1530/REP-11-0450>.
- Jeyendran, RS, Van der Ven, HH, Perez-Pelaez, M, Crabo, BG, et Zaneveld, LJ. 1984. « Development of an Assay to Assess the Functional Integrity of the Human Sperm Membrane and Its Relationship to Other Semen Characteristics ». *Journal of Reproduction and Fertility* 70 (1):219-28.
- Jin, B, Tao, Q, Peng, J, Soo, HM, Wu, W, Ying, J, Fields, CR, et al. 2008. « DNA Methyltransferase 3B (DNMT3B) Mutations in ICF Syndrome Lead to Altered Epigenetic Modifications and Aberrant Expression of Genes Regulating Development, Neurogenesis and Immune Function ». *Human Molecular Genetics* 17 (5):690-709. <https://doi.org/10.1093/hmg/ddm341>.
- Jin, SG, Kadam, S, et Pfeifer, GP. 2010. « Examination of the Specificity of DNA Methylation Profiling Techniques towards 5-Methylcytosine and 5-Hydroxymethylcytosine ». *Nucleic Acids Research* 38 (11):e125-e125. <https://doi.org/10.1093/nar/gkq223>.
- Jing-bo, L, Ying, Y, Bing, Y, Xiang-bing, M, Zhi-qing, H, Guo-quan, H, Hong, C, et Dai-wen, C. 2012. « Folic acid supplementation prevents the changes in hepatic promoter methylation status and gene expression in intrauterine growth-retarded piglets during early weaning period ». *Journal of Animal Physiology and Animal Nutrition* 97 878-886 DOI: 10.1111/j.1439-0396.2012.01333.x.
- Jost, A, Vigier, B, Prépin, J, et Perchellet, JP. 1973. « Studies on Sex Differentiation in Mammals ». *Recent Progress in Hormone Research* 29:1-41.
- Józków, P, et Rossato, M. 2017. « The Impact of Intense Exercise on Semen Quality ». *American Journal of Men's Health* 11 (3):654-62. <https://doi.org/10.1177/1557988316669045>.
- Junien, C, Panchenko, P, Fneich, S, Pirola, L, Chriett, S, Amarger, V, Kaeffler, B, Parnet, P, Torrisani, J, Bolaños Jimenez, F, Jammes, H, Gabory, A. 2016. « Épigénétique et réponses transgénérationnelles aux impacts de l'environnement : Des faits aux lacunes ». *médecine/sciences* 2016 ; 32 : 35-44.
- Kaneda, M, Okano, M, Hata, K, Sado, T, Tsujimoto, N, Li, E, et Sasaki, H. 2004. « Essential Role for de Novo DNA Methyltransferase Dnmt3a in Paternal and Maternal Imprinting ». *Nature* 429 (6994):900-903. <https://doi.org/10.1038/nature02633>.
- Karimi, M, Johansson, S, et Ekström, TJ. 2006. « Using LUMA : A Luminometric-Based Assay for Global DNA-Methylation ». *Epigenetics* 1 (1):46-49. <https://doi.org/10.4161/epi.1.1.2587>.
- Karimi, M, Johansson, S, Stach, D, Corcoran, M, Grandér, D, Schalling, M, Bakalkin, G, Lyko, F, Larsson, C, et Ekström, TJ. 2006. « LUMA (Luminometric Methylation Assay)—A High Throughput Method to the Analysis of Genomic DNA Methylation ». *Experimental Cell Research* 312 (11):1989-95. <https://doi.org/10.1016/j.yexcr.2006.03.006>.
- Karolchik, D, Kuhn, RM, Baertsch, R, Barber, GP, Clawson, H, Diekhans, M, Giardine, B, Harte, RA, Hinrichs, AS, Hsu, F, Kober, KM, Miller, W, Pedersen, JS, Pohl, A, Raney, BJ, Rhead, B, Rosenbloom, KR, Smith, KE, Stanke, M, Thakkapallayil, A, Trumbower, H, Wang, T, Zweig, AS, Haussler, D, et Kent, W. J. 2007. « The UCSC Genome Browser Database: 2008 update ». *Nucleic Acids Research*, 2008, Vol. 36, Database issue D773–D779 doi:10.1093/nar/gkm966.
- Karweina, D, Kreuzer-Redmer, S, Müller, U, Franken, T, Pieper, R, Baron, U, et al. 2015. « The Zinc Concentration in the Diet and the Length of the Feeding Period Affect the Methylation Status of the ZIP4 Zinc Transporter Gene in Piglets ». *PLoS ONE* 10 (11): e0143098. doi:10.1371/journal.pone.0143098.
- Kasimanickam, R, Kasimanickam, V, Thatcher, CD, Nebel, RL, et Cassell, BG. 2007. « Relationships among Lipid Peroxidation, Glutathione Peroxidase, Superoxide Dismutase, Sperm Parameters, and Competitive Index in Dairy Bulls ». *Theriogenology* 67 (5):1004-12. <https://doi.org/10.1016/j.theriogenology.2006.11.013>.
- Kastelic, JP, Cook, RB, Coulter, GH. 2000. « Scrotal / testicular thermoregulation in bulls ». In: Chenoweth PJ (ed.), Topics in Bull Fertility. International Veterinary Information Service.
- Kastelic, JP, et Thundathil, JC. 2008. « Breeding Soundness Evaluation and Semen Analysis for Predicting Bull Fertility ». *Reproduction in Domestic Animals* 43 (juillet):368-73. <https://doi.org/10.1111/j.1439-0531.2008.01186.x>.
- Katari, S, Nahid, T, Bibikova, M, Erinle, O, Chalian, R, 2009. « DNA methylation and gene expression differences in children conceived in vitro or in vivo ». *Human Molecular Genetics* vol 18 3769-3778.
- Kato, Y, Kaneda, M, Hata, K, Kumaki, K, Hisano, M, Kohara, Y, Okano, M, Li, E, Nozaki, M, et Sasaki, H. 2007. « Role of the Dnmt3 Family in de Novo Methylation of Imprinted and Repetitive Sequences during Male Germ Cell Development in the Mouse ». *Human Molecular Genetics* 16 (19):2272-80. <https://doi.org/10.1093/hmg/ddm179>.
- Kearney, JF, Wall, E, Villanueva, B, et Coffey, MP. 2004. « Inbreeding Trends and Application of Optimized Selection in the UK Holstein Population ». *Journal of Dairy Science* 87 (10):3503-9. [https://doi.org/10.3168/jds.S0022-0302\(04\)73485-2](https://doi.org/10.3168/jds.S0022-0302(04)73485-2).

- Keefer, CL. 2015. « Artificial Cloning of Domestic Animals ». *Proceedings of the National Academy of Sciences* 112 (29):8874-78. <https://doi.org/10.1073/pnas.1501718112>.
- Keeney, S. 2008. « Spo11 and the Formation of DNA Double-Strand Breaks in Meiosis ». *Genome dynamics and stability* 2 (janvier):81-123. https://doi.org/10.1007/7050_2007_026.
- Kiefer, H, Jouneau, L, Campion, E, Rousseau-Ralliard, D, Larcher, T, Martin-Magniette, ML, Balzergue, S, et al. 2016. « Altered DNA methylation associated with an abnormal liver phenotype in a cattle model with a high incidence of perinatal pathologies ». *Scientific Reports* 6 (décembre):38869. <https://doi.org/10.1038/srep38869>.
- Kile, ML, Baccarelli, A, Hoffman, E, Tarantini, L, Quamruzzaman, Q, Rahman, M, Mahiuddin, G, et al. 2012. « Prenatal Arsenic Exposure and DNA Methylation in Maternal and Umbilical Cord Blood Leukocytes ». *Environmental Health Perspectives* 120 (7):1061-66. <https://doi.org/10.1289/ehp.1104173>.
- Kläver, R, Sánchez, V, Damm, OS, Redmann, K, Lahrmann, E, Sandhowe-Klaverkamp, R, Rohde, C, et al. 2015. « Direct but No Transgenerational Effects of Decitabine and Vorinostat on Male Fertility ». *PLoS One* 10 (2):e0117839. <https://doi.org/10.1371/journal.pone.0117839>.
- Kobayashi, H, Hiura, H, John, RM, Sato, A, Otsu, E, Kobayashi, N, Suzuki, R, et al. 2009. « DNA Methylation Errors at Imprinted Loci after Assisted Conception Originate in the Parental Sperm ». *European Journal of Human Genetics: EJHG* 17 (12):1582-91. <https://doi.org/10.1038/ejhg.2009.68>.
- Kobayashi, H, Sakurai, T, Miura, F, Imai, M, Mochiduki, K, Yanagisawa, E, Sakashita, E, et al. 2013. « High-Resolution DNA Methylome Analysis of Primordial Germ Cells Identifies Gender-Specific Reprogramming in Mice ». *Genome Research* 23 (4):616-27. <https://doi.org/10.1101/gr.148023.112>.
- Kobayashi, H, Sato, A, Otsu, E, Hiura, H, Tomatsu, C, Utsunomiya, T, Sasaki, H, Yaegashi, N, et Arima, T. 2007. « Aberrant DNA Methylation of Imprinted Loci in Sperm from Oligospermic Patients ». *Human Molecular Genetics* 16 (21):2542-51. <https://doi.org/10.1093/hmg/ddm187>.
- Kokocinski, F, Murphy, DN, Nag, R, Ruffier, M, Schuster, M, Tang, YA, Vogel, JH, White, S, Zadissa, A, Flicek, P et Searle, SM. 2016. « The Ensembl gene annotation system ». *Database*. baw093. doi: 10.1093/database/baw093.
- Kondo, E, Gu, Z, Horii, A, et Fukushige, S. 2005. « The Thymine DNA Glycosylase MBD4 Represses Transcription and Is Associated with Methylated P16INK4a and HMLH1 Genes ». *Molecular and Cellular Biology* 25 (11):4388-96. <https://doi.org/10.1128/MCB.25.11.4388-4396.2005>.
- Krausz, C, Bonaccorsi, L, Luconi, M, Fuzzi, B, Criscuoli, L, Pellegrini, S, Forti, G, et Baldi, E. 1995. « Intracellular Calcium Increase and Acrosome Reaction in Response to Progesterone in Human Spermatozoa Are Correlated with In-Vitro Fertilization ». *Human Reproduction (Oxford, England)* 10 (1):120-24.
- Krausz, C, Bonaccorsi, L, Maggio, P, Luconi, M, Criscuoli, L, Fuzzi, B, Pellegrini, S, Forti, G, et Baldi, E. 1996. « Two Functional Assays of Sperm Responsiveness to Progesterone and Their Predictive Values in In-Vitro Fertilization ». *Human Reproduction (Oxford, England)* 11 (8):1661-67.
- Krausz, C, Sandoval, J, Sayols, S, Chianese, C, Giachini, C, Heyn, H, et Esteller, M. 2012. « Novel Insights into DNA Methylation Features in Spermatozoa: Stability and Peculiarities ». Édité par Joel R. Drevet. *PLoS ONE* 7 (10):e44479. <https://doi.org/10.1371/journal.pone.0044479>.
- Kriaucionis, S, et Heintz, N. 2009. « The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain ». *Science (New York, N.Y.)* 324 (5929):929-30. <https://doi.org/10.1126/science.1169786>.
- Kropp, J, Carrillo, JA, Namous, H, Daniels, A, Salih, SM, Song, J, et Khatib, H. 2017. « Male Fertility Status Is Associated with DNA Methylation Signatures in Sperm and Transcriptomic Profiles of Bovine Preimplantation Embryos ». *BMC Genomics* 18 (1). <https://doi.org/10.1186/s12864-017-3673-y>.
- Krueger, Felix, et Simon R. Andrews. 2011. « Bismark: A Flexible Aligner and Methylation Caller for Bisulfite-Seq Applications ». *Bioinformatics* 27 (11):1571-72. <https://doi.org/10.1093/bioinformatics/btr167>.
- Ku, Hsueh-Yen, et Haifan Lin. 2014. « PIWI Proteins and Their Interactors in PiRNA Biogenesis, Germline Development and Gene Expression ». *National Science Review* 1 (2):205-18. <https://doi.org/10.1093/nsr/nwu014>.
- Kuhn, MT, Hutchison, JL, et Wiggans, GR. 2006. « Characterization of Holstein Heifer Fertility in the United States ». *Journal of Dairy Science* 89 (12):4907-20. [https://doi.org/10.3168/jds.S0022-0302\(06\)72541-3](https://doi.org/10.3168/jds.S0022-0302(06)72541-3).
- Kuhn, MT, et Hutchison, JL. 2008. « Prediction of Dairy Bull Fertility from Field Data: Use of Multiple Services and Identification and Utilization of Factors Affecting Bull Fertility ». *Journal of Dairy Science* 91 (6):2481-92. <https://doi.org/10.3168/jds.2007-0743>.
- Kuhn, MT, Hutchison, JL, et Norman, HD. 2008. « Modeling Nuisance Variables for Prediction of Service Sire Fertility ». *Journal of Dairy Science* 91 (7):2823-35. <https://doi.org/10.3168/jds.2007-0946>.
- Kuramochi-Miyagawa, S, Watanabe, T, Gotoh, K, Totoki, Y, Toyoda, A, Ikawa, M, Asada, N, et al. 2008. « DNA Methylation of Retrotransposon Genes Is Regulated by Piwi Family Members MIL1 and MIW12 in Murine Fetal Testes ». *Genes & Development* 22 (7):908-17. <https://doi.org/10.1101/gad.1640708>.
- Kurdyukov, S, et Bullock, M. 2016. « DNA Methylation Analysis: Choosing the Right Method ». *Biology* 5 (1):3. <https://doi.org/10.3390/biology5010003>.
- La Salle, S, Oakes, CC, Neaga, OR, Bourc'his, D, Bestor, TH, et Trasler, JM. 2007. « Loss of Spermatogonia and Wide-Spread DNA Methylation Defects in Newborn Male Mice Deficient in DNMT3L ». *BMC Developmental Biology* 7 (1):104. <https://doi.org/10.1186/1471-213X-7-104>.
- Laget, S, Miotto, B, Chin, HG, Estève, PO, Roberts, RJ, Pradhan, S, et Defossez, PA. 2014. « MBD4 Cooperates with DNMT1 to Mediate Methyl-DNA Repression and Protects Mammalian Cells from Oxidative Stress ». *Epigenetics* 9 (4):546-56. <https://doi.org/10.4161/epi.27695>.
- Lander, ES, Linton, LM, Birren, B, Nusbaum, C, Zody, MC, Baldwin, J, Devon, K, et al. 2001. « Initial Sequencing and Analysis of the Human Genome ». *Nature* 409 (6822):860-921. <https://doi.org/10.1038/35057062>.
- Langmead, B, Trapnell, C, Pop, M, et Salzberg, SL. 2009. « Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome ». *Genome Biology* 10 (3):R25. <https://doi.org/10.1186/gb-2009-10-3-r25>.
- Larson, JL, et Miller, DJ. 2000. « Can Relative Spermatozoal Galactosyltransferase Activity Be Predictive of Dairy Bull Fertility ? » *Journal of Dairy Science* 83 (11):2473-79. [https://doi.org/10.3168/jds.S0022-0302\(00\)75139-3](https://doi.org/10.3168/jds.S0022-0302(00)75139-3).

- Laurent, L, Wong, E, Li, G, Huynh, T, Tsirigos, A, Ong, CT, Low, HM, et al. 2010. « Dynamic Changes in the Human Methylome during Differentiation ». *Genome Research* 20 (3):320-31. <https://doi.org/10.1101/gr.101907.109>.
- Lauvie, A, et Couix, N, 2012. « Diversité des formes de valorisation des populations animales locales et gestion des ressources génétiques animales ». *INRA Prod. Anim.*, 25, 5, 431-440.
- Leblond, CP, et Clermont, Y. 1952. « Definition of the Stages of the Cycle of the Seminiferous Epithelium in the Rat ». *Annals of the New York Academy of Sciences* 55 (4):548-73.
- Lees-Murdock, DJ, De Felici, M, et Walsh, CP. 2003. « Methylation Dynamics of Repetitive DNA Elements in the Mouse Germ Cell Lineage ». *Genomics* 82 (2):230-37. [https://doi.org/10.1016/S0888-7543\(03\)00105-8](https://doi.org/10.1016/S0888-7543(03)00105-8).
- Le Mezec, P, Barbat-Leterrier, A, Barbier, S, Gion, A, et Ponsart, C 2010. « La fertilité des vaches laitières en 2008 : des perturbations mais des espoirs côté génétique ». Institut de l'Élevage.
- Le Mezec, P, Verrier, E. 2009. « Les programmes de sélection d'hier à aujourd'hui : efficacité, limites, coûts et délais ». Journée CSAGAD / Institut de l'Élevage.
- Leonhardt, H, Page, AW, Weier, HU, et Bestor, TH. 1992. « A Targeting Sequence Directs DNA Methyltransferase to Sites of DNA Replication in Mammalian Nuclei ». *Cell* 71 (5):865-73.
- Leontiou, CA, Hadjidaniel, MD, Mina, P, Antoniou, P, Ioannides, M, et Patsalis, PC. 2015. « Bisulfite Conversion of DNA: Performance Comparison of Different Kits and Methylation Quantitation of Epigenetic Biomarkers That Have the Potential to Be Used in Non-Invasive Prenatal Testing ». Édité par Osman El-Maarri. *PLOS ONE* 10 (8):e0135058. <https://doi.org/10.1371/journal.pone.0135058>.
- Lesch, BJ, Dokshin, GA, Young, RA, McCarrey, JR, et Page, DC. 2013. « A Set of Genes Critical to Development Is Epigenetically Poised in Mouse Germ Cells from Fetal Stages through Completion of Meiosis ». *Proceedings of the National Academy of Sciences* 110 (40):16061-66. <https://doi.org/10.1073/pnas.1315204110>.
- Lewis, SEM, et Aitken, RJ. 2005. « DNA Damage to Spermatozoa Has Impacts on Fertilization and Pregnancy ». *Cell and Tissue Research* 322 (1):33-41. <https://doi.org/10.1007/s00441-005-1097-5>.
- Li, LC, et Dahiya, R. 2002. « MethPrimer: Designing Primers for Methylation PCRs ». *Bioinformatics (Oxford, England)* 18 (11):1427-31.
- Li, N, Ye, M, Li, Y, Yan, Z, Butcher, LM, Sun, J, Han, X, Chen, Q, Zhang, X, et Wang, J. 2010. « Whole Genome DNA Methylation Analysis Based on High Throughput Sequencing Technology ». *Methods* 52 (3):203-12. <https://doi.org/10.1016/j.ymeth.2010.04.009>.
- Li, X, Ito, M, Zhou, F, Youngson, N, Zuo, X, Leder, P, et Ferguson-Smith, AC. 2008. « A Maternal-Zygotic Effect Gene, Zfp57, Maintains Both Maternal and Paternal Imprints ». *Developmental Cell* 15 (4):547-57. <https://doi.org/10.1016/j.devcel.2008.08.014>.
- Li, XZ, Roy, CK, Dong, X, Bolcun-Filas, E, Wang, J, Han, BW, Xu, J, et al. 2013. « An Ancient Transcription Factor Initiates the Burst of PiRNA Production during Early Meiosis in Mouse Testes ». *Molecular Cell* 50 (1):67-81. <https://doi.org/10.1016/j.molcel.2013.02.016>.
- Li, Y, Lin, H, Li, Y, et Cao, J. 2011. « Association between Socio-Psycho-Behavioral Factors and Male Semen Quality: Systematic Review and Meta-Analyses ». *Fertility and Sterility* 95 (1):116-23. <https://doi.org/10.1016/j.fertnstert.2010.06.031>.
- Li, C, Guo, S, Gao, J, Guo, Y, Du, E, Lv, Z, Zhang, B. 2015. « Maternal high-zinc diet attenuates intestinal inflammation by reducing DNA methylation and elevating H3K9 acetylation in the A20 promoter of offspring chicks ». *Journal of Nutritional Biochemistry* 26 173-183.
- Lidegaard, Ø, Pinborg, A, et Andersen, AN. 2005. « Imprinting Diseases and IVF: Danish National IVF Cohort Study ». *Human Reproduction* 20 (4):950-54. <https://doi.org/10.1093/humrep/deh714>.
- Lisanti, S, Omar, WA, Tomaszewski, B, De Prins, S, Jacobs, G, Koppen, G, Mathers, JC, et Langie, SA. 2013. « Comparison of Methods for Quantification of Global DNA Methylation in Human Cells and Tissues ». Édité par Keith William Brown. *PLoS ONE* 8 (11):e79044. <https://doi.org/10.1371/journal.pone.0079044>.
- Lister, R, Mukamel, EA, Nery, JR, Urich, M, Puddifoot, CA, Johnson, ND, Lucero, J, et al. 2013. « Global epigenomic reconfiguration during mammalian brain development ». *Science* 341 (6146):1237905.
- Lister, R, Pelizzola, M, Downen, RH, Hawkins, RD, Hon, G, Tonti-Filippini, J, Nery, JR, et al. 2009. « Human DNA methylomes at base resolution show widespread epigenomic differences ». *Nature* 462 (7271):315-22. <https://doi.org/10.1038/nature08514>.
- Liu, DY, Lopata, A, Johnston, WI, et Baker, HW. 1988. « A Human Sperm-Zona Pellucida Binding Test Using Oocytes That Failed to Fertilize in Vitro ». *Fertility and Sterility* 50 (5):782-88.
- Liu, Z, Li, Q, Pan, Z, Qu, X, Zhang, C, et Xie, Z. 2011. « Comparative Analysis on mRNA Expression Level and Methylation Status of DAZL Gene between Cattle-Yaks and Their Parents ». *Animal Reproduction Science* 126 (3-4):258-64. <https://doi.org/10.1016/j.anireprosci.2011.05.013>.
- Locatelli, Y, 2012. « Creation of frozen arks for mammalian species conservation ». *Ciência Animal*, 22 (1), 106-123.
- Longo, L, Camacho Vanegas, O, Patel, M, Rosti, V, Li, H, Waka, J, Merghoub, T, et al. 2002. « Maternally Transmitted Severe Glucose 6-Phosphate Dehydrogenase Deficiency Is an Embryonic Lethal ». *The EMBO Journal* 21 (16):4229-39.
- López-Gatiús, F. 2003. « Is Fertility Declining in Dairy Cattle? ». *Theriogenology* 60 (1):89-99. [https://doi.org/10.1016/S0093-691X\(02\)01359-6](https://doi.org/10.1016/S0093-691X(02)01359-6).
- Lucio, AC, Alves, BG, Alves, KA, Martins, MC, Braga, LS, Miglio, L, Alves, BG, Silva, TH, Jacomini, JO, et Beletti, ME. 2016. « Selected Sperm Traits Are Simultaneously Altered after Scrotal Heat Stress and Play Specific Roles in in Vitro Fertilization and Embryonic Development ». *Theriogenology* 86 (4):924-33. <https://doi.org/10.1016/j.theriogenology.2016.03.015>.
- Lucy, MC. 2001. « Reproductive Loss in High-Producing Dairy Cattle: Where Will It End? ». *Journal of Dairy Science* 84 (6):1277-93. [https://doi.org/10.3168/jds.S0022-0302\(01\)70158-0](https://doi.org/10.3168/jds.S0022-0302(01)70158-0).
- Luense, LJ, Carletti, MZ, et Christenson, LK. 2009. « Role of Dicer in Female Fertility ». *Trends Endocrinol Metab.* 2009 August ; 20(6): 265-272. doi:10.1016/j.tem.2009.05.001.

- Lunstra, DD, et Coulter, GH. 1997. « Relationship between Scrotal Infrared Temperature Patterns and Natural-Mating Fertility in Beef Bulls ». *Journal of Animal Science* 75 (3):767-74.
- Lunstra, DD, Ford, JJ, et Echternkamp, SE. 1978. « Puberty in Beef Bulls : Hormone Concentrations, Growth, Testicular Development, Sperm Production and Sexual Aggressiveness in Bulls of Different Breeds ». *Journal of Animal Science* 46 (4):1054-62.
- Maatouk, DM. 2006. « DNA Methylation Is a Primary Mechanism for Silencing Postmigratory Primordial Germ Cell Genes in Both Germ Cell and Somatic Cell Lineages ». *Development* 133 (17):3411-18. <https://doi.org/10.1242/dev.02500>.
- MacMillan, KL, Lean, IJ, et Westwood, CT. 1996. « The Effects of Lactation on the Fertility of Dairy Cows ». *Australian Veterinary Journal* 73 (4):141-47.
- MacMillan, KL, et Watson, JD. 1975. « Fertility Differences between Groups of Sires Relative to the Stage of Oestrus at the Time of Insemination ». *Animal Production* 21 (03):243-49. <https://doi.org/10.1017/S0003356100030701>.
- Magee, DA, Spillane, C, Berkowicz, EW, Sikora, KM, et MacHugh, DE. 2014. « Imprinted loci in domestic livestock species as epigenomic targets for artificial selection of complex traits ». *Stichting International Foundation for Animal Genetics*, 45 (Suppl. 1), 25–39 doi: 10.1111/age.12168.
- Maier, WM, Nussbaum, G, Domenjoud, L, Klemm, U, et Engel, W. 1990. « The Lack of Protamine 2 (P2) in Boar and Bull Spermatozoa Is Due to Mutations within the P2 Gene ». *Nucleic Acids Research* 18 (5):1249-54.
- Malo, AF, Gomendio, M, Garde, J, Lang-Lenton, B, Soler, AJ, et Roldan, ER. 2006. « Sperm Design and Sperm Function ». *Biology Letters* 2 (2):246-49. <https://doi.org/10.1098/rsbl.2006.0449>.
- Mamo, S, Mehta, JP, McGettigan, P, Fair, T, Spencer, TE, Bazer, FW, et Lonergan, P. 2011. « RNA Sequencing Reveals Novel Gene Clusters in Bovine Conceptuses Associated with Maternal Recognition of Pregnancy and Implantation 1 ». *Biology of Reproduction* 85 (6):1143-51. <https://doi.org/10.1095/biolreprod.111.092643>.
- Mansour, AA, Gafni, O, Weinberger, L, Zviran, A, Ayyash, M, Rais, Y, Krupalnik, V, et al. 2012. « The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming ». *Nature* 488 (7411):409-13. <https://doi.org/10.1038/nature11272>.
- Marques, CJ, Costa, P, Vaz, B, Carvalho, F, Fernandes, S, Barros, A, et Sousa, M. 2008. « Abnormal Methylation of Imprinted Genes in Human Sperm Is Associated with Oligozoospermia ». *MHR: Basic Science of Reproductive Medicine* 14 (2):67-74. <https://doi.org/10.1093/molehr/gam093>.
- Marques, CJ, Pinho, MJ, Carvalho, J, Bièche, I, Barros, A, et Sousa, M. 2011. « DNA Methylation Imprinting Marks and DNA Methyltransferase Expression in Human Spermatogenic Cell Stages ». *Epigenetics* 6 (11):1354-61. <https://doi.org/10.4161/epi.6.11.17993>.
- Martinez-Arguelles, DB, Lee, S, et Papadopoulos, V. 2014. « In Silico Analysis Identifies Novel Restriction Enzyme Combinations That Expand Reduced Representation Bisulfite Sequencing CpG Coverage ». *BMC Research Notes* 7 (août):534. <https://doi.org/10.1186/1756-0500-7-534>.
- Matoba, S, Liu, Y, Lu, F, Iwabuchi, KA, Shen, L, Inoue, A, et Zhang, Y. 2014. « Embryonic Development Following Somatic Cell Nuclear Transfer Impeded by Persisting Histone Methylation ». *Cell* 159 (4):884-95. <https://doi.org/10.1016/j.cell.2014.09.055>.
- Matzuk, MM, Lamb, DJ. 2008. « The biology of infertility: research advances and clinical challenges ». *Nat Med.* 2008 November; 14(11): 1197–1213. doi:10.1038/nm.f.1895.
- Maunakea, AK, Chepelev, I, Cui, K, et Zhao, K. 2013. « Intragenic DNA Methylation Modulates Alternative Splicing by Recruiting MeCP2 to Promote Exon Recognition ». *Cell Research* 23 (11):1256-69. <https://doi.org/10.1038/cr.2013.110>.
- Maunakea, AK, Nagarajan, RP, Bilenky, M, Ballinger, TJ, D'Souza, C, Fouse, SH, Johnson, BE, et al. 2010. « Conserved role of intragenic DNA methylation in regulating alternative promoters ». *Nature* 466 (7303):253-57. <https://doi.org/10.1038/nature09165>.
- Mc Parland, S, Kearney, JF, Rath, M, et Berry, DP. 2007. « Inbreeding Trends and Pedigree Analysis of Irish Dairy and Beef Cattle Populations ». *Journal of Animal Science* 85 (2):322-31. <https://doi.org/10.2527/jas.2006-367>.
- McGrath, J, et Solter, D. 1984. « Completion of Mouse Embryogenesis Requires Both the Maternal and Paternal Genomes ». *Cell* 37 (1):179-83.
- Meissner, A. 2005. « Reduced Representation Bisulfite Sequencing for Comparative High-Resolution DNA Methylation Analysis ». *Nucleic Acids Research* 33 (18):5868-77. <https://doi.org/10.1093/nar/gki901>.
- Meuwissen, TH, Hayes, BJ, et Goddard, ME. 2001. « Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps ». *Genetics* 157 (4):1819-29.
- Michot, P, Fritz, S, Barbat, A, Boussaha, M, Deloche, MC, Grohs, C, Hoze, C, et al. 2017. « A Missense Mutation in PFAS (Phosphoribosylformylglycinamide Synthase) Is Likely Causal for Embryonic Lethality Associated with the MH1 Haplotype in Montbéliarde Dairy Cattle ». *Journal of Dairy Science* 100 (10):8176-87. <https://doi.org/10.3168/jds.2017-12579>.
- Mikkelsen, TS, Ku, M, Jaffe, DB, Issac, B, Lieberman, E, Giannoukos, G, Alvarez, P, et al. 2007. « Genome-wide maps of chromatin state in pluripotent and lineage-committed cells ». *Nature* 448 (7153):553-60. <https://doi.org/10.1038/nature06008>.
- Mohn, F, et Schübeler, D. 2009. « Genetics and Epigenetics: Stability and Plasticity during Cellular Differentiation ». *Trends in Genetics* 25 (3):129-36. <https://doi.org/10.1016/j.tig.2008.12.005>.
- Molaro, A, Hodges, E, Fang, F, Song, Q, McCombie, WR, Hannon, GJ, et Smith, AD. 2011. « Sperm Methylation Profiles Reveal Features of Epigenetic Inheritance and Evolution in Primates ». *Cell* 146 (6):1029-41. <https://doi.org/10.1016/j.cell.2011.08.016>.
- Moore, LD, Le, T, et Fan, G. 2012. « DNA methylation and its basic function ». *Neuropsychopharmacology* 38 (1):23–38.
- Mortusewicz, Oliver, Lothar Schermelleh, Joachim Walter, M. Cristina Cardoso, et Heinrich Leonhardt. 2005. « Recruitment of DNA Methyltransferase I to DNA Repair Sites ». *Proceedings of the National Academy of Sciences of the United States of America* 102 (25):8905-9. <https://doi.org/10.1073/pnas.0501034102>.

- Moura, AA. 2006a. « Identification of Proteins in the Accessory Sex Gland Fluid Associated With Fertility Indexes of Dairy Bulls: A Proteomic Approach ». *Journal of Andrology* 27 (2):201-11. <https://doi.org/10.2164/jandrol.05089>.
- Moura, AA. 2006b. « Proteins of the Cauda Epididymal Fluid Associated With Fertility of Mature Dairy Bulls ». *Journal of Andrology* 27 (4):534-41. <https://doi.org/10.2164/jandrol.05201>.
- Murphy, SK, Adigun, A, Huang, Z, Overcash, F, Wang, F, Jirtle, RL, Schildkraut, JM, Murtha, AP, Iversen, ES, et Hoyo, C. 2012. « Gender-Specific Methylation Differences in Relation to Prenatal Exposure to Cigarette Smoke ». *Gene* 494 (1):36-43. <https://doi.org/10.1016/j.gene.2011.11.062>.
- Muthusami, K, et Chinnaswamy, P. 2005. « Effect of Chronic Alcoholism on Male Fertility Hormones and Semen Quality ». *Fertility and Sterility* 84 (4):919-24. <https://doi.org/10.1016/j.fertnstert.2005.04.025>.
- Nair, SS, Coolen, MW, Stirzaker, C, Song, JZ, Statham, AL, Strbenac, D, Robinson, MD, et Clark, SJ. 2011. « Comparison of Methyl-DNA Immunoprecipitation (MeDIP) and Methyl-CpG Binding Domain (MBD) Protein Capture for Genome-Wide DNA Methylation Analysis Reveal CpG Sequence Coverage Bias ». *Epigenetics* 6 (1):34-44. <https://doi.org/10.4161/epi.6.1.13313>.
- Nakamura, T, Arai, Y, Umehara, H, Masuhara, M, Kimura, T, Taniguchi, H, Sekimoto, T, et al. 2007. « PGC7/Stella protects against DNA demethylation in early embryogenesis ». *Nature Cell Biology* 9 (1):64-71. <https://doi.org/10.1038/ncb1519>.
- Nakamura, T, Liu, YJ, Nakashima, H, Umehara, H, Inoue, Matoba, S, Tachibana, M, Ogura, A, Shinkai, Y, et Nakano, T. 2012. « PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos ». *Nature*, juin. <https://doi.org/10.1038/nature11093>.
- Nan, X, Hou, J, Maclean, A, Nasir, J, Lafuente, MJ, Shu, X, Kriaucionis, S, et Bird, A. 2007. « Interaction between Chromatin Proteins MECP2 and ATRX Is Disrupted by Mutations That Cause Inherited Mental Retardation ». *Proceedings of the National Academy of Sciences of the United States of America* 104 (8):2709-14. <https://doi.org/10.1073/pnas.0608056104>.
- Nanassy, L, et Carrell, DT. 2011a. « Analysis of the Methylation Pattern of Six Gene Promoters in Sperm of Men with Abnormal Protamination ». *Asian Journal of Andrology* 13 (2):342-46. <https://doi.org/10.1038/aja.2010.160>.
- Nanassay, L, et Carrell, DT. 2011b. « Abnormal Methylation of the Promoter of CREM Is Broadly Associated with Male Factor Infertility and Poor Sperm Quality but Is Improved in Sperm Selected by Density Gradient Centrifugation ». *Fertility and Sterility* 95 (7):2310-14. <https://doi.org/10.1016/j.fertnstert.2011.03.096>.
- Nascimento, AV, Matos, MC, Seno, LO, Romero, AR, Garcia, JF, et Grisolia, AB. 2016. « Genome Wide Association Study on Early Puberty in Bos Indicus ». *Genetics and Molecular Research: GMR* 15 (1). <https://doi.org/10.4238/gmr.15017548>.
- Navarro-Costa, P, Nogueira, P, Carvalho, M, Leal, F, Cordeiro, I, Calhaz-Jorge, C, Goncalves, J, et Plancha, CE. 2010. « Incorrect DNA Methylation of the DAZL Promoter CpG Island Associates with Defective Human Sperm ». *Human Reproduction* 25 (10):2647-54. <https://doi.org/10.1093/humrep/deq200>.
- Nebendahla, C, Görsa, S, Albrecht, E, Krügera, R, Martensa, K, Gillerc, K, Hammona, HM, Rimbachc, G, Metgesa, CC. 2016. « Early postnatal feed restriction reduces liver connective tissue levels and affects H3K9 acetylation state of regulated genes associated with protein metabolism in low birth weight pigs ». *Journal of Nutritional Biochemistry* 29 41–55
- Nestor, CE, Ruzov, A, Meehan, R, et Dunican, D. 2010. « Enzymatic Approaches and Bisulfite Sequencing Cannot Distinguish between 5-Methylcytosine and 5-Hydroxymethylcytosine in DNA ». *BioTechniques* 48 (4):317-19. <https://doi.org/10.2144/000113403>.
- Nishiyama, A, Yamaguchi, L, Sharif, J, Johmura, Y, Kawamura, T, Nakanishi, K, Shimamura, S, et al. 2013. « Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication ». *Nature* 502 (7470):249-53. <https://doi.org/10.1038/nature12488>.
- Nishiyama, A, Yamaguchi, L, et Nakanishi, M. 2015. « Regulation of maintenance DNA methylation via histone ubiquitylation ». *J Biochem*. 2016 Jan;159(1):9-15. doi: 10.1093/jb/mvv113.
- Noh, KM, Wang, H, Kim, HR, Wenderski, W, Fang, F, Li, CH, Dewell, S, et al. 2015. « Engineering of a Histone-Recognition Domain in Dnmt3a Alters the Epigenetic Landscape and Phenotypic Features of Mouse ESCs ». *Molecular Cell* 59 (1):89-103. <https://doi.org/10.1016/j.molcel.2015.05.017>.
- Norman, HD, Hutchison, JL, et Miller, RH. 2010. « Use of Sexed Semen and Its Effect on Conception Rate, Calf Sex, Dystocia, and Stillbirth of Holsteins in the United States ». *Journal of Dairy Science* 93 (8):3880-90. <https://doi.org/10.3168/jds.2009-2781>.
- Norman, HD, Wright, JR, Hubbard, SM, Miller, RH, et Hutchison, JL. 2009. « Reproductive Status of Holstein and Jersey Cows in the United States ». *Journal of Dairy Science* 92 (7):3517-28. <https://doi.org/10.3168/jds.2008-1768>.
- Oakes, CC, La Salle, S, Smiraglia, DJ, Robaire, B, et Trasler, JM. 2007. « Developmental Acquisition of Genome-Wide DNA Methylation Occurs Prior to Meiosis in Male Germ Cells ». *Developmental Biology* 307 (2):368-79. <https://doi.org/10.1016/j.ydbio.2007.05.002>.
- Oh, HJ, Lee, TH, Lee, JH, et Lee, BC. 2012. « Trichostatin a Improves Preimplantation Development of Bovine Cloned Embryos and Alters Expression of Epigenetic and Pluripotency Genes in Cloned Blastocysts ». *The Journal of Veterinary Medical Science* 74 (11):1409-15.
- Ohno, R, Nakayama, M, Naruse, C, Okashita, N, Takano, O, Tachibana, M, Asano, M, Saitou, M, et Seki, Y. 2013. « A Replication-Dependent Passive Mechanism Modulates DNA Demethylation in Mouse Primordial Germ Cells ». *Development* 140 (14):2892-2903. <https://doi.org/10.1242/dev.093229>.
- Okamoto, I, Patrat, C, Thépot, D, Peynot, N, Fauque, P, Daniel, N, Diabangouaya, P, et al. 2011. « Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development ». *Nature* 472 (7343):370-74. <https://doi.org/10.1038/nature09872>.
- Okamoto, Y, Yoshida, N, Suzuki, T, Shimozawa, N, Asami, M, Matsuda, T, Kojima, N, Perry, AC, et Takada, T. 2016. « DNA Methylation Dynamics in Mouse Preimplantation Embryos Revealed by Mass Spectrometry ». *Scientific Reports* 6 (1). <https://doi.org/10.1038/srep19134>.

- Okamura, N, Tajima, Y, Soejima, A, Masuda, H, et Sugita, Y. 1985. « Sodium Bicarbonate in Seminal Plasma Stimulates the Motility of Mammalian Spermatozoa through Direct Activation of Adenylate Cyclase ». *The Journal of Biological Chemistry* 260 (17):9699-9705.
- Okano, M, Bell, DW, Haber, DA, et Li, E. 1999. « DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for de Novo Methylation and Mammalian Development ». *Cell* 99 (3):247-57.
- Oldenbroek, K, Van der Waaij, L. 2015. « Textbook Animal Breeding and Genetics for BSc students ». Centre for Genetic Resources The Netherlands and Animal Breeding and Genomics Centre, Groen Kennisnet: <https://wiki.groenkennisnet.nl/display/TAB/>.
- Olkhov-Mitsel, E, et Bapat, B. 2012. « Strategies for Discovery and Validation of Methylated and Hydroxymethylated DNA Biomarkers ». *Cancer Medicine* 1 (2):237-60. <https://doi.org/10.1002/cam4.22>.
- Ooi, SK, Qiu, C, Bernstein, E, Li, K, Jia, D, Yang, Z, Erdjument-Bromage, H, et al. 2007. « DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA ». *Nature* 448 (7154):714-17. <https://doi.org/10.1038/nature05987>.
- Orgebin, MC. 1961. « Etude du transit épидидymaire des spermatozoïdes de taureau marqués à l'aide du 32P ». *Annales de biologie animale, biochimie, biophysique* 1 (2):117-20.
- Orth, JM. 1993. « Cell biology of testicular development in the fetus and neonate ». In: Desjardins C, Ewing LL (Eds). *Cell and Molecular Biology of the Testis*. Oxford Uni. Press, Oxford.
- Pacheco, SE, Houseman, EA, Christensen, BC, Marsit, CJ, Kelsey, KT, Sigman, M, et Boekelheide, K. 2011. « Integrative DNA Methylation and Gene Expression Analyses Identify DNA Packaging and Epigenetic Regulatory Genes Associated with Low Motility Sperm ». Édité par Brian P. Chadwick. *PLoS ONE* 6 (6):e20280. <https://doi.org/10.1371/journal.pone.0020280>.
- Palidwor, GA, Andrade-Navarro, MA. 2010. « MLTrends : Graphing MEDLINE term usage over time ». *Journal of Biomedical Discovery and Collaboration* (2010) 5:1-6.
- Parkinson, TJ. 2004. « Evaluation of Fertility and Infertility in Natural Service Bulls ». *The Veterinary Journal* 168 (3):215-29. <https://doi.org/10.1016/j.tvjl.2003.10.017>.
- Pfaffeneder, T, Hackner, B, Truß, M, Münzel, M, Müller, M, Deiml, CA, Hagemeyer, C, et Carell, T. 2011. « The Discovery of 5-Formylcytosine in Embryonic Stem Cell DNA ». *Angewandte Chemie International Edition* 50 (31):7008-12. <https://doi.org/10.1002/anie.201103899>.
- Pilsner, JR, Hall, MN, Liu, X, Ilievski, V, Slavkovich, V, Levy, D, Factor-Litvak, P, et al. 2012. « Influence of Prenatal Arsenic Exposure and Newborn Sex on Global Methylation of Cord Blood DNA ». Édité par Cees Oudejans. *PLoS ONE* 7 (5):e37147. <https://doi.org/10.1371/journal.pone.0037147>.
- Pilsner, JR, Liu, X, Ahsan, H, Ilievski, V, Slavkovich, V, Levy, D, Factor-Litvak, P, Graziano, JH, et Gamble, MV. 2007. « Genomic Methylation of Peripheral Blood Leukocyte DNA: Influences of Arsenic and Folate in Bangladeshi Adults ». *The American Journal of Clinical Nutrition* 86 (4):1179-86.
- Ponger, L, et Mouchiroud, D. 2002. « CpGProD: identifying CpG islands associated with transcription start sites in large genomic mammalian sequences ». Availability: <http://pbil.univ-lyon1.fr/software/cpgprod.html>.
- Poplinski, A, Tüttelmann, F, Kanber, D, Horsthemke, B, et Gromoll, J. 2009. « Idiopathic Male Infertility Is Strongly Associated with Aberrant Methylation of *MEST* and *IGF2/H19 ICR1* ». *International Journal of Andrology*, octobre. <https://doi.org/10.1111/j.1365-2605.2009.01000.x>.
- Popp, C, Dean, W, Feng, S, Cokus, SJ, Andrews, S, Pellegrini, M, Jacobsen, SE, et Reik, W. 2010. « Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency ». *Nature* 463 (7284):1101-5. <https://doi.org/10.1038/nature08829>.
- Prentki Santos, E, López-Costa, S, Chenlo, P, Pugliese, MN, Curi, S, Ariagno, J, Repetto, H, Sardi, M, Palaoro, L, et Mendeluk, G. 2011. « Impact of Spontaneous Smoking Cessation on Sperm Quality: Case Report: Smoking Cessation and Sperm Quality ». *Andrologia* 43 (6):431-35. <https://doi.org/10.1111/j.1439-0272.2010.01089.x>.
- Probst, AV, Okamoto, I, Casanova, M, El Marjou, F, Le Baccon, P, et Almouzni, G. 2010. « A Strand-Specific Burst in Transcription of Pericentric Satellites Is Required for Chromocenter Formation and Early Mouse Development ». *Developmental Cell* 19 (4):625-38. <https://doi.org/10.1016/j.devcel.2010.09.002>.
- Pronina, IV, Loginov, VI, Burdenny, AM, Fridman, MV, Kazubskaya, TP, Dmitriev, AA, et Braga, EA. 2016. « Expression and DNA Methylation Alterations of Seven Cancer-Associated 3p Genes and Their Predicted Regulator MiRNAs (MiR-129-2, MiR-9-1) in Breast and Ovarian Cancers ». *Gene* 576 (1):483-91. <https://doi.org/10.1016/j.gene.2015.10.059>.
- Puumala, SE, Nelson, HH, Ross, JA, Nguyen, RH, Damaro, MA, et Spector, LG. 2012. « Similar DNA Methylation Levels in Specific Imprinting Control Regions in Children Conceived with and without Assisted Reproductive Technology : A Cross-Sectional Study ». *BMC Pediatrics* 12 (mars):33. <https://doi.org/10.1186/1471-2431-12-33>.
- Qu, X, Davison, J, Du, L, Storer, B, Stirewalt, DL, Heimfeld, S, Estey, E, Appelbaum, FR, et Fang, M. 2015. « Identification of Differentially Methylated Markers among Cytogenetic Risk Groups of Acute Myeloid Leukemia ». *Epigenetics* 10 (6):526-35. <https://doi.org/10.1080/15592294.2015.1048060>.
- Quenneville, S, Verde, G, Corsinotti, A, Kapopoulou, A, Jakobsson, J, Offner, S, Baglivo, I, et al. 2011. « In Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions ». *Molecular Cell* 44 (3):361-72. <https://doi.org/10.1016/j.molcel.2011.08.032>.
- Ramsahoye, BH, Biniszkiwicz, D, Lyko, F, Clark, V, Bird, et Jaenisch, R. 2000. « Non-CpG Methylation Is Prevalent in Embryonic Stem Cells and May Be Mediated by DNA Methyltransferase 3a ». *Proceedings of the National Academy of Sciences of the United States of America* 97 (10):5237-42.
- Rawlings, N, Evans, A, Chandolia, RK, et Bagu, ET. 2008. « Sexual Maturation in the Bull ». *Reproduction in Domestic Animals* 43 (juillet):295-301. <https://doi.org/10.1111/j.1439-0531.2008.01177.x>.
- Reis e Silva, AR, Bruno, C, Fleurot, R, Daniel, N, Archilla, C, Peynot, N, Lucci, CM, Beaujean, N, et Duranthon, V. 2012. « Alteration of DNA Demethylation Dynamics by in Vitro Culture Conditions in Rabbit Pre-Implantation Embryos ». *Epigenetics* 7 (5):440-46. <https://doi.org/10.4161/epi.19563>.

- Rhind, SM, Taylor, JE, De Sousa, PA, King, TJ, McGarry, M, Wilmut, I. 2003. « Human cloning: can it be made safe? » *Nat Rev Genet.* 2003 Nov;4(11):855-64.
- Rice, P. 2000. « EMBOSS: The European Molecular Biology Open Software Suite ». *TIG* June 2000, volume 16, No. 6.
- Robinson, MD, Storzaker, C, Statham, AL, Coolen, MW, Song, JZ, Nair, SS, Strbenac, D, Speed, TP, et Clark, SJ. 2010. « Evaluation of Affinity-Based Genome-Wide DNA Methylation Data: Effects of CpG Density, Amplification Bias, and Copy Number Variation ». *Genome Research* 20 (12):1719-29. <https://doi.org/10.1101/gr.110601.110>.
- Roche, JF, Mackey, D, et Diskin, MD. 2000. « Reproductive Management of Postpartum Cows ». *Animal Reproduction Science* 60-61 (juillet):703-12.
- Roelofs, J, López-Gatius, F, Hunter, RH, Van Eerdenburg, FJ, et Hanzen, CH. 2010. « When Is a Cow in Estrus? Clinical and Practical Aspects ». *Theriogenology* 74 (3):327-44. <https://doi.org/10.1016/j.theriogenology.2010.02.016>.
- Ronaghi, M, Karamohamed, S, Pettersson, B, Uhlén, M, et Nyrén, P. 1996. « Real-Time DNA Sequencing Using Detection of Pyrophosphate Release ». *Analytical Biochemistry* 242 (1):84-89. <https://doi.org/10.1006/abio.1996.0432>.
- Royal, M, Mann, GE, et Flint, AP. 2000. « Strategies for Reversing the Trend Towards Subfertility in Dairy Cattle ». *The Veterinary Journal* 160 (1):53-60. <https://doi.org/10.1053/tvjl.1999.0450>.
- Rubes, J, Selevan, SG, Evenson, DP, Zudova, D, Vozdova, M, Zudova, Z, Robbins, WA, et Perreault, SD. 2005. « Episodic Air Pollution Is Associated with Increased DNA Fragmentation in Human Sperm without Other Changes in Semen Quality ». *Human Reproduction* 20 (10):2776-83. <https://doi.org/10.1093/humrep/dei122>.
- Ryan, DP, Prichard, JF, Kopel, E, et Godke, RA. 1993. « Comparing Early Embryo Mortality in Dairy Cows during Hot and Cool Seasons of the Year ». *Theriogenology* 39 (3):719-37.
- Saacke, RG. 2008. « Sperm Morphology: Its Relevance to Compensable and Uncompensable Traits in Semen ». *Theriogenology* 70 (3):473-78. <https://doi.org/10.1016/j.theriogenology.2008.04.012>.
- Sachs, M, Onodera, C, Blaschke, K, Ebata, KT, Song, JS, et Ramalho-Santos, M. 2013. « Bivalent Chromatin Marks Developmental Regulatory Genes in the Mouse Embryonic Germline In Vivo ». *Cell Reports* 3 (6):1777-84. <https://doi.org/10.1016/j.celrep.2013.04.032>.
- Sadakerska-Chudy, A, et Filip, M. 2015. « A Comprehensive View of the Epigenetic Landscape. Part II: Histone Post-Translational Modification, Nucleosome Level, and Chromatin Regulation by NcRNAs ». *Neurotoxicity Research* 27 (2):172-97. <https://doi.org/10.1007/s12640-014-9508-6>.
- Sadakerska-Chudy, A, Kostrzewa, RM, et Filip, M. 2014. « A Comprehensive View of the Epigenetic Landscape Part I: DNA Methylation, Passive and Active DNA Demethylation Pathways and Histone Variants ». *Neurotoxicity Research*, novembre. <https://doi.org/10.1007/s12640-014-9497-5>.
- Sailer, BL, Jost, LK, Erickson, KR, Tajiran, MA, et Evenson, DP. 1995. « Effects of X-Irradiation on Mouse Testicular Cells and Sperm Chromatin Structure ». *Environmental and Molecular Mutagenesis* 25 (1):23-30.
- Saito, Y, et Jones, PM. 2006. « Epigenetic Activation of Tumor Suppressor MicroRNAs in Human Cancer Cells ». *Cell Cycle* 5 (19):2220-22. <https://doi.org/10.4161/cc.5.19.3340>.
- Sakaguchi, H, Muramatsu, H, Okuno, Y, Makishima, H, Xu, Y, Furukawa-Hibi, Y, Wang, X, et al. 2015. « Aberrant DNA Methylation Is Associated with a Poor Outcome in Juvenile Myelomonocytic Leukemia ». *PloS One* 10 (12):e0145394. <https://doi.org/10.1371/journal.pone.0145394>.
- Sakai, Y, Suetake, I, Shinozaki, F, Yamashina, S, et Tajima, S. 2004. « Co-Expression of de Novo DNA Methyltransferases Dnmt3a2 and Dnmt3L in Gonocytes of Mouse Embryos ». *Gene Expression Patterns* 5 (2):231-37. <https://doi.org/10.1016/j.modgep.2004.07.011>.
- Salisbury, GW, VanDemark, NL. 1961. « Significance of semen quality ». In: *Physiology of reproduction and artificial insemination in cattle*. 1st edition. WH Freeman and Co., San Francisco (CA); 359–379.
- Samans, B, Yang, Y, Krebs, S, Sarode, GV, Blum, H, Reichenbach, M, Wolf, E, Steger, K, Dansranjav, T, et Schagdarsurengin, U. 2014. « Uniformity of Nucleosome Preservation Pattern in Mammalian Sperm and Its Connection to Repetitive DNA Elements ». *Developmental Cell* 30 (1):23-35. <https://doi.org/10.1016/j.devcel.2014.05.023>.
- Sanchez-Albisua, I, Borell-Kost, S, Mau-Holzmann, UA, Licht, P, et Krägeloh-Mann, I. 2007. « Increased Frequency of Severe Major Anomalies in Children Conceived by Intracytoplasmic Sperm Injection ». *Developmental Medicine and Child Neurology* 49 (2):129-34. <https://doi.org/10.1111/j.1469-8749.2007.00129.x>.
- Sánchez-Peña, LC, Reyes, BE, López-Carrillo, L, Recio, R, Morán-Martínez, J, Cebrián, ME, et Quintanilla-Vega, B. 2004. « Organophosphorous Pesticide Exposure Alters Sperm Chromatin Structure in Mexican Agricultural Workers ». *Toxicology and Applied Pharmacology* 196 (1):108-13. <https://doi.org/10.1016/j.taap.2003.11.023>.
- Santos, F, et Dean, W. 2004. « Epigenetic Reprogramming during Early Development in Mammals ». *Reproduction* 127 (6):643-51. <https://doi.org/10.1530/rep.1.00221>.
- Santos, JE, Thatcher, WW, Chebel, RC, Cerri, RL, et Galvão, KN. 2004. « The Effect of Embryonic Death Rates in Cattle on the Efficacy of Estrus Synchronization Programs ». *Animal Reproduction Science* 82-83 (juillet):513-35. <https://doi.org/10.1016/j.anireprosci.2004.04.015>.
- Sarraf, SA, et Stancheva, I. 2004. « Methyl-CpG Binding Protein MBD1 Couples Histone H3 Methylation at Lysine 9 by SETDB1 to DNA Replication and Chromatin Assembly ». *Molecular Cell* 15 (4):595-605. <https://doi.org/10.1016/j.molcel.2004.06.043>.
- Sartori, R, Sartor-Bergfeldt, R, Mertens, SA, Guenther, JN, Parrish, JJ, et Wiltbank, MC. 2002. « Fertilization and Early Embryonic Development in Heifers and Lactating Cows in Summer and Lactating and Dry Cows in Winter ». *Journal of Dairy Science* 85 (11) :2803-12. [https://doi.org/10.3168/jds.S0022-0302\(02\)74367-1](https://doi.org/10.3168/jds.S0022-0302(02)74367-1).
- Sasai, N, et Defossez, PA. 2009. « Many paths to one goal ? The proteins that recognize methylated DNA in eukaryotes ». *Int. J. Dev. Biol.* 53: 323-334 doi: 10.1387/ijdb.082652ns.
- Sasaki, H. 2017. « Roles and Regulations of Hippo Signaling during Preimplantation Mouse Development ». *Development, Growth & Differentiation* 59 (1):12-20. <https://doi.org/10.1111/dgd.12335>.

- Sato, A, Hiura, H, Okae, H, Miyauchi, N, Abe, Y, Utsunomiya, T, Yaegashi, N, et Arima, T. 2011. « Assessing Loss of Imprint Methylation in Sperm from Subfertile Men Using Novel Methylation Polymerase Chain Reaction Luminex Analysis ». *Fertility and Sterility* 95 (1):129-134.e4. <https://doi.org/10.1016/j.fertnstert.2010.06.076>.
- Saxonov, S, Berg, P, et Brutlag, DL. 2006. « A Genome-Wide Analysis of CpG Dinucleotides in the Human Genome Distinguishes Two Distinct Classes of Promoters ». *Proceedings of the National Academy of Sciences of the United States of America* 103 (5):1412-17. <https://doi.org/10.1073/pnas.0510310103>.
- Schefers, JM, et Weigel, KA. 2012. « Genomic Selection in Dairy Cattle: Integration of DNA Testing into Breeding Programs ». *Animal Frontiers* 2 (1):4-9. <https://doi.org/10.2527/af.2011-0032>.
- Schultz, MD, He, Y, Whitaker, JW, Hariharan, M, Mukamel, EA, Leung, D, Rajagopal, N, et al. 2015. « Human body epigenome maps reveal noncanonical DNA methylation variation ». *Nature* 523 (7559):212-16. <https://doi.org/10.1038/nature14465>.
- Schütte, B, El Hajj, N, Kuhtz, J, Nanda, I, Gromoll, J, Hahn, T, Dittrich, M, Schorsch, M, Müller, T, et Haaf, T. 2013. « Broad DNA Methylation Changes of Spermatogenesis, Inflammation and Immune Response-Related Genes in a Subgroup of Sperm Samples for Assisted Reproduction ». *Andrology* 1 (6):822-29. <https://doi.org/10.1111/j.2047-2927.2013.00122.x>.
- Sedgwick, B. 2004. « Repairing DNA-methylation damage ». *Nature Reviews Molecular Cell Biology* 5 (2):148-57. <https://doi.org/10.1038/nrm1312>.
- Seisenberger, S, Peat, JR, Hore, TA, Santos, F, Dean, W, et Reik, W. 2012. « Reprogramming DNA Methylation in the Mammalian Life Cycle: Building and Breaking Epigenetic Barriers ». *Philosophical Transactions of the Royal Society B: Biological Sciences* 368 (1609):20110330-20110330. <https://doi.org/10.1098/rstb.2011.0330>.
- Seisenberger, S, Andrews, S, Krueger, F, Arand, J, Walter, J, Santos, F, Popp, C, Thienpont, B, Dean, W, et Reik, W. 2012. « The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells ». *Molecular Cell* 48 (6):849-62. <https://doi.org/10.1016/j.molcel.2012.11.001>.
- Seki, Y, Yamaji, M, Yabuta, Y, Sano, M, Shigeta, M, Matsui, Y, Saga, Y, Tachibana, M, Shinkai, Y, et Saitou, M. 2007. « Cellular Dynamics Associated with the Genome-Wide Epigenetic Reprogramming in Migrating Primordial Germ Cells in Mice ». *Development* 134 (14):2627-38. <https://doi.org/10.1242/dev.005611>.
- Sekido, R, et Lovell-Badge, R. 2008. « Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer ». *Nature* 453 (7197):930-34. <https://doi.org/10.1038/nature06944>.
- Sellem, E, Broekhuijse, ML, Chevrier, L, Camugli, S, Schmitt, E, Schibler, L, et Koenen, EP. 2015. « Use of Combinations of in Vitro Quality Assessments to Predict Fertility of Bovine Semen ». *Theriogenology* 84 (9):1447-1454.e5. <https://doi.org/10.1016/j.theriogenology.2015.07.035>.
- Senger, PL. 2004. « Pathways to Pregnancy and Parturition ». 2nd edition, Current Conceptions, Inc., Washington.
- Serre, D, Lee, BH, et Ting, AH. 2010. « MBD-Isolated Genome Sequencing Provides a High-Throughput and Comprehensive Survey of DNA Methylation in the Human Genome ». *Nucleic Acids Research* 38 (2):391-99. <https://doi.org/10.1093/nar/gkp992>.
- Shen, Y, Yue, F, McCleary, DF, Ye, Z, Edsall, L, Kuan, S, Wagner, U, et al. 2012. « A map of the cis-regulatory sequences in the mouse genome ». *Nature* 488 (7409):116-20. <https://doi.org/10.1038/nature11243>.
- Sienski, G, Dönertas, D, et Brennecke, J. 2012. « Transcriptional Silencing of Transposons by Piwi and Maelstrom and Its Impact on Chromatin State and Gene Expression ». *Cell* 151 (5):964-80. <https://doi.org/10.1016/j.cell.2012.10.040>.
- Silke, V, Diskin, MG, Kenny, DA, Boland, MP, Dillon, P, Mee, JF, et Sreenan, JM. 2002. « Extent, Pattern and Factors Associated with Late Embryonic Loss in Dairy Cows ». *Animal Reproduction Science* 71 (1-2):1-12.
- Sillaste, G, Kaplinski, L, Meier, R, Jaakma, U, Eriste, E, et Salumets, A. 2017. « A Novel Hypothesis for Histone-to-Protamine Transition in *Bos Taurus* Spermatozoa ». *Reproduction* 153 (3):241-51. <https://doi.org/10.1530/REP-16-0441>.
- Simon, L, et Lewis, SE. 2011. « Sperm DNA Damage or Progressive Motility : Which One Is the Better Predictor of Fertilization in Vitro ? ». *Systems Biology in Reproductive Medicine* 57 (3):133-38. <https://doi.org/10.3109/19396368.2011.553984>.
- Sinclair, KD, Allegrucci, C, Singh, R, Gardner, DS, Sebastian, S, Bispham, J, Thurston, A, Huntley, JF, Rees, WD, Maloney, CA, Lea, RG, Craigon, J, McEvoy, TG, et Young, LE. 2007. « DNA methylation, insulin resistance, and blood pressure in offspring determined by maternal periconceptional B vitamin and methionine status ». *PNAS* vol. 104 no.49 19351-19356.
- Singh, P, Li, AX, Tran, DA, Oates, N, Kang, ER, Wu, X, et Szabó, PE. 2013. « De Novo DNA Methylation in the Male Germ Line Occurs by Default but Is Excluded at Sites of H3K4 Methylation ». *Cell Reports* 4 (1):205-19. <https://doi.org/10.1016/j.celrep.2013.06.004>.
- Sinkkonen, L, Hugaschmidt, T, Berninger, P, Gaidatzis, D, Mohn, F, Artus-Revel, CG, Zavolan, M, Svoboda, P, et Filipowicz, W. 2008. « MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells ». *Nature Structural & Molecular Biology* 15 (3):259-67. <https://doi.org/10.1038/nsmb.1391>.
- Skvortsova, K, Zotenko, E, Luu, PL, Gould, CM, Nair, SS, Clark, SJ, et Stirzaker, C. 2017. « Comprehensive evaluation of genome-wide 5-hydroxymethylcytosine profiling approaches in human DNA ». *Epigenetics & Chromatin* 10:16. DOI 10.1186/s13072-017-0123-7.
- Smith, ZD, et Meissner, A. 2013. « DNA methylation: roles in mammalian development ». *Nat Rev Genet.* 2013 Mar;14(3):204-20. doi: 10.1038/nrg3354.
- Smith, LC, Therrien, J, Filion, F, Bressan, F, et Meirelles, FV. 2015. « Epigenetic consequences of artificial reproductive technologies to the bovine imprinted genes SNRPN, H19/IGF2, and IGF2R ». *Frontiers in Genetics* 6 (février). <https://doi.org/10.3389/fgene.2015.00058>.
- Song, N, Endo, D, Song, B, Shibata, Y, et Koji, T. 2016. « 5-Aza-2'-Deoxycytidine Impairs Mouse Spermatogenesis at Multiple Stages through Different Usage of DNA Methyltransferases ». *Toxicology* 361-362 (juin):62-72. <https://doi.org/10.1016/j.tox.2016.07.005>.

- Stadler, MB, Murr, R, Burger, L, Ivanek, R, Lienert, F, Schöler, A, Wirbelauer, C, et al. 2011. « DNA-binding factors shape the mouse methylome at distal regulatory regions ». *Nature*, décembre. <https://doi.org/10.1038/nature10716>.
- Stephens, DT, Hickman, R, et Hoskins, DD. 1988. « Description, Validation, and Performance Characteristics of a New Computer-Automated Sperm Motility Analysis System ». *Biology of Reproduction* 38 (3):577-86.
- Stouder, C, Somm, E, et Paoloni-Giacobino, A. 2011. « Prenatal Exposure to Ethanol: A Specific Effect on the H19 Gene in Sperm ». *Reproductive Toxicology* 31 (4):507-12. <https://doi.org/10.1016/j.reprotox.2011.02.009>.
- Suh, TK, Schenk, JL, et Seidel, GE. 2005. « High Pressure Flow Cytometric Sorting Damages Sperm ». *Theriogenology* 64 (5):1035-48. <https://doi.org/10.1016/j.theriogenology.2005.02.002>.
- Sutcliffe, AG, Peters, CJ, Bowdin, S, Temple, K, Reardon, W, Wilson, L, Clayton-Smith, J, Brueton, LA, Bannister, W, et Maher, ER. 2006. « Assisted Reproductive Therapies and Imprinting Disorders—a Preliminary British Survey ». *Human Reproduction* 21 (4):1009-11. <https://doi.org/10.1093/humrep/dei405>.
- Talbert, PB, et Henikoff, S. 2010. « Histone variants — ancient wrap artists of the epigenome ». *Nature Reviews Molecular Cell Biology* 11 (4):264-75. <https://doi.org/10.1038/nrm2861>.
- Tang, WW, Dietmann, S, Irie, N, Leitch, HG, Floros, VI, Bradshaw, CR, Hackett, JA, Chinnery, PF, et Surani, MA. 2015. « A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development ». *Cell* 161 (6):1453-67. <https://doi.org/10.1016/j.cell.2015.04.053>.
- Tarasova, GV, Nayakshina, TN, et Degtyarev, SK. 2008. « Substrate Specificity of New Methyl-Directed DNA Endonuclease Glal ». *BMC Molecular Biology* 9 (1):7. <https://doi.org/10.1186/1471-2199-9-7>.
- Taş, M, Bacinoglu, S, Cirit, U, Özgümüş, S, Kaşgöz, H, et Pabuccuoğlu, S. 2007. « Estimation of the Potential Fertility Based upon Non-Return Rates of Bulls: Using Polyacrylamide Gel Instead of Cervical Mucus in the Sperm Penetration Test ». *Theriogenology* 68 (7):981-87. <https://doi.org/10.1016/j.theriogenology.2007.07.011>.
- Taş, M, Bacinoglu, S, Cirit, U, Özdaş, OB, et Ak, K. 2007. « Relationship between Bovine Fertility and the Number of Spermatozoa Penetrating the Cervical Mucus within Straws ». *Animal Reproduction Science* 101 (1-2):18-27. <https://doi.org/10.1016/j.anireprosci.2006.08.020>.
- Thompson, GJ, Kucharski, R, Maleszka, R, et Oldroyd, BP. 2008. « Genome-Wide Analysis of Genes Related to Ovary Activation in Worker Honey Bees ». *Insect Molecular Biology* 17 (6):657-65. <https://doi.org/10.1111/j.1365-2583.2008.00838.x>.
- Thomsen, B. 2005. « A Missense Mutation in the Bovine SLC35A3 Gene, Encoding a UDP-N-Acetylglucosamine Transporter, Causes Complex Vertebral Malformation ». *Genome Research* 16 (1):97-105. <https://doi.org/10.1101/gr.3690506>.
- Thundathil, JC, Dance, AL, et Kastelic, JP. 2016. « Fertility Management of Bulls to Improve Beef Cattle Productivity ». *Theriogenology* 86 (1):397-405. <https://doi.org/10.1016/j.theriogenology.2016.04.054>.
- Tóth, KF, Pezic, D, Stuwe, E, et Webster, A. 2016. « The piRNA Pathway Guards the Germline Genome Against Transposable Elements ». In *Non-coding RNA and the Reproductive System*, édité par Dagmar Wilhelm et Pascal Bernard, 886:51-77. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-7417-8_4.
- Urduingio, RG, Bayón, GF, Dmitrijeva, M, Torano, EG, Bravo, C, Fraga, MF, Bassas, L, Larriba, S, et Fernández, AF. 2015. « Aberrant DNA Methylation Patterns of Spermatozoa in Men with Unexplained Infertility ». *Human Reproduction* 30 (5):1014-28. <https://doi.org/10.1093/humrep/dev053>.
- Ushizawa, K, Herath, CB, Kaneyama, K, Shiojima, S, Hirasawa, A, Takahashi, T, Imai, K, et al. 2004. « CDNA Microarray Analysis of Bovine Embryo Gene Expression Profiles during the Pre-Implantation Period ». *Reproductive Biology and Endocrinology: RB&E* 2 (novembre):77. <https://doi.org/10.1186/1477-7827-2-77>.
- VanRaden, PM, et Miller, RH. 2006. « Effects of Nonadditive Genetic Interactions, Inbreeding, and Recessive Defects on Embryo and Fetal Loss by Seventy Days ». *Journal of Dairy Science* 89 (7):2716-21. [https://doi.org/10.3168/jds.S0022-0302\(06\)72347-5](https://doi.org/10.3168/jds.S0022-0302(06)72347-5).
- VanRaden, PM, Olson, KM, Null, DJ, et Hutchison, JL. 2011. « Harmful Recessive Effects on Fertility Detected by Absence of Homozygous Haplotypes ». *Journal of Dairy Science* 94 (12):6153-61. <https://doi.org/10.3168/jds.2011-4624>.
- VanRaden, PM, Sanders, AH, Tooker, ME, Miller, RH, Norman, HD, Kuhn, MT, et Wiggans, GR. 2004. « Development of a National Genetic Evaluation for Cow Fertility ». *Journal of Dairy Science* 87 (7):2285-92. [https://doi.org/10.3168/jds.S0022-0302\(04\)70049-1](https://doi.org/10.3168/jds.S0022-0302(04)70049-1).
- Van Vleck, LD, et Norman, HD. 1972. « Association of Type Traits with Reasons for Disposal ». *Journal of Dairy Science* 55 (12):1698-1705. [https://doi.org/10.3168/jds.S0022-0302\(72\)85747-3](https://doi.org/10.3168/jds.S0022-0302(72)85747-3).
- Vaquerizas, JM, Kummerfeld, SK, Teichmann, SA, et Luscombe, NM. 2009. « A census of human transcription factors: function, expression and evolution ». *Nature Reviews Genetics* 10 (4):252-63. <https://doi.org/10.1038/nrg2538>.
- Varley, KE, Gertz, J, Bowling, KM, Parker, SL, Reddy, TE, Pauli-Behn, F, Cross, MK, et al. 2013. « Dynamic DNA Methylation across Diverse Human Cell Lines and Tissues ». *Genome Research* 23 (3):555-67. <https://doi.org/10.1101/gr.147942.112>.
- Verma, A, Rajput, S, De, S, Kumar, R, Chakravarty, AK, et Datta, TK. 2014. « Genome-Wide Profiling of Sperm DNA Methylation in Relation to Buffalo (*Bubalus Bubalis*) Bull Fertility ». *Theriogenology* 82 (5):750-759.e1. <https://doi.org/10.1016/j.theriogenology.2014.06.012>.
- Vincent, JJ, Huang, Y, Chen, PY, Feng, S, Calvopiña, JH, Nee, K, Lee, SA, et al. 2013. « Stage-Specific Roles for Tet1 and Tet2 in DNA Demethylation in Primordial Germ Cells ». *Cell Stem Cell* 12 (4):470-78. <https://doi.org/10.1016/j.stem.2013.01.016>.
- Vishwanath, R. 2003. « Artificial Insemination: The State of the Art ». *Theriogenology* 59 (2):571-84.
- Vlachogiannis, G, Niederhuth, CE, Tuna, S, Stathopoulou, A, Viiri, K, de Rooij, DG, Jenner, RG, Schmitz, RJ, et Ooi, SK. 2015. « The Dnmt3L ADD Domain Controls Cytosine Methylation Establishment during Spermatogenesis ». *Cell Reports* 10 (6):944-56. <https://doi.org/10.1016/j.celrep.2015.01.021>.
- Vogiatzi, P, Chrelias, C, Cahill, DJ, Creata, M, Vrachnis, N, Iliodromiti, Z, Kassanos, D, et Siristatidis, C. 2013. « Hemizona Assay and Sperm Penetration Assay in the Prediction of IVF Outcome: A Systematic Review ». *BioMed Research International* 2013:1-10. <https://doi.org/10.1155/2013/945825>.

- Waddington, CH. 1957. « The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology ». George Allen and Unwin, London.
- Walters, AH, Saacke, RG, Pearson, RE, et Gwazdauskas, FC. 2006. « Assessment of Pronuclear Formation Following in Vitro Fertilization with Bovine Spermatozoa Obtained after Thermal Insulation of the Testis ». *Theriogenology* 65 (6):1016-28. <https://doi.org/10.1016/j.theriogenology.2005.07.005>.
- Wang, J, Xia, Y, Li, L, Gong, D, Yao, Y, Luo, H, Lu, H, et al. 2013. « Double Restriction-Enzyme Digestion Improves the Coverage and Accuracy of Genome-Wide CpG Methylation Profiling by Reduced Representation Bisulfite Sequencing ». *BMC Genomics* 14 (janvier):11. <https://doi.org/10.1186/1471-2164-14-11>.
- Wang, K, Li, X, Dong, S, Liang, J, Mao, F, Zeng, C, Wu, H, Wu, J, Cai, J, et Sun, ZS. 2015. « Q-RRBS: A Quantitative Reduced Representation Bisulfite Sequencing Method for Single-Cell Methylome Analyses ». *Epigenetics* 10 (9):775-83. <https://doi.org/10.1080/15592294.2015.1075690>.
- Wang, L, Sun, J, Wu, H, Liu, S, Wang, J, Wu, B, Huang, S, Li, N, Wang, J, et Zhang, X. 2012. « Systematic Assessment of Reduced Representation Bisulfite Sequencing to Human Blood Samples: A Promising Method for Large-Sample-Scale Epigenomic Studies ». *Journal of Biotechnology* 157 (1):1-6. <https://doi.org/10.1016/j.jbiotec.2011.06.034>.
- Ward, F, Rizos, D, Corridan, D, Quinn, K, Boland, M, et Lonergan, P. 2001. « Paternal Influence on the Time of First Embryonic Cleavage Post Insemination and the Implications for Subsequent Bovine Embryo Development in Vitro and Fertility in Vivo ». *Molecular Reproduction and Development* 60 (1):47-55. <https://doi.org/10.1002/mrd.1060>.
- Waterland, RA, et Jirtle, RL. 2004. « Early Nutrition, Epigenetic Changes at Transposons and Imprinted Genes, and Enhanced Susceptibility to Adult Chronic Diseases ». *Nutrition* 20 (1):63-68. <https://doi.org/10.1016/j.nut.2003.09.011>.
- Way, AL, Griel, LC, et Killian, GJ. 2000. « Effects of Accessory Sex Gland Fluid on Viability, Capacitation, and the Acrosome Reaction of Cauda Epididymal Bull Spermatozoa ». *Journal of Andrology* 21 (2):213-19.
- Weber, M, Davies, JJ, Wittig, D, Oakeley, EJ, Haase, M, Lam, WL, et Schübeler, D. 2005. « Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells ». *Nature Genetics* 37 (8):853-62. <https://doi.org/10.1038/ng1598>.
- Wierda, RJ, et Van den Elsen, PJ. 2012. « Genetic and Epigenetic Regulation of CCR5 Transcription ». *Biology* 2012, 1, 869-879; doi:10.3390/biology1030869.
- Wiggans, GR, Cooper, TA, VanRaden, PM, Van Tassell, CP, Bickhart, DM, et Sonstegard, TS. 2016. « Increasing the Number of Single Nucleotide Polymorphisms Used in Genomic Evaluation of Dairy Cattle 1 ». *Journal of Dairy Science* 99 (6):4504-11. <https://doi.org/10.3168/jds.2015-10456>.
- Wojdacz, TK, et Dobrovic, A. 2007. « Methylation-Sensitive High Resolution Melting (MS-HRM): A New Approach for Sensitive and High-Throughput Assessment of Methylation ». *Nucleic Acids Research* 35 (6):e41-e41. <https://doi.org/10.1093/nar/gkm013>.
- Wolf, FR, Almquist, JO, et Hale, EB. 1965. « Prepuberal Behaviour and Puberal Characteristics of Beef Bulls on High Nutrient Allowance ». *Journal of Animal Science* 24 (août):761-65.
- Wooding, FB. 1992. « Current Topic: The Synepitheliochorial Placenta of Ruminants: Binucleate Cell Fusions and Hormone Production ». *Placenta* 13 (2):101-13.
- Wossidlo, M, Nakamura, T, Lepikhov, K, Marques, CJ, Zakhartchenko, V, Boiani, M, Arand, J, Nakano, T, Reik, W, et Walter, J. 2011. « 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming ». *Nature Communications* 2 (mars):241. <https://doi.org/10.1038/ncomms1240>.
- Wrobel, KH. 2000. « Prespermatogenesis and Spermatogoniogenesis in the Bovine Testis ». *Anatomy and Embryology* 202 (3):209-22.
- Wrobel, KH, et Süß, F. 1998. « Identification and Temporospatial Distribution of Bovine Primordial Germ Cells Prior to Gonadal Sexual Differentiation ». *Anatomy and Embryology* 197 (6):451-67.
- Wu, TP, Wang, T, Seetin, MG, Lai, Y, Zhu, S, Lin, K, Liu, Y, et al. 2016. « DNA methylation on N6-adenine in mammalian embryonic stem cells ». *Nature* 532 (7599):329-33. <https://doi.org/10.1038/nature17640>.
- Wu, W, Shen, O, Qin, Y, Niu, X, Lu, C, Xia, Y, Song, L, Wang, S, et Wang, X. 2010. « Idiopathic Male Infertility Is Strongly Associated with Aberrant Promoter Methylation of Methylenetetrahydrofolate Reductase (MTHFR) ». Édité par Robert Feil. *PLoS ONE* 5 (11):e13884. <https://doi.org/10.1371/journal.pone.0013884>.
- Yamada, O, Todoroki, J, Kizaki, K, Takahashi, T, Imai, K, Patel, OV, Schuler, LA, et Hashizume, K. 2002. « Expression of Prolactin-Related Protein I at the Fetomaternal Interface during the Implantation Period in Cows ». *Reproduction (Cambridge, England)* 124 (3):427-37.

Annexes

Annexes des résultats n°1

Supplementary Table S1: reference genomes used for in silico analyses and origin of the files used for annotation.

Supplementary Table S2: in silico characterization of bovine reduced restriction (RR) genomes generated using different size selection criteria.

Supplementary Table S3: comparison of RR genomes obtained with a 40-290 bp selection size window in different species.

Supplementary Table S4: primers and PCR conditions used to generate the pyrosequencing templates.

Supplementary Table S5: pyrosequencing primers.

Supplementary Table S6: library characterization, mapping efficiency on the bovine genome (UMD3.1), coverage and average methylation in RRBS libraries.

Supplementary Table S7: results of comparisons between tissues by RRBS.

Supplementary Table S8: mapping efficiency on a Rebase artificial bovine genome, coverage and average methylation in RRBS libraries.

Supplementary Figure S1: Average methylation percentages for CpGs 5-500 and CpGs >500 in each cell type, in reads uniquely aligned on a Rebase bovine artificial genome.

Annexes des résultats n°2

Tableau n°1. Résultats des analyses fonctionnelles au CASA

Tableau n°2. Liste des analyses de double digestion enzymatique in silico pour le RRBS

Tableau n°3. Amorces et conditions de PCR pour générer les matrices de pyroséquençage

Tableau n°4. Amorces de pyroséquençage

Tableau n°5. Données générales issues des banques RRBS

Tableau n°6. Listes des gènes associés à une DMR pour la race Holstein

Tableau n°7. Listes et répartition des DMR identifiées entre les catégories de fertilité pour la race Montbéliarde

Liste des communications présentées à des congrès

Supplementary Tables and Figure

Supplementary Table S1. Reference genomes used for *in silico* analyses and origin of the files used for annotation

Species	Reference genome	Gtf file	Biomart file	CpG island file	Repeat file
Cattle	UMD3.1	ftp://ftp.ensembl.org/pub/release-81/gtf/bos/taurus/Bos_taurus.UMD3.1.81.gtf.gz	http://juil2015.archive.ensembl.org/biomart/martview/Ensembl_Genes_81_Bos_taurus_genes(UMD3.1)	http://hgdownload.cse.ucsc.edu/goldenPath/bosTau6/database/cpgIslandExt.txt.gz	http://hgdownload.cse.ucsc.edu/goldenPath/bosTau6/database/rmsk.txt.gz
Sheep	Oar_v3.1	ftp://ftp.ensembl.org/pub/release-87/gtf/ovis/aries/Ovis_aries.Oar_v3.1.87.gtf.gz	http://dec2016.archive.ensembl.org/biomart/martview/Ensembl_Genes_87_Sheep_genes(Oar_v3.1)	http://hgdownload.cse.ucsc.edu/goldenPath/oviAri3/database/cpgIslandExt.txt.gz	http://hgdownload.cse.ucsc.edu/goldenPath/oviAri3/database/rmsk.txt.gz
Horse	EquCab2	ftp://ftp.ensembl.org/pub/release-87/gtf/equus/caballus/Equus_caballus.EquCab2.87.gtf.gz	http://dec2016.archive.ensembl.org/biomart/martview/Ensembl_Genes_87_Horse_genes(EquCab2)	NA	http://hgdownload.cse.ucsc.edu/goldenPath/equCab2/database/chri1_rmsk.txt.gz http://hgdownload.cse.ucsc.edu/goldenPath/equCab2/database/chrX_rmsk.txt.gz
Pig	Scrofa10.2	ftp://ftp.ensembl.org/pub/release-84/gtf/sus/scrofa/Sus_scrofa.Scrofa10.2.84.gtf.gz	http://mar2016.archive.ensembl.org/biomart/martview/Ensembl_Genes_84_Sus_scrofa_genes(Scrofa10.2)	http://hgdownload.cse.ucsc.edu/goldenPath/susScr3/database/cpgIslandExt.txt.gz	http://hgdownload.cse.ucsc.edu/goldenPath/susScr3/database/rmsk.txt.gz
Mouse	GRCm38	ftp://ftp.ensembl.org/pub/release-83/gtf/mus/musculus/Mus_musculus.GRCm38.83.gtf.gz	http://dec2015.archive.ensembl.org/biomart/martview/Ensembl_Genes_83_Mouse_genes(GRCm38.p4)	http://hgdownload.cse.ucsc.edu/goldenPath/mm10/database/cpgIslandExt.txt.gz	http://hgdownload.cse.ucsc.edu/goldenPath/mm10/database/rmsk.txt.gz
Human	GRCh38	ftp://ftp.ensembl.org/pub/release-87/gtf/homo/sapiens/Homo_sapiens.GRCh38.87.gtf.gz	http://dec2016.archive.ensembl.org/biomart/martview/Ensembl_Genes_87_Human_genes(GRCh38.p7)	http://hgdownload.cse.ucsc.edu/goldenPath/hg38/database/cpgIslandExt.txt.gz	http://hgdownload.cse.ucsc.edu/goldenPath/hg38/database/rmsk.txt.gz

NA: not available

Supplementary Table S2. *In silico* characterization of bovine reduced restriction (RR) genomes generated using different size selection criteria

Size selection (bp)	No	0-250	40-290	80-330	120-370	160-410		
RR genome size (Mb)	2,670	71	79	83	85	86	Regions targeted by the microarray	
Per cent of whole genome	100.0	2.7	3.0	3.1	3.2	3.2		
Number of MspI fragments	1,990,837	810,994	585,584	456,504	376,199	317,974		
Number of CpG sites	27,540,276	3,588,657	3,454,028	3,127,005	2,854,907	2,666,727		
Percent of total genomic CpG sites	100.0	13.0	12.5	11.4	10.4	9.7		
Percent in 3'UTR	0.4	0.6	0.6	0.6	0.6	0.6		0.1
Percent in 5'UTR	0.2	0.5	0.4	0.3	0.2	0.2		0.1
Percent in exon	4.6	9.2	8.7	7.6	6.4	5.4	0.4	
Percent in intron	31.1	30.7	32.0	33.0	33.8	34.5	5.4	
Percent in intergenic	57.0	46.9	48.3	50.1	51.7	52.7	16.9	
Percent in promoter-TSS	4.8	9.7	7.7	6.1	5.0	4.3	76.6	
Percent in TTS	1.9	2.4	2.4	2.4	2.3	2.3	0.6	
Percent in CpG islands	13.4	31.4	22.9	16.1	11.5	8.3	51.0	
Percent in overlapping repeats	61.9	26.2	31.2	37.8	43.4	47.9	48.0	

For comparison with RR genomes, the regions targeted by the microarray used for MeDIP-chip were annotated using the same pipeline. The genome features are shaded in gray for these regions.

Supplementary Table S3. Comparison of RR genomes obtained with a 40-290 bp selection size window in different species

Species	Human	Mouse	Pig	Sheep	Horse	Cattle
Whole genome size (Gb)	3.1	2.7	2.8	2.6	2.5	2.7
Total number of CpG sites	29,345,332	21,867,837	30,460,432	26,376,870	29,873,125	27,203,575
CpG sites per Mb whole genome	9,466	8,099	10,840	10,068	12,094	10,075
RR genome size (Mb)	99	56	106	75	109	79
Number of MspI fragments in RR genome	738,215	400,739	782,851	538,822	844,221	585,584
Percent of whole genome	3.2	2.0	3.8	2.9	4.4	3.0
Number of CpG sites in RR genome	4,140,424	1,914,962	5,132,906	3,310,947	4,854,816	3,454,028
Percent of total CpG sites	14.1	8.8	16.9	12.6	16.3	12.5
CpG sites per Mb RR genome	41,992	34,442	48,424	43,912	44,540	43,611
CpG enrichment ratio	4.4	4.3	4.5	4.4	3.7	4.2

The CpG enrichment ratio was obtained by dividing the number of CpG sites per Mb RR genome by the number of CpG sites per Mb whole genome.

Supplementary Table S4. Primers and PCR conditions used to generate the pyrosequencing templates

Primer name	Gene symbol (EMBL accession number)	Primer sequence (5'-3')	Size of the product	Hybridization temperature	MgCl ₂ concentration
bLSM4_bis_F1	<i>LSM4</i> (ENSBTAG00000008578)	GTTTTGGTGGTTAGTTTTTTG	266 bp	60°C	2 mM
bLSM4_bio_R1		AATTAATACTCCTAACCTTATCCCTC			
bDDX4_bis_F1	<i>DDX4</i> (ENSBTAG00000008871)	GTTGGGATGATTTTTGTATTGGGAAAAG	324 bp	58°C	1.5 mM
bDDX4_bio_R1		CCAC <u>C</u> ATCAACCTTATACCCCAAAC			
bSYCP3_bis_F1	<i>SYCP3</i> (ENSBTAG00000002492)	GGTTAAGAGTAGTTTTGGTTTAGAT	318°C	56°C	1.5 mM
bSYCP3_bio_R1		ATCAACAACCTCACAAAATTCTTC			
bBTSAT4_bis_F1	NA (bovine satellite)	TGTAGATTGGGGATAGGAGAGTTAG	380 bp	60°C	3 mM
bBTSAT4_bio_R1		CCCTCCTAATCTAAACAAAAAATC			

The primer bDDX4_bio_R1 contains a CpG site. To avoid the biased amplification of the methylated/unmethylated allele, the underlined base, originally a G, has been mutated. NA: not available.

Supplementary Table S5. Pyrosequencing primers

Name	Sequence (5'-3')	Gene symbol (EMBL accession number)	Template	CpGs
bLSM4_pyr1	GAGTAGTTTGTTTGG	<i>LSM4</i> (ENSBTAG00000008578)	bLSM4_bis_F1 x bLSM4_bio_R1	#1-9
bLSM4_pyr3	TTGTTGTTAAAGAG			#10-15
bLSM4_pyr5	GGAGGTGAATTAAGG			#16
bDDX4_pyr2	TTTATTTT <u>A</u> GTTTTTTTATTTTA	<i>DDX4</i> (ENSBTAG00000008871)	bDDX4_bis_F1 x bDDX4_bio_R1	#1-8
bSYCP3_pyr1	GAGGAT <u>A</u> GTAGTTAATGTTTT	<i>SYCP3</i> (ENSBTAG00000002492)	bSYCP3_bis_F1 x bSYCP3_bio_R1	#1-2
bSYCP3_pyr2	TTGAAGTGTATTT			#3-6
bSYCP3_pyr4	GGGAGAAAAGTTAGTTT			#9-12
bBTSAT4_pyr3	ATTTATAGGTTGGAG	Bovine satellite	bBTSAT4_bis_F1 x bBTSAT4_bio_R1	#1
bBTSAT4_pyr4	TTTTATTAAGAGGGG			#4-6
bBTSAT4_pyr5	GTTTGAATGTTTT			#9-12

The last column refers to the CpG positions indicated Figure 6B. The primers bDDX4_pyr2 and bSYCP3_pyr1 contain a CpG site. To avoid the biased elongation of the methylated/unmethylated allele, the underlined base, originally a C, has been mutated.

Supplementary Table S6. Library characterization, mapping efficiency on the bovine genome (UMD3.1), coverage and average methylation in RRBS libraries

Sample	spz32	spz34	mono1	mono2	F029	F5538
Number of read pairs (million)	26.8	29.9	27.7	22.6	40.8	33.7
Uniquely mapped reads (%)	35.5	36.9	38.5	36.8	35.8	36.4
Ambiguous reads (%)	55.6	54.3	53.1	52.5	50.0	48.6
Unmapped reads (%)	8.9	8.8	8.4	10.7	14.2	15.0
Bisulfite conversion rate (%)	99.7	99.8	99.8	99.9	99.5	99.6
Methylation, all CpGs (%)	54.0	53.8	60.6	59.6	49.7	49.0
Methylation, CpGs 5-500 (%)	51.4	52.1	58.3	56.9	48.3	47.6
Methylation, CpGs >500 (%)	23.9	21.2	81.6	81.2	67.0	66.3
Number of CpGs 5-500	2,310,551	2,438,658	2,537,411	2,318,185	2,673,984	2,609,798
Number of CpGs >500	8,254	10,258	8,168	6,807	11,364	8,949
Hypomethylated CpGs 5-500 fraction	0.45	0.44	0.35	0.36	0.38	0.39
Intermediate CpG 5-500 fraction	0.06	0.06	0.11	0.11	0.27	0.28
Hypermethylated CpGs 5-500 fraction	0.49	0.50	0.54	0.53	0.35	0.34

Bisulfite conversion rate was estimated from the unmethylated cytosine artificially added during the end-repair step. CpGs 5-500 comprise all CpGs covered by 5 to 500 reads for each sample, and CpGs >500 comprise all CpGs covered by more than 500 reads for each sample. Hypomethylated, intermediate and hypermethylated CpGs 5-500 fractions: fractions of CpGs 5-500 with average methylation <20%, [20%; 80%], and >80%, respectively.

Supplementary Table S7. Results of comparisons between tissues by RRBS

Comparison	Sperm vs Fibroblasts	Sperm vs Monocytes	Fibroblasts vs Monocytes
Number of methylKit DMCs	450,959	298,874	239,017
Number of obvious DMCs	100,686	128,477	32,883
Number of obvious DMCs added	12	27	19

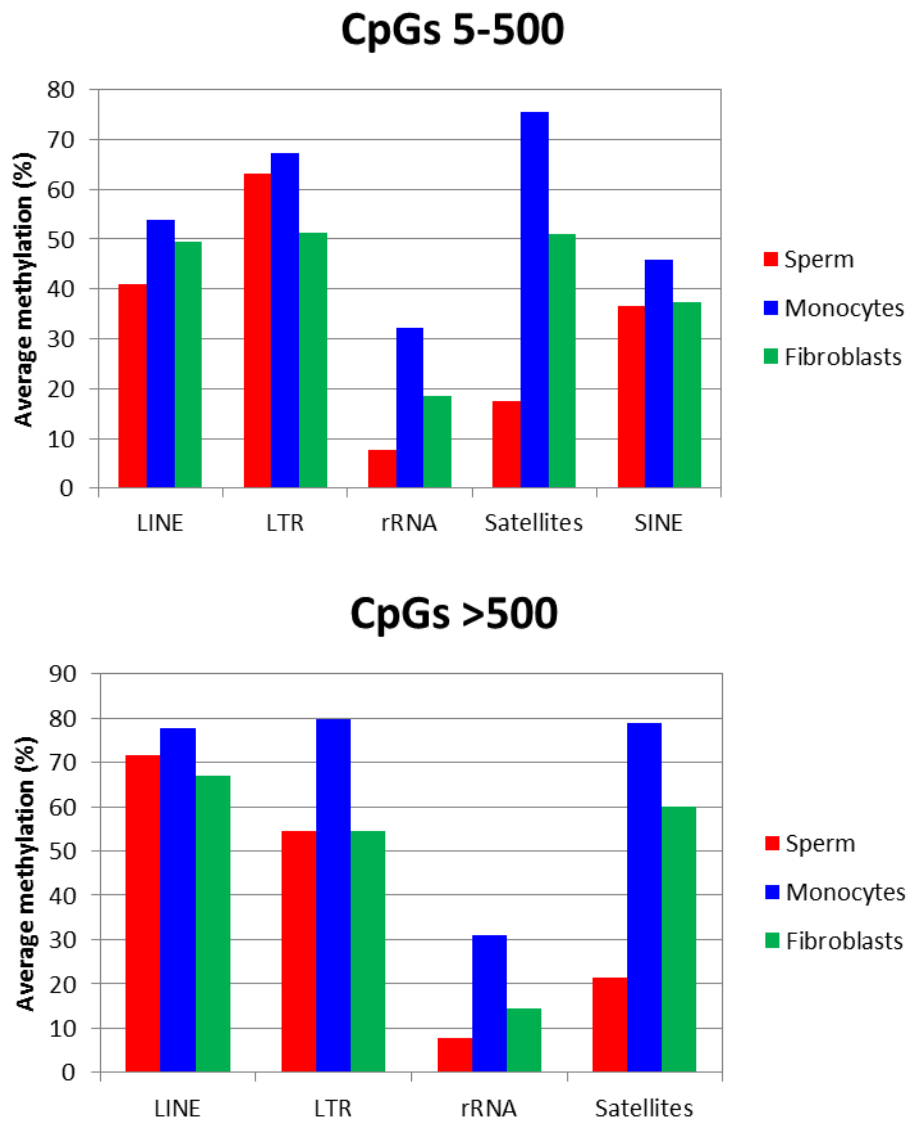
The number of obvious DMC added corresponds to the obvious DMCs that were not identified using methylKit.

Supplementary Table S8. Mapping efficiency on a Repbase artificial bovine genome, coverage and average methylation in RRBS libraries

Sample	spz32	spz34	mono1	mono2	F029	F5538
Uniquely mapped reads (%)	25.8	25.3	25.0	23.5	22.8	21.2
Unmapped reads (%)	74.2	74.7	75.0	76.5	77.2	78.8
Methylation, all CpGs (%)	39.1	38.0	56.2	58.4	44.7	46.0
Methylation, CpGs 5-500 (%)	47.2	45.5	64.0	65.8	49.1	51.2
Methylation, CpGs >500 (%)	25.7	25.3	56.7	60.2	39.1	44.2
Number of CpGs 5-500	557	541	569	513	559	536
Number of CpGs >500	370	383	391	344	389	382
Hypomethylated CpGs 5-500 fraction	0.29	0.30	0.16	0.12	0.25	0.19
Intermediate CpG 5-500 fraction	0.49	0.49	0.40	0.44	0.54	0.61
Hypermethylated CpGs 5-500 fraction	0.23	0.21	0.44	0.44	0.21	0.21
Hypomethylated CpGs >500 fraction	0.65	0.64	0.14	0.02	0.42	0.25
Intermediate CpG >500 fraction	0.28	0.31	0.50	0.65	0.44	0.61
Hypermethylated CpGs >500 fraction	0.07	0.05	0.36	0.33	0.13	0.14

CpGs 5-500 comprise all CpGs covered by 5 to 500 reads for each sample, and CpGs >500 comprise all CpGs covered by more than 500 reads for each sample. Hypomethylated, intermediate and hypermethylated CpGs 5-500 (resp. CpGs >500) fractions: fractions of CpGs 5-500 (resp. CpGs >500) with average methylation <20%, [20%; 80%], and >80%, respectively.

Supplementary Figure S1



Average methylation percentage for CpGs 5-500 and CpGs >500 in each cell type, in reads uniquely aligned on a Repbase bovine artificial genome. For each RRBS sample, reads were mapped on an artificial genome containing one copy of each bovine repeat as defined in the Repbase database. The average percentage of methylation was then calculated individually for each consensus repeat, from either CpGs 5-500 or CpGs >500. The figures shows this average methylation percentages in each family of repeat.

Tableau n°2. Liste des analyses de double digestion enzymatique *in silico* pour le RRBS

Les colonnes correspondent aux fenêtres de sélection de taille évaluées (taille de fragment minimale - taille de fragment maximale)

MspI (C[^]CGG) seul												
Aucune	0 - 250	40 - 290	80 - 330	120 - 370	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRgenome (pb)	2 670 422 117	71 343 283	79 166 662	82 526 978	84 828 041	85 756 656	86 827 054	87 215 301	87 472 807	88 299 045		
Fraction du génome (%)	100,0	2,7	3,0	3,1	3,2	3,2	3,3	3,3	3,3	3,3		
Nombre de fragments générés	1 990 837	810 994	585 584	456 504	376 199	317 974	278 770	247 536	241 349	224 718		
Nombre de CpG analysables	27 540 276	3 588 657	3 454 028	3 127 005	2 854 907	2 666 727	2 476 921	2 307 125	2 274 752	2 184 879		
Fraction des CpG totaux analysables (%)	100,0	13,0	12,5	11,4	10,4	9,7	9,0	8,4	8,3	7,9		

BstEII (G[^]GTGACC) puis MspI (C[^]CGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 370	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRgenome (pb)	2 670 088 148	75 884 655	84 357 604	88 157 512	90 834 979	91 820 456	93 567 096	94 017 576	94 304 162	95 279 388		
Fraction du génome (%)	100,0	2,8	3,2	3,3	3,4	3,4	3,5	3,5	3,5	3,6		
Nombre de fragments générés	2 099 515	856259	621 673	486 677	402 744	372 939	300 288	266 486	259 795	242 107		
Nombre de CpG analysables	27 532 641	3 694 669	3 558 600	3 228 191	2 953 736	2 836 743	2 582 915	2 404 855	2 371 467	2 280 037		
Fraction des CpG totaux analysables (%)	100,0	13,4	12,9	11,7	10,7	10,3	9,4	8,7	8,6	8,3		

BglII (A[^]GATCT) puis MspI (C[^]CGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 370	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRgenome (pb)	2 668 038 656	98 011 995	117 485 888	126 961 699	134 844 739	137 120 042	144 058 281	148 374 039	149 394 712	151 827 154		
Fraction du génome (%)	99,9	3,7	4,4	4,8	5,0	5,1	5,4	5,6	5,6	5,7		
Nombre de fragments générés	2785323	103351	813 138	677 777	591 895	553 069	457 824	420 547	412 637	388 904		
Nombre de CpG analysables	27 540 276	3 938 291	3 886 598	3 607 642	3 381 812	3 280 424	3 049 882	2 896 744	2 867 336	2 782 819		
Fraction des CpG totaux analysables (%)	100,0	14,3	14,1	13,1	12,3	11,9	11,1	10,5	10,4	10,1		

TaqII (T[^]CGA) puis MspI (C[^]CGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRgenome (pb)	2 665 860 386	136 074 767	153 800 810	165 780 775	177 048 048	178 878 369	184 286 553	188 858 400	189 849 081	195 190 880		
Fraction du génome (%)	99,8	5,1	5,8	6,2	7,1	6,6	6,9	7,1	7,1	7,3		
Nombre de fragments générés	3 511 411	1513930	1 100 755	896 551	800 487	711 221	659 699	534 534	521 546	492 905		
Nombre de CpG analysables	26 019 699	4 913 833	4 740 382	4 352 268	4 245 079	3 846 170	3 447 662	3 250 065	3 203 807	3 104 236		
Fraction des CpG totaux analysables (%)	94,5	17,8	17,2	15,8	15,4	14,0	12,5	11,8	11,6	11,3		

BamHI (G[^]ATCC) puis MspI (C[^]CGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRgenome (pb)	2 669 376 701	83 403 607	93 742 247	99 075 942	111 988 840	104 867 281	107 873 752	109 211 923	109 658 846	111 201 186		
Fraction du génome (%)	100,0	3,1	3,5	3,7	4,2	3,9	4,0	4,1	4,1	4,2		
Nombre de fragments générés	2332416	932298	682 734	542 337	477 596	423 298	345 623	309 288	301 878	282 437		
Nombre de CpG analysables	27 524 680	3 832 053	3 711 094	3 392 464	3 322 582	3 014 088	2 749 888	2 578 263	2 545 237	2 455 317		
Fraction des CpG totaux analysables (%)	99,9	13,9	13,5	12,3	12,1	10,9	10,0	9,4	9,2	8,9		

AflII (C[^]TTAAG) puis MspI (C[^]CGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 370	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRgenome (pb)	2 668 801 292	88 408 059	100 887 484	108 322 825	124 207 320	116 737 870	121 922 180	124 876 523	125 549 810	127 920 585		
Fraction du génome (%)	99,9	3,3	3,8	4,1	4,7	4,4	4,6	4,7	4,7	4,8		
Nombre de fragments générés	2531111	949520	720 173	585 862	525 672	468 721	389 073	352 543	344 452	323 940		
Nombre de CpG analysables	27 540 276	3 832 801	3 747 440	3 454 569	3 413 004	3 104 655	2 859 562	2 699 269	2 667 227	2 579 309		
Fraction des CpG totaux analysables (%)	100,0	13,9	13,6	12,5	12,4	11,3	10,4	9,8	9,7	9,4		

AvrII (C[^]CTAGG) puis MspI (C[^]CGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 370	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRgenome (pb)	2 669 237 327	85 291 085	96 231 707	102 213 949	115 917 954	108 668 046	112 474 384	114 264 859	114 781 557	116 566 181		
Fraction du génome (%)	100,0	3,2	3,6	3,8	4,3	4,1	4,2	4,3	4,3	4,4		
Nombre de fragments générés	2385767	988233	698 222	557 744	493 596	438 310	359 441	323 087	315 635	295 897		
Nombre de CpG analysables	27 540 276	3 850 462	3 741 881	3 427 271	3 363 179	3 053 713	2 797 048	2 628 010	2 593 690	2 502 572		
Fraction des CpG totaux analysables (%)	100,0	14,0	13,6	12,4	12,2	11,1	10,2	9,5	9,4	9,1		

BcII (T⁺GAICA) puis MspI (C⁺CGG)											
Aucune	0 - 250	40 - 290	80 - 330	120 - 370	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530	
Taille RRgenome (pb)	2 668 124 342	97 212 894	109 538 281	117 724 144	135 808 152	128 085 736	131 627 816	135 536 338	136 430 835	139 416 765	
Fraction du géome (%)	99,9	3,6	4,1	4,4	5,1	4,8	4,9	5,1	5,1	5,2	
Nombre de fragments générés	27 56744	1065584	773 837	632 210	573 572	514 709	484 538	381 527	373 573	353 084	
Nombre de CpG analysables	27 540 276	3 964 103	3 835 881	3 527 616	3 494 820	3 184 080	3 125 912	2 778 800	2 747 912	2 665 915	
Fraction des CpG totaux analysables (%)	100,0	14,4	13,9	12,8	12,7	11,6	11,4	10,1	10,0	9,7	

BfaI (C⁺TAG) puis MspI (C⁺CGG)											
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530	
Taille RRgenome (pb)	2 649 364 256	535 550 784	631 546 339	685 543 634	799 386 871	753 511 648	754 249 912	742 100 565	739 108 944	732 151 345	
Fraction du géome (%)	99,2	20,1	23,6	25,7	29,9	28,2	28,2	27,8	27,7	27,4	
Nombre de fragments générés	9 010 119	5 256 223	4 436 256	3 654 065	3 389 596	3 038 439	2 811 845	2 130 388	2 066 378	1 893 512	
Nombre de CpG analysables	27 540 276	8 473 230	8 962 374	8 826 091	9 358 311	8 653 363	8 438 046	7 676 039	7 561 076	7 293 340	
Fraction des CpG totaux analysables (%)	100,0	30,8	32,5	32,0	34,0	31,4	30,6	27,9	27,5	26,5	

BsrGI (T⁺GTACA) puis MspI (C⁺CGG)											
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530	
Taille RRgenome (pb)	2 667 969 839	96 760 266	111 036 493	120 715 687	146 830 932	138 637 997	141 420 806	150 108 882	151 198 749	154 730 305	
Fraction du géome (%)	99,9	3,6	4,2	4,5	5,5	5,2	5,3	5,6	5,7	5,8	
Nombre de fragments générés	2808262	1023556	783 695	647 992	607 445	544 959	511 237	421 842	413 719	392 636	
Nombre de CpG analysables	27 540 276	3 929 357	3 854 887	3 575 833	3 588 395	3 270 777	3 210 950	2 889 317	2 860 794	2 785 305	
Fraction des CpG totaux analysables (%)	100,0	14,3	14,0	13,0	13,0	11,9	11,7	10,5	10,4	10,1	

CviAI (C⁺ATG) puis MspI (C⁺CGG)											
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530	
Taille RRgenome (pb)	2 630 068 532	1 066 422 793	1 171 969 354	1 218 655 577	1 253 882 225	1 144 643 971	1 115 453 776	1 016 963 338	911 529 088	841 246 212	
Fraction du géome (%)	98,5	39,9	43,9	45,6	47,0	42,9	41,8	38,1	34,1	31,5	
Nombre de fragments générés	15441999	11872031	8 769 922	6 928 664	5 650 116	4 807 818	4 337 712	3 403 459	2 630 284	2 244 929	
Nombre de CpG analysables	27 540 276	14 421 325	14 421 700	13 460 396	12 968 227	11 545 750	10 919 285	9 563 690	8 216 919	7 424 220	
Fraction des CpG totaux analysables (%)	100,0	52,4	52,4	48,9	47,1	41,9	39,6	30,8	29,8	27,0	

CviQI (G⁺TAC) puis MspI (C⁺CGG)											
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530	
Taille RRgenome (pb)	2 655 454 079	349 715 583	416 107 382	462 678 173	546 961 238	515 651 825	522 319 201	551 579 962	551 894 814	552 036 519	
Fraction du géome (%)	99,4	13,1	15,6	17,3	20,5	19,3	19,6	20,7	20,7	20,7	
Nombre de fragments générés	6826653	3403123	2 872 462	2 461 899	2 302 916	2 062 423	1 913 602	1 737 671	1 524 856	1 411 159	
Nombre de CpG analysables	27 106 257	6 749 171	7 048 772	6 981 868	7 350 717	6 787 065	6 565 310	6 236 345	6 151 764	5 963 260	
Fraction des CpG totaux analysables (%)	98,4	24,5	25,6	25,4	26,7	24,6	23,8	22,6	22,3	21,7	

HindIII (A⁺AGCTT) puis MspI (C⁺CGG)											
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530	
Taille RRgenome (pb)	2 668 186 406	95 307 587	109 644 642	118 670 024	137 395 367	129 268 943	131 579 628	140 347 622	141 389 252	144 214 908	
Fraction du géome (%)	99,9	3,6	4,1	4,4	5,1	4,8	4,9	5,3	5,3	5,4	
Nombre de fragments générés	2736072	1 012 495	778 299	638 658	578 764	516 784	481 661	395 701	387 730	365 176	
Nombre de CpG analysables	27 540 276	3 941 393	3 868 044	3 585 074	3 565 293	3 247 802	3 180 349	2 850 554	2 820 660	2 737 756	
Fraction des CpG totaux analysables (%)	100,0	14,3	14,0	13,0	12,9	11,8	11,5	10,4	10,2	9,9	

MfeI (C⁺AATTG) puis MspI (C⁺CGG)											
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530	
Taille RRgenome (pb)	2 669 098 196	82 356 633	93 467 879	99 823 650	113 876 342	106 852 539	107 982 949	120 336 281	121 211 480	123 238 031	
Fraction du géome (%)	100,0	3,1	3,5	3,7	4,3	4,0	4,0	4,5	4,5	4,6	
Nombre de fragments générés	2432144	901840	675 327	542 334	483 188	429 644	396 662	335 320	328 788	309 647	
Nombre de CpG analysables	27 540 276	3 742 838	3 638 371	3 330 449	3 269 598	2 968 694	2 898 757	2 621 853	2 593 204	2 510 071	
Fraction des CpG totaux analysables (%)	100,0	13,6	13,2	12,1	11,9	10,8	10,5	9,5	9,4	9,1	

MseI (T⁺TAA) puis MspI (C⁺CGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRgènome (pb)	2 621 185 955	1 194 208 795	1 248 258 409	1 197 377 177	1 056 413 026	1 002 190 827	906 031 152	824 238 835	804 835 545	746 088 743		
Fraction du gènome (%)	98,2	44,7	46,7	44,8	39,6	37,5	33,9	30,9	30,1	27,9		
Nombre de fragments gènérés	18 402 884	15 206 621	10 065 589	7 153 484	4 584 597	3 946 365	3 047 692	2 428 212	2 301 259	1 966 102		
Nombre de CpG analysables	27 540 276	11 668 545	12 262 196	11 987 764	10 968 784	10 595 830	9 802 878	9 042 894	8 841 147	8 229 742		
Fraction des CpG totaux analysables (%)	100,0	42,4	44,5	43,5	39,8	38,5	35,6	32,8	32,1	29,9		

NcoI (C⁺CATGG) puis MspI (C⁺CGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRgènome (pb)	2 665 289 549	1 35 554 463	1 49 715 249	1 63 824 337	1 82 175 832	1 87 012 093	1 95 948 840	2 02 700 003	2 04 350 635	2 08 992 226		
Fraction du gènome (%)	99,8	5,1	5,6	6,1	6,8	7,0	7,3	7,6	7,7	7,8		
Nombre de fragments gènérés	3 701 688	1 602 569	1 052 283	872 757	719 811	680 715	622 784	570 764	559 528	528 665		
Nombre de CpG analysables	27 540 276	4 652 261	4 580 404	4 341 517	4 071 262	3 994 862	3 849 433	3 690 839	3 658 556	3 565 491		
Fraction des CpG totaux analysables (%)	100,0	16,9	16,6	15,8	16,2	14,5	14,0	13,4	13,3	12,9		

PciI (A⁺CATGT) puis MspI (C⁺CGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRgènome (pb)	2 667 823 073	1 01 516 734	1 16 843 377	1 27 009 740	1 37 615 877	1 40 452 882	1 46 193 075	1 50 620 308	1 51 903 512	1 55 659 154		
Fraction du gènome (%)	99,9	3,8	4,4	4,8	5,2	5,3	5,5	5,6	5,7	5,8		
Nombre de fragments gènérés	2 857 184	1 064 889	825 791	687 357	546 623	511 610	464 152	424 016	415 989	393 703		
Nombre de CpG analysables	27 540 276	4 016 503	3 953 964	3 682 486	3 311 334	3 252 978	3 090 690	2 939 833	2 912 184	2 833 868		
Fraction des CpG totaux analysables (%)	100,0	14,6	14,4	13,4	12,0	11,8	11,2	10,7	10,6	10,3		

SpeI (A⁺CTAGT) puis MspI (C⁺CGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRgènome (pb)	2 669 255 399	81 485 969	92 332 170	98 472 387	105 241 145	107 013 451	110 434 325	112 761 581	113 573 844	115 568 886		
Fraction du gènome (%)	100,0	3,1	3,5	3,7	3,9	4,0	4,1	4,2	4,3	4,3		
Nombre de fragments gènérés	2 379 743	895 603	666 468	535 716	423 065	392 652	352 387	318 093	311 573	292 596		
Nombre de CpG analysables	27 540 276	3 733 600	3 633 290	3 332 152	2 971 690	2 908 288	2 736 219	2 577 132	2 547 364	2 458 740		
Fraction des CpG totaux analysables (%)	100,0	13,6	13,2	12,1	11,9	10,6	9,9	9,4	9,2	8,9		

XbaI (T⁺CTAGA) puis MspI (C⁺CGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRgènome (pb)	2 668 356 719	93 017 711	1 06 593 769	1 15 649 472	1 25 802 186	1 27 993 617	1 32 720 899	1 36 286 630	1 37 306 651	1 40 243 566		
Fraction du gènome (%)	99,9	3,5	4,0	4,3	4,7	4,8	5,0	5,1	5,1	5,3		
Nombre de fragments gènérés	2 679 303	993 118	757 981	621 938	504 058	468 412	422 345	384 207	376 509	355 359		
Nombre de CpG analysables	27 540 276	3 897 412	3 817 576	3 533 644	3 192 959	3 128 865	2 961 226	2 805 992	2 777 147	2 694 883		
Fraction des CpG totaux analysables (%)	100,0	14,2	13,9	12,8	12,7	11,4	10,8	10,2	10,1	9,8		

BseRI (G⁺AGGAG) puis MspI (C⁺CGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 370	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRgènome (pb)	2 666 266 202	1 30 077 229	1 48 293 185	1 54 614 763	1 61 832 602	1 67 642 248	1 73 564 349	1 81 897 402	1 83 263 496	1 86 934 604		
Fraction du gènome (%)	99,8	4,9	5,6	5,8	6,1	6,3	6,5	6,8	6,9	7,0		
Nombre de fragments gènérés	3 376 141	1 450 559	1 106 839	842 450	707 909	615 429	554 831	511 226	500 551	470 922		
Nombre de CpG analysables	27 540 276	4 762 573	4 677 075	4 273 398	3 968 238	3 755 271	3 563 946	3 408 801	3 371 592	3 269 773		
Fraction des CpG totaux analysables (%)	100,0	17,3	17,0	15,5	14,4	13,6	12,9	12,4	12,2	11,9		

BamHI (G⁺GATCC) puis MspI (C⁺CGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRgènome (pb)	2 666 266 202	83 403 607	93 742 247	99 075 942	1 11 988 840	1 05 741 047	1 07 873 752	1 09 211 923	1 09 658 846	1 11 201 186		
Fraction du gènome (%)	99,8	3,1	3,5	3,7	4,2	4,0	4,0	4,1	4,1	4,2		
Nombre de fragments gènérés	3 376 141	932 298	682 734	542 337	477 596	389 808	345 623	309 288	301 878	282 437		
Nombre de CpG analysables	27 540 276	3 832 053	3 711 094	3 392 464	3 322 582	2 939 011	2 749 888	2 578 263	2 545 237	2 455 317		
Fraction des CpG totaux analysables (%)	100,0	13,9	13,5	12,3	12,1	10,7	10,0	9,4	9,2	8,9		

BspHI (T¹CAIGA) puis MspI (C¹CGG)												
	Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530	
Taille RRénone (pb)	2 667 756 242	101 082 150	115 975 386	123 829 326	143 564 407	135 019 934	139 189 079	142 444 161	145 984 011	147 190 001	149 961 783	
Fraction du génome (%)	99,9	3,8	4,3	4,6	5,4	5,1	5,2	5,3	5,5	5,5	5,6	
Nombre de fragments générés	2 879 460	1 077 486	832 009	665 181	608 464	543 293	512 584	453 230	410 808	402 930	379 105	
Nombre de CpG analysables	27 540 276	3 982 056	3 912 267	3 587 462	3 573 519	3 250 929	3 192 961	3 015 439	2 856 465	2 829 318	2 738 196	
Fraction des CpG totaux analysables (%)	100,0	14,5	14,2	13,0	13,0	11,8	11,6	10,9	10,4	10,3	9,9	

AluI (AG¹CT) puis MspI (C¹CGG)												
	Aucune	0 - 250	40 - 290	80 - 330	120 - 370	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530	
Taille RRénone (pb)	2 637 633 236	858 955 562	964 879 356	1 010 471 107	1 085 003 979	987 191 174	970 945 702	935 286 248	885 812 713	863 356 743	815 393 014	
Fraction du génome (%)	98,8	32,2	36,1	37,8	40,6	37,0	36,4	35,0	33,2	32,3	30,5	
Nombre de fragments générés	12 920 444	9 229 131	7 021 712	5 658 718	4 840 542	4 094 161	3 684 545	3 082 162	2 598 993	2 456 608	2 147 733	
Nombre de CpG analysables	27 540 276	12 560 571	12 688 080	11 951 511	11 648 315	10 485 407	9 925 177	9 098 390	8 266 908	7 991 564	7 373 636	
Fraction des CpG totaux analysables (%)	100,0	45,6	46,1	43,4	42,3	38,1	36,0	33,0	30,0	29,0	26,8	

AluI (GGATC(4/5)) puis MspI (C¹CGG)												
	Aucune	0 - 250	40 - 290	80 - 330	120 - 370	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530	
Taille RRénone (pb)	2 664 817 466	153 497 221	175 687 959	189 883 403	225 441 872	213 740 384	219 559 457	228 490 401	231 863 829	231 648 933	236 903 386	
Fraction du génome (%)	99,8	5,7	6,6	7,1	8,4	8,0	8,2	8,6	8,7	8,7	8,9	
Nombre de fragments générés	3 854 143	1 641 913	1 214 596	1 000 174	938 708	849 340	804 013	736 985	652 823	632 072	599 446	
Nombre de CpG analysables	27 407 420	4 832 501	4 781 449	4 516 838	4 615 066	4 238 377	4 186 184	4 022 799	3 769 867	3 726 009	3 639 712	
Fraction des CpG totaux analysables (%)	99,5	17,5	17,4	16,4	16,8	15,4	15,2	14,6	13,7	13,5	13,2	

AseI (AT¹TAAT) puis MspI (C¹CGG)												
	Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530	
Taille RRénone (pb)	2 666 968 613	111 107 910	129 546 783	142 300 713	167 660 829	158 216 519	161 959 194	169 125 420	175 726 479	177 387 741	182 271 394	
Fraction du génome (%)	99,9	4,2	4,9	5,3	6,3	5,9	6,1	6,3	6,6	6,6	6,8	
Nombre de fragments générés	3 142 005	1 151 430	899 070	756 429	701 593	629 561	591 680	536 770	493 511	484 583	459 940	
Nombre de CpG analysables	27 540 276	3 960 672	3 911 138	3 651 688	3 669 547	3 349 839	3 298 635	3 143 604	3 016 505	2 993 472	2 932 726	
Fraction des CpG totaux analysables (%)	100,0	14,4	14,2	13,3	13,3	12,2	12,0	11,4	11,0	10,9	10,6	

BbsI (GAAGAC(2/6)) puis MspI (C¹CGG)												
	Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530	
Taille RRénone (pb)	2 668 028 210	101 354 909	120 852 929	132 912 042	150 392 205	142 041 944	143 999 207	148 426 957	154 750 192	149 002 892	150 889 929	
Fraction du génome (%)	99,9	3,8	4,5	5,0	5,6	5,3	5,4	5,6	5,8	5,6	5,7	
Nombre de fragments générés	2 787 260	1 061 403	834 728	700 165	633 036	569 305	532 031	481 636	443 927	408 426	382 679	
Nombre de CpG analysables	27 498 013	4 058 234	4 033 781	3 775 030	3 737 403	3 411 195	3 341 102	3 164 007	3 034 114	2 932 375	2 835 968	
Fraction des CpG totaux analysables (%)	99,8	14,7	14,6	13,7	13,6	12,4	12,1	11,5	11,0	10,6	10,3	

BclI (CCATC(4/5)) puis MspI (C¹CGG)												
	Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530	
Taille RRénone (pb)	2 660 257 229	235 933 318	266 767 942	285 868 767	338 377 994	319 771 509	327 655 961	362 305 624	367 718 085	369 817 799	380 229 755	
Fraction du génome (%)	99,6	8,8	10,0	10,7	12,7	12,0	12,3	13,6	13,8	13,8	14,2	
Nombre de fragments générés	5 372 650	2 654 045	1 911 455	1 531 478	1 428 200	1 286 404	1 207 451	1 137 120	1 020 194	1 000 862	956 689	
Nombre de CpG analysables	27 377 165	5 876 275	5 855 575	5 555 511	5 722 370	5 243 692	5 167 001	5 139 386	4 891 822	4 847 437	4 754 944	
Fraction des CpG totaux analysables (%)	99,4	21,3	21,3	20,2	20,8	19,0	18,8	18,7	17,8	17,6	17,3	

BclII (GTATCC(6/5)) puis MspI (C¹CGG)												
	Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530	
Taille RRénone (pb)	2 669 514 119	80 004 866	90 052 205	95 476 420	108 135 100	101 338 775	102 365 725	104 723 563	106 516 271	107 180 159	108 933 678	
Fraction du génome (%)	100,0	3,0	3,4	3,6	4,0	3,8	3,8	3,9	4,0	4,0	4,1	
Nombre de fragments générés	2 290 432	884 734	654 596	521 601	460 514	408 714	376 859	334 896	301 040	294 473	276 323	
Nombre de CpG analysables	27 525 128	3 736 151	3 622 352	3 310 592	3 241 008	2 941 288	2 871 165	2 690 921	2 527 291	2 497 365	2 413 641	
Fraction des CpG totaux analysables (%)	99,9	13,6	13,2	12,0	11,8	10,7	10,4	9,8	9,2	9,1	8,8	

Bmrl (ACTGGG(5/4)) puis Mspl (C^CGG)

	Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530
Taille RRgenome (pb)	2 668 241 045	100 909 269	114 666 083	122 548 733	139 319 026	130 675 942	132 128 322	135 623 938	138 047 624	138 774 783	140 849 273
Fraction du génome (%)	99,9	3,8	4,3	4,6	5,2	4,9	4,9	5,1	5,2	5,2	5,3
Nombre de fragments générés	2 717 861	1 086 946	825 237	667 029	592 298	526 345	487 710	434 038	390 324	381 410	357 081
Nombre de CpG analysables	27 540 276	4 183 350	4 096 589	3 797 337	3 732 530	3 395 766	3 290 955	3 095 277	2 906 599	2 870 589	2 775 820
Fraction des CpG totaux analysables (%)	100,0	15,2	14,9	13,8	13,6	12,3	11,9	11,2	10,6	10,4	10,1

Bmrl (GCTAG^C) puis Mspl (C^CGG)

	Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530
Taille RRgenome (pb)	2 669 848 088	77 118 794	86 338 462	91 192 578	102 105 455	95 535 514	96 086 545	98 226 639	99 572 682	99 979 120	100 983 668
Fraction du génome (%)	100,0	2,9	3,2	3,4	3,8	3,6	3,6	3,7	3,7	3,7	3,8
Nombre de fragments générés	2 181 444	865 371	633 248	500 984	436 433	386 339	354 434	314 425	282 026	275 346	256 824
Nombre de CpG analysables	27 528 542	3 703 019	3 579 013	3 265 420	3 171 203	2 874 516	2 795 461	2 613 200	2 449 296	2 417 157	2 322 039
Fraction des CpG totaux analysables (%)	100,0	13,4	13	11,9	11,5	10,4	10,2	9,5	8,9	8,8	8,4

BtsCI (GGATG(2/0)) puis Mspl (C^CGG)

	Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530
Taille RRgenome (pb)	2 661 974 291	197 458 072	246 866 357	268 509 298	316 840 138	299 871 401	309 039 558	322 403 771	325 796 507	327 434 164	311 455 331
Fraction du génome (%)	99,7	7,4	9,2	10,1	11,9	11,2	11,6	12,1	12,2	12,3	11,7
Nombre de fragments générés	4 806 771	2 223 592	1 629 734	1 393 537	1 315 449	1 186 234	1 130 705	1 039 282	937 408	918 598	791 742
Nombre de CpG analysables	27 540 276	5 576 355	5 673 581	5 411 812	5 549 857	5 089 346	5 029 808	4 830 497	4 586 286	4 546 940	4 265 603
Fraction des CpG totaux analysables (%)	100,0	20,2	20,6	19,7	20,2	18,5	18,3	17,5	16,7	16,5	15,5

BtsI (GCGATG(2/0)) puis Mspl (C^CGG)

	Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530
Taille RRgenome (pb)	2 667 936 869	104 460 557	118 631 605	126 135 041	144 013 276	135 403 616	136 950 076	140 294 416	149 596 909	150 573 473	153 284 454
Fraction du génome (%)	99,9	3,9	4,4	4,7	5,4	5,1	5,1	5,3	5,6	5,6	5,7
Nombre de fragments générés	2 819 252	1 122 905	851 128	681 495	610 380	544 710	505 598	451 292	417 006	408 336	384 626
Nombre de CpG analysables	27 540 276	4 209 893	4 117 979	3 806 000	3 767 165	3 428 255	3 339 525	3 139 935	3 059 112	3 025 757	2 930 326
Fraction des CpG totaux analysables (%)	100,0	15,3	15,0	13,8	13,7	12,4	12,1	11,4	11,1	11,0	10,6

BtsImspl (CAGTIG(2/0)) puis Mspl (C^CGG)

	Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530
Taille RRgenome (pb)	2 658 411 065	292 006 044	343 160 193	376 344 068	439 161 644	414 241 546	422 765 195	436 186 042	445 683 340	447 239 308	454 556 835
Fraction du génome (%)	99,6	10,9	12,9	14,1	16,4	15,5	15,8	16,3	16,7	16,7	17
Nombre de fragments générés	5 994 515	2 926 505	2 352 208	1 998 202	1 856 494	1 666 064	1 569 037	1 391 253	1 259 579	1 230 218	1 149 942
Nombre de CpG analysables	27 540 276	6 823 898	6 931 503	6 713 460	6 903 671	6 321 567	6 168 513	5 926 974	5 680 771	5 620 019	5 499 768
Fraction des CpG totaux analysables (%)	100,0	24,8	25,2	24,4	25,1	23,0	22,4	21,5	20,6	20,4	20,0

Dral (TTT^AAA) puis Mspl (C^CGG)

	Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530
Taille RRgenome (pb)	2 662 048 889	210 393 804	249 101 598	278 293 717	327 022 518	308 670 109	315 658 533	327 422 474	336 722 382	337 969 109	342 517 491
Fraction du génome (%)	99,7	7,9	9,3	10,4	12,2	11,6	11,8	12,3	12,6	12,7	12,8
Nombre de fragments générés	4 781 910	2 078 089	1 697 472	1 469 340	1 369 554	1 229 253	1 158 162	1 045 494	952 759	929 294	869 830
Nombre de CpG analysables	27 540 276	4 729 018	4 848 466	4 741 646	4 949 283	4 558 062	4 527 582	4 404 626	4 297 100	4 273 603	4 202 938
Fraction des CpG totaux analysables (%)	100,0	17,2	17,6	17,2	18,0	16,6	16,4	16,0	15,6	15,5	15,3

HaeIII (GG^CC) puis Mspl (C^CGG)

	Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530
Taille RRgenome (pb)	2 649 920 779	431 598 420	469 553 140	478 960 055	524 869 243	472 156 191	466 831 710	462 034 257	458 513 245	457 033 358	451 374 342
Fraction du génome (%)	99,2	16,2	17,6	17,9	19,7	17,7	17,5	17,3	17,2	17,1	16,9
Nombre de fragments générés	8 182 289	5 220 287	3 549 369	2 732 825	2 355 755	1 955 432	1 773 760	1 489 170	1 307 844	1 267 953	1 154 802
Nombre de CpG analysables	26 607 056	9 233 989	8 630 781	7 651 101	7 274 010	6 316 294	5 967 629	5 400 338	5 014 830	4 926 989	4 664 338
Fraction des CpG totaux analysables (%)	96,6	33,5	31,3	27,8	26,4	22,9	21,7	19,6	18,2	17,9	16,9

HpyCH4V (TG^CA) puis MspI (C^CGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRgènome (pb)	2 631 414 836	1 076 847 247	1 211 242 645	1 243 785 416	1 157 271 458	1 132 119 997	1 051 375 772	918 992 250	893 763 694	812 102 207		
Fraction du gènome (%)	98,5	40,3	45,4	46,6	43,3	42,4	39,4	34,4	33,5	30,4		
Nombre de fragments gènérés	14 993 240	11 447 312	8 837 613	7 042 309	4 872 542	4 424 753	3 595 338	2 743 243	2 588 057	2 156 769		
Nombre de CpG analysables	27 540 276	14 706 618	15 014 834	13 852 259	11 842 548	11 230 779	9 911 313	8 329 239	29,2	7 060 211		
Fraction des CpG totaux analysables (%)	100,0	53,4	54,5	50,3	43,0	40,8	36,0	30,2	8 035 139,0	25,6		

KpnI (GGTAC^C) puis MspI (C^CGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRgènome (pb)	2 669 728 277	78 777 809	87 917 884	92 619 692	104 136 269	97 497 033	98 098 659	100 808 178	101 269 639	103 187 023		
Fraction du gènome (%)	100,0	3,0	3,3	3,5	3,9	3,7	3,7	3,8	3,8	3,9		
Nombre de fragments gènérés	2 217 928	882 536	643 896	509 527	445 422	394 823	362 348	285 354	278 636	261 578		
Nombre de CpG analysables	27 528 991,0	3 745 726	3 611 834	3 291 742	3 203 801	2 904 891	2 829 739	2 467 737	2 435 286	2 350 298		
Fraction des CpG totaux analysables (%)	100,0	13,6	13,1	12,0	11,6	10,5	10,3	9,0	8,8	8,5		

MluCI (A^AATT) puis MspI (C^CGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRgènome (pb)	2 618 123 831	1 290 719 590	1 369 697 324	1 311 117 090	1 258 369 935	1 138 255 082	1 085 373 808	860 978 819	829 617 351	745 321 620		
Fraction du gènome (%)	98,0	48,3	51,3	49,1	47,1	42,6	40,6	32,2	31,1	27,9		
Nombre de fragments gènérés	19 423 585	16 284 322	10 954 770	7 841 301	5 843 627	4 922 911	4 314 553	2 581 819	2 409 457	1 992 838		
Nombre de CpG analysables	27 540 276	12 389 141	13 209 384	12 950 202	12 804 469	11 682 637	11 291 059	9 311 028	9 050 718	8 286 819		
Fraction des CpG totaux analysables (%)	100,0	45,0	48,0	47,0	46,5	42,4	41,0	33,8	32,9	30,1		

MspI (TGG^CCA) puis MspI (C^CGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRgènome (pb)	2 667 388 790	114 025 444	129 503 283	138 744 581	158 033 973	143 192 419	142 713 721	160 497 763	161 563 726	165 011 780		
Fraction du gènome (%)	99,9	4,3	4,8	5,2	5,9	5,4	5,3	6	6,1	6,2		
Nombre de fragments gènérés	3 001 946	1 187 843	926 029	767 822	697 932	586 338	521 576	490 877	439 946	415 823		
Nombre de CpG analysables	27 540 276	4 409 239	4 338 635	4 037 632	4 007 026	3 553 691	3 388 259	3 278 455	3 083 306	2 998 218		
Fraction des CpG totaux analysables (%)	100,0	16,0	15,8	14,7	14,5	12,9	12,3	11,3	11,2	10,9		

NdeI (CA^TATG) puis MspI (C^CGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRgènome (pb)	2 668 303 271	91 795 602	105 771 215	114 739 925	133 143 294	125 186 806	127 348 629	137 684 725	138 734 694	141 669 342		
Fraction du gènome (%)	99,9	3,4	4,0	4,3	5,0	4,7	4,8	5,2	5,2	5,3		
Nombre de fragments gènérés	2 697 119	980 359	749 012	616 072	560 248	499 495	465 329	387 453	379 724	358 100		
Nombre de CpG analysables	27 540 276	3 835 267	3 758 065	3 478 862	3 456 583	3 146 556	3 082 549	2 777 981	2 750 419	2 670 718		
Fraction des CpG totaux analysables (%)	100,0	13,9	13,6	12,6	12,6	11,4	11,2	10,1	10,0	9,7		

NsiI (ATGCA^T) puis MspI (C^CGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRgènome (pb)	2 667 783 110	100 115 950	116 277 857	127 195 751	148 525 624	139 622 897	142 343 514	152 943 516	154 230 881	157 762 110		
Fraction du gènome (%)	99,9	3,7	4,4	4,8	5,6	5,2	5,3	5,7	5,8	5,9		
Nombre de fragments gènérés	2 870 505	1 045 588	813 193	678 857	624 279	556 628	520 665	430 880	422 500	398 788		
Nombre de CpG analysables	27 540 276	4 008 690	3 967 598	3 715 026	3 728 226	3 391 686	3 326 560	3 011 702	2 981 956	2 899 673		
Fraction des CpG totaux analysables (%)	100,0	14,6	14,4	13,5	13,5	12,3	12,1	10,9	10,8	10,5		

PacI (TTAAT^TAA) puis MspI (C^CGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRgènome (pb)	2 670 018 926	74 707 734	83 400 767	87 568 735	98 058 740	91 755 230	92 149 243	94 690 791	95 115 637	96 498 115		
Fraction du gènome (%)	100,0	2,8	3,1	3,3	3,7	3,4	3,5	3,5	3,6	3,6		
Nombre de fragments gènérés	2 125 234	839 823	612 114	481 405	420 024	371 985	340 557	268 081	261 787	245 058		
Nombre de CpG analysables	27 540 276	3 643 884	3 515 138	3 188 231	3 098 202	2 807 259	2 731 013	2 380 342	2 349 198	2 266 355		
Fraction des CpG totaux analysables (%)	100,0	13,2	12,8	11,6	11,2	10,2	9,9	8,6	8,5	8,2		

PstII (TTA^TAA) puis MspI (C^CGG)												
Taille RRénone (pb)	Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530	
	2 666 766 623	114 144 366	133 411 453	147 181 614	173 467 690	163 635 448	167 627 951	175 988 057	182 451 928	183 979 843	189 111 524	
Fraction du génome (%)	99,9	4,3	5,0	5,5	6,5	6,1	6,3	6,6	6,8	6,9	7,1	
Nombre de fragments générés	3 209 335	1 173 142	925 013	781 666	725 008	649 991	611 017	558 513	513 202	502 963	477 386	
Nombre de CpG analysables	27 540 276	3 963 361	3 922 633	3 678 794	3 706 335	3 384 533	3 334 917	3 196 569	3 066 356	3 041 078	2 981 825	
Fraction des CpG totaux analysables (%)	100,0	14,4	14,2	13,4	13,5	12,3	12,1	11,6	11,1	11,0	10,8	

PstII (CTGCA^G) puis MspI (C^CGG)												
Taille RRénone (pb)	Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530	
	2 665 654 469	140 072 885	159 249 348	170 245 635	192 571 812	180 916 002	184 464 994	191 361 833	214 078 272	220 371 355	230 602 951	
Fraction du génome (%)	99,8	5,2	6,0	6,4	7,2	6,8	6,9	7,2	8,0	8,3	8,6	
Nombre de fragments générés	3 580 051	1 539 239	1 126 109	918 735	813 144	724 311	680 738	618 336	590 960	587 289	566 507	
Nombre de CpG analysables	27 540 276	4 900 863	4 851 660	4 503 895	4 426 576	4 024 917	3 948 967	3 747 267	3 770 037	3 765 634	3 675 268	
Fraction des CpG totaux analysables (%)	100,0	17,8	17,6	16,4	16,1	14,6	14,3	13,6	13,7	13,7	13,3	

PvuII (CAG^CTG) puis MspI (C^CGG)												
Taille RRénone (pb)	Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530	
	2 667 414 224	116 066 031	131 206 498	138 938 211	156 598 017	146 529 192	147 955 949	151 185 663	153 380 133	153 989 519	155 802 412	
Fraction du génome (%)	99,9	4,3	4,9	5,2	5,9	5,5	5,5	5,7	5,7	5,8	5,8	
Nombre de fragments générés	2 993 461	1 277 501	966 555	761 126	667 514	590 688	546 403	484 392	434 737	424 335	396 244	
Nombre de CpG analysables	27 540 276	4 633 778	4 532 624	4 113 716	4 007 345	3 625 189	3 533 528	3 306 042	3 106 061	3 065 837	2 948 135	
Fraction des CpG totaux analysables (%)	100,0	16,8	16,5	14,9	14,6	13,2	12,8	12,0	11,3	11,1	10,7	

SacI-HF (GAGCT^C) puis MspI (C^CGG)												
Taille RRénone (pb)	Aucune	0 - 250	40 - 290	80 - 330	120 - 370	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530	
	2 668 919 171	91 506 070	103 124 766	109 426 727	113 526 977	115 300 852	115 994 629	118 366 434	120 142 983	120 909 202	122 273 538	
Fraction du génome (%)	99,9	3,4	3,9	4,1	4,3	4,3	4,3	4,4	4,5	4,5	4,6	
Nombre de fragments générés	2 488 646	1 014 075	751 953	598 987	499 412	465 167	428 123	379 649	340 512	332 974	310 898	
Nombre de CpG analysables	27 512 803	4 043 206	3 921 729	3 603 574	3 312 538	3 190 938	3 080 223	2 883 975	2 708 235	2 676 132	2 575 969	
Fraction des CpG totaux analysables (%)	99,9	14,7	14,2	13,1	12,0	11,6	11,2	10,5	9,8	9,7	9,4	

SbfI (CCTGCA^GG) puis MspI (C^CGG)												
Taille RRénone (pb)	Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530	
	2 670 209 441	74 724 541	82 988 924	86 569 101	96 247 885	89 918 405	90 126 471	91 025 835	91 446 602	91 729 166	92 583 567	
Fraction du génome (%)	100,0	2,8	3,1	3,2	3,6	3,4	3,4	3,4	3,4	3,4	3,5	
Nombre de fragments générés	2 061 729	848 142	613 743	478 782	413 619	365 373	333 557	292 411	259 618	253 153	235 627	
Nombre de CpG analysables	27 540 276	3 681 239	3 546 007	3 211 221	3 108 734	2 814 024	2 735 087	2 540 302	2 365 894	2 332 811	2 241 517	
Fraction des CpG totaux analysables (%)	100,0	13,4	12,9	11,7	11,3	10,2	9,9	9,2	8,6	8,5	8,1	

ScaI (AGT^ACT) puis MspI (C^CGG)												
Taille RRénone (pb)	Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530	
	2 668 153 523	97 283 640	110 754 360	119 456 216	138 382 418	130 217 521	127 773 448	138 353 002	143 419 633	144 560 571	148 132 441	
Fraction du génome (%)	99,9	3,6	4,1	4,5	5,2	4,9	4,8	5,2	5,4	5,4	5,5	
Nombre de fragments générés	2 747 035	1 021 266	786 312	651 422	599 617	537 607	464 610	433 596	400 649	393 316	373 390	
Nombre de CpG analysables	27 540 276	4 027 955	3 962 754	3 692 948	3 699 474	3 376 224	3 187 607	3 073 856	2 947 061	2 919 340	2 841 127	
Fraction des CpG totaux analysables (%)	100,0	14,6	14,4	13,4	13,4	12,3	11,6	11,2	10,7	10,6	10,3	

SphI (GCATG^C) puis MspI (C^CGG)												
Taille RRénone (pb)	Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530	
	2 668 959 218	87 418 392	98 155 683	104 519 545	118 714 059	111 098 569	112 302 538	115 045 850	117 309 803	117 970 112	120 005 720	
Fraction du génome (%)	99,9	3,3	3,7	3,9	4,4	4,2	4,2	4,3	4,4	4,4	4,5	
Nombre de fragments générés	2 476 966	976 801	710 617	570 513	505 617	447 550	413 158	367 594	331 586	324 194	304 416	
Nombre de CpG analysables	27 512 049	3 924 655	3 809 291	3 499 072	3 433 761	3 108 387	3 030 676	2 837 911	2 675 332	2 642 518	2 553 891	
Fraction des CpG totaux analysables (%)	99,9	14,3	13,8	12,7	12,5	11,3	11,0	10,3	9,7	9,6	9,3	

MspI (AATⁿAITT) puis MspI (CⁿCGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRegenome (pb)	157 898 386	187 033 580	210 509 691	252 715 142	239 179 094	246 747 624	259 428 780	270 339 879	273 322 215	282 973 045		
Fraction du génome (%)	5,9	7,0	7,9	9,5	9,0	9,2	9,7	10,1	10,2	10,6		
Nombre de fragments générés	1 552 619	1 269 994	1 107 365	1 048 632	945 207	897 421	819 313	758 010	745 834	713 666		
Nombre de CpG analysables	4 272 227	4 300 465	4 126 333	4 271 540	3 924 820	3 897 788	3 777 316	3 679 340	3 669 505	3 637 045		
Fraction des CpG totaux analysables (%)	15,5	15,6	15,0	15,5	14,3	14,2	13,7	13,4	13,3	13,2		

StuI (AGGⁿCCT) puis MspI (CⁿCGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRegenome (pb)	100 712 211	112 391 490	119 098 069	135 186 485	126 762 344	128 205 513	131 212 336	132 404 504	132 962 050	134 177 923		
Fraction du génome (%)	3,8	4,2	4,5	5,1	4,7	4,8	4,9	5,0	5,0	5,0		
Nombre de fragments générés	1 156 340	811 580	649 469	574 992	510 714	472 923	420 587	375 200	366 284	340 210		
Nombre de CpG analysables	4 241 797	4 092 956	3 751 683	3 685 371	3 346 789	3 263 292	3 063 441	2 861 159	2 822 912	2 722 208		
Fraction des CpG totaux analysables (%)	15,4	14,9	13,6	13,4	12,2	11,8	11,1	10,4	10,3	9,9		

SwaI (ATTⁿAAAT) puis MspI (CⁿCGG)												
Aucune	0 - 250	40 - 290	80 - 330	120 - 390	140-390	160 - 410	200 - 450	240 - 490	250-500	280 - 530		
Taille RRegenome (pb)	75 331 979	84 343 221	88 832 286	100 135 834	93 764 427	94 463 910	96 216 433	97 663 652	98 192 405	99 681 047		
Fraction du génome (%)	2,8	3,2	3,3	3,7	3,5	3,5	3,6	3,7	3,7	3,7		
Nombre de fragments générés	843 013	617 466	487 548	427 663	379 109	348 199	307 779	276 193	270 068	253 008		
Nombre de CpG analysables	3 639 022	3 515 584	3 196 005	3 115 148	2 823 945	2 752 197	2 566 266	2 404 314	2 375 673	2 290 134		
Fraction des CpG totaux analysables (%)	13,2	12,8	11,6	11,3	10,3	10,0	9,3	8,7	8,6	8,3		

Tableau n° 3. Amorceurs et conditions de PCR pour générer les matrices de pyroséquenceage

Nom de l'amorce	Nom du gène (numéro d'accèsion EMBL)	Chromosome	Élément génomique	Logiciel utilisé	Sens de l'amorce	Séquence de l'amorce (5' - 3')	Taille de l'amplicon	Température d'hybridation	Concentration en MgCl2
bRARB_bis_F1	RARB (ENSBTAG00000011518)	27	Intron	MethPrimer	sens	GGGAAGGAGGGATATTGTTATTAG	199 pb	60°C	2 mM
bRARB_bio_R1					antisens	AAAATAAAACTCCTAATCCCAACAC			
bSNIP1_bis_F1	SNIP1 (ENSBTAG00000016852)	3	Exon	MethPrimer	sens	AATTTTTGGGGTATTTTTGAG	325 pb	60°C	2 mM
bSNIP1_bio_R1					antisens	CAAATCTCAATAATCCTAACTAACC			
bWWC2_bis_F1	WWC2 (ENSBTAG00000012191)	27	Exon	Qiagen	sens	GTTTGGAGAATTAATGGGTAGAG	323 pb	58°C	1,5 mM
bWWC2_bio_R1					antisens	AAATAAAACAAAAACACAAACCACCTC			
bTLE3_bis_F1	TLE3 (ENSBTAG00000015580)	10	Intron	MethPrimer	sens	AAGTAAATTAATAATGGTAGAGTTAATAAGT	343 pb	60°C	2 mM
bTLE3_bio_R1					antisens	AAACTTTACAACCTAACTACAAAAAAA			
bLAMB2_bis_F1	LAMB2 (ENSBTAG00000022635)	22	Promoteur	Qiagen	sens	TTGGGTTTGTTAATTTAGGGTATTT	230 pb	56°C	1,5 mM
bLAMB2_bio_R1					antisens	CCAAAATCTCCCACTATTCATCCACTC			
bSMARCD3_bis_F1	SMARCD3 (ENSBTAG00000014372)	4	Intron	Qiagen	sens	ATTTTTGTGGGTTGGTTAGTTAAG	145 pb	56°C	1,5 mM
bSMARCD3_bio_R1					antisens	TCCTCCAAATCCAACCTCACCA			

Tableau n° 4. Amorceurs de pyroséquenceage

Nom du gène (numéro d'accèsion EMBL)	Amorce de pyroséquenceage	Séquence de l'amorce (5' - 3')	Produit de PCR cible	Positions des CpG
RARB (ENSBTAG00000011518)	pyr2 pyr3	GGAGTTTGTGTTT GGGTTATTTTTGA	bRARB_bis_F1 X bRARB_bio_R1	1 à 3 4 à 5
SNIP1 (ENSBTAG00000016852)	pyr3	GTGATGTATATTAT	bSNIP1_bis_F1 X bSNIP1_bio_R1	1 à 7
WWC2 (ENSBTAG00000012191)	pyr1 pyr2 pyr3	AGGTAGATAGTAAATTATGTTTT ATGAGTAGGTTTAGTGAGGT GGAGGAAGTAGTGGTTATGGAGA	bWWC2_bis_F1 X bWWC2_bio_R1	1 à 5 6 à 11 12 à 18
TLE3 (ENSBTAG00000015580)	pyr2 pyr3 pyr5	GTAATTTAATTAGT GTTATTTTATTTTAT GGTAGTAAGGAGAGG	bTLE3_bis_F1 X bTLE3_bio_R1	1 à 5 6 à 9 13 à 21
LAMB2 (ENSBTAG00000022635)	pyr1	ATGGATTTAGGAGTTTGGT	bLAMB2_bis_F1 X bLAMB2_bio_R1	1 à 6
SMARCD3 (ENSBTAG00000014372)	pyr1	TTTGAGGTAGGGAT	bSMARCD3_bis_F1 X bSMARCD3_bio_R1	1 à 2

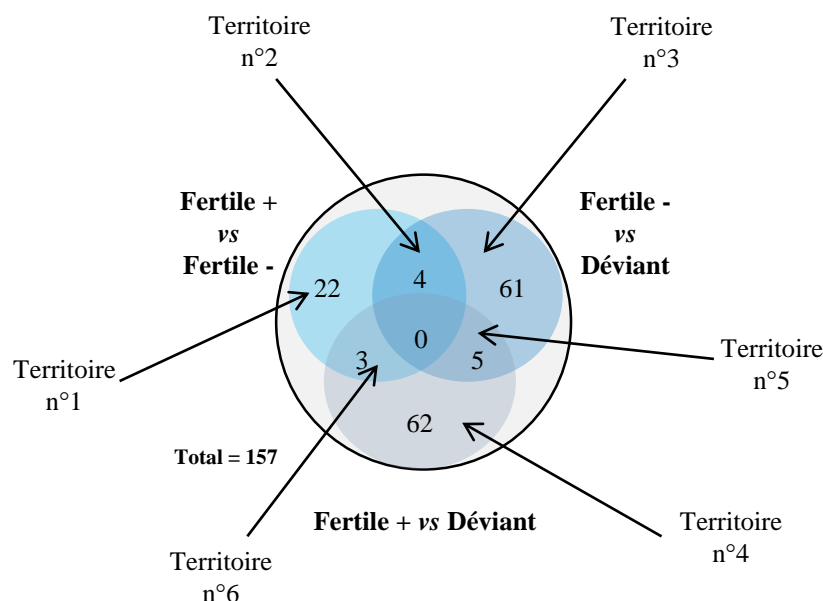
Tableau n° 5. Données générales issues des banques RRBS

Informations sur le taureau		Préparation des librairies et séquençage				Efficacité d'alignement			Couverture des CpG						
Race	Catégorie	Número du taureau	Série de séquençage	Nombre de paires de séquences analysées	Taux de conversion au bisulfite (%)	Séquence R1 tronquée car Pired-20 (%)	Séquence R2 tronquée car Pired-20 (%)	Nombre d'alignements uniques	Alignement uniques (%)	Alignement multiples (%)	Pas d'alignement (%)	Nombre de CpG couverts	Nombre de CpGs sur le chrY	Nombre de CpG avec couverture ≥ 10	Nombre de CpG avec couverture ≥ 500
Fertiles +		12	1	43 616 478	98	2,2	4	15 626 940	35,8	54,7	9,5	3 107 246	3 559	65,1	0,4
		14	4	34 227 891	98,9	1,9	5,5	13 609 835	39,8	51,1	9,1	3 236 731	4 017	61,5	0,4
		16	1	32 826 830	99,2	3,1	5	12 876 937	39,2	50,7	10,1	3 177 554	3 942	60,6	0,4
		21	1	37 441 800	99,2	3,1	5,4	14 588 346	39	50,7	10,3	3 190 027	3 996	62,9	0,4
		54	4	33 255 865	99,4	2,1	5,6	13 568 275	40,8	49,8	9,4	3 302 468	4 172	60,4	0,4
		56	4	27 400 349	99,5	1,9	6,1	11 145 437	40,7	49,7	9,6	3 359 379	4 189	54,8	0,3
		62	1	35 562 850	99,4	2,2	4,1	14 111 658	39,7	50,7	9,6	3 135 541	3 896	63,1	0,4
		63	4	29 070 976	99,6	3,1	3,8	11 981 894	41,2	49,4	9,4	3 202 874	3 764	58,4	0,3
Fertiles -		76	1	42 920 692	99,7	2,4	3,6	16 900 280	39,4	51,1	9,5	3 166 562	3 920	66,4	0,4
		86	4	31 274 197	99,4	3,2	3,9	12 263 934	39,2	51,2	9,6	3 201 913	3 793	59,1	0,4
		17	2	34 750 604	98,4	2,6	4	13 708 811	39,4	51,1	9,5	3 434 024	4 079	61,6	0,4
		18	4	30 237 945	99,4	2	5,8	11 949 738	39,5	51,3	9,2	3 235 554	3 781	58,1	0,4
		20	1	31 982 782	98,9	2,4	3,7	12 167 798	38	52,1	9,9	3 172 806	3 833	58,8	0,3
		22	4	39 401 532	99,5	1,9	5,7	16 089 490	40,8	50	9,2	3 458 347	4 642	62,6	0,4
		28	4	35 385 579	99,7	2,9	3,5	13 127 092	37,1	53,7	9,2	3 207 133	3 997	59,3	0,4
		36	5	32 220 820	99,2	2,3	6,9	13 362 572	41,5	48,9	9,6	3 327 022	4 323	60,3	0,4
Déviants		66	5	31 046 833	99,3	2,1	7,2	13 387 429	43,1	47,3	9,6	3 389 512	4 523	59,2	0,4
		77	5	37 591 375	99,3	2,3	7	15 796 160	42	48,2	9,8	3 350 040	4 519	63,9	0,4
		23	4	28 873 710	98,9	3	3,8	11 719 480	40,6	50,2	9,2	3 292 512	4 128	57,2	0,3
		24	1	42 360 913	99	3	5,2	16 562 633	39,1	50,4	10,5	3 275 324	4 226	64,4	0,4
		37	4	32 720 099	99,6	2,6	4,8	13 682 857	41,8	48,8	9,4	3 342 782	4 278	60,7	0,4
		41	5	37 790 131	98,9	2,1	7,3	16 582 065	43,9	46,4	9,7	3 424 954	4 738	64,7	0,4
		45	1	32 155 456	99,3	3,2	5,1	12 071 301	37,5	52,4	10,1	3 103 676	3 729	59,2	0,4
		48	4	34 981 631	99,4	1,9	5,8	14 205 847	40,6	49,7	9,7	3 309 305	4 220	62,2	0,4
	57	5	35 255 444	99	2	6,9	16 038 617	45,5	44,8	9,7	3 416 277	4 693	64	0,4	
	83	1	39 859 836	99,6	3,3	5	14 379 116	36,1	53,9	10	3 153 036	3 717	62,3	0,4	
	91	1	37 440 964	99,3	2,3	4,2	14 361 092	38,4	51,7	9,9	3 162 909	3 817	63,4	0,4	
	93	1	41 783 780	99,4	2,2	4	16 071 781	38,5	52	9,5	3 175 425	3 924	65,5	0,4	

Race	Informations sur le tumeur		Méthylation moyenne sur alignement unique (%)				Distribution de la méthylation (%)										
	Catégorie	Numéro du tumeur	Série de séquençage	Sans filtre sur la couverture	Couverture ≥ 10	Couverture ≥ 500	Méthylation en contexte CHG (%)	Méthylation en contexte CHH (%)	Sans filtre sur la couverture - hypométhylé (<20%)	Sans filtre sur la couverture - intermédiaire (>20% et <80%)	Sans filtre sur la couverture - hyperméthylé (>80%)	Couverture ≥ 10 - hypométhylé (<20%)	Couverture ≥ 10 - intermédiaire (>20% et <80%)	Couverture ≥ 10 - hyperméthylé (>80%)	Couverture ≥ 500 - hypométhylé (<20%)	Couverture ≥ 500 - intermédiaire (>20% et <80%)	Couverture ≥ 500 - hyperméthylé (>80%)
Femles +		12	1	53	50,3	22,6	3,5	3,8	42,7	8,9	48,4	46,1	7,9	46	50,7	48,9	0,4
		14	4	53,2	51,3	21,6	1,4	1,4	42,7	7,9	49,4	44,9	7	48,1	54,1	45,4	0,5
		16	1	52,4	50,2	18,6	1,1	1,1	43,3	8	48,7	45,7	7,4	46,9	58,6	41,1	0,3
		21	1	51,7	50,1	15,7	1,2	1,1	44,3	7,5	48,2	46,2	6,7	47,1	73	26,6	0,4
		54	4	54,1	52,8	21,5	0,9	0,8	41,9	7,2	50,9	43,4	6,6	50	56	43,2	0,8
		56	4	53,8	52	18,2	0,7	0,6	42,2	7,3	50,5	44,1	6,8	49,1	61,5	38,2	0,3
		62	1	53,1	51,3	20	0,8	0,7	43,2	6,8	50	45,2	6,2	48,6	58,1	41,3	0,6
		63	4	52,7	51,9	18	0,6	0,6	43,1	7,5	49,4	44,3	6,5	49,2	67,2	32,3	0,5
		76	1	53	50,5	23,4	0,5	0,5	43,2	6,8	50	45,7	6,4	47,9	47,2	52,4	0,4
		86	4	52,6	51,2	18,4	0,9	0,9	43,2	7,7	49,1	45	6,6	48,4	65,4	34,1	0,5
Femles -		17	2	54,3	51,5	21,2	2,7	2,9	41,5	8,6	49,9	44,8	7,7	47,5	55,3	44,4	0,3
		18	4	53,3	51,3	22,2	0,8	0,7	42,7	7,2	50,1	44,8	6,5	48,7	53,8	45,5	0,7
		20	1	53,8	49,4	26,4	1,7	1,7	42,1	8	49,9	46,8	7,4	45,8	39,2	60,4	0,4
		22	4	53,7	52,2	21,6	0,6	0,6	41,8	8,2	50	43,4	7,6	49	57	42,2	0,8
		28	4	53,1	47,6	24,3	0,5	0,5	42,8	7,5	49,7	48,4	7,1	44,5	45	54,2	0,8
		36	5	55,4	54,2	27,5	1,1	1,1	40,3	7,9	51,8	41,6	7,4	51	36,2	62,9	0,9
		66	5	53,7	53,6	15,9	0,9	0,9	42,2	7,6	50,2	42,4	7,1	50,5	72,8	26,7	0,5
		77	5	53,1	52,2	16,3	1	1	42,8	7,5	49,7	44	6,8	49,2	70,9	28,8	0,3
		23	4	53,3	51,2	22,2	1,6	1,7	42,2	8,7	49,1	44,7	7,5	47,8	52,3	47,4	0,3
		24	1	52,3	50,3	16,3	1,4	1,4	43,8	7,5	48,7	46	6,8	47,2	70,4	29,1	0,5
Déviants		37	4	53,7	51,8	16	0,6	0,6	42,3	7,3	50,4	44,3	7	48,7	72	27,8	0,2
		41	5	53,9	53	17,4	1,7	1,7	42,1	7,7	50,2	43,2	6,9	49,9	67	32,4	0,6
		45	1	51,8	49,4	15,7	1	0,9	44,2	7,5	48,3	46,8	6,8	46,4	70,1	29,7	0,2
		48	4	52,7	51,7	17,9	0,8	0,7	42,7	8,4	48,9	43,7	7,9	48,4	64	35,7	0,3
		57	5	53,8	52,8	19	1,6	1,7	41,8	8,5	49,7	43	7,8	49,2	60,6	39	0,4
		83	1	51,3	49,1	16,1	0,7	0,6	44,5	7,8	47,7	46,9	7,2	45,9	67,6	32	0,4
		91	1	51,4	49,3	14,8	1	1	44,5	7,6	47,9	46,9	6,9	46,2	74,1	25,6	0,3
		93	1	51,9	49,9	18,1	0,9	0,9	43,9	7,8	48,3	46	7,2	46,8	61,2	38,4	0,4

Informations sur le tableau		Méthylation moyenne sur alignement unique (%)				Distribution de la méthylation (%)											
Race	Catégorie	Nombre du tableau	Série de séquençage	Sans filtre sur la couverture	Couverture ≥ 10	Couverture ≥ 500	Méthylation en contexte CHG (%)	Méthylation en contexte CHH (%)	Sans filtre sur la couverture - intermédiaire (>20% et <=80%)	Sans filtre sur la couverture - hypométhylé (<20%)	Sans filtre sur la couverture - intermédiaire (>20% et <=80%)	Couverture ≥ 10 - intermédiaire (>20% et <=80%)	Couverture ≥ 10 - hypométhylé (<20%)	Couverture ≥ 500 - intermédiaire (>20% et <=80%)	Couverture ≥ 500 - hypométhylé (<20%)	Couverture ≥ 500 - intermédiaire (>20% et <=80%)	Couverture ≥ 500 - hypométhylé (<20%)
Mombéland	Femmes +	26	3	51,3	49,1	16,6	0,6	0,6	7,4	44,7	47,9	47	6,8	46,2	64,3	35,3	0,4
		33	5	52,2	51,1	22	0,6	0,5	8,9	42,9	48,2	44	8,5	47,5	53,9	45,8	0,3
		35	3	51,1	48	16,4	0,8	0,7	7,2	45,1	47,7	48,4	6,5	45,1	69,2	30,4	0,4
		38	3	52	49,6	18,2	2,8	2,9	8,1	44	47,9	47	7,1	45,9	61,1	38,5	0,4
		49	3	52,2	49,7	19	1,9	2	7,8	43,8	48,4	46,7	7	46,3	58,3	41,3	0,4
		50	3	51,2	48,6	15,4	1	0,9	7	45,1	47,9	47,9	6,3	45,8	72,3	27,3	0,4
		55	5	51,9	50,7	17,1	0,5	0,4	7,8	43,8	48,4	45,2	7,2	47,6	65,7	33,8	0,5
		78	6	53	51,8	18,2	0,2	0,2	6,8	43,2	50	44,4	6,4	49,2	62,6	36,9	0,5
		80	1	52,4	49,5	16,7	1,5	1,5	7	44	49	47,3	6,1	46,6	67,7	31,8	0,5
		81	3	52	49,6	16,8	2,2	2,3	7,7	44,1	48,2	47	6,8	46,2	66,6	33,1	0,3
Mombéland	Femmes -	9	3	52,3	50,2	18,2	2	2,1	8	43,6	48,4	46	7,3	46,7	60,4	39,1	0,5
		30	3	51,2	48,3	16,1	0,7	0,6	7,3	44,8	47,9	47,9	6,6	45,5	69,6	29,9	0,5
		32	6	52,5	50,6	17,2	0,5	0,4	7,8	43,3	48,9	45,1	7,4	47,5	67,3	32,3	0,4
		44	6	52,4	50,7	15,4	0,3	0,2	7,1	43,7	49,2	45,4	6,6	48	71,9	27,6	0,5
		71	3	51,1	48,1	15,5	1,6	1,6	7,6	45	47,4	48,4	6,7	44,9	73,7	25,8	0,5
		79	6	54	53,3	19	0,4	0,3	7	42,1	50,9	42,8	6,5	50,7	62,5	36,8	0,7
		82	3	52,1	49,5	16,5	0,6	0,5	6,4	44,3	49,3	47,1	5,7	47,2	67,6	32	0,4
		85	5	53,4	51,6	26,4	2,2	2,2	8,9	42	49,1	44,2	8,2	47,6	40	59,5	0,5
		90	5	53,1	51,4	22,3	1,8	1,9	8,3	42,6	49,1	44,5	7,7	47,8	52,6	47	0,4
		92	6	52,4	51,4	16,5	0,3	0,3	7,5	43,4	49,1	44,4	7,2	48,4	67,7	31,7	0,6
Dixvants	Dixvants	10	5	53,1	51,9	23,2	0,8	0,8	8,4	42,4	49,2	43,6	8	48,4	51,2	48,3	0,5
		11	3	52	49,4	16,8	0,9	0,8	6,6	44,4	49	47,3	5,8	46,9	68,1	31,4	0,5
		13	6	52,9	51,8	15,2	0,2	0,2	7,3	43,1	49,6	44	7,1	48,9	73,7	25,9	0,4
		15	6	52,6	51,6	15,5	0,5	0,4	7,1	43,5	49,4	44,5	6,7	48,8	72,7	26,7	0,6
		29	5	51,3	48,7	15,3	0,5	0,5	6,7	45,1	48,2	47,8	6,1	46,1	71,8	27,7	0,5
		39	6	52,5	51,2	19,4	0,2	0,2	7,8	43,2	49	44,4	7,5	48,1	59,9	39,7	0,4
		40	3	52,5	49,2	17,3	0,9	0,8	7,1	43,6	49,3	47,2	6,4	46,4	64	35,5	0,5
		61	6	53,1	52,1	18,6	0,3	0,2	7,4	42,7	49,9	43,7	7,1	49,2	62,5	37	0,5
		64	6	53,1	52,1	16,4	0,5	0,4	7,3	42,9	49,8	44	6,8	49,2	68,5	31,1	0,4
		72	5	52,7	51,6	16,9	0,8	0,7	7,5	43,3	49,2	44,4	7,1	48,5	64,8	34,9	0,3

Tableau n°6. Listes des gènes associés à une DMR pour la race Holstein



Territoire n°1

Gène	Nom Complet	Identifiant Ensembl
/	/	ENSBTAG00000046749
/	/	ENSBTAG00000043771
ARHGAP44	Rho GTPase activating protein 44	ENSBTAG00000021938
CCDC30	Coiled-Coil Domain Containing 30	ENSBTAG00000004585
CELSR1	Cadherin, EGF LAG Seven-pass G-type Receptor 1	ENSBTAG00000008036
COL4A1	Collagen type IV alpha 1	ENSBTAG00000012849
CORO2B	Coronin, actin binding protein 2B	ENSBTAG00000012681
EHMT2	Euchromatic Histone-lysine N-Methyltransferase 2	ENSBTAG00000005676
FANK1	Fibronectin type III and Ankyrin repeat domains 1	ENSBTAG00000022917
FUT7	Fucosyltransferase 7	ENSBTAG00000046516
LY86	Lymphocyte antigen 86	ENSBTAG00000031444
PLXND1	Plexin D1	ENSBTAG00000001814
PRKCZ	Protein Kinase C zeta	ENSBTAG00000014119
PTPRE	Protein Tyrosine Phosphatase Receptor type E	ENSBTAG00000002435
RIMS4	Regulating synaptic membrane exocytosis 4	ENSBTAG00000003382
RPP40	Ribonuclease P/MRP 40kDa subunit	ENSBTAG00000003332
SH2D4A	SH2 domain containing 4A	ENSBTAG00000001186
SLC22A23	Solute Carrier family 22 member 23	ENSBTAG00000010943
SPECC1L	Sperm antigen with Calponin homology and Coiled-Coil domains 1 Like	ENSBTAG00000021656
SSR1	Signal Sequence Receptor alpha	ENSBTAG00000022731
TRIM26	Tripartite Motif containing 26	ENSBTAG00000035744
WWOX	WW domain containing Oxidoreductase	ENSBTAG00000015894

Territoire n°2

Gène	Nom Complet	Identifiant Ensembl
BOLA	MHC class I heavy chain	ENSBTAG00000002069
EEFSEC	Eukaryotic Elongation Factor Selenocysteine-tRNA-specific	ENSBTAG00000030962
GAN	Gigaxonin	ENSBTAG00000009428
PTPRE	Protein Tyrosine Phosphatase Receptor type E	ENSBTAG00000002435

Territoire n°3

Gène	Nom Complet	Identifiant Ensembl
/	/	ENSBTAG00000026070
/	/	ENSBTAG00000025812
ACACA	Acetyl-CoA Carboxylase Alpha	ENSBTAG00000017567
ADAM12	ADAM metallopeptidase domain 12	ENSBTAG00000012444
ADORA2A	Adenosine A2a receptor	ENSBTAG00000016944
AMZ1	Archaealysin family metallopeptidase 1	ENSBTAG00000026194
ARHGEF1	Rho guanine nucleotide exchange factor 1	ENSBTAG00000018912
ARHGEF10	Rho guanine nucleotide exchange factor 10	ENSBTAG00000008585
BAHCC1	BAH domain and Coiled-Coil containing 1	ENSBTAG00000016776
BARHL2	BarH-Like homeobox 2	ENSBTAG00000005938
bta-mir-759	bta-mir-759	ENSBTAG00000030110
CCDC88B	Coiled-Coil Domain Containing 88B	ENSBTAG00000040190
CD46	Membrane cofactor protein isoform 2 precursor	ENSBTAG00000005397
CELF3	CUGBP Elav-Like Family member 3	ENSBTAG00000019131
CELSR1	Cadherin EGF LAG Seven-pass G-type Receptor 1	ENSBTAG00000008036
COL4A2	Collagen type IV alpha 2	ENSBTAG00000025210
CTBP1	C-Terminal Binding Protein 1	ENSBTAG00000008480
CXCR7	Chemokine (C-X-C motif) Receptor 7	ENSBTAG00000018424
EBF3	Early B-cell Factor 3	ENSBTAG00000008272
EFHD1	EF-Hand Domain family member D1	ENSBTAG00000014596
EXT2	Exostosin 2	ENSBTAG00000019644
FBXL22	F-Box and Leucine-rich repeat protein 22	ENSBTAG00000008807
GABRP	Gamma-Aminobutyric acid (GABA) A Receptor Pi	ENSBTAG00000002779
GGNBP1	Gametogenin Binding Protein 1	ENSBTAG00000011057
HIST3H2A	Histone cluster 3 H2a	ENSBTAG00000037964
HPS4	Hermansky-Pudlak Syndrome 4	ENSBTAG00000017038
HRNR	Hornerin	ENSBTAG00000011756
IGF2	Insulin-like Growth Factor 2	ENSBTAG00000013066
KCNQ2	Potassium Channel voltage gated KQT-like subfamily Q member 2	ENSBTAG00000003030
LARS2	Leucyl-tRNA Synthetase 2, mitochondrial	ENSBTAG00000007315
LRFN3	Leucine Rich repeat and Fibronectin type III domain containing 3	ENSBTAG00000000153
MATN2	Matrilin 2	ENSBTAG00000017165
MYO19	Myosin XIX	ENSBTAG00000012675
NFE2	Nuclear Factor Erythroid-derived 2	ENSBTAG00000001562
NRP2	Neuropilin 2	ENSBTAG00000011971
NSG1	Neuron-Specific protein family member 1	ENSBTAG00000005711
NUP210	Nucleoporin 210kDa	ENSBTAG00000003665
PADI2	Peptidyl Arginine Deiminase type II	ENSBTAG00000003403
PALLD	Palladin cytoskeletal associated protein	ENSBTAG00000001081
PLEKHA6	Pleckstrin Homology domain containing family A member 6	ENSBTAG00000004641
PLXND1	Plexin D1	ENSBTAG00000001814
PPFIA1	Protein Tyrosine Phosphatase receptor type F polypeptide (PTPRF) Interacting protein (liprin) alpha 1	ENSBTAG00000002188
PPP1R12B	Protein Phosphatase 1 Regulatory subunit 12B	ENSBTAG00000011772
REN	Renin	ENSBTAG00000021333
RHBDL2	Rhomboid veinlet-like 2	ENSBTAG00000008023
RIMS4	Regulating synaptic membrane exocytosis 4	ENSBTAG00000003382
RNF43	Ring Finger protein 43	ENSBTAG00000020119
RPP40	Ribonuclease P/MRP 40kDa subunit	ENSBTAG00000003332
RRP1	Ribosomal RNA Processing 1 homolog	ENSBTAG00000003981
SHANK3	SH3 and multiple Ankyrin repeat domains 3	ENSBTAG00000030180
SLC41A1	Solute Carrier family 41 member 1	ENSBTAG00000005379
SLC6A11	Solute Carrier family 6 member 11	ENSBTAG000000038962
SNIP1	Smad Nuclear Interacting Protein 1	ENSBTAG00000016852
SPSB4	splA/ryanodine receptor domain and SOCS Box containing 4	ENSBTAG00000009115
SRBD1	S1 RNA Binding Domain 1	ENSBTAG00000007508
TBC1D22A	TBC1 domain family member 22A	ENSBTAG00000012291
TLE3	Transducin-Like Enhancer of split 3	ENSBTAG00000015580
TRAPPC9	Trafficking Protein Particle Complex 9	ENSBTAG00000013955
VPS18	Vacuolar Protein Sorting 18 homolog	ENSBTAG00000011567
VSTM2B	V-set and transmembrane domain containing 2B	ENSBTAG00000018244
VWC2	Von Willebrand factor C domain containing 2	ENSBTAG00000044068

Territoire n°4

Gène	Nom Complet	Identifiant Ensembl
/	/	ENSBTAG00000015759
/	/	ENSBTAG00000026070
/	/	ENSBTAG00000047148
/	/	ENSBTAG00000019547
ABCG2	ATP Binding Cassette, sub-family G (WHITE) member 2	ENSBTAG00000017704
ACACA	Acetyl-CoA Carboxylase Alpha	ENSBTAG00000017567
ACSL6	Acyl-CoA Synthetase Long-chain family member 6	ENSBTAG00000019708
ADORA2A	Adenosine A2a Receptor	ENSBTAG00000016944
BARHL2	BarH-like Homeobox 2	ENSBTAG00000005938
BRD3	Bromodomain containing 3	ENSBTAG00000010582
C2CD2L	C2CD2-Like	ENSBTAG00000015013
CAMSAP3	Calmodulin regulated Spectrin-Associated Protein family member 3	ENSBTAG00000011356
CCDC141	Coiled-Coil Domain Containing 141	ENSBTAG00000027875
CCDC88B	Coiled-Coil Domain Containing 88B	ENSBTAG00000040190
CHD7	Chromodomain Helicase DNA binding protein 7	ENSBTAG00000021841
CXXC5	CXXC finger protein 5	ENSBTAG00000003986
DSP	Desmoplakin	ENSBTAG00000015106
EBF3	Early B-cell Factor 3	ENSBTAG00000008272
EFHD1	EF-hand domain family member D1	ENSBTAG00000014596
ETS2	v-ets Erythroblastosis virus E26 oncogene homolog 2 (avian)	ENSBTAG00000009214
F13A1	Coagulation Factor XIII A1 polypeptide	ENSBTAG00000007268
FAM174B	Family with sequence similarity 174 member B	ENSBTAG00000022989
GGNBP1	Gametogenetin Binding Protein 1	ENSBTAG00000011057
GMDS	GDP-Mannose 4,6-Dehydratase	ENSBTAG00000012058
GPD1L	Glycerol-3-Phosphate Dehydrogenase 1-Like	ENSBTAG00000009826
GSX2	GS homeobox 2	ENSBTAG00000045812
HDAC4	Histone Deacetylase 4	ENSBTAG00000017764
IRF7	Interferon Regulatory Factor 7	ENSBTAG00000047680
KLHDC7B	Kelch Domain Containing 7B	ENSBTAG00000046800
KNDC1	Kinase Non-catalytic C-lobe domain (KIND) containing 1	ENSBTAG00000011092
LAMB2	Laminin beta 2	ENSBTAG00000022635
LRFN3	Leucine Rich repeat and Fibronectin type III domain containing 3	ENSBTAG00000000153
NRP1	Neuropilin 1	ENSBTAG00000019866
NUDCD1	NudC Domain Containing 1	ENSBTAG00000007386
PARVG	Parvin gamma	ENSBTAG00000008057
PEX14	Peroxisomal biogenesis factor 14	ENSBTAG00000013538
PLA2G2D4	Phospholipase A2	ENSBTAG00000046977
PLA2G4D	Phospholipase A2 group IV D	ENSBTAG00000005268
PLCG2	Phospholipase C Gamma 2	ENSBTAG00000002103
PLCL2	Phospholipase C-Like 2	ENSBTAG00000009942
PLEKHM3	Pleckstrin Homology domain containing, family M member 3	ENSBTAG00000007830
PTPRN2	Protein Tyrosine Phosphatase Receptor type N polypeptide 2	ENSBTAG00000004958
RBM20	RNA Binding Motif protein 20	ENSBTAG00000023891
REXO4	REX4 RNA exonuclease 4 homolog (S. cerevisiae)	ENSBTAG00000019721
RRP1	Ribosomal RNA Processing 1 homolog (S. cerevisiae)	ENSBTAG00000003981
SEMA3F	Sema domain immunoglobulin domain (Ig) short basic domain,secreted (semaphorin) 3F	ENSBTAG00000018307
SERPIND1	Serpin peptidase inhibitor clade D (heparin cofactor) member 1	ENSBTAG00000013973
SH2B1	SH2B adaptor protein 1	ENSBTAG00000019219
SIPAIL1	Signal-Induced Proliferation-Associated 1 Like 1	ENSBTAG00000015805
SLC16A6	Solute Carrier family 16 member 6	ENSBTAG00000003074
SNIP1	Smad Nuclear Interacting Protein 1	ENSBTAG00000016852
SSH1	Slingshot protein phosphatase 1	ENSBTAG00000018825
TEKT5	Tektin-5	ENSBTAG00000017015
TENM3	Teneurin transmembrane protein 3	ENSBTAG00000021133
TFIP11	Tuftelin Interacting Protein 11	ENSBTAG00000012288
TLE3	Transducin-Like Enhancer of split 3	ENSBTAG00000015580
TP53I3	Tumor Protein p53 inducible protein 3	ENSBTAG00000003405
TPPP	Tubulin Polymerization Promoting Protein	ENSBTAG00000047116
URB1	Nucleolar pre-ribosomal-associated protein 1	ENSBTAG00000012412
VWC2	Von Willebrand factor C domain containing 2	ENSBTAG00000044068
ZDHHC5	Zinc finger, DHHC-type Containing 5	ENSBTAG00000002381
ZNF469	Zinc Finger protein 469	ENSBTAG00000045507

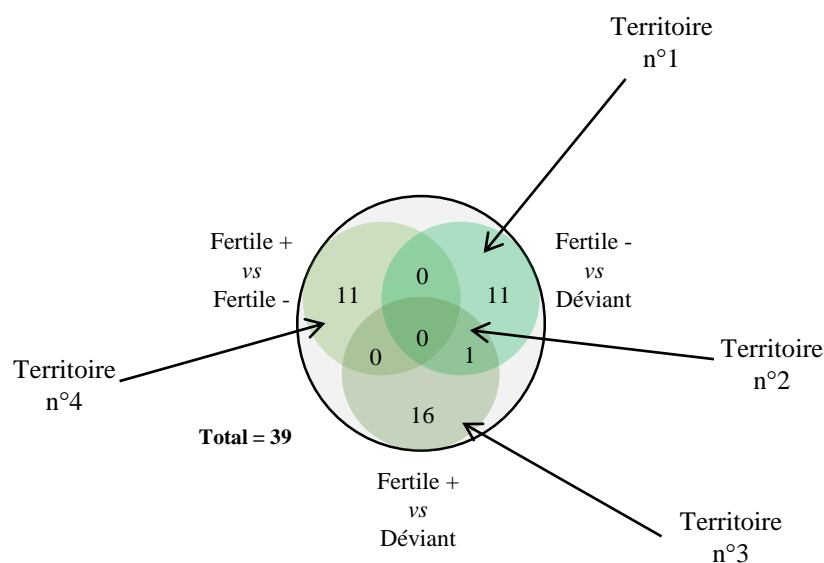
Territoire n°5

Gène	Nom Complet	Identifiant Ensembl
AHCTF1	AT Hook Containing Transcription Factor 1	ENSBTAG00000013949
C6orf222	Chromosome 6 Open Reading Frame 222	ENSBTAG00000017617
KCNIP1	Kv Channel Interacting Protein 1	ENSBTAG00000002865
RARB	Retinoic Acid Receptor Beta	ENSBTAG00000011518
WWC2	WW and C2 domain containing 2	ENSBTAG00000012191

Territoire n°6

Gène	Nom Complet	Identifiant Ensembl
COL25A1	Collagen, type XXV alpha 1	ENSBTAG00000017282
IGLL1	ImmunoGlobulin Lambda-Like polypeptide 1	ENSBTAG00000031160
MAD1L1	Mitotic Arrest Deficient-Like 1	ENSBTAG00000002474

Tableau n°7. Listes et répartition des DMR identifiées entre les catégories de fertilité pour la race Montbéliarde



Territoire n°1

Gène	Nom Complet	Identifiant Ensembl
/	/	ENSBTAG00000037661
DGAT2	Diacylglycerol O-Acyltransférase 2	ENSBTAG00000001154
GABRP	Gamma-Aminobutyric Acid (GABA) A Receptor Pi	ENSBTAG00000002779
LRP5	LDL Receptor related Protein 5	ENSBTAG00000005903
MYLK4	Myosin Light chain Kinase family member 4	ENSBTAG00000012079
MYO18B	Myosin XVIIIIB	ENSBTAG00000000698
NRDE2	NRDE-2 necessary for RNA interference domain containing	ENSBTAG00000005427
PDCD1	Programmed Cell Death 1	ENSBTAG00000011543
PSD3	Pleckstrin and Sec7 Domain containing 3	ENSBTAG00000033137
RFTN1	Raftlin lipid raft linker 1	ENSBTAG00000030593
SFRP1	Secreted Frizzled-Related Protein 1	ENSBTAG00000027625

Territoire n°2

Gène	Nom Complet	Identifiant Ensembl
COL4A1	Collagen type IV Alpha 1	ENSBTAG00000012849

Territoire n°3

Gène	Nom Complet	Identifiant Ensembl
/	/	ENSBTAG00000047148
/	/	ENSBTAG00000033143
CACNA1A	voltage-dependent P/Q-type calcium channel subunit alpha-1A	ENSBTAG00000014828
F13A1	Coagulation Factor XIII A1 polypeptide	ENSBTAG00000007268
GFRAL	GDNF family Receptor Alpha Like	ENSBTAG00000012973
HTRA3	HtrA serine peptidase 3	ENSBTAG00000047613
IGF2	Insulin-like Growth Factor 2	ENSBTAG00000013066
PPHLN1	Periplin 1	ENSBTAG00000008595
PPP1R26	Protein Phosphatase 1 Regulatory subunit 26	ENSBTAG00000030317
RFTN1	Raftlin lipid raft linker 1	ENSBTAG00000030593
SPATA13	Spermatogenesis Associated 13	ENSBTAG00000019545
SPECC1L	Sperm antigen with calponin homology and Coiled-Coil domains 1 Like	ENSBTAG00000021656
SSR1	Signal Sequence Receptor alpha	ENSBTAG00000022731
TBC1D22A	TBC1 Domain family member 22A	ENSBTAG00000012291
TMC6	Transmembrane Channel-like 6	ENSBTAG00000018661
WWC1	WW and C2 domain containing 1	ENSBTAG00000013880

Territoire n°4

Gène	Nom Complet	Identifiant Ensembl
ARHGEF10	Rho guanine nucleotide exchange factor 10	ENSBTAG00000008585
GUCA2A	Guanylate Cyclase Activator 2A	ENSBTAG00000007195
NRDE2	NRDE-2 necessary for RNA interference domain containing	ENSBTAG00000005427
PADI2	Peptidyl Arginine Deiminase type II	ENSBTAG00000003403
RPN1	Ribophorin I	ENSBTAG00000005191
SFRP1	Secreted Frizzled-Related Protein 1	ENSBTAG00000027625
SHANK2	SH3 and multiple Ankyrin repeat domains 2	ENSBTAG00000003171
SYNJ2	Synaptojanin 2	ENSBTAG00000000817
SYTL3	Synaptotagmin-Like 3	ENSBTAG00000010334
TSC1	Tuberous Sclerosis 1	ENSBTAG00000005190
ZNF423	Zinc Finger protein 423	ENSBTAG00000017397

Liste des communications présentées à des congrès

Communications orales :

- **Colloque Doc’J** (association des jeunes chercheurs du centre INRA de Jouy-en-Josas), présentation de thèse en 180 secondes, Jouy-en-Josas, France (2014)
« *Bovine semen epigenetics: molecular analysis of semen quality and impact on embryo development* »
- **Journée des mastères et des doctorants de l’unité BDR**, Jouy-en-Josas, France (2015)
« Epigénétique de la semence bovine : Rôle de la méthylation de l’ADN comme marqueur de la qualité de la semence et impact des facteurs épigénétiques sur le développement embryonnaire »
- **66^{ème} congrès de l’European Federation of Animal Science (EAAP)**, Varsovie, Pologne (2015)
« *Epigenetics of bovine semen: tools to DNA methylation analysis* »
- **Livestock Genomic Meeting**, Cambridge, Angleterre (2016)
« *Bull spermatozoa methylation landscape displays breed-specific features revealed by RRBS analysis* »
- **CRB-ANIM**, Jouy-en-Josas, France (2016)
« Analyse multi-échelle de la méthylation de l’ADN dans la semence bovine et identification de régions génomiques hypométhylées dans les spermatozoïdes »
- **Journée des mastères et des doctorants de l’unité BDR**, Jouy-en-Josas, France (2017)
« Identification de nouveaux biomarqueurs de la fertilité mâle dans le méthylome de la semence bovine : une analyse par RRBS »
- **Epiphase**, Jouy-en-Josas, France (2017)
« Identification de nouveaux biomarqueurs de la fertilité mâle dans le méthylome de la semence bovine : une analyse par RRBS »

Communications affichées :

- **11^{ème} Cours d'Epigénétique**, Institut Curie, Paris, France (2015)
« *Bovine Semen Epigenetics: molecular analysis of semen quality and impact on embryo development* »
- **Epiconcept Workshop**, Dubrovnik, Croatie (2015)
« *Epigenetics of bovine semen: tools for DNA methylation analysis* »
- **Conférences Jacques Monod « Méthylation et déméthylation de l'ADN »**, Roscoff, France (2015)
« *Epigenetics of bovine semen: tools for DNA methylation analysis* »
- **Plant and Animal Genome Conference (PAG)**, San Diego, Etats-Unis (2016)
« *Bull spermatozoa DNA methylation landscape displays unique features revealed by multi-scale analysis* »
- **Colloque Doc'J** (association des jeunes chercheurs du centre INRA de Jouy-en-Josas), présentation de thèse en 180 secondes, Jouy-en-Josas, France (2017)
« *Bull spermatozoa DNA methylation landscape displays unique features revealed by multi-scale analysis* »

Titre : Epigénétique de la semence bovine : analyse moléculaire de la qualité de la semence et impact potentiel sur le développement embryonnaire

Mots clés : Epigénétique, Reproduction, Fertilité mâle, Développement embryonnaire, Effets environnementaux, Effets transgénérationnels

Résumé : La présence de taureaux sub-fertiles sur le marché de l'Insémination Animale (IA) influence négativement l'efficacité des élevages. L'évaluation de la fertilité des taureaux, basée sur l'analyse combinée de marqueurs génétiques, morphologiques, cinétiques et métaboliques de la semence, n'est pas suffisante pour identifier les taureaux sub-fertiles avant leur entrée en production. Cela suggère l'implication d'autres facteurs, notamment d'origines épigénétiques. Parmi les marques épigénétiques, la méthylation de l'ADN tient une place essentielle : son remaniement est en effet indispensable aux processus fondamentaux que sont la différenciation des cellules germinales, la spermatogénèse et le développement embryonnaire précoce. Alors que chez l'Homme, de nombreuses études ont montré que des altérations des profils de méthylation spermatiques sont associées à la sub-fertilité, très peu de données existent chez le bovin. L'objectif de ce travail de thèse est la caractérisation du méthylome de la semence bovine et l'identification de nouveaux marqueurs fiables de la fertilité mâle à un stade précoce. Cette étude s'inscrit dans un large projet intitulé SeQuaMol (pour « Qualité Moléculaire de la Semence »), mis en place au sein d'un laboratoire commun entre l'INRA et la fédération ALLICE. La caractérisation du méthylome spermatique bovin a été réalisée en utilisant une approche multi-échelle (globale, pangénomique et séquence-spécifique). Les analyses ont permis de révéler l'hypométhylation très marquée du spermatozoïde bovin. L'hypométhylation affecte des gènes importants pour la différenciation de la lignée germinale, les fonctions spermatiques, ainsi que des séquences satellites.

L'identification de biomarqueurs de la fertilité a été réalisée en utilisant une approche pangénomique (*Reduced Representation Bisulfite Sequencing*, RRBS). L'analyse a été effectuée sur une cohorte de 94 taureaux, dont les individus de races Holstein et Montbéliard ont été catégorisés en fonction de l'adéquation entre un indicateur génétique de la fertilité et leur fertilité réelle. Le dispositif inclus également des taureaux de 4 autres races pour obtenir une estimation de la variabilité épigénétique liée à la race. Cette analyse a nécessité l'optimisation et l'automatisation du protocole pour la préparation des banques RRBS à haut-débit, ainsi que la mise au point de l'ensemble de la procédure de traitement bio-informatique et statistique des données. Plusieurs milliers de biomarqueurs de la fertilité ont été identifiés et permettent de prédire de façon robuste le statut de fertilité. De plus, une démarche d'intégration des données de génotype et d'épigénotype a été amorcée, soulignant les interactions potentielles en ces strates d'informations. Enfin, nous avons mis en évidence l'influence de l'âge à la production de semence sur le méthylome. L'ensemble de ces données souligne que les modifications du méthylome spermatique peuvent affecter des gènes impliqués dans les processus précoces et plus tardifs du développement, ainsi que quelques voies du fonctionnement du spermatozoïde. L'ensemble des biomarqueurs identifiés serviront de base à la poursuite du projet SeQuaMol, dont l'aboutissement sera le développement d'outils technologiques et statistiques pouvant être utilisés en routine pour améliorer la prédiction de la fertilité mâle.

Title: Bovine Semen Epigenetics: molecular analysis of semen quality and potential impact on embryo development

Keywords: Epigenetics, Reproduction, Male fertility, Embryo development, Environmental effects, transgenerational effects

Abstract: The presence of subfertile bulls in the market of Animal Insemination negatively influences the efficacy of breeding farms. The evaluation of bull fertility, based on the combined analysis of genetic, morphological, kinetic and metabolic markers of the semen, is not sufficient to identify subfertile bulls before entering semen production. It suggests the implication of other factors, especially epigenetic ones. Within the epigenetic markers, DNA methylation holds a crucial position: indeed, its reprogramming is indispensable to the fundamental processes that are germinal cell differentiation, spermatogenesis and the embryonic development. Whereas several studies have showed that the alteration of spermatid methylation profiles are associated with subfertility in humans, only a little data exists concerning bovines. The aim of this thesis is the characterization of the methylome of bovine semen and the identification of new reliable markers of male fertility at an early stage. This study is part of a larger project called SeQuaMol (Molecular Quality of Semen) which rely on a common laboratory between INRA and the ALLICE federation. The characterization of the bovine sperm methylome was achieved by using a multiscale approach (global, pangénomique and sequence-specific). The analyses made it possible to observe the hypomethylation of the bovine semen. Hypomethylation affects genes that are crucial for the differentiation of the germline, spermatid functions, and also on satellite sequences.

The identification fertility biomarkers was carried out by using a pangenomic approach (*Reduced Representation Bisulfite Sequencing*, RRBS). The analysis has been performed on a cohort of 94 bulls, including subjects of the Holstein race and Montbéliard race, categorized according to the adequacy between a genetic indicator of fertility and their actual fertility. The process also includes bulls from 4 other breeds in order to obtain an estimate of the genetic variability linked to the breed. This analysis has required the optimization and the automation of the protocol for the high throughput preparation of RRBS libraries, as well as the development of the whole bioinformatic pipeline and data statistics. Several thousands of biomarkers of fertility have been identified, which allow to predict in a robust way the fertility status. Furthermore, a process of integration of genotype and epigenotype data has been started, which underlines the potential interaction between these levels of information. Finally, we have highlighted the influence of the age of semen production on the methylome. Altogether, these data suggest that modifications of the semen methylome can affect genes that are involved in the process of early and late embryo development, and parts of the functioning of the sperm cell. The biomarkers identified will form a basis for the pursuit of the SeQuaMol project, of which the completion will be the development of technological and statistical tools that may be used routinely to improve male's fertility prediction.

