



**HAL**  
open science

# Epigénétique transgénérationnelle : l'environnement embryonnaire impacte-t-il les performances des générations ultérieures via l'épigénome chez la caille ?

Chloé Cerutti

## ► To cite this version:

Chloé Cerutti. Epigénétique transgénérationnelle : l'environnement embryonnaire impacte-t-il les performances des générations ultérieures via l'épigénome chez la caille?. Toxicologie. Institut National Polytechnique de Toulouse - INPT, 2023. Français. NNT : 2023INPT0133 . tel-04521219

**HAL Id: tel-04521219**

**<https://theses.hal.science/tel-04521219>**

Submitted on 29 Apr 2024

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

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



Université  
de Toulouse

# THÈSE

En vue de l'obtention du

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

**Délivré par :**

Institut National Polytechnique de Toulouse (Toulouse INP)

**Discipline ou spécialité :**

Infectiologie, Physio-pathologie, Toxicologie, Génétique et Nutrition

---

**Présentée et soutenue par :**

Mme CHLOÉ CERUTTI

le mardi 19 décembre 2023

**Titre :**

Epigénétique transgénérationnelle : l'environnement embryonnaire impacte-t-il les performances des générations ultérieures via l'épigénome chez la caille ?

---

**École doctorale :**

Sciences Écologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB)

**Unité de recherche :**

Génétique, Physiologie et Systèmes d'Élevage (GENPHYSE)

**Directeur(s) de Thèse :**

MME FREDERIQUE PITEL

**Rapporteurs :**

M. DOMINIQUE ROCHA, INRAE JOUY-EN-JOSAS

MME ANNE GABORY, INRAE JOUY-EN-JOSAS

**Membre(s) du jury :**

MME GWENNAELE FICHANT, UNIVERSITE TOULOUSE 3, Présidente

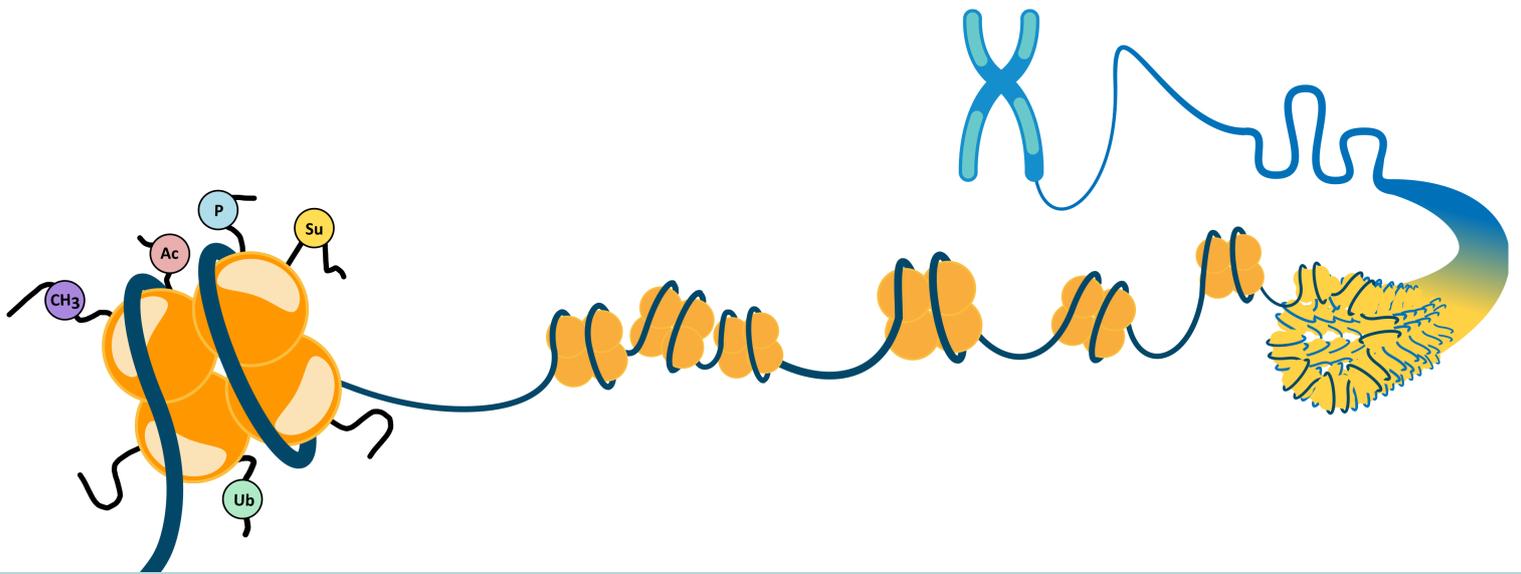
M. GUILLAUME DEVAILLY, INRAE TOULOUSE, Membre

MME CELINE COSSEAU, UNIVERSITE DE PERPIGNAN, Membre

MME FREDERIQUE PITEL, INRAE TOULOUSE, Membre

MME MARIA MANZANARES-DAULEUX, INRAE RENNES, Membre





# Epigénétique transgénérationnelle : l'environnement embryonnaire impacte-t-il les performances des générations ultérieures via l'épigénome chez la caille ?

2020-2023

CERUTTI Chloé





# Remerciements

Je tiens tout d'abord à remercier mes encadrants, Pitou et Guillaume, pour m'avoir permis d'effectuer cette thèse et pour leur accompagnement tout au long de ces trois années. Merci également pour leur soutien et pour leurs encouragements dans les moments difficiles. Et enfin, un grand merci pour le temps consacré à la relecture du manuscrit, jusqu'à la dernière minute. Je remercie également tous les membres de l'équipe GenEpi (Julie, Sophie, Magui, Katia, Nathalie, Juliette, Annie, Denis, Jean-Noël, Elena, Keyvan, Mathilde et Sonia) pour la bonne ambiance, leur implication et tout particulièrement pour leur aide dans la préparation de toutes les étapes importantes rencontrées au cours d'une thèse.

Une pensée pour mes encadrants de stage de M2, Annie et Thomas, qui m'ont donné l'opportunité et l'envie de poursuivre.

Merci aussi à l'équipe GenROC pour leur accueil et leur gentillesse (Laurence, Stéphane, Julien, Florence, Carine, Gwenola et Florent).

Je tiens à remercier les membres de mon comité de suivi de thèse (Laurence Drouilhet, Sylvain Foissac et Sandrine Lagarrigue) pour leurs conseils.

Merci à tous ceux qui ont contribué aux analyses effectuées dans le cadre de ce travail (Sophie, Paul, Yann, Ingrid, Tatiana, Christophe, ...) pour leur implication et leurs conseils.

Je remercie Vincent Coustham de m'avoir donné l'opportunité de collaborer avec lui dans le cadre de son projet (QuailHeatE).

Un grand merci à Maxime, qui partage ma vie, de m'avoir soutenue et encouragé tout au long de la thèse sans laquelle je ne l'aurais pas rencontré. Merci également à ma belle-famille pour son accueil chaleureux.

Merci à mes parents, mes frères (Franck et Joël), ma soeur (Aude) pour le soutien et la relecture du manuscrit, sans oublier Lilou.

Merci à la région Occitanie et au département de génétique animale pour leur financement.

Je suis enfin très reconnaissante envers Anne Gabory, Dominique Rocha, Céline Cosseau, Gwen-naële Fichant et Maria Manzanara-Dauleux, qui ont accepté de prendre part au jury qui évaluera ce travail de thèse.



## Résumé

Le changement climatique, les contraintes économiques et les préoccupations sociétales pour une agriculture durable vont fortement influencer les systèmes de production animale et nécessitent une caractérisation plus fine de la capacité d'adaptation du génome des animaux à des fluctuations environnementales tout au long de leur vie, dès le stade embryonnaire. L'impact environnemental au niveau du génome est notamment caractérisé par la variabilité de marques épigénétiques (notamment de la méthylation de l'ADN). Elles interviennent dans la régulation de l'expression des gènes et peuvent également être transmissibles à la descendance. L'analyse des marques épigénétiques est rendue possible grâce au développement de technologies de séquençage haut-débit innovantes et d'outils bioinformatiques adaptés.

La question scientifique majeure de cette thèse est l'étude chez l'oiseau des effets de l'environnement (injection de molécules contaminantes) pendant le développement embryonnaire, et de leur transmission entre générations. Les travaux ont été réalisés chez la caille, excellente espèce modèle d'intérêt agronomique pour l'étude de ces phénomènes.

Le premier objectif consiste à caractériser la réponse de l'organisme face aux variations de l'environnement embryonnaire par l'analyse de (i) données phénotypiques (ponte et croissance) et (ii) de méthylation de l'ADN (RRBS et WGBS) issues de cailles ayant reçu directement une injection dans l'œuf de génistéine, bisphénol A (perturbateurs endocriniens) ou 5Aza (inhibiteur de l'ADN méthyltransférase). Un effet significatif des différentes molécules injectées sur la croissance des cailles a été mis en évidence, les individus injectés ayant un poids plus faible que les individus non injectés, sans impacter les caractères de ponte. L'analyse de méthylation a été réalisée au moyen d'un pipeline bioinformatique permettant d'effectuer une analyse différentielle du taux de méthylation entre plusieurs conditions expérimentales. Ces analyses ont permis d'identifier des régions différenciellement méthylées entre les conditions expérimentales, mettant en évidence un effet des molécules sur des gènes liés aux voies métaboliques (croissance), au comportement (développement du système nerveux central, transport des lipides et réponse au cholestérol) et à la reproduction (maturité sexuelle).

La deuxième partie consiste à caractériser la réponse de l'épigénome aux effets de l'environnement trois générations après l'injection dans l'œufs de la molécule génistéine (phyto-œstrogène présent notamment dans le soja). Deux 'épilignées' de caille (Epi+ et Epi-) ont été produites en utilisant des œufs fécondés provenant de la même population fondatrice. L'analyse des données de méthylation de l'ADN, issues de deux techniques de séquençage différentes (WGBS et ONT)

réalisées sur des échantillons de la troisième génération de ce dispositif, a pu mettre en évidence que l'exposition à la génistéine, lors du développement embryonnaire, impacte la méthylation de l'ADN de la descendance trois générations après injection. L'analyse de méthylation différentielle entre les deux épilignées a permis d'identifier des régions différentiellement méthylées en lien avec des gènes impliqués dans des voies métaboliques, dans le développement du cerveau et dans la reproduction. Ces résultats sont en accord avec les effets observés sur les caractères de reproduction et de comportement. Même si les effets génétiques ont été minimisés, une cause génétique à certains différentiels de méthylation observés n'a pas pu être écartée. Cependant, parmi les régions identifiées, certaines semblent bien être liées à une transmission épigénétique transgénérationnelle des effets de l'environnement.

L'information de ces marques épigénétiques pourrait être prise en compte dans les schémas de sélection par l'utilisation de modèles de "transmissibilité" incluant les informations génétiques et épigénétiques de chaque individu.

**Mots-clés :** *Coturnix japonica*, épigénétique transgénérationnelle, méthylation de l'ADN, perturbateurs endocriniens, bisulfite, nanopore.

# Abstract

Climate change, economic constraints and societal concerns about sustainable agriculture are going to have a major impact on animal production systems, requiring more detailed characterisation of the capacity of the animal genome to adapt to environmental fluctuations throughout an individual's life, from the embryonic stage onwards. The environmental impact on the genome is characterized in particular by the variability of epigenetic marks (particularly DNA methylation). These marks play a role in regulating gene expression and could also be passed on to offspring. The analysis of epigenetic marks has been made possible by the development of innovative high-throughput sequencing technologies and appropriate bioinformatics tools.

The major scientific question of this thesis is the study in birds of the effects of the environment (injection of contaminating molecules) during embryonic development, and their intergenerational transmission. This work has been done in quail, which is a good model species of agronomic interest for the study of these phenomena.

The first objective was to characterise the organism's response to variations in the embryonic environment by analysing (i) phenotypic data (laying and growth) and (ii) DNA methylation data (RRBS and WGBS) from quails injected directly into the egg with genistein, bisphenol A (endocrine disruptors) or 5Aza (a DNA methyltransferase inhibitor). A significant effect of the different molecules injected on quail growth was demonstrated, with injected individuals weighing less than non-injected individuals, without any impact on egg-laying characteristics. Differential methylation analysis was made possible by the development of a bioinformatics pipeline enabling differential analysis of the methylation rate between several experimental conditions. These analyses enabled us to identify regions that were differentially methylated between experimental conditions, highlighting an effect of the molecules on genes linked to metabolic pathways (growth), behaviour (development of the central nervous system, lipid transport and response to cholesterol) and reproduction (sexual maturity).

The second part consists of characterising the response of the epigenome to the effects of the environment three generations after the injection into the egg of genistein (a phytoestrogen found notably in soya). Two quail 'epilines' (Epi+ and Epi-) were produced using fertilised eggs from the same founder population. Analysis of DNA methylation data from two different sequencing techniques (WGBS and ONT) from the third generation of this animal design showed that exposure to genistein during embryonic development has an impact on DNA methylation in the offspring three generations after injection. Differential methylation analysis between the

two epilines identified differentially methylated regions linked to genes involved in metabolic pathways, brain development and reproduction. These results are consistent with the effects observed on reproductive and behavioural traits. Although genetic effects were minimised, a genetic cause for some of the observed methylation differentials could not be ruled out. However, some of the regions identified appear to be linked to transgenerational epigenetic transmission of environmental effects.

Information from these epigenetic marks could be taken into account in selection schemes by using a 'transmissibility' model that includes the genetic and epigenetic information of each individual.

**Keywords :** *Coturnix japonica*, transgenerational epigenetics, DNA methylation, endocrine disruptors, bisulfite, nanopore.

# Table des matières

I	<b>Introduction générale</b>	
<b>1</b>	<b>De la génétique à l'épigénétique</b>	<b>23</b>
1.1	<b>Génétique et variabilité phénotypique</b>	<b>24</b>
1.1.1	L'ADN : le support de l'information génétique	24
1.1.2	Variations de la séquence d'ADN	27
1.2	<b>Environnement et expression des gènes : avènement de l'épigénétique</b>	<b>29</b>
<b>2</b>	<b>Mécanismes épigénétiques</b>	<b>31</b>
2.1	<b>Méthylation de l'ADN</b>	<b>32</b>
2.1.1	Dynamique de méthylation : ADN méthyltransférases ( <i>writers</i> )	35
2.1.2	Dynamique de déméthylation : ADN déméthylases ( <i>erasers</i> )	37
2.1.3	Ilots CpG et régulation de l'expression des gènes	38
2.1.4	Contrôles génétiques de la méthylation de l'ADN	40
2.2	<b>Modifications post-traductionnelles des histones</b>	<b>42</b>
2.3	<b>Autres mécanismes épigénétiques</b>	<b>44</b>
2.3.1	ARN non codants	44
2.3.2	Remaniement de la chromatine	46
2.4	<b>Des phénomènes en interaction</b>	<b>48</b>
2.4.1	Interaction entre les marques épigénétiques	48
2.4.2	Empreinte génomique parentale	50
2.4.3	Inactivation du chromosome X	52
<b>3</b>	<b>Exposition aux effets de l'environnement</b>	<b>53</b>
3.1	<b>Phases critiques du développement embryonnaire</b>	<b>54</b>

<b>3.2</b>	<b>Facteurs environnementaux</b>	<b>56</b>
3.2.1	Température	57
3.2.2	Alimentation	59
3.2.3	Interactions sociales	61
3.2.4	Molécules biologiquement actives	62
3.2.4.1	<i>Principaux perturbateurs endocriniens</i>	63
3.2.4.2	<i>Mode d'action des perturbateurs endocriniens</i>	63
<b>4</b>	<b>Transmission des effets environnementaux</b>	<b>69</b>
<b>4.1</b>	<b>Transmission intergénérationnelle</b>	<b>71</b>
<b>4.2</b>	<b>Transmission transgénérationnelle</b>	<b>75</b>
<b>5</b>	<b>Technologies de séquençage pour l'étude du méthylome</b>	<b>81</b>
<b>5.1</b>	<b>Traitement de l'ADN au bisulfite de sodium</b>	<b>84</b>
5.1.1	Séquençage haut-débit de l'ADN par la technologie Illumina	84
5.1.2	Séquençage au bisulfite tout génome (WGBS)	86
5.1.3	Séquençage au bisulfite à représentation réduite (RRBS)	88
5.1.4	Analyse du méthylome et biais d'analyse	90
<b>5.2</b>	<b>Technologie Nanopore</b>	<b>92</b>
5.2.1	Séquençage nanopore tout génome	92
5.2.2	Analyse du méthylome et biais d'analyse	94
<b>6</b>	<b>Approches pour la détection et l'analyse d'effets épigénétiques</b>	<b>95</b>
<b>6.1</b>	<b>Détection des marques de méthylation</b>	<b>95</b>
6.1.1	Contrôle qualité des lectures avec FastQC	98
6.1.2	Nettoyage des séquences avec TrimGalore	98
6.1.3	Alignement des lectures sur le génome de référence avec Bismark	98
6.1.4	Extraction des niveaux de méthylation et mise en forme des résultats	99
<b>6.2</b>	<b>Outils d'analyse de méthylation différentielle</b>	<b>99</b>

II

## Situation du sujet et objectifs de la thèse

III

## Développement bioinformatique : pipeline SeqOccin

<b>7</b>	<b>Pré-traitement des données</b>	<b>113</b>
----------	-----------------------------------	------------

<b>8</b>	<b>Analyse de méthylation différentielle</b>	<b>115</b>
<b>8.1</b>	<b>DSS</b>	<b>115</b>
8.1.1	Ajustement des données	115
8.1.2	Comparaison entre deux groupes	115
8.1.3	Détection de DMC à partir d'un design expérimental complexe	116
<b>8.2</b>	<b>edgeR</b>	<b>116</b>
8.2.1	Ajustement des données	116
8.2.2	Détection de DMC à partir d'un design expérimental	117

## IV

### Caractérisation de la réponse de l'organisme aux variations de l'environnement embryonnaire

<b>9</b>	<b>Impacts sur les phénotypes de croissance et de ponte</b>	<b>123</b>
9.1	Dispositif expérimental	123
9.2	Article 1 : Effets de l'injection <i>in-ovo</i> de perturbateurs endocriniens et d'un inhibiteur de l'ADN méthyltransférase sur la croissance et les performances de ponte chez la caille japonaise.	124
<b>10</b>	<b>Impacts sur la méthylation de l'ADN</b>	<b>139</b>
10.1	Dispositif expérimental	139
10.2	Article 2 : Effets de l'injection <i>in-ovo</i> de perturbateurs endocriniens et d'un inhibiteur de l'ADN méthyltransférase sur la méthylation de l'ADN chez la caille adulte.	141

## V

### Caractérisation de la réponse de l'épigénome de la descendance aux effets de l'environnement embryonnaire

<b>11</b>	<b>Des effets visibles au niveau du méthylome de la descendance</b>	<b>185</b>
11.1	Dispositif expérimental	185
11.2	Article 3 : Effets transgénérationnels de l'injection <i>in-ovo</i> de génistéine sur la méthylation de l'ADN de la descendance chez la caille	186
<b>12</b>	<b>Programmation thermique transgénérationnelle</b>	<b>219</b>
12.1	Présentation du projet QuailHeatE	219

<b>13</b>	<b>Rappels des principaux résultats</b> .....	<b>225</b>
<b>14</b>	<b>Dispositif expérimental</b> .....	<b>227</b>
14.1	Choix et intérêt du modèle de la Caille	227
14.2	Forces et limitations des dispositifs expérimentaux étudiés	227
14.3	Comparaison avec le dispositif de manipulation thermique	228
14.4	Recommandations pour la création d'un dispositif expérimental pour étudier les transmissions épigénétiques inter et transgénérationnelles	229
<b>15</b>	<b>Avantages et inconvénients des différentes technologies de séquençage haut-débit pour l'étude de la méthylation de l'ADN chez la caille</b> ...	<b>231</b>
15.1	Préparation des échantillons	231
15.2	Nombre de CpG/DMC détectés à partir des différentes technologies	231
15.3	La technologie ONT est-elle la meilleure méthode ?	231
15.4	Puce de méthylation	232
<b>16</b>	<b>Développement du pipeline <i>SeqOccin</i> et pistes d'améliorations</b> ....	<b>233</b>
16.1	Rappel de nouveaux éléments implémentés par rapport à l'ancienne version	233
16.2	Normalisation des données	233
16.3	Filtrage des SNP	234
16.4	Nouvelles améliorations possibles du pipeline <i>SeqOccin</i>	234
<b>17</b>	<b>Validation des DMC et impacts fonctionnels</b> .....	<b>237</b>
17.1	Pyroséquençage	237
17.2	Edition de l'épigénome par transfection	237
17.3	Lien entre l'épigénétique et les ARN non codants : rôle dans la régulation de l'expression des gènes	238
17.4	Projets internationaux contribuant à l'acquisition de données et à l'amélioration des connaissances de l'annotation des génomes	238
<b>18</b>	<b>Prise en compte des variations épigénétiques dans les schémas de sélection</b> .....	<b>241</b>

VII

## Annexes

19	Epigénétique .....	247
20	Continuité du stage de M2 : analyse des circRNA .....	283

VIII

## Références bibliographiques

Références

287



# Abréviations

**5Aza** 5-aza-2'deoxyctidine

**A/T/C/G** Adénine/Thymine/Cytosine/Guanine

**AA** Acide aminé

**ADN** Acide DésoxyriboNucléique

**AIC** Critère d'information d'Akaike

**ARNlnc** ARN long non codant

**ARNmi** micro ARN

**ARNm** ARN messenger

**ARNnc** ARN non codant

**ARNpi** ARN interagissants avec la protéine PIWI

**ARNr** ARN ribosomique

**ARNsi** ARN interférent

**ARNsno** ARN nucléolaire

**ARNsn** ARN nucléaire

**ARNt** ARN de transfert

**ARN** Acide RiboNucléique

**ASM** Méthylation allèle-spécifique

**ATP/ADP** Adénosine triphosphate/diphosphate

**BER** *Base excision repair*

**BH** Benjamini-Hochberg

**BPA** Bisphénol A

**C/IC** Contrôlé injecté/non-injecté

**CGI** Ilôt CpG

**chr** chromosome

**circRNA** ARN circulaire

**CNV** *Copy Number Variation*

**DMC** Cytosine différentiellement méthylée

**DMR** Région différentiellement méthylée

**DNMT** ADN méthyltransférase

**dNTP** Désoxyribonucléotide triphosphate

**DSS** *Dispersion Shrinkage for Sequencing data*

**epi** epilignée

**FDR** *False Discovery Rate*

**G0,1,2,3** Génération 0,1,2,3

**GWAS** *Genome-Wide association studies*

**HMT** Histones méthyltransférases

**HSD** Différence significative honnête

**IAP** Particule intracisternale de type A

**MBD** *methyl CpG binding domain*

**MHM** (*Male Hypermethylated*)

**MPTH** Adénine,Thymine,Cytosine,Guanine

**MT** Manipulation thermique

**O/N/H** Oxygène/Azote/Hydrogène

**ONT** *Oxford Nanopore Technologies*

**OT/OB** *Original Top/Original Bottom*

**OTC/OBC** *Original Top Complementary/Original Bottom Complementary*

**PCR** *Polymerase Chain Reaction*

**PE** Perturbateur endocrinien

**PGC** Cellules germinales primordiales

**PTGS** *Post-Transcriptional Gene Silencing*

**QTL** *Quantitative Trait Locus*

**RA** Récepteur des androgènes

**RE** Récepteur aux œstrogènes

**RISC** *RNA-induced silencing complex*

---

**RRBS** *Reduced Representation Bisulfite Sequencing*

**RSC** *Chromatin structure remodeling*

**SAH** *S-adénylhomocystéine*

**SAM** *S-adénylméthionine*

**SNP** *Single Nucleotide Polymorphism*

**SPM** *Stress prénatal maternel*

**SRA** *SET and Ring finger Associated*

**SWI/SNF** *SWItch/Sucrose Non-Fermentable*

**TAD** *Topological associated domain*

**TDG** *Thymine DNA glycosylase*

**TET** *Ten-Eleven Translocation*

**TGS** *Transcriptional Gene Silencing*

**TSS** *Site d'initiation de la transcription*

**WGBS** *Whole Genome Bisulfite Sequencing*



# Introduction générale

<b>1</b>	<b>De la génétique à l'épigénétique</b> .....	<b>23</b>
1.1	Génétique et variabilité phénotypique	
1.2	Environnement et expression des gènes : avènement de l'épigénétique	
<b>2</b>	<b>Mécanismes épigénétiques</b> .....	<b>31</b>
2.1	Méthylation de l'ADN	
2.2	Modifications post-traductionnelles des histones	
2.3	Autres mécanismes épigénétiques	
2.4	Des phénomènes en interaction	
<b>3</b>	<b>Exposition aux effets de l'environnement</b> .....	<b>53</b>
3.1	Phases critiques du développement embryonnaire	
3.2	Facteurs environnementaux	
<b>4</b>	<b>Transmission des effets environnementaux</b> .....	<b>69</b>
4.1	Transmission intergénérationnelle	
4.2	Transmission transgénérationnelle	
<b>5</b>	<b>Technologies de séquençage pour l'étude du méthylome</b> ...	<b>81</b>
5.1	Traitement de l'ADN au bisulfite de sodium	
5.2	Technologie Nanopore	
<b>6</b>	<b>Approches pour la détection et l'analyse d'effets épigénétiques</b>	<b>95</b>
6.1	Détection des marques de méthylation	
6.2	Outils d'analyse de méthylation différentielle	



# 1. De la génétique à l'épigénétique

L'objectif de ce chapitre est de présenter les connaissances actuelles sur les facteurs à l'origine de la variabilité des phénotypes observable au sein d'une population.

La construction du phénotype d'un individu, c'est-à-dire l'ensemble des caractères apparents d'un individu, résulte de son génotype (la séquence qu'il porte en chaque position), de l'environnement dans lequel il évolue et de ceux dans lequel il a précédemment évolué, ainsi que de l'interaction entre son génotype et son environnement. L'effet de l'environnement (au sens large, y compris l'âge, le type cellulaire, ...) sur les phénotypes est en partie médié par des phénomènes épigénétiques (Figure 1.1).

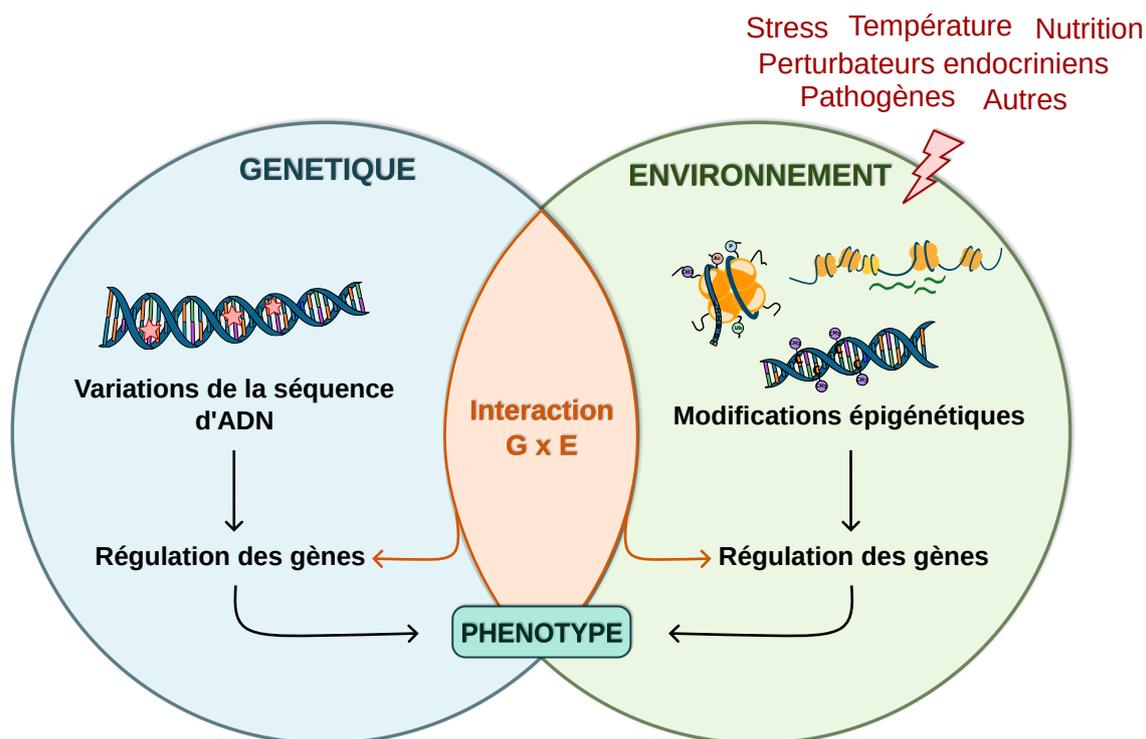


FIGURE 1.1 – Interaction entre génétique et environnement dans la construction des phénotypes. Abréviations : G = génétique, E = environnement. Figure adaptée de Coustham et al., 2023 (Annexe 19, en cours de publication).

## 1.1 Génétique et variabilité phénotypique

### 1.1.1 L'ADN : le support de l'information génétique

L'étude de l'information génétique d'un organisme peut permettre de mieux comprendre les mécanismes génétiques à l'origine d'une partie de la variabilité des phénotypes, sous l'hypothèse que la variation d'origine génétique est imputable à des différences au niveau de la séquence d'ADN. L'information génétique d'un individu est contenue dans des molécules d'acide désoxyribonucléique (ADN) formé de deux brins antiparallèles enroulés formant une double hélice appelée ADN double brin (Watson & Crick, 1953). Chaque brin d'ADN est constitué d'une succession de nucléotides, eux-mêmes constitués d'une base nucléique : adénine (A), cytosine (C), guanine (G) et thymine (T) liée à un sucre (désoxyribose), lui-même lié à un groupe phosphate (**Figure 1.2**).

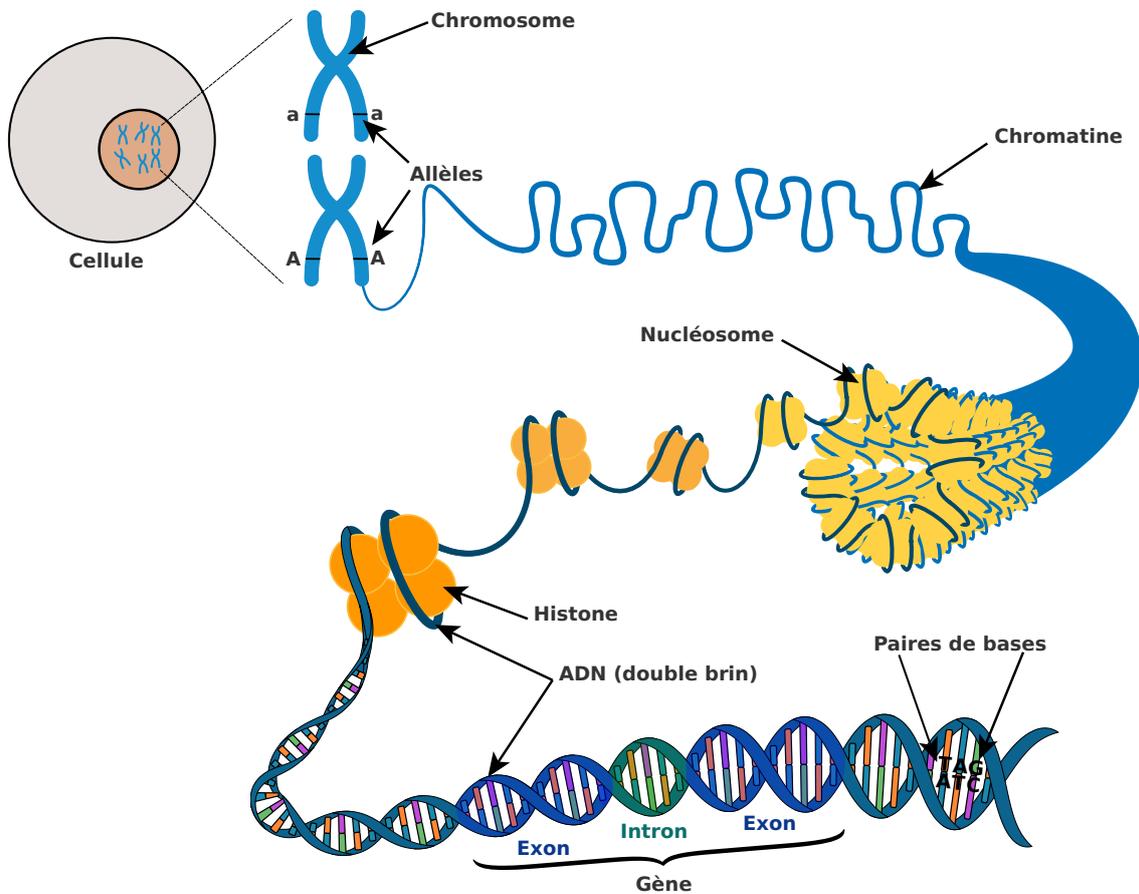


FIGURE 1.2 – **Compaction de l'ADN.** Dans les cellules eucaryotes, l'ADN s'enroule autour de protéines structurales (histones) qui sont empaquetées dans des nucléosomes. Les nucléosomes et l'ADN qui les entoure forment la chromatine qui se condense pour former des chromosomes. A et a représentent les deux allèles d'un même gène. *Figure adaptée de <https://images.nagwa.com>.*

Chez les eucaryotes, les protéines sont produites par traduction d'ARN messagers (ARNm). Ces ARNm sont issus de l'épissage d'ARN pré-messager (pré-ARNm) consistant à enlever les séquences introniques. Les pré-ARNm sont produits par transcription de l'ADN par l'ARN polymérase II (**Figure 1.3**). La succession des bases nucléiques sur l'ADN détermine la succession des acides aminés qui constituent les protéines issues de ces gènes. La correspondance entre bases nucléiques et acides aminés est le code génétique. Il existe également des portions d'ADN transcrites mais non traduites en protéines, correspondant aux ARN non codants. L'ensemble des régions codantes et non codantes d'un organisme constitue son génome.

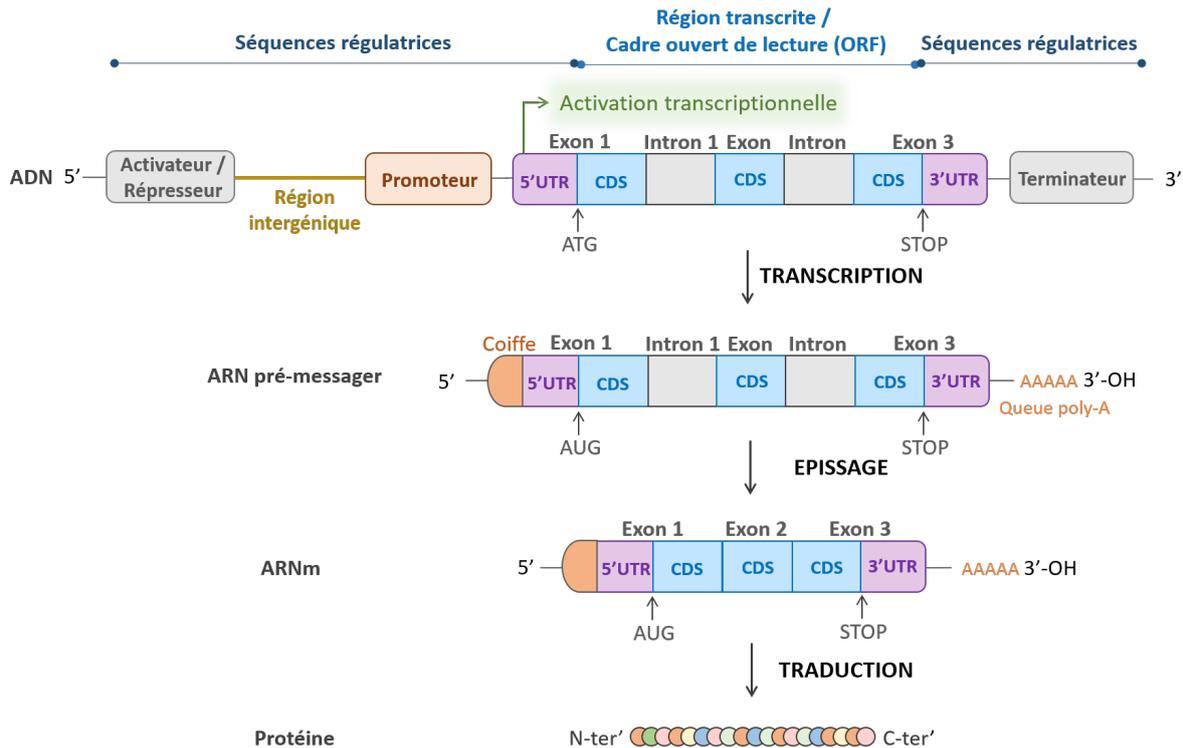


FIGURE 1.3 – **Étapes de l'expression d'un gène protéique chez les eucaryotes (exemple d'un gène codant à 3 exons)**. Seul le brin sens de l'ADN est représenté (5' → 3'). La région génique est constituée de séquences régulatrices et de régions transcrites ou cadre ouvert de lecture (ORF). Elle comprend un activateur (ou répresseur) situé, dans cet exemple, avant le promoteur et un terminateur. La séquence codante est constituée d'exons. L'exon 1 est constitué d'une région non traduite 5'UTR, d'un codon d'initiation ATG et d'une séquence codante (CDS), l'exon 2 est constitué d'un CDS et l'exon 3 d'un CDS, d'un codon STOP et d'une région non traduite 3'UTR. L'ADN est transcrit par l'ARN polymérase II en ARN pré-messager constitué d'une coiffe à l'extrémité 5'. Au cours de la transcription, les thymine (T) sont remplacées en uraciles (U), formant un codon d'initiation AUG. Une queue poly-A est formée à l'extrémité 3'. L'ARN pré-messager subit un phénomène d'épissage qui enlève les introns. L'ARN messager (ARNm) formé est constitué d'exons. Cet ARNm est ensuite traduit en protéines. Les régions 5'UTR et 3'UTR ne sont pas traduites en protéines. Les protéines ont des extrémités N-terminale (N-ter') et C-terminale (C-ter').

Chez la plupart des vertébrés, la taille du génome peut varier entre environ 0,9 Gb et 3,2 Gb (D. Li et al., 2022) (**Table.1.1**). Alors que le génome humain mesure 3,1 Gb, le génome de la caille est plus petit, soit 0,93 Gb, avec un total de 927 656 957 paires de bases dans l'assemblage actuel (GCA\_001577835.1, Ensembl release 103). Il est structuré en 32 chromosomes dont 28 autosomes et 2 chromosomes sexuels (ZZ pour les mâles et ZW pour les femelles). Le nombre de gènes codants pour des protéines identifiés chez la caille japonaise est de 15 732 contre 17 007 chez la poule, 23 217 chez l'Homme et 21 684 chez la souris. Parmi ces 15 732 gènes, 12 031 sont retrouvés chez l'Homme (orthologues 1:1), 10 685 chez la souris et 12 911 chez la poule. La caille et la poule sont deux espèces très proches (Thayer et al., 2018), toutes deux faisant partie de l'ordre des Galliformes. Outre un nombre de chromosomes identique ( $2n=78$ ), l'ordre des gènes entre ces deux espèces est assez bien conservé (Kayang et al., 2006).

TABLE 1.1 – **Taille des génomes chez certaines espèces de vertébrés.** 1 giga base (Gb) = 1 million de bases. *Table adaptée de D. Li et al. (2022).*

Espèces	Assemblage du génome	Taille du génome (Gb)	Nombre d'autosomes
Caille ( <i>Coturnix japonica</i> )	GCA_001577835.1	0,93	28
Poule ( <i>Gallus gallus</i> )	GCA_016699485.1	1,23	32
Poisson zèbre ( <i>Danio rerio</i> )	GCA_000002035.4	1,37	25
Vache ( <i>Bos taurus</i> )	GCA_002263795.2	2,67	29
Mouton ( <i>Ovis aries</i> )	GCA_000298735.1	2,62	26
Porc ( <i>Sus scrofa</i> )	GCA_000003025.6	2,50	18
Chat ( <i>Felis catus</i> )	GCA_000181335.4	2,46	18
Chien ( <i>Canis lupus familiaris</i> )	GCA_014441545.1	2,41	38
Homme ( <i>Homo sapiens</i> )	GCA_000001405.28	3,10	22
Macaque ( <i>Macaca mulatta</i> )	GCA_003339765.3	3,24	20
Souris ( <i>Mus musculus</i> )	GCA_000001635.9	2,73	19
Rat ( <i>Rattus norvegicus</i> )	GCA_015227675.2	2,87	20
Lapin ( <i>Oryctolagus cuniculus</i> )	GCA_000003625.1	2,74	21

### 1.1.2 Variations de la séquence d'ADN

Au sein d'une population, une grande variabilité de phénotypes peut exister, celle-ci étant en partie causée par une variabilité génétique. Les variations alléliques sont soit héritées, soit nouvellement acquises par mutation *de novo*. En fonction de la période de la vie touchée, du type de mutation mais aussi des cellules concernées, ces mutations ont des conséquences différentes. Il peut s'agir de mutations germinales ou bien de mutations somatiques (**Figure 1.4**).

Les mutations germinales (ou mutations *de novo*) surviennent durant la formation des gamètes (gamétogenèse) et touchent les cellules reproductrices (ovocytes et spermatozoïdes) (**Figure 1.4.A**). Ces mutations, présentes dans toutes les cellules de l'organisme, sont transmises à la descendance et sont connues pour jouer un rôle dans l'évolution des espèces (Bergeron et al., 2023). Elles sont également connues pour être à l'origine de maladies génétiques, telle que la trisomie 21 (cause du syndrome de Down chez l'Homme) et peuvent modifier la prédisposition d'un individu à développer un cancer au cours de sa vie (Acuna-Hidalgo et al., 2016).

Les mutations somatiques affectent certaines cellules en dehors des gamètes (cellules somatiques) et peuvent survenir à n'importe quel stade du développement d'un individu (**Figure 1.4.B**). Elles peuvent être acquises par hasard, au cours d'erreurs de réplication ou de réparation de l'ADN, ou faire suite à des expositions exogènes et s'accumulent tout au long de la vie (Manders et al., 2021). Ces mutations peuvent survenir durant le stade précoce du développement (stade embryonnaire) ou plus tardivement. A un stade précoce, toutes les cellules descendantes de la cellule mutée sont porteuses de la mutation. Une partie plus ou moins étendue de l'organisme est génétiquement différente (mosaïcisme). A un stade plus tardif du développement, l'exposition à des mutagènes peut engendrer une augmentation du taux de mutations. La mutation peut altérer le contrôle du cycle cellulaire induisant une prolifération incontrôlée de la cellule mutée, pouvant ainsi induire la formation de tumeurs (Martincorena & Campbell, 2015).

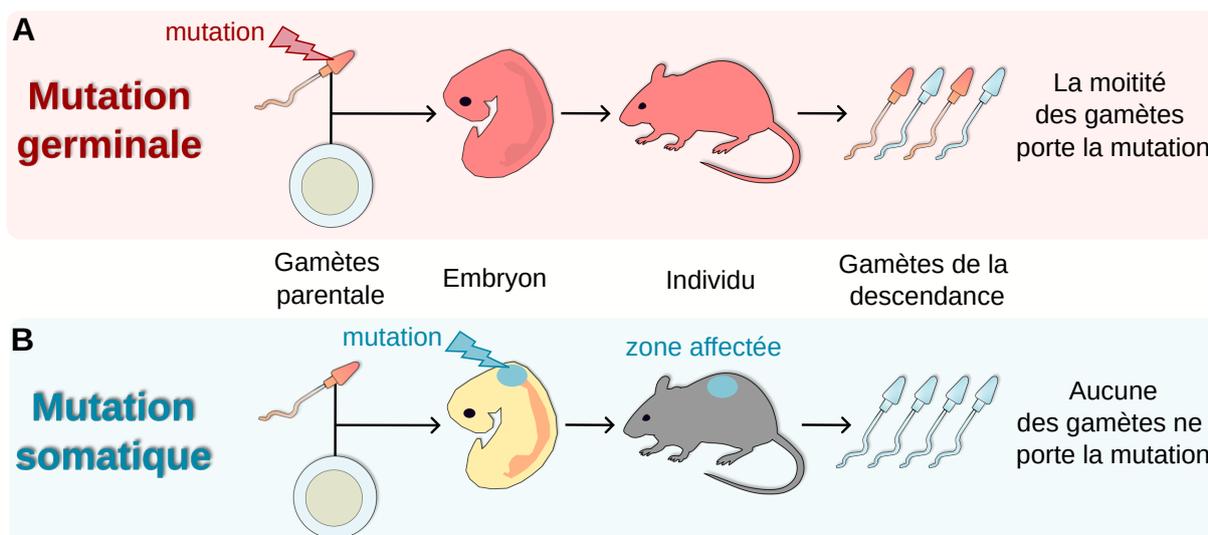


FIGURE 1.4 – Mutations germinales et somatiques. (A) Mutation germinale. (B) Mutation somatique.

Au cours du temps, les génomes évoluent sous l'influence de mutations pouvant être d'origine environnementale (mutations induites) ou bien intervenant de manière aléatoire (mutations spontanées). Dans une espèce ou population, on retrouve des séquences ADN invariantes et des différences de séquences appelées allèles. Si un individu est porteur de deux allèles identiques, il est homozygote (AA) pour cette position et s'ils sont différents, il est hétérozygote (Aa). En fonction de leurs répercussions fonctionnelles, ces mutations peuvent être conservées au sein de la population. Les mutations induites peuvent être dues à l'exposition à des éléments mutagènes, tels que les rayonnements ultraviolets (UV) ou des molécules chimiques (Ségurel et al., 2014). Ces altérations de l'ADN peuvent être prises en charge par des mécanismes de réparation de l'ADN au cours desquels des enzymes sont capables de se fixer à l'ADN endommagé et de le réparer. Dans le cas où l'ADN n'est pas réparé, ces mutations sont transmises lors de la division cellulaire. Les mutations spontanées peuvent survenir de manière aléatoire suite à des erreurs de synthèse lors de la réplication de l'ADN au cours de laquelle l'ADN polymérase va copier l'ADN pour synthétiser le nouveau brin. Ces erreurs peuvent conduire à une insertion (ajout), une substitution (remplacement) ou bien une délétion (perte) d'une ou plusieurs bases (**Figure 1.5**). Les petites mutations (une à quelques dizaines de paires de bases) regroupent les SNP (*Single Nucleotide Polymorphism*), variation d'une base de l'ADN, les petites insertions et délétions (indels) et les petites séquences répétées (microsatellites) (**Figure 1.5.A**). Les grandes mutations (dizaines à millions de nucléotides) correspondent à des variants de structure qui regroupent les grandes insertions/délétions, les duplications, les inversions et les translocations chromosomiques (Hollox et al., 2022). Parmi ces grandes mutations, les insertions, délétions et duplications, dont le nombre de copies change dans la séquence d'ADN, sont appelées des CNV (*Copy Number Variants*) (**Figure 1.5.B**).

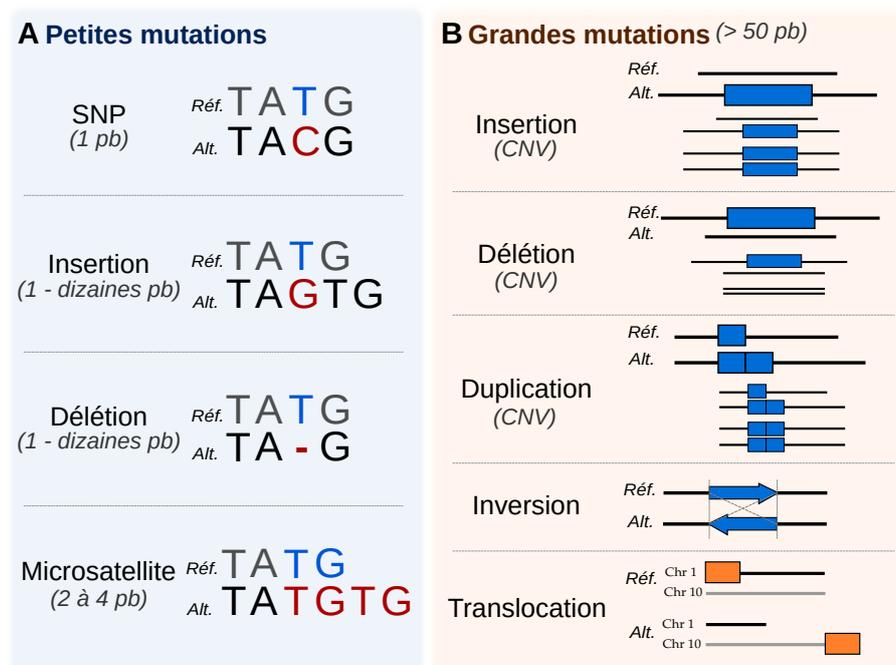


FIGURE 1.5 – Les différents polymorphismes génétiques. (A) Grandes et (B) petites mutations. Abréviations : SNP = *Single Nucleotide Polymorphism*, CNV = *Copy number Variant*. Table adaptée de la thèse de Ben Braiek, 2022 et Hollox et al., (2022).

En production animale, l'amélioration des schémas de sélection repose en partie sur l'identification de polymorphismes responsables de la variabilité des caractères d'intérêt. Les régions du génome (*loci*) pour lesquelles la variation allélique est associée à ces phénotypes d'intérêt sont appelées des QTL (*Quantitative Trait Locus* ou locus de caractère quantitatif) (Ytournal et al., 2008). Les régions QTL sont principalement détectées par analyse d'association à l'échelle du génome ou GWAS (*Genome-Wide association studies*) permettant de relier chaque marqueur génétique associé à un phénotype d'intérêt (Uffelmann et al., 2021). Les GWAS consistent donc à identifier des marqueurs moléculaires, tels que les microsatellites ou encore les SNP, corrélés à un phénotype observé. Les analyses des phénotypes et des génotypes ont ainsi permis de sélectionner des caractères d'intérêts en agronomie tels que la production et la qualité des œufs chez la poule (Becot et al., 2021) ainsi que la production laitière chez les bovins (Boichard et al., 2000). Ainsi, des différences au niveau de la séquence d'ADN sous-tendent les mécanismes génétiques à l'origine de la variation des phénotypes. Même si la séquence du génome est bien connue, la fonction d'une part importante de la séquence d'ADN est toujours méconnue. Les variations de séquence n'expliquent pas toute la variabilité observée des caractères, cette dernière étant en partie d'origine environnementale.

## 1.2 Environnement et expression des gènes : avènement de l'épigénétique

De nombreuses études ont pu mettre en évidence un effet de l'environnement sur l'expression des gènes chez certaines espèces de plantes, d'insectes et d'animaux, induisant une diversité phénotypique (Mazzio & Soliman, 2012) (**Figure 1.6**). Des individus de même génotype peuvent développer des phénotypes différents.

Un exemple classique est la différenciation cellulaire : les cellules partagent le même génome (à quelques exceptions près) mais ont des phénotypes cellulaires très variés. Chez les plantes, l'exemple bien connu de la Linaria commune (*Linaria vulgaris*) a permis de mettre en évidence que, malgré un génotype identique, ces dernières peuvent développer une conformation florale variée. Certaines auront une fleur à symétrie bilatérale (**Figure 1.6.A**, gauche), d'autres vont développer une fleur anormale à symétrie radiale (**Figure 1.6.A**, droite), due à l'inactivation du gène *Lcyc* par des mécanismes épigénétiques (Cubas et al., 1999). Nous pourrions également donner l'exemple, pour les vertébrés, des souris agoutis. Pour un ensemble de souris génétiquement identiques (génotype  $A^{vy}/a$  au locus Agouti), ces dernières ont une distribution de couleur du pelage assez variée, allant du jaune clair (**Figure 1.6.B**, gauche) au gris foncé (**Figure 1.6.B**, droite) (Morgan et al., 1999). Cette variation serait due à une expression plus ou moins importante de l'allèle  $A^{vy}$  du gène Agouti, sous contrôle de la méthylation d'une région régulatrice (Dolinoy et al., 2007). Chez les abeilles, en fonction de l'alimentation que ces dernières recevront au stade larvaire, elles ne se développeront pas de la même manière (Cridge et al., 2015 ; Kucharski et al., 2008b). Un même œuf fécondé d'abeille donnera soit une ouvrière, si la larve qui en résulte est nourrie principalement à la bouillie larvaire (**Figure 1.6.C**, bas), soit une reine, si la larve est nourrie exclusivement à la gelée royale (**Figure 1.6.C**, haut).

Ainsi, des individus génétiquement identiques peuvent donc présenter des caractères distincts liés à une expression différente de mêmes gènes.

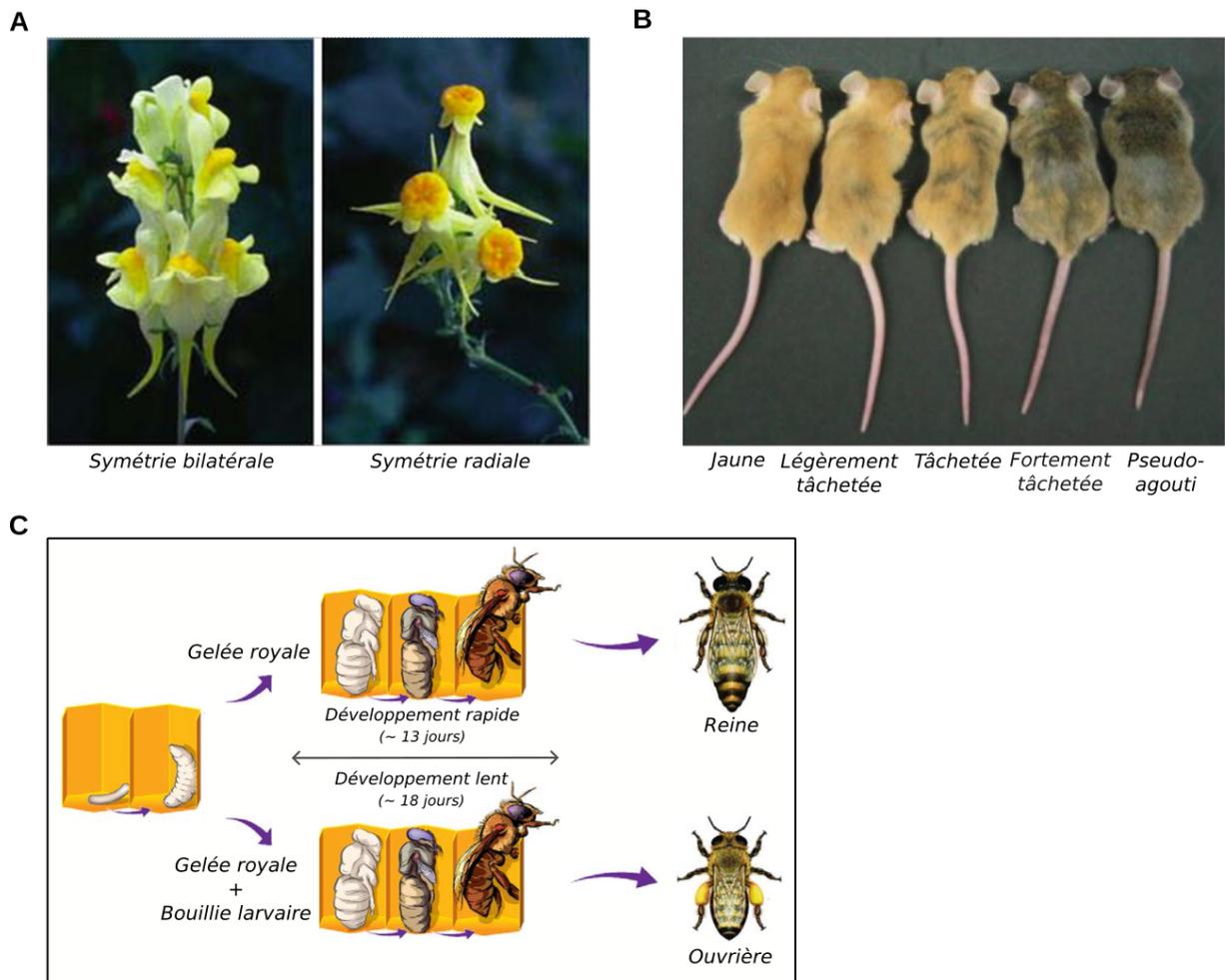


FIGURE 1.6 – Effet de l'environnement sur la variabilité phénotypique. Exemples de (A) la Linaire commune ou *Linaria vulgaris* (Photographies issues de Hitchins (2010)), (B) la souris Agouti (Photographies issues de Dolinoy et al. (2007)), et (C) des abeilles (Figure adaptée de Cridge et al. (2015)).

Les exemples cités précédemment indiquent que les effets de l'environnement sur le phénotype ne passent pas forcément (et même très rarement) par une modification de la séquence d'ADN mais aussi par d'autres voies, en particulier les phénomènes épigénétiques (Waddington, 1942). Les variations génétiques n'expliquent donc pas toute la variabilité des caractères, cette dernière n'étant pas non plus directement expliquée par une variabilité environnementale. D'autres paramètres interviennent dans la variabilité phénotypique, cette variation pouvant être d'origine génétique, environnementale, mais aussi épigénétique.

## 2. Mécanismes épigénétiques

Alors que la génétique correspond à l'étude des gènes, l'épigénétique désigne l'étude des changements de la fonction des gènes hérités par voie mitotique et/ou méiotique qui ne peuvent pas être expliqués par des changements de la séquence d'ADN (Riggs, 1996). Les marques épigénétiques peuvent impacter l'expression des gènes et être en partie modulées par l'environnement dans lequel un individu évolue (Feil & Fraga, 2012). Il existe plusieurs types de marques épigénétiques telles que la méthylation de l'ADN (Yong et al., 2016), les modifications post-traductionnelles des histones et des variants d'histones (Bannister & Kouzarides, 2011), la régulation de l'expression des gènes par les acides ribonucléiques (ARN) non codants (X. Li et al., 2016) ainsi que le remaniement de la chromatine (Triantaphyllopoulos et al., 2016) (**Figure 2.1**). L'épigénome désigne l'ensemble des marques épigénétiques d'une cellule (cellules d'un organe, d'un tissu). La méthylation de l'ADN étant le sujet principal de cette thèse, nous insisteront plus particulièrement sur ce type de mécanisme dans la suite de l'exposé.

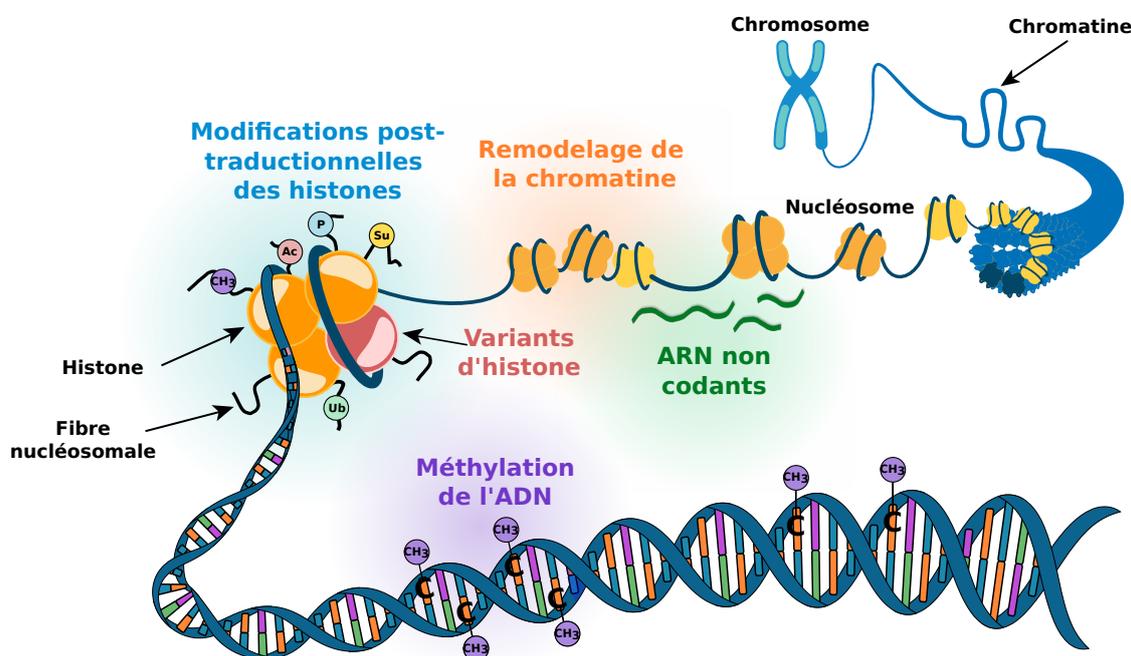


FIGURE 2.1 – Principales marques épigénétiques. Le remodelage de la chromatine, les modifications post-traductionnelles des histones, la méthylation de l'ADN et les ARN non codants sont représentés respectivement en orange, bleu, violet et vert. Abréviations : P = phosphorylation, Ac = acétylation, Ub = ubiquitination, CH<sub>3</sub> = méthylation, Su = sumoylation. *Figure adaptée de <https://images.nagwa.com>.*

## 2.1 Méthylation de l'ADN

La méthylation de l'ADN désigne l'ajout d'un groupement méthyle ( $\text{CH}_3$ ) au niveau des bases de l'ADN. Une famille d'enzymes, appelée ADN méthyltransférase (DNMT), peut méthyler les cytosines (C) et les adénines (A) de l'ADN par l'ajout d'un groupement méthyle ( $\text{CH}_3$ ) (**Figure 2.2**). Ces modifications se traduisent par une liaison covalente entre un groupement méthyle ( $\text{CH}_3$ ) et une C ou A de l'ADN. La méthylation des C en position 4 ou en position 5 donne respectivement une  $N^4$ -méthylcytosine (4mC) (**Figure 2.2.A**) et une 5-méthylcytosine (5mC) (**Figure 2.2.B**). La 4mC a été détectée chez les procaryotes et les eucaryotes (Janulaitis et al., 1983) mais est principalement étudiée chez les bactéries, jouant notamment un rôle dans la réparation de l'ADN (Pukkila et al., 1983). Chez les eucaryotes, le rôle de la 4mC est moins bien connu, les technologies disponibles limitant sa détection (Khanal et al., 2021). La 5mC est majoritairement retrouvée chez de nombreuses espèces d'animaux, de plantes et de champignons (Breiling & Lyko, 2015) mais a également été détectée chez les bactéries (Huo et al., 2015). Lorsque cette méthylation a lieu au niveau d'une A, une  $N^6$ -méthyladénine (6mA) est produite (**Figure 2.2.C**). La 6mA est surtout retrouvée chez les procaryotes (Vanyushin et al., 1968) mais a également été détectée chez les eucaryotes (Wu, 2020).

Chez les eucaryotes, les 5mC sont retrouvées au niveau de différents contextes génomiques tels que les C suivies d'une guanine (G), appelés sites CpG (ou dinucléotides CG), les sites CHG et les sites CHH, où H peut représenter les bases A, T ou C (**Figure 2.3**). Chez les vertébrés, le 5mC est la modification la plus étudiée et a majoritairement lieu au niveau des sites CpG (Gardiner-Garden & Frommer, 1987; Lister et al., 2009b) (**Figure 2.2.B**).

Les sites CpG sont des sites palindromiques, c'est-à-dire qu'ils peuvent être lus de la même manière dans les deux sens. Ils peuvent être hémi-méthylés (méthylé sur un brin et non méthylé sur l'autre) ou bien être méthylés sur les deux brins, ce qui est généralement le cas.

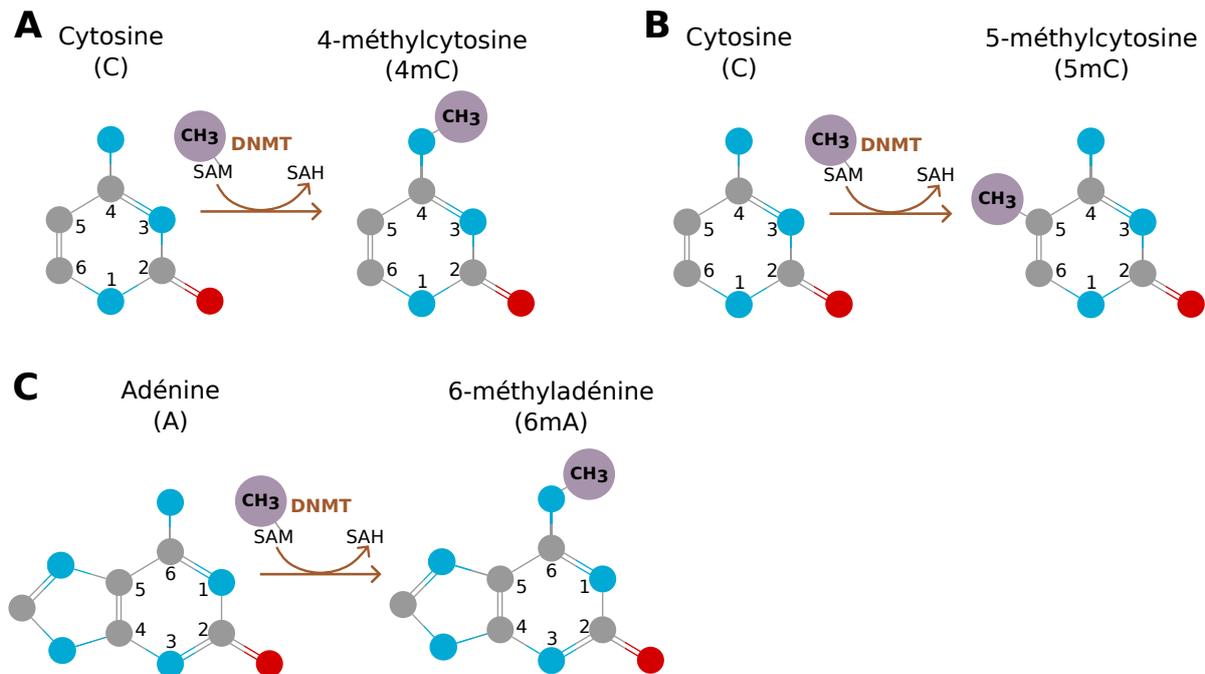


FIGURE 2.2 – Méthylation des bases de l'ADN. Les atomes d'azote (N), carbone, et oxygène (O) sont représentés respectivement en bleu, gris et rouge. Les atomes d'hydrogène ne sont pas représentés. Les groupements méthyles ( $\text{CH}_3$ ) sont représentés en violet. L'ADN méthyltransférase (DNMT) à la position 4 d'une cytosine (C), 5 d'une C ou 6 d'une adénine (A), transfère un  $\text{CH}_3$  d'une SAM (S- adénosylméthionine) à une SAH (S- adénosylhomocystéine) formant respectivement une  $N^4$ -méthylcytosine (4mC), 5-méthylcytosine (5mC) et  $N^6$ -méthyladénine (6mA).

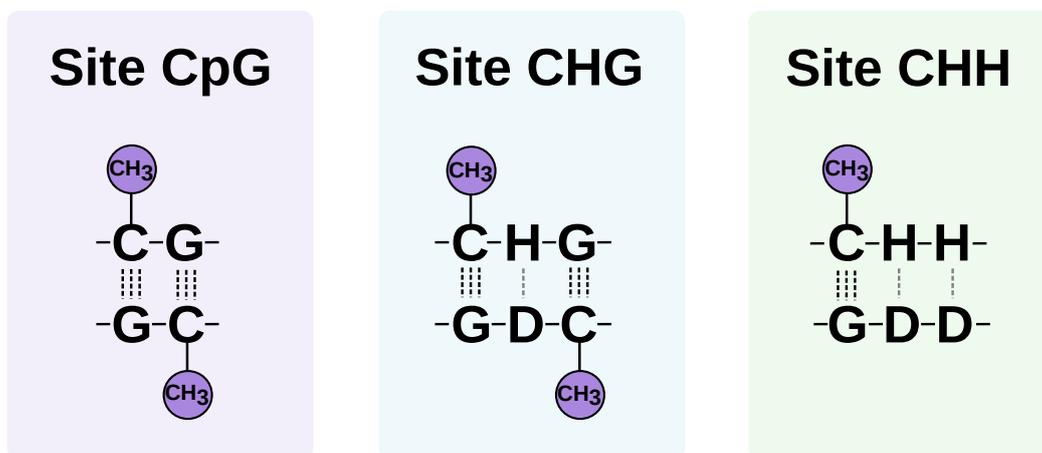


FIGURE 2.3 – Contextes de méthylation de l'ADN. De gauche à droite, méthylation de l'ADN au niveau des sites CpG, CHG et CHH. H peut représenter les bases A, T ou C et D peut représenter les bases A, T ou G. Les groupements méthyles ( $\text{CH}_3$ ) sont représentés en violet. *Figure adaptée de Erdmann & Picard (2020).*

La méthylation de l'ADN est majoritairement régulée par des protéines regroupées en trois grandes catégories : les méthyltransférases (*writers*), les déméthylases (*erasers*) et les protéines de liaison à l'ADN méthylé (*readers*) (**Figure 2.4**) (Moore et al., 2013 ; Biswas & Rao, 2018). Chacune de ces trois classes regroupe de nombreuses protéines (**Table 2.1**) (Moore et al., 2013 ; Zhao et al., 2022). La classe des *writers* regroupe la famille des DNMT (ADN méthyltransférases) constituée des enzymes DNMT1, DNMT3A et DNMT3B et du co-facteur DNMT3L (Lyko, 2018 ; Z. Chen & Zhang, 2020), qui peuvent méthyler les cytosines de l'ADN par l'ajout d'un CH<sub>3</sub>. Les *erasers* regroupent des enzymes de la famille TET (*ten-eleven translocation*) constituée des enzymes TET1, TET2 et TET3. Ces dernières sont des ADN déméthylases qui vont induire la perte d'un CH<sub>3</sub> au niveau des cytosines méthylées de l'ADN (Rasmussen & Helin, 2016). Enfin, la classe des *readers* regroupe les familles des MBD (*methyl CpG binding domain*), constituée des protéines MeCP2, MBD1, MBD2 et MBD4 (Du et al., 2015), des protéines contenant le domaine SRA (UHRF1 et UHRF2) (Shimbo & Wade, 2016) et des doigts de zinc liant le méthyl-CpG (Kaiso, ZBTB4, ZBTB38, KLF4) (Hudson & Buck-Koehntop, 2018). Ces protéines vont reconnaître et se lier aux marques de méthylation de l'ADN et ainsi influencer l'expression des gènes.

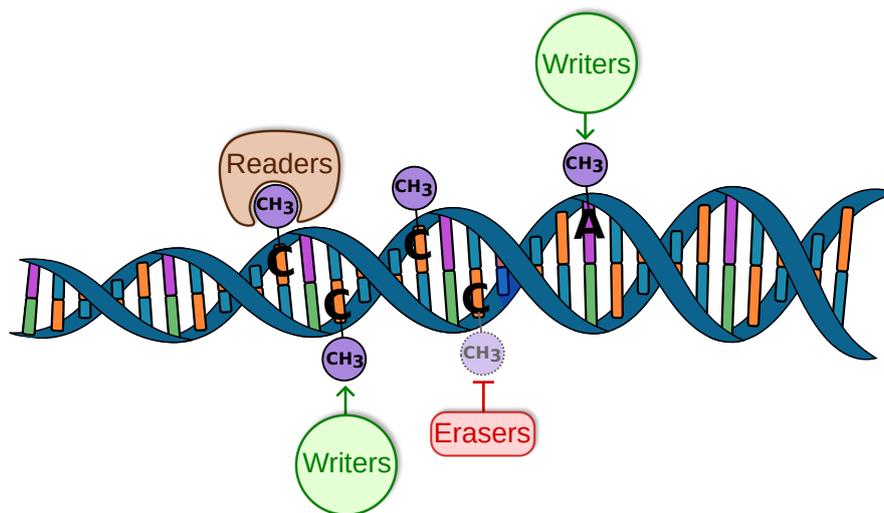


FIGURE 2.4 – **Modifications de la méthylation de l'ADN.** Les protéines impliquées dans la méthylation de l'ADN sont regroupées en trois catégories : les *writers*, qui correspondent à des méthyltransférases ajoutant un groupement méthyle (CH<sub>3</sub>) au niveau des bases de l'ADN (vert) ; les *erasers*, qui regroupent des déméthylases induisant la perte d'un CH<sub>3</sub> (rouge) ; et enfin, les *readers*, des protéines de liaison à l'ADN (marron).

TABLE 2.1 – **Protéines régulatrices de la méthylation de l'ADN.** Abréviations : DNMT = ADN méthyltransférases ; TET = *ten-eleven translocation* ; MBD = *methyl CpG binding domain* ; MeCP2 = *methyl CpG binding protein 2* ; SRA = *SET and Ring finger Associated* ; UHRF1 = *Ubiquitin-like, containing PHD and RING finger domains, 1* ; KLF4 = *Kruppel-like factor 4*. Table adaptée de Zhao et al. (2022) et Moore et al. (2013).

Classes	Familles	Noms des protéines
<b>Writers</b>	DNMT	DNMT1, DNMT3A, DNMT3B, DNMT3L
<b>Erasers</b>	TET	TET1, TET2, TET3
<b>Readers</b>	MBD	MeCP2, MBD1, MBD2, MBD4
	Protéines contenant le domaine SRA	UHRF1, UHRF2
	Doigts de zinc liant le méthyle-CpG	Kaiso, KLF4, ZBTB4, ZBTB38

### 2.1.1 Dynamique de méthylation : ADN méthyltransférases (*writers*)

La famille des DNMT comprend les DNMT1, DNMT3A, DNMT3B et DNMT3L (Goll & Bestor, 2005 ; Jeltsch, 2006). Les DNMT interviennent dans la méthylation *de novo* de l'ADN ainsi que dans son maintien (**Figure 2.5**).

Les enzymes DNMT3A et DNMT3B sont connues pour être impliquées dans le mécanisme de méthylation *de novo*. La DNMT3L n'a pas d'activité catalytique mais faciliterait la liaison des autres méthyltransférases *de novo* à l'ADN, telles que les DNMT3, en stimulant leur activité enzymatique (Lyko, 2018). De plus, elle est majoritairement conservée chez les mammifères et certains reptiles, excepté les monotrèmes (animaux pondant des œufs et allaitant leurs petits), les amphibiens, les poissons et les oiseaux (Raddatz et al., 2021).

L'enzyme DNMT1 est considérée comme étant l'enzyme majoritairement impliquée dans le maintien de la méthylation de l'ADN et méthyle préférentiellement l'ADN hémi-méthylé (Moore et al., 2013). Lors de la réplication de l'ADN, cette enzyme reproduit le motif de méthylation du brin d'origine sur le nouveau brin, passant d'un CpG hémi-méthylé à un CpG dont les deux brins sont méthylés (méthylation symétrique). En l'absence de DNMT1, la méthylation ne serait pas conservée aboutissant à des CpG non méthylés après réplication. L'absence de DNMT1 peut être en partie compensée par les enzymes DNMT3A et DNMT3B (Liang et al., 2002). Ce maintien permet donc de préserver la méthylation de l'ADN après chaque cycle cellulaire lors de la réplication de l'ADN.

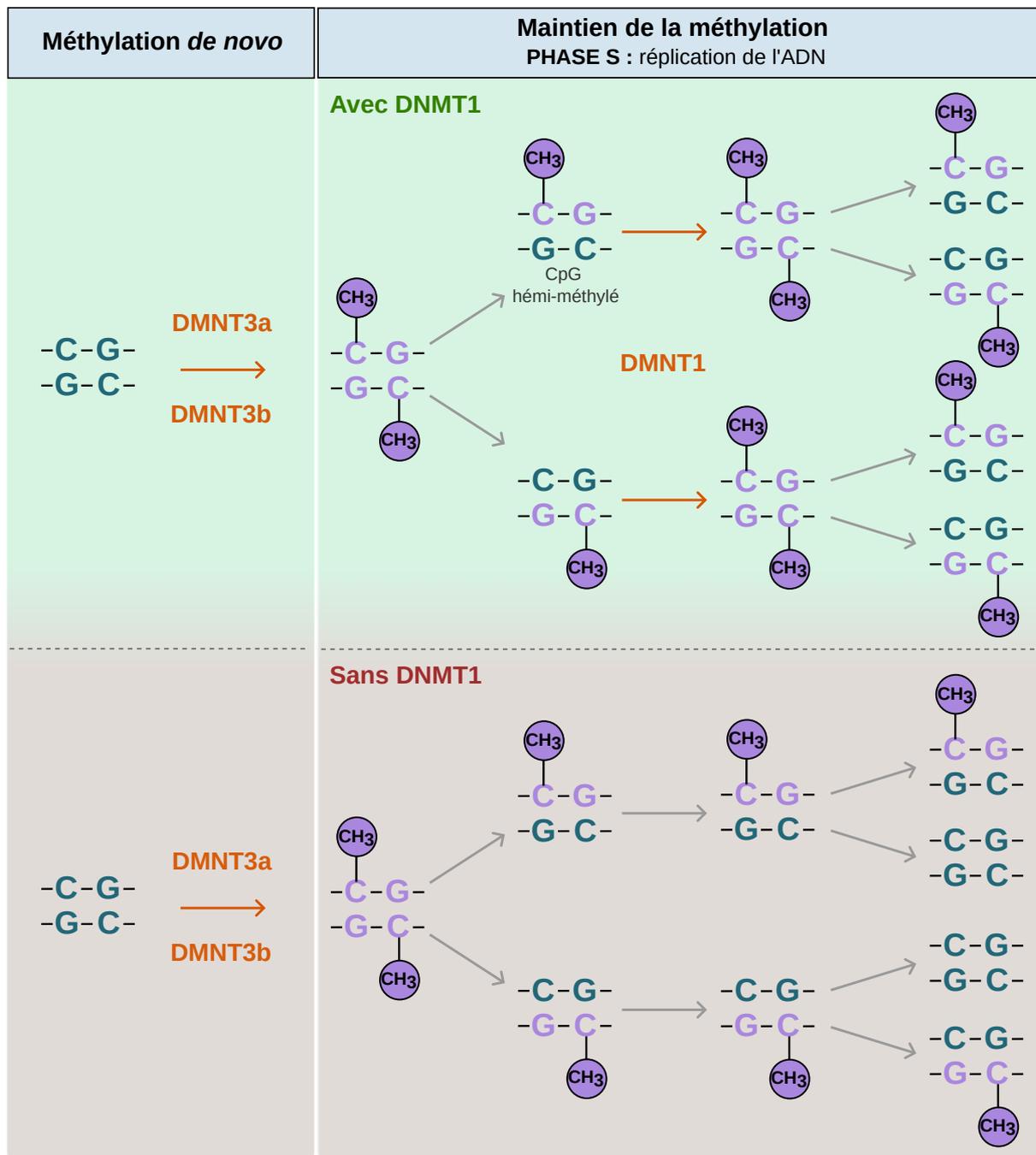


FIGURE 2.5 – Méthylation *de novo* et son maintien. Les groupes CH<sub>3</sub> sont représentés en violet. A gauche, méthylation *de novo* induite par les ADN méthyltransférases 3 (DNMT3A, DNMT3B). A droite, maintien de la méthylation induite par la DNMT1 lors de la réplication de l'ADN (phase S) : en présence de DNMT1 (en haut à droite), la méthylation est maintenue, en l'absence de DNMT1 (en bas à droite), la méthylation n'est pas maintenue.

### 2.1.2 Dynamique de déméthylation : ADN déméthylases (*erasers*)

La déméthylation de l'ADN désigne la perte d'un  $\text{CH}_3$  au niveau des bases de l'ADN (Sacks et al., 2018). On distingue la déméthylation passive et la déméthylation active (**Figure 2.6**) (Bochtler et al., 2017).

La déméthylation passive se produit lors de la réplication de l'ADN, lorsque le nouveau brin synthétisé n'est pas méthylé et que la marque de méthylation n'est pas conservée au cours du temps (dilution) (**Figure 2.5, bas**).

Contrairement à la déméthylation passive, la déméthylation active fait intervenir des enzymes (**Figure 2.5, haut**). Il peut s'agir d'une modification des bases de l'ADN, en partie médiée par les déméthylases de la famille TET (*ten-eleven translocation*). Ces protéines oxydent les bases méthylées, notamment les 5-méthylcytosines (5mC) pour former des 5-hydroxyméthylcytosines (5hmC) (Rasmussen & Helin, 2016). L'oxydation de la 5hmC va former une 5-formylcytosine (5fC) qui va ensuite être convertie en 5-carboxylcytosine (5caC) (Shen & Zhang, 2012). La liaison glycosidique des bases intermédiaires 5fC et 5caC peut être directement clivée par l'enzyme TDG (thymine ADN glycosylases), impliquée dans la réparation de l'ADN. Les sites abasiques qui en résultent vont être pris en charge par la voie de réparation par excision de bases (BER) et être remplacés par des C non méthylées.

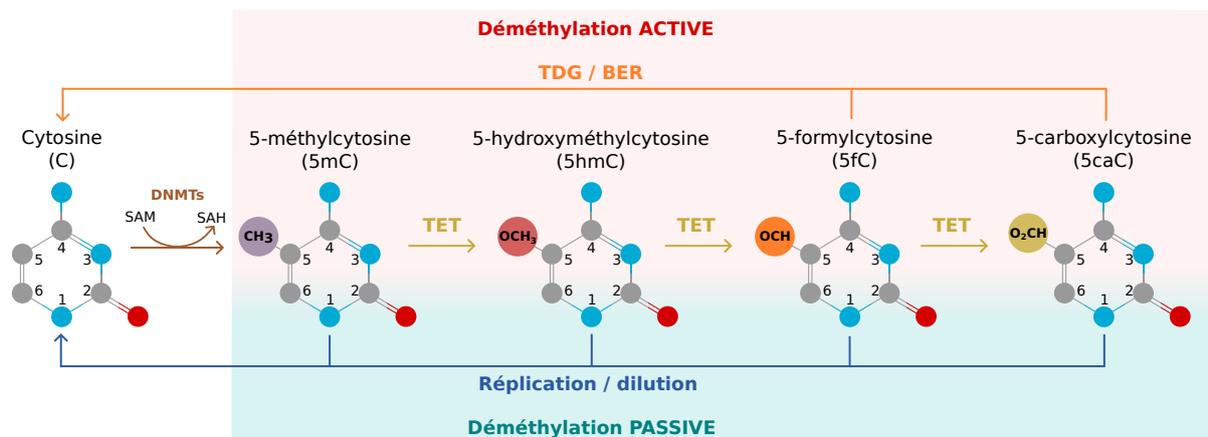


FIGURE 2.6 – Cycle de déméthylation des cytosines. Une 5mC peut être déméthylée par une voie (A) active ou (B) passive. Abréviations : DNMT = ADN méthyltransférases ; 5mC = 5-méthylcytosine ; 5hmC = 5-hydroxyméthylcytosine, 5fC = 5-formylcytosine et 5caC = 5-carboxylcytosine ; TET = *ten-eleven translocation*, BER = réparation par excision de bases ; SAH : S-adénylhomocystéine ; SAM = S-adénylméthionine ; TDG = thymine ADN glycosylase. *Figure adaptée de Ecsedi et al. (2018).*

### 2.1.3 Îlots CpG et régulation de l'expression des gènes

La méthylation de l'ADN est souvent associée à une régulation négative de l'expression des gènes (A. P. Bird, 1986 ; Jones, 2012). Certaines régions du génome sont relativement riches en sites CpG (ou dinucléotides CG) et sont appelées des îlots CpG (CGI). Les CGI sont majoritairement retrouvés au niveau des régions promotrices des gènes et sont souvent peu méthylés (**Figure 2.7.A**, haut). Lorsque ces sites CpG sont méthylés au niveau du promoteur, cela peut engendrer la répression de la transcription du gène qu'il régule (**Figure 2.7.A**, bas). Cette marque épigénétique est également impliquée dans la régulation de la compaction de la chromatine et réciproquement. En effet, dans le cas d'un promoteur non méthylé, la chromatine est dite relâchée, l'ARN polymérase et les facteurs de transcriptions pourront se lier à l'ADN et activer la transcription (**Figure 2.7.B**, haut). En revanche, lorsque ces sites CpG sont méthylés au niveau du promoteur, la chromatine est dans un état condensé, ce qui va empêcher l'ARN polymérase ainsi que certains facteurs de transcription de se lier à l'ADN (**Figure 2.7.B**, bas). Certaines facteurs de transcription peuvent être gênés par la méthylation de l'ADN et ne vont donc pas pouvoir se fixer à l'ADN. Cependant, d'autres facteurs de transcription, appelés des facteurs pionniers (*pioneer factors*), peuvent se lier directement à la chromatine condensée et ainsi initier la transcription sans être gênés par la méthylation de l'ADN (Cirillo et al., 2002).

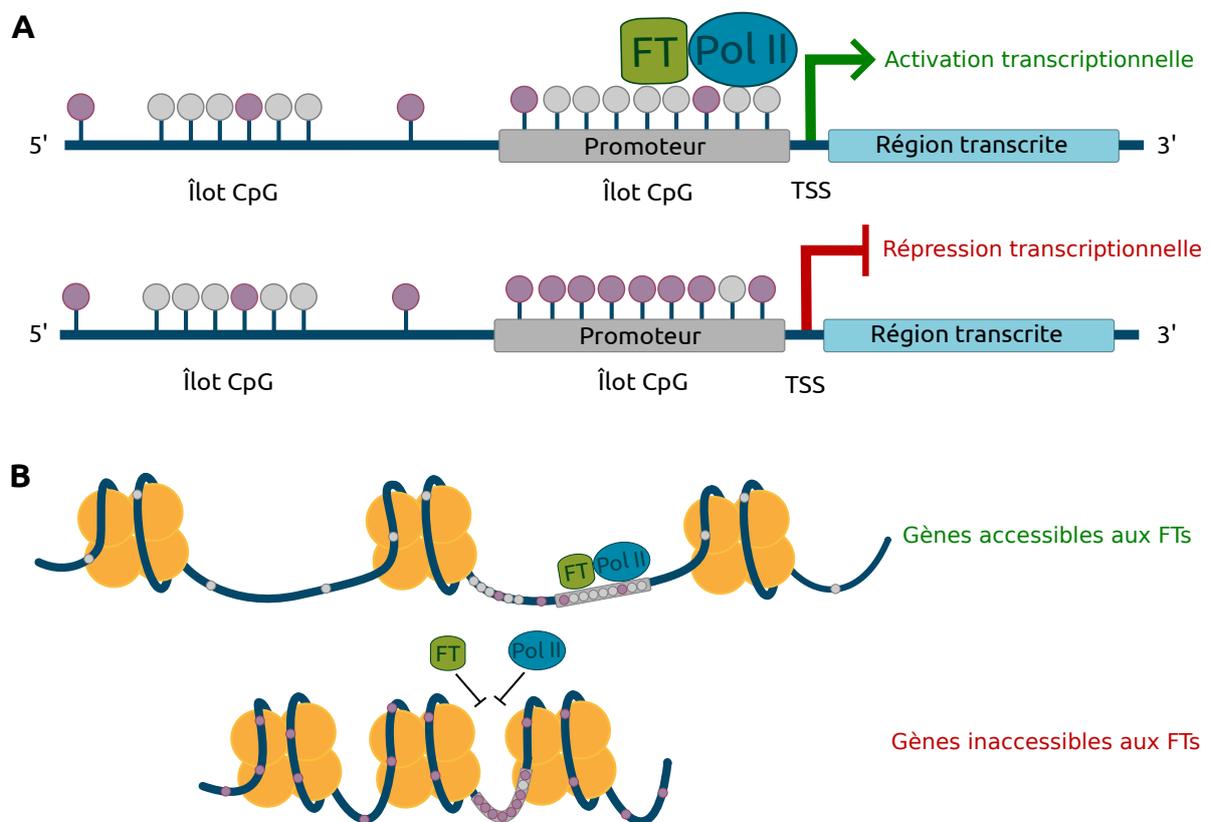


FIGURE 2.7 – Îlots CpG et expression des gènes. (A) Région promotrice riche en sites CpG (îlots CpG ou CGI) peu méthylée (en haut) et très méthylée (en bas). (B) Effet de la méthylation de l'ADN sur la compaction de la chromatine : chromatine relâchée (en haut) et condensée (en bas). Abréviations : TSS = site d'initiation de la transcription ; Pol II = ARN polymérase II ; FT = facteurs de transcription. Figure adaptée de Beaujean et al. (2020).

Le navigateur UCSC fournit la liste des CGI pour de nombreuses espèces mais pas dans le cas de la caille japonaise. Au cours de cette thèse, les CGI ont donc été directement détectés le long du génome de référence de la caille à l'aide de la commande *newcpgreport* de l'outil EMBOSS (*European Molecular Biology Open Software Suite*) (Rice et al., 2000) (<https://forgemia.inra.fr/genepi/tools/cpg-island-detection>). Le génome de référence de la caille japonaise comporte un total de 8 974 565 sites CpG dont 14 191 CGI. Parmi ces derniers, 7 808 CGI sont associés à un promoteur (55%), dans l'état actuel de l'annotation du génome. Sur les 21 561 sites d'initiation de la transcription (TSS) identifiés chez la caille, 7 793 promoteurs sont associés à un CGI (36,1%).

Un CGI est défini par sa taille ( $\geq 400$  pb), le ratio du nombre de CpG observé sur le nombre de CpG théorique (observé/théorique  $\geq 0,6$ ) et son pourcentage minimal en C+G ( $\geq 50\%$ ). Le ratio du nombre de CpG observé/théorique est calculé sur une fenêtre (région) glissante de 100 pb le long de la séquence génomique, ce dernier étant recalculé jusqu'à ce que la séquence soit entièrement parcourue. Le nombre observé de CpG dans une fenêtre correspond au nombre de fois qu'un C suivi d'un G est identifié. Pour chaque fenêtre, le nombre théorique de CpG correspond au nombre de C multiplié par le nombre de G dans la fenêtre, divisé par la longueur de cette dernière (équation 2.1) :

$$\text{Nombre de CpG théorique} = \frac{\text{Nombre de C} \times \text{Nombre de G}}{\text{Taille de la fenêtre}} \quad (2.1)$$

L'association entre un promoteur et un CGI n'est considérée que si une distance de moins de 1000 pb les sépare.

Parmi les exemples cités dans la **Chapitre 1.2**, la variabilité phénotypique des Linaires communes ainsi que des souris Agouti serait en partie médiée par la méthylation de l'ADN. La différence de conformation de la Linaire commune serait liée à une différence d'expression du gène *CYCLOIDEA* qui contrôle la symétrie de la fleur, cette différence d'expression étant elle-même médiée par la méthylation de l'ADN (Cubas et al., 1999). Lorsque le promoteur du gène n'est pas méthylé, il y a expression du gène et obtention d'une fleur normale (**Figure 2.8.A**, gauche), alors que si ce dernier est méthylé, l'expression du gène est inhibée et la fleur développe une conformation anormale (**Figure 2.8.A**, droite). Concernant les souris Agouti (*Avy*, *Agouti viable yellow*), la variabilité de la couleur du pelage serait liée à une expression plus ou moins importante du gène *Agouti* (**Figure 2.8.B**). L'allèle *Avy* résulte de l'insertion d'un rétrotransposon IAP (particule intracisternale de type A) en amont du TSS du gène *Agouti*. Ce dernier code pour la protéine ASIP (*Agouti Signaling Protein*), une molécule de signalisation paracrine, qui pousse les mélanocytes du follicule pileux à synthétiser de la phéomélanine (pigment jaune) au lieu de l'eumélanine (pigment noir ou brun). Le niveau d'expression du gène *Agouti* est corrélé au niveau de méthylation de l'IAP (Morgan et al., 1999). Plus la méthylation de l'IAP est faible, plus la protéine ASIP est produite et plus les follicules pileux produiront de phéomélanine. Ainsi, l'absence de méthylation donnera une fourrure jaune, une méthylation partielle induira des marbrures alors qu'une méthylation complète donnera une fourrure foncée (pseudo-Agouti).

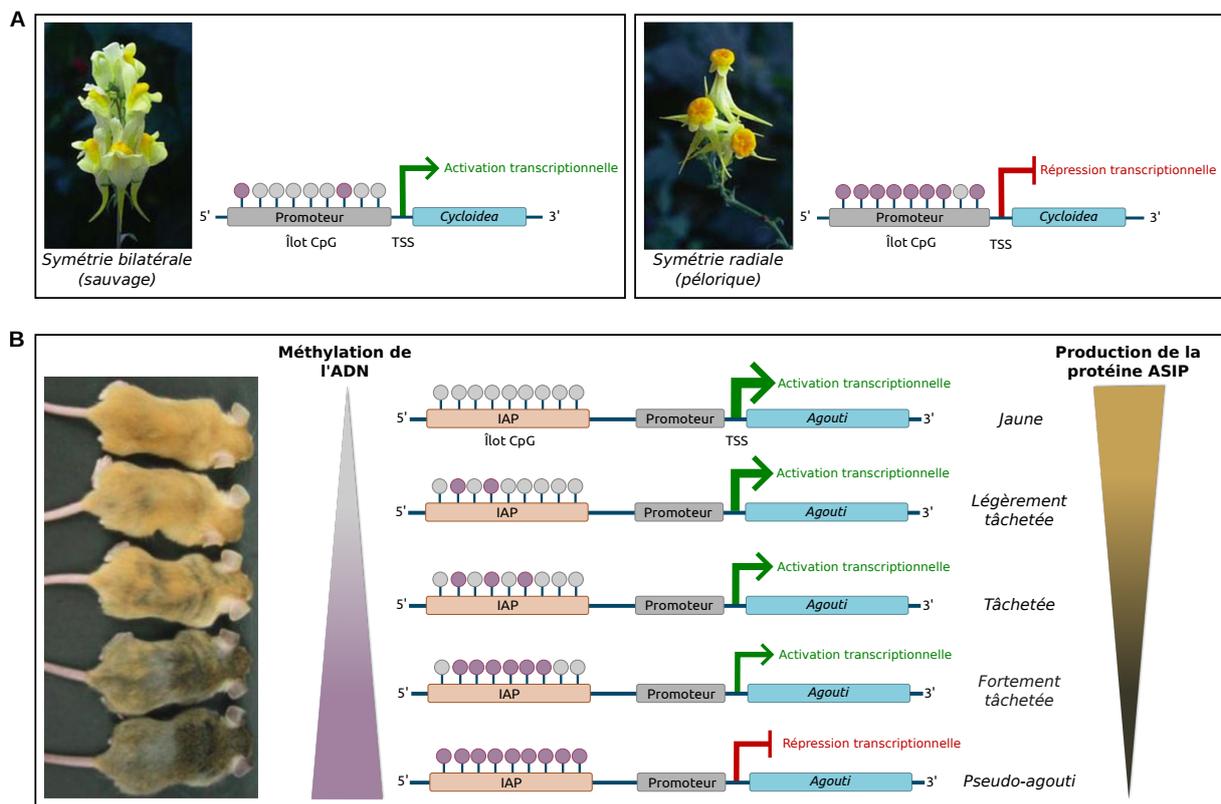


FIGURE 2.8 – Méthylation de l'ADN et variabilité phénotypique. Régulation (A) du gène *CYCLOIDEA* chez la Linaria commune (*Linaria vulgaris*) et (B) du gène *Agouti* (*Avy*, *Agouti viable yellow*) chez la souris par la méthylation de l'ADN. L'allèle *Avy* résulte de l'insertion d'un rétrotransposon IAP (particule intracisternale de type A) en amont (100 kb) du site d'initiation de la transcription (TSS) du gène *Agouti*. Plus la flèche est grande, plus l'expression du gène est importante. Photographies issues de Hitchins (2010) et Dolinoy et al. (2007).

### 2.1.4 Contrôles génétiques de la méthylation de l'ADN

Les marques épigénétiques, telles que la méthylation de l'ADN, peuvent être sous influence génétique (Bell et al., 2011). Chez l'Homme, le génotype expliquerait une grande partie de la variation de la méthylation de l'ADN (Gertz et al., 2011). Le niveau de méthylation d'un site CpG peut être spécifique à chaque allèle d'un locus, phénomène appelé méthylation allèle-spécifique (ASM, *Allele-Specific Methylation*). L'ASM peut intervenir dans les mécanismes dépendant de la séquence (Kerkel et al., 2008) (**Figure 2.9**), où la méthylation d'un site CpG peut être influencée par la séquence d'ADN environnante, notamment par la présence de polymorphismes, tels que les SNP. Par exemple, la présence d'un SNP hétérozygote (AT) à proximité d'un site CpG peut modifier la méthylation de ce site, entraînant une différence du taux de méthylation entre les deux allèles, chaque allèle ayant un taux de méthylation distinct. Les régions du génome où ces SNP sont associés à des variations du niveau de méthylation de sites CpG voisins, sont appelés des loci quantitatifs de méthylation (methQTL), en cis.

D'autres phénomènes peuvent aboutir à une méthylation différente entre les deux chromosomes d'une paire pour un locus, notamment dans le cas de gènes soumis à empreinte parentale (Barlow, 2011) et lors de l'inactivation de gènes d'un chromosome X chez les femelles (Wutz, 2011).

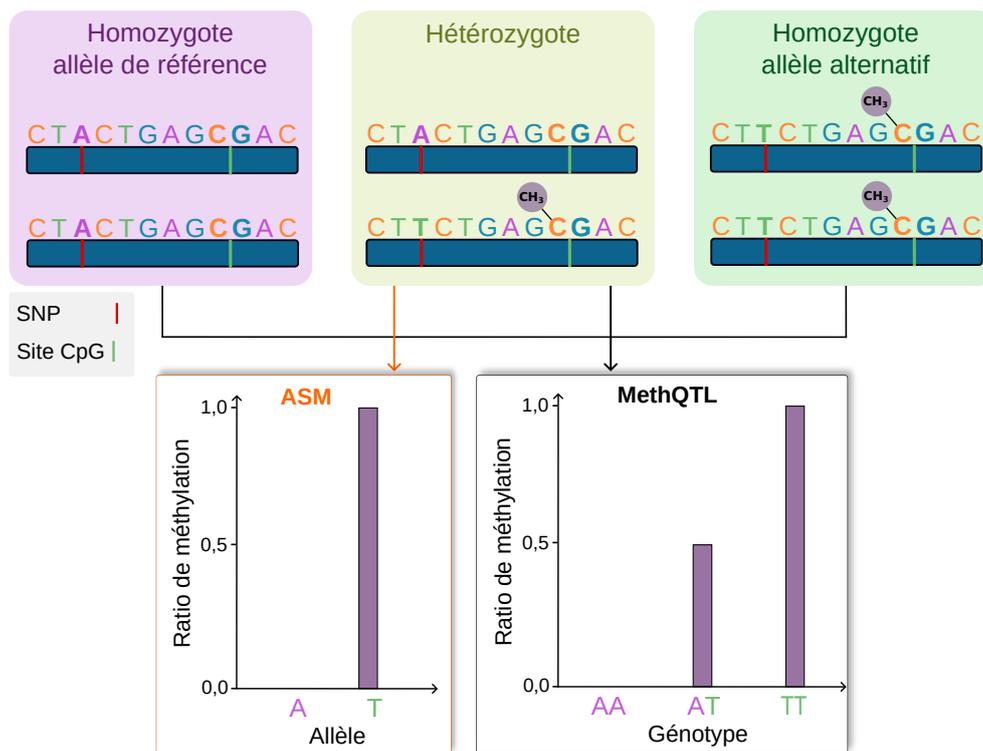


FIGURE 2.9 – **Méthylation allèle-spécifique.** Chez un individu homozygote pour l'allèle de référence (AA, gauche), le site CpG voisin de l'allèle A n'est pas méthylé. Pour un individu hétérozygote (AT, milieu), ce même site CpG n'est pas méthylé en présence de l'allèle A alors qu'il est méthylé en présence de l'allèle T. Pour un individu homozygote pour l'allèle alternatif (TT, droite), le site CpG est méthylé dans les deux cas. Les individus homozygotes AA, hétérozygotes AT et homozygotes TT ont donc respectivement un ratio de méthylation de 0, 0.5 et 1. L'ASM est caractérisée par des taux de méthylation spécifiques à chaque allèle (panneau de gauche), pouvant donner lieu à la détection de methQTL (panneau de droite), permettant d'identifier des régions du génome où des variations génétiques sont associées à des différences du taux de méthylation. *Figure adaptée de Fan et al. (2019).*

Des études chez l'Homme ont permis d'identifier des SNP associés à la fois à l'expression de gènes proches et à la méthylation de sites CpG proches (Van Eijk et al., 2012; Wagner et al., 2014). Les methQTL sont souvent colocalisés avec des QTL d'expression (eQTL) (Grundberg et al., 2013), suggérant que les effets génétiques sur l'expression des gènes peuvent être médiés par la méthylation de l'ADN. L'ASM peut avoir des implications importantes dans la régulation de l'expression allèle-spécifique (ASE, *Allele-Specific Expression*) des gènes ainsi que dans la variabilité phénotypique (Hutchinson et al., 2014).

Les profils de méthylation peuvent donc jouer un rôle dans la régulation de l'expression des gènes, mais d'autres phénomènes épigénétiques peuvent également être impliqués.

## 2.2 Modifications post-traductionnelles des histones

Les modifications post-traductionnelles des histones (MPTH) font également partie des marques épigénétiques très étudiées et sont connues pour avoir une influence sur le niveau de l'expression des gènes (Bannister & Kouzarides, 2011). Chez les eucaryotes, l'ADN s'enroule autour des nucléosomes, chacun étant constitué d'un octamère d'histones canoniques (2 protéines H2A, H2B, H3 et H4) et d'une histone de liaison à l'ADN H1, formant ainsi la chromatine (Kornberg & Lorch, 1999). Les queues d'histones étant accessibles, les acides aminés (AA) qui les composent peuvent être ciblés par des enzymes influençant la structure de la chromatine et ainsi subir des MPTH telles que la méthylation, l'ubiquitinylation, l'acétylation, la phosphorylation et la sumoylation (**Figure 2.10.A**). La lysine, arginine, thréonine et sérine sont les AA principalement sujets à des MPTH (Cheung & Lau, 2005). Les AA touchés donneront des modifications post-traductionnelles dont le rôle fonctionnel sera différent.

Ces MPTH peuvent avoir un impact au niveau de la structure de la chromatine et de la régulation de l'expression des gènes, en modifiant l'accès à l'ADN des facteurs de transcription (**Figure 2.10.B**). Comme pour la méthylation de l'ADN, certaines protéines vont ajouter (*writers*), retirer (*erasers*) ou bien lire (*readers*) ces modifications (Zhao et al., 2022). Certaines MPTH vont favoriser ou bien au contraire compliquer le recrutement de facteurs de transcription. Certaines modifications, comme H3K27ac (acétylation) et H3K4me3 (triméthylation), peuvent accentuer l'ouverture de la chromatine, favorisant le recrutement de facteurs de transcription (**Figure 2.10.B**, haut). D'autres modifications, comme H3K9me3, H3K36me3 et H3K27me3, peuvent favoriser la compaction de la chromatine et ainsi être associées à une répression transcriptionnelle (T. Zhang et al., 2015; Zhao & Garcia, 2015) (**Figure 2.10.B**, bas).

Les MPTH sont très étudiées chez les plantes et les animaux. L'influence de certaines MPTH, telles que H3K4me3, H3K27me3 et H3K27ac, sur l'expression des gènes a notamment été mise en évidence chez la poule (Jahan et al., 2016). Chez la plante *Arabidopsis thaliana*, la marque H3K27me3 est impliquée dans le phénomène de vernalisation, qui correspond à l'évocation florale déclenchée par le froid. Cette marque inhibe un répresseur floral, favorisant ainsi la floraison (Lämke et al., 2016; Coustham et al., 2012).

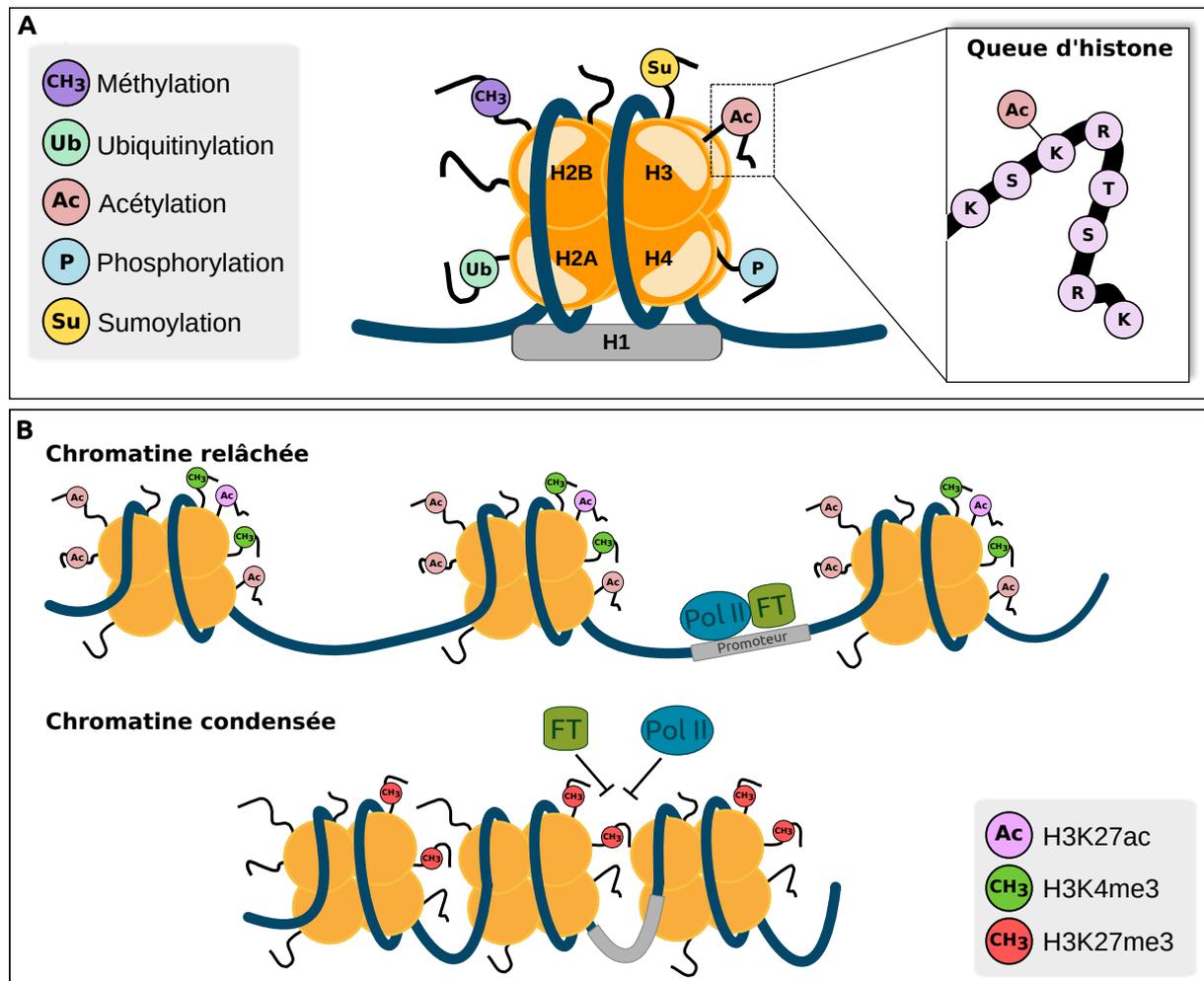


FIGURE 2.10 – **Modifications post-traductionnelles des histones (MPTH)**. (A) Représentation d'un nucléosome constitué d'histones canoniques (2 H2A, 2 H2B, 2 H3 et 2 H4) et d'une histone de liaison H1. L'ADN (en bleu) s'enroule autour du nucléosome formant la chromatine. Les queues des histones (traits noirs) peuvent subir des modifications telles qu'une méthylation (CH<sub>3</sub>), ubiquitylation (Ub), acétylation (Ac), phosphorylation (P) et sumoylation (Su), représentées respectivement en violet, vert, rouge, bleu et jaune. (B) Effet des MPTH sur la conformation de la chromatine (chromatine relâchée en haut, et condensée en bas). La H3K4me<sub>3</sub> (vert) correspond à la triméthylation de la lysine en position 4 sur l'histone H3. H3K27me<sub>3</sub> (rouge) et H3K27ac (rose) correspondent respectivement à la triméthylation et acétylation de la lysine en position 27 de l'histone H3. K, S, R et T correspondent respectivement à la lysine, sérine, arginine et thréonine *Figure adaptée de Torres-Perez et al. (2021) et Etier et al. (2022)*.

## 2.3 Autres mécanismes épigénétiques

### 2.3.1 ARN non codants

La régulation de l'expression des gènes par les acides ribonucléiques (ARN) non codants est un autre phénomène épigénétique (X. Li et al., 2016). Comme décrit précédemment (**Chapitre 1.1.1**), l'ADN peut être transcrit en ARN classés en deux grandes catégories : les ARN codants et les ARN non codants (ARNnc). Les ARN codants comprennent les ARN messagers (ARNm) qui sont traduits en protéines, tandis que les ARNnc ne subissent pas de traduction. Le groupe des ARNnc est constitué de nombreuses sous-classes subdivisées en fonction de leur taille et de leur fonction, telles que les grands ARNnc (taille supérieure à 200 nucléotides) et les petits ARNnc (taille inférieure à 200 nucléotides). Les grands ARNnc regroupent notamment les longs ARN non-codants (ARNlnc) et les ARN ribosomiques (ARNr). Les petits ARNnc regroupent entre autres les micro ARN (ARNmi), les petits ARN interférents (ARNsi), les petits ARN nucléolaires (ARNsno) et nucléaires (ARNsn), les ARN interagissant avec la protéine PIWI (ARNpi) et les ARN de transfert (ARNt). Les ARNsn sont impliqués dans le transfert des pré-ARNm dans le noyau et permettent donc l'épissage des ARNm. Les ARNsno, présents dans le nucléole des cellules eucaryotes, aident à la maturation des ARN ribosomiques (ARNr) (**Figure 2.11**).

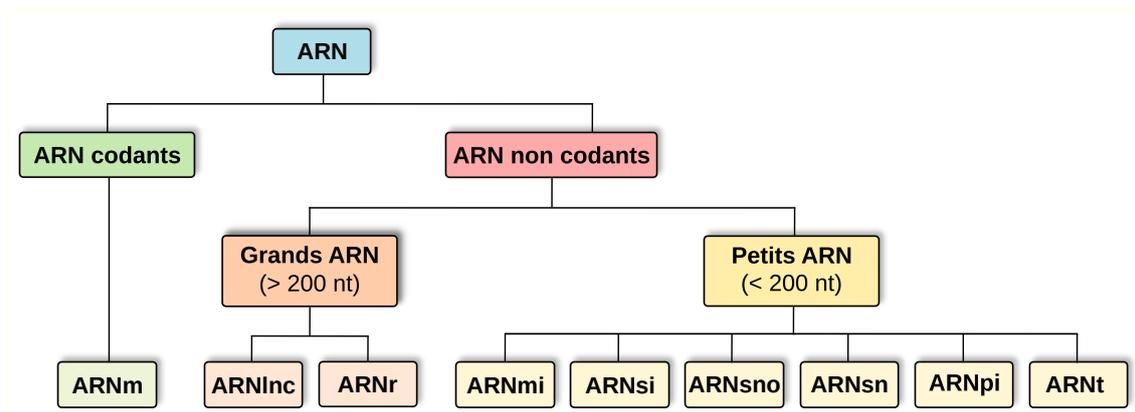


FIGURE 2.11 – Les différentes catégories d'ARN. ARNm = ARN messager, ARNr = ARN ribosomique, ARNmi = micro ARN, ARNsi = petits ARN interférents (*small interfering RNA*), ARNsno = petit ARN nucléolaire (*small nucleolar RNA*), ARNsn = petit ARN nucléaire *small nuclear RNA*, ARNpi = ARN interagissant avec PIWI, ARNt = ARN de transfert.

Les principaux petits ARNnc impliqués dans les mécanismes épigénétiques sont les ARNmi (Gangaraju & Lin, 2009), ARNsi (Verdel et al., 2009) et ARNpi (C. Li et al., 2009) (**Figure 2.12**). Ces petits ARNnc peuvent se lier à l'ARNm et ainsi influencer l'expression des gènes. Cette liaison va venir perturber le traitement normal de l'ARNm, pouvant ainsi induire une modulation de l'activité de cet ARNm (inhibition, activation, déstabilisation). Les ARNmi, produits de la transcription ou de l'épissage, se replient pour former une boucle (pré-ARNmi). Un des deux brins de l'ARN (le moins stable) va se lier au complexe RISC (*RNA-induced silencing complex*), qui va ensuite recruter à un ARNm cible spécifique contenant une séquence complémentaire du ARNmi. Cette liaison peut induire la répression ou l'activation traductionnelle de l'ARNm (**Figure 2.12.A**). Les ARNsi peuvent aussi se lier au complexe RISC et induire une dégradation de l'ARNm via la voie PTGS (*Post-Transcriptional Gene Silencing*) (**Figure 2.12.B**). Ces ARNnc

peuvent également induire des modifications de la chromatine via la voie TGS (*Transcriptional Gene Silencing*). Les ARNpi sont de petits fragments d'ARN qui se lient à la protéine PIWI, protéine régulatrice responsable de la différenciation des cellules souches et des cellules germinales, et qui vont induire une régulation de la chromatine et la répression de l'activité des transposons (séquences d'ADN mobiles pouvant induire des mutations) (**Figure 2.12.C**).

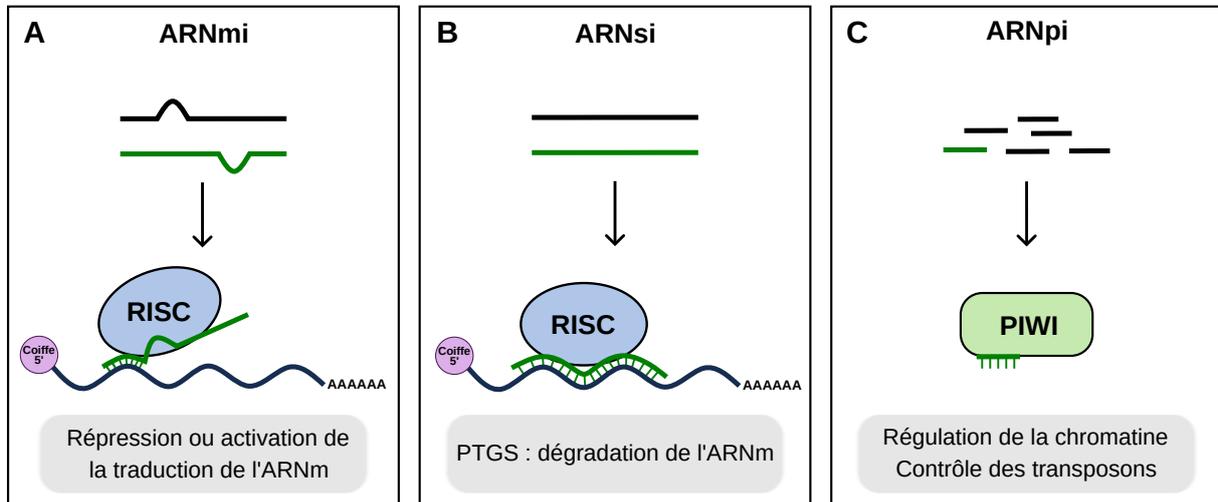


FIGURE 2.12 – Mode d'action des principaux petits ARN non codants. (A) ARNmi, (B) ARNsi et (C) ARNpi. ARNmi = micro ARN, ARNsi = petits ARN interférents, ARNpi = ARN interagissant avec PIWI, ARNm = ARN messager, RISC = *RNA-induced silencing complex*, PTGS = *Post-Transcriptional Gene Silencing*. Figure adaptée de Collins et al. (2011).

Les ARNnc ainsi que les ARNm peuvent également subir des modifications chimiques telles que la N6,2'-O-diméthyladénosine (m6Am), N6-méthyladénosine (m6A), l'inosine (I), la 5-méthylcytidine (m5C), la N1-méthyladénosine (m1A), la pseudouridine ( $\Psi$ ), et la 5-hydroxyméthylcytidine (hm5C), modifiant ainsi l'expression des gènes (X. Li et al., 2017, 2021) (**Figure 2.13**).

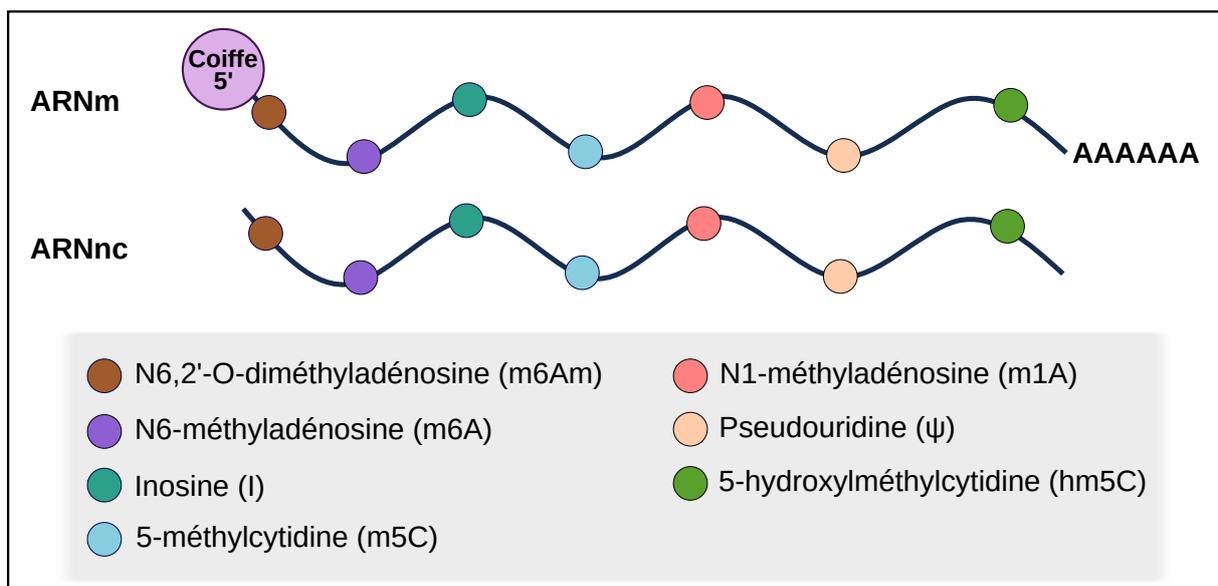


FIGURE 2.13 – Modifications chimiques des ARN. Figure adaptée de X. Li et al. (2017).

### 2.3.2 Remaniement de la chromatine

La conformation de la chromatine peut évoluer passant d'un état condensé à un état relâché et inversement (Lin et al., 2020). Cette restructuration de la chromatine va faciliter ou bien limiter l'accès à l'ADN à la machinerie transcriptionnelle. Les complexes de remaniement de la chromatine ATP (adénosine triphosphate) dépendants sont la principale cause de ces modifications structurales. Les complexes de remodelage les plus connus sont les complexes SWI/SNF (*SWItch/Sucrose Non-Fermentable*) et RSC (*Chromatin structure remodeling*) (Quina et al., 2006). Ces complexes impactent la liaison entre l'ADN et les histones en utilisant l'énergie de l'hydrolyse de l'ATP en ADP (adénosine diphosphate). L'énergie produite va permettre le mouvement des nucléosomes le long de l'ADN. Ces derniers peuvent glisser le long de l'ADN (**Figure 2.14.A**, haut), être éjectés (**Figure 2.14.A**, milieu) ou bien être remplacés (variants d'histone) (**Figure 2.14.A**, bas). Le glissement des nucléosomes va favoriser l'ouverture de la chromatine et par conséquent favoriser la transcription. Certains complexes de remodelage équilibrent la distance irrégulière entre les nucléosomes, phénomène appelé espacement des nucléosomes (**Figure 2.14.B**). Des chaperons d'histones peuvent collaborer avec un complexe de remodelage permettant ainsi d'enrouler l'ADN autour des octamères d'histones afin de former de nouveaux nucléosomes (**Figure 2.14.C**). Parmi les chaperons d'histone, *Asf1* est connue pour intervenir dans le processus de transcription. Chez les vertébrés, la *Spt2* est connue pour faciliter la transcription de l'ADN ribosomique par le remodelage de la chromatine pendant la transcription (Osakabe et al., 2013). D'autres chaperons telles que SPT6 et FACT sont connues pour s'associer à l'ARN polymérase II et ainsi faciliter la transcription (De Koning et al., 2007).

A plus large échelle, la chromatine est organisée en domaines topologiquement associés (TAD - *Topological associated domain*). Au sein d'un TAD, les régions proches interagissent plus fréquemment entre elles qu'avec des régions situées en dehors du domaine (Dixon et al., 2012). Les gènes situés dans un même TAD sont souvent corégulés, coexprimés et peuvent présenter des marques épigénétiques corrélées (Nora et al., 2012; Lazar et al., 2018). Un pan de la littérature scientifique s'intéresse à ces domaines mais ce sujet n'a pas été développé dans cette partie.

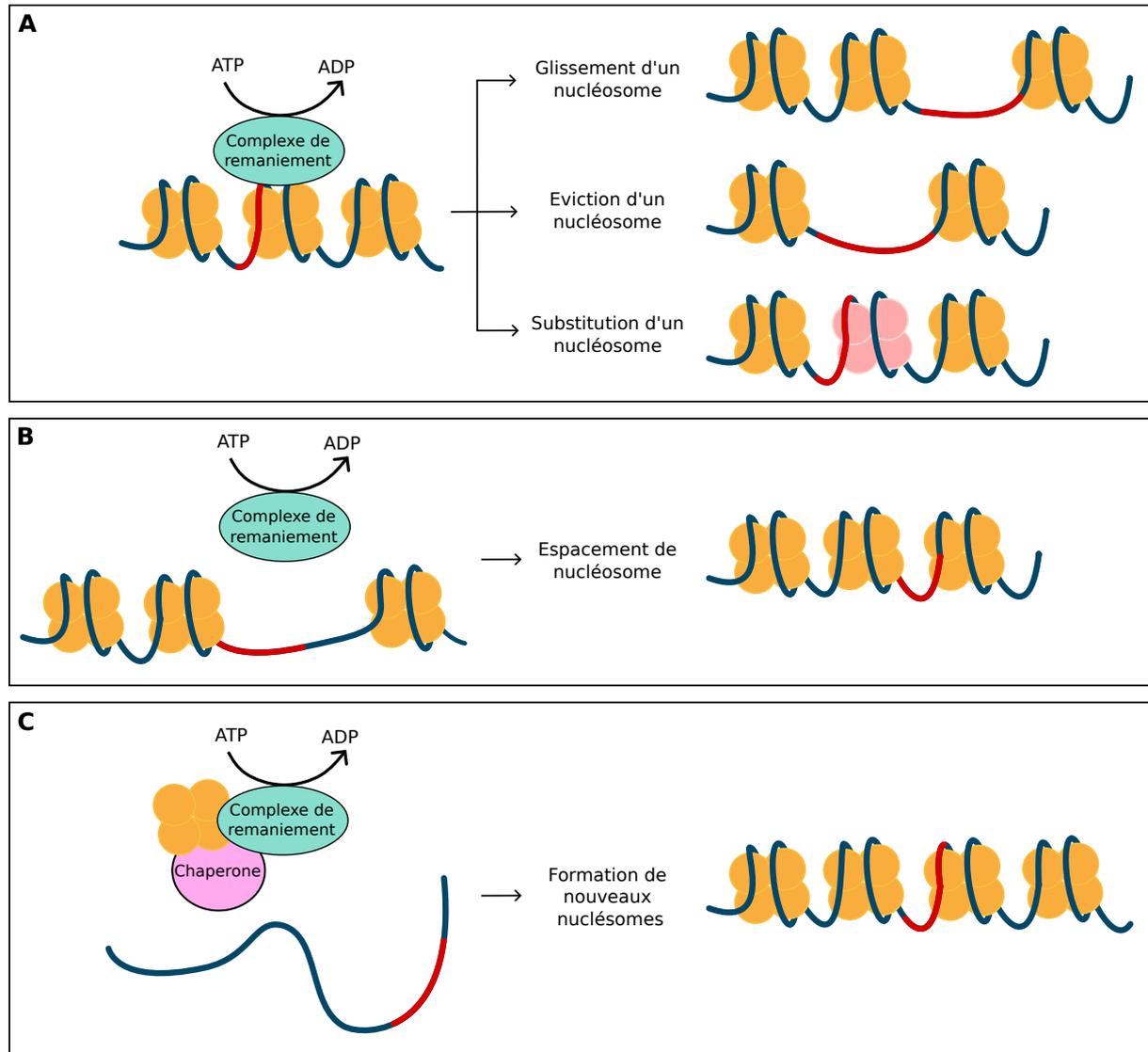


FIGURE 2.14 – Remodelage de la chromatine par le complexe de remodelage dépendant de l'ATP. (A) Conséquences du mouvement des nucléosomes : glissement d'un nucléosome pour exposer une région qui a été précédemment occultée (en haut) ; éjection d'un nucléosome pour exposer l'ADN correspondant (au milieu) ; substitution d'un nucléosome standard par une variante d'histone (en bas). (B) Certains complexes de remodelage équilibrent la distance irrégulière entre les nucléosomes, appelée espacement des nucléosomes. (C) Certains chaperons d'histones collaborent avec un complexe de remodelage pour enrouler l'ADN autour des octamères d'histones afin de former de nouveaux nucléosomes. La ligne rouge représente l'ADN de liaison ou l'ADN nucléosomal. Les variants d'histones sont représentés en rouge. *Figure adaptée de Lin et al. (2020).*

## 2.4 Des phénomènes en interaction

### 2.4.1 Interaction entre les marques épigénétiques

Les marques épigénétiques ne sont pas indépendantes et peuvent s'influencer mutuellement (**Figure 2.15**).

La méthylation de l'ADN peut être guidée par les modifications au niveau des histones (MPTH). Par exemple, la lecture de la méthylation de la lysine 36 de l'histone H3 (H3K36) par l'enzyme DNMT3A est nécessaire pour la méthylation de l'ADN (Chantalat et al., 2011 ; Martin-Herranz et al., 2019) (**Figure 2.15.A**). La méthylation de l'ADN peut également favoriser des MPTH. Des complexes d'histones méthyltransférases (HMT) dotées du domaine de liaison méthyl-CpG (MBD) couplé à la protéine SET peuvent se lier directement aux CpG méthylés et ainsi méthyliser les queues d'histones (Y. Zhang & Reinberg, 2001) (**Figure 2.15.B**).

La méthylation de l'ADN peut être aussi médiée par les ARNnc (Wang et al., 2019). Une étude portant sur l'ARNnc H19 a pu mettre en évidence, qu'en l'absence de cet ARNnc, la S-adénosyl homocystéine hydrolase (SAHH) hydrolyse la S-adénosyl homocystéine (SAH), ce qui induit la méthylation de la région promotrice de *Beclin1* (protéine intervenant dans l'autophagie) facilitée par l'enzyme DNMT3B (**Figure 2.15.C**, haut). Cependant, en présence de l'ARNnc H19, ce dernier se lie à la SAHH, ce qui favorise l'accumulation de la SAH et empêche l'enzyme DNMT3B de méthyliser la région promotrice de *Beclin1* (**Figure 2.15.C**, bas). Les ARNnc peuvent aussi être impactés par la méthylation de l'ADN (L. Yu et al., 2020). La méthylation des sites CpG du promoteur de miR-149 inhibe l'expression de miR-149 (**Figure 2.15.D**).

Certains ARNnc sont connus pour réguler les MPTH, notamment l'ARNnc HOTAIR (*HOX Transcript antisense RNA*) (Wang et al., 2022). Les extrémités 3' et 5' de cet ARNnc peuvent se lier à des complexes protéiques. La liaison entre l'extrémité 3' de l'ARNnc et du complexe de la déméthylase spécifique à la lysine 1 (LSD1) induit la déméthylation de H3K4 (H3K4me2) ce qui inactive l'expression des gènes cibles. La liaison entre l'extrémité 5' de l'ARNnc et le complexe PRC2 (*Polycomb Repressive Complex 2*) induit la méthylation de H3K27 (**Figure 2.15.E**).

L'expression des ARNnc tels que les miARN peut être contrôlée par les MPTH (Bure et al., 2022). Les marques épigénétiques telles que la méthylation de l'ADN et les MPTH peuvent contribuer au contrôle transcriptionnel de l'expression des ARNmi. Dans le cas du miR-127, la méthylation des sites CpG et la désacétylation des histones autour de sa région promotrice induisent une répression transcriptionnelle dans les lignées cellulaires tumorales. Un traitement contenant un inhibiteur de l'histone désacéthylase (acide 4-phénylbutyrique ou Pba) induit une réduction de la méthylation de l'ADN et une augmentation de l'acétylation des histones, ce qui permet aux ARNmi d'être exprimés (Chuang & Jones, 2007) (**Figure 2.15.F**).

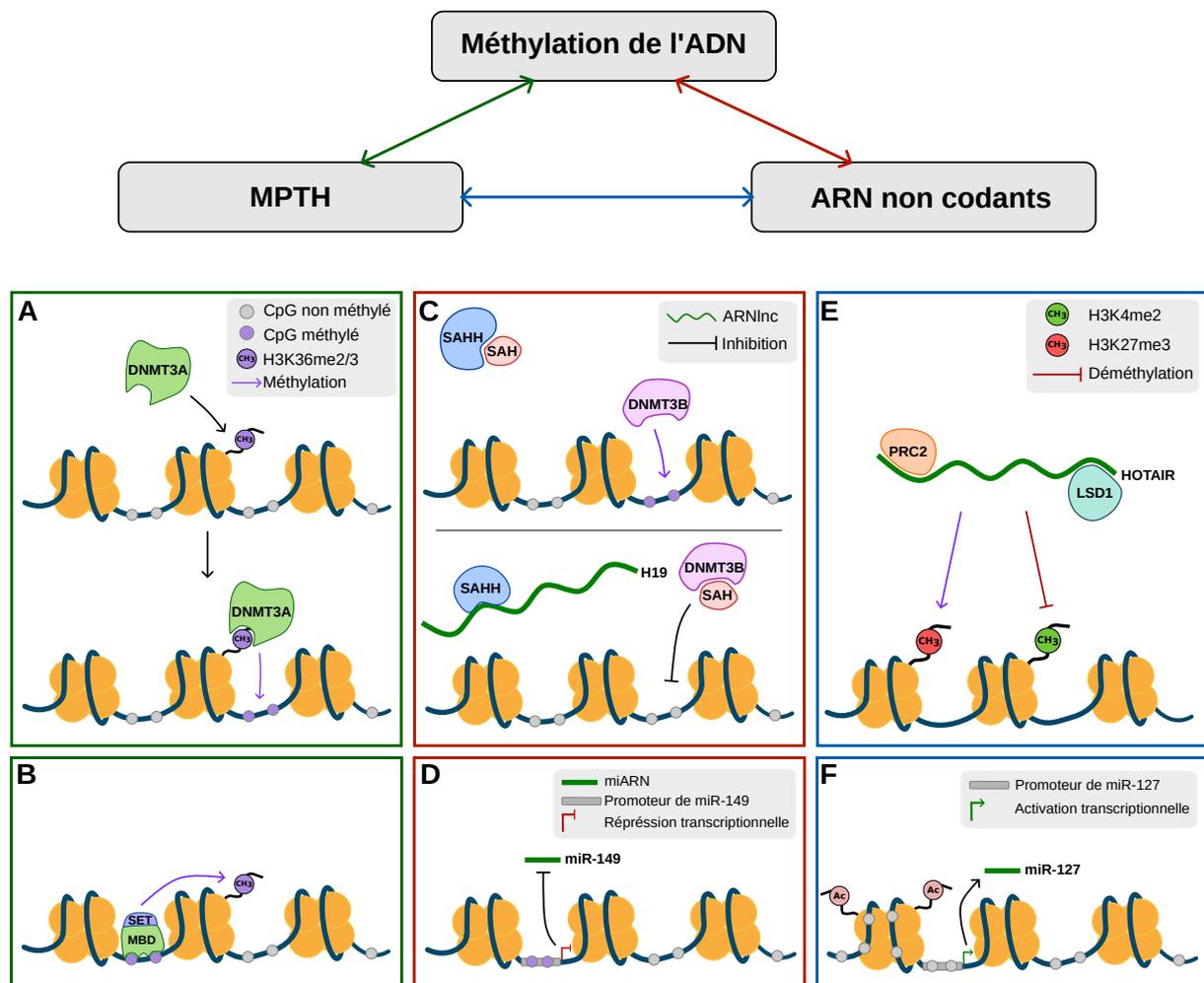


FIGURE 2.15 – **Interaction entre les marques épigénétiques.** (A) Effet des MPTH sur la méthylation de l'ADN : la H3K36me2/3 permet le recrutement de la DNMT3A (ADN méthyltransférase 3A) ce qui induit la méthylation de l'ADN environnante (*figure adaptée de Martin-Herranz et al. (2019)*); (B) Effet de la méthylation de l'ADN sur les MPTH : les histones méthyltransférases (HMT) contenant le domaine de liaison méthyl-CpG (MBD) couplé à la protéine SET se lient aux CpG méthylés et méthylent les queues d'histones (Y. Zhang & Reinberg, 2001); (C) Effet des ARNnc sur la méthylation de l'ADN : en absence de l'ARNnc H19 (haut), la S-adenosyl homocystéine hydrolase (SAHH) hydrolyse S-adenosyl homocystéine (SAH), provoquant la méthylation facilitée par l'enzyme DNMT3B de la région promotrice de *Beclin1* (*figure adaptée de Wang et al. (2019)*); En présence de H19 (bas), SAHH est inhibée ce qui engendre l'accumulation de SAH empêchant l'enzyme DNMT3B de méthyler la région promotrice de *Beclin1* (*figure adaptée de Wang et al. (2019)*); (D) Effet de la méthylation de l'ADN sur les ARNnc : la méthylation de la région promotrice de l'ARNnc-194 va induire une répression transcriptionnelle du gène codant pour cet ARNnc (*figure adaptée de L. Yu et al. (2020)*); (E) Effet des ARNnc sur les MPTH : la liaison entre l'extrémité 3' de l'ARNnc HOTAIR (*HOX Transcript antisense RNA*) et du complexe de la déméthylase spécifique à la lysine 1 (LSD1) induit la déméthylation de H3K4 (H3K4me2) ce qui inactive l'expression des gènes cibles. La liaison entre l'extrémité 5' de l'ARNnc et le complexe PRC2 (*Polycomb Repressive Complex 2*) induit la méthylation de H3K27 (*figure adaptée de Wang et al. (2022)*); (F) Effet des MPTH sur les ARNnc : l'acétylation des histones et la réduction de la méthylation de l'ADN permet l'expression de miR-127 (*Figure adaptée de Chuang et Jones (2007)*).

Comme mentionné précédemment (**Chapitre 2.1.4**), les marques épigénétiques jouent un rôle dans les phénomènes d’empreinte parentale (Barlow, 2011) et d’inactivation du chromosome X (Wutz, 2011).

### 2.4.2 Empreinte génomique parentale

La majorité des gènes est exprimée à partir de deux copies parentales (ou allèles), alors que pour certains gènes, une seule copie est exprimée. Ces gènes sont soumis à l’empreinte génomique parentale, c’est-à-dire qu’ils peuvent être exprimés à partir d’un seul allèle en fonction de son origine parentale (s’il est issu de la mère ou du père). Dans ce cas, l’allèle hérité de la mère et celui hérité du père ne sont pas exprimés de la même manière, la descendance portant donc un allèle actif et un allèle inactif. La méthylation de l’ADN peut intervenir dans l’empreinte parentale et donc sur le niveau d’expression de ces gènes. Dans les deux exemples présentés en **Figure 2.16.A-B**, l’allèle actif, non méthylé, est hérité du père, alors que l’allèle inactif, méthylé, est hérité de la mère. Dans les deux cas, la descendance exprime l’allèle issu du père, la méthylation de l’ADN dépendant de l’empreinte génomique parentale, non pas de la séquence. Chez la souris, le locus *H19-IGF2* est constitué d’une région de contrôle de l’empreinte méthylée sur l’allèle paternel. Sur l’allèle maternel, cette région n’est pas méthylée et est liée au répresseur transcriptionnel CTCF réduisant l’expression d’*IGF2* et stimulant la transcription de *H19* (**Figure 2.16.C**). C’est donc l’allèle paternel du gène *IGF2* qui est exprimé au cours du développement (Pidsley et al., 2012). Les gènes soumis à l’empreinte parentale jouent un rôle important dans le développement ainsi que le comportement de certains mammifères (Barton et al., 1984; Ho-Shing & Dulac, 2019; Hubert & Demars, 2022). Le phénomène d’empreinte n’a pas été mis en évidence chez les oiseaux et semble spécifique des mammifères (Fresard et al., 2014; Zhuo et al., 2017).

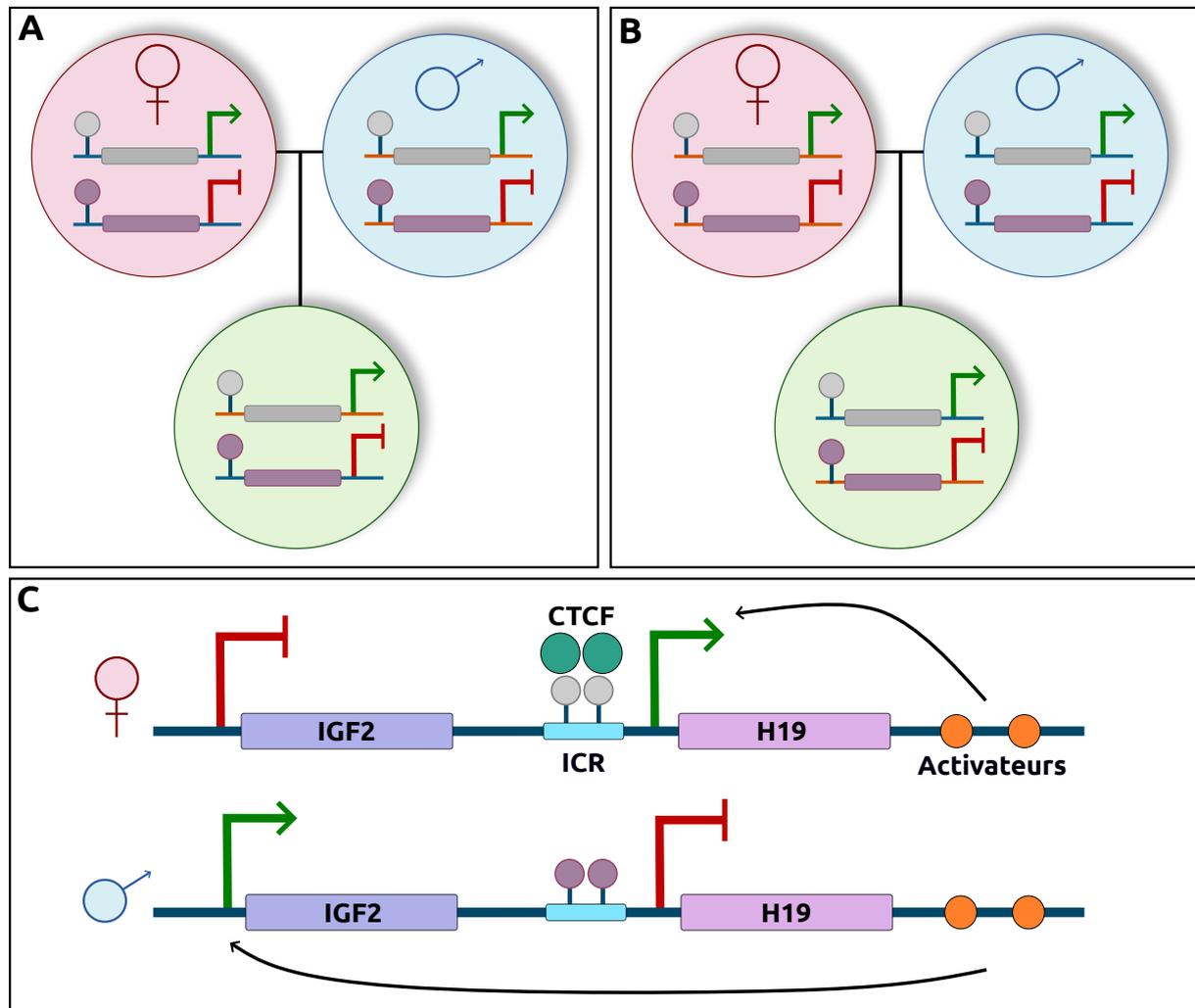


FIGURE 2.16 – Méthylation allèle-spécifique liée à l’empreinte génomique parentale. (A) Croisement entre une femelle (♀, cercle rouge) et un mâle (♂, cercle bleu) dont la descendance (cercle vert) est porteuse de l’allèle non méthylé issu du père (en orange) et de l’allèle méthylé issu de la mère (en bleu). (B) Croisement entre une ♀ et un mâle ♂ dont la descendance est porteuse de l’allèle non méthylé issu du père (en bleu) et de l’allèle méthylé issu de la mère (en orange). Dans les deux cas, A et B, l’allèle maternel est réprimé, quelle que soit sa séquence. (C) Empreinte parentale au locus *H19-IGF2* : le locus *H19-IGF2* est constitué d’une région de contrôle de l’empreinte (ICR) méthylée sur l’allèle paternel. Sur l’allèle maternel, cette ICR n’est pas méthylée et est liée à la protéine insulatrice CTCF qui empêche l’accès du promoteur de l’*IGF2* aux activateurs en aval (*Figure adaptée de la thèse de Kameswaran et Kaestner (2014)*). La méthylation de l’allèle dépend de l’empreinte génomique parentale, non pas de la séquence. Les CpG non méthylés et méthylés sont représentés respectivement en gris et violet. La répression et l’inactivation transcriptionnelle sont représentées respectivement par les flèches rouges et vertes.

### 2.4.3 Inactivation du chromosome X

Chez les oiseaux et les mammifères, le sexe d'un individu est déterminé par la paire de chromosomes sexuels. Chez la plupart des mammifères et certains insectes (coléoptères et drosophiles), les femelles possèdent deux chromosomes X (XX, homogamétique) et les mâles possèdent un chromosome X et un chromosome Y (XY, hétérogamétique), appelé le système XY. Chez les oiseaux, certains reptiles (serpents, lézards), poissons et insectes (papillons), le déterminisme sexuel est régi par le système ZW : les mâles possèdent deux chromosomes Z et les femelles possèdent un chromosome Z et un chromosome W (Irwin, 2018). Le sexe hétérogamétique (mâle chez les mammifères - XY, femelle chez les oiseaux - ZW) devrait donc avoir un niveau d'expression des gènes (dose des gènes) plus faible. Pour équilibrer l'expression de ces gènes entre les deux sexes, un mécanisme de compensation de dose, en système XY, peut intervenir (Disteche, 2012). Chez les mammifères dotés du système XY, cette compensation consiste à inactiver l'un des deux chromosomes X chez les femelles (compensation globale) (**Figure 2.17**). Un exemple bien connu de ce phénomène est celui du chat calico dont le pelage est tricolore (variégation) chez les femelles hétérozygotes, les gènes de coloration étant portés par le chromosome X (Osgood, 1994). Chez les oiseaux, la compensation de dose a bien été mise en évidence mais cette dernière est incomplète, la plupart des gènes liés au chromosome Z ayant des niveaux d'expression variables entre les deux sexes (Ellegren et al., 2007). Cette compensation de dose serait également incomplète chez le Monstre de Gila (lézard) (Webster et al., 2023). La compensation de dose peut être en partie médiée par des marques épigénétiques. Chez les oiseaux mâles, la région MHM (*Male Hypermethylated*) est un locus du chromosome Z fortement méthylé (hyperméthylé) dont l'expression est réprimée (Itoh et al., 2010). Chez les femelles, ce locus, transcrit en ARNlnc, est très peu méthylé (hypométhylé) et fortement acétylé (hyperacétylé) au niveau de la lysine 16 de l'histone H4 (H4K16ac) (Bisoni et al., 2005).

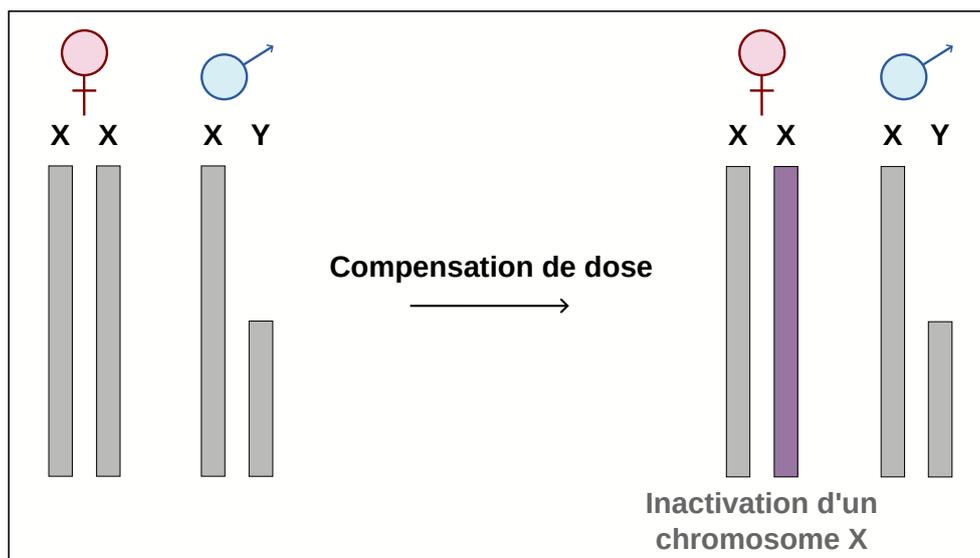


FIGURE 2.17 – **Inactivation du chromosome X.** Les chromosomes XX (femelles, ♀) et XY (mâles, ♂) sont représentés par les rectangles. Les rectangles gris et violets représentent respectivement les chromosomes actifs et inactifs.

### 3. Exposition aux effets de l'environnement

Tout au long de sa vie, un individu est exposé à des facteurs environnementaux, auxquels il doit s'adapter, et qui peuvent avoir des répercussions sur son développement (**Figure 3.1**). L'exposition aux effets de l'environnement joue un rôle crucial dans la construction des phénotypes des individus. Une exposition précoce, pendant les stades critiques du développement (embryonnaire, fœtal, et post-natal), peut avoir des effets durables sur les phénotypes des individus. L'exposition continue à certains facteurs environnementaux, notamment les perturbateurs endocriniens, peut avoir des effets cumulatifs au fil du temps, entraînant des modifications des phénotypes. Ces expositions peuvent affecter divers aspects d'un individu, tels que la croissance, le métabolisme, la réponse au stress, qu'elle soit physiologique ou comportementale, et la susceptibilité de développer certaines maladies. Chez de nombreuses espèces, cette influence de l'environnement sur la construction du phénotype est en partie médiée par l'intermédiaire de marques épigénétiques, telles que la méthylation de l'ADN. L'interaction complexe entre l'environnement et ces phénomènes épigénétiques contribue ainsi à la construction des phénotypes.

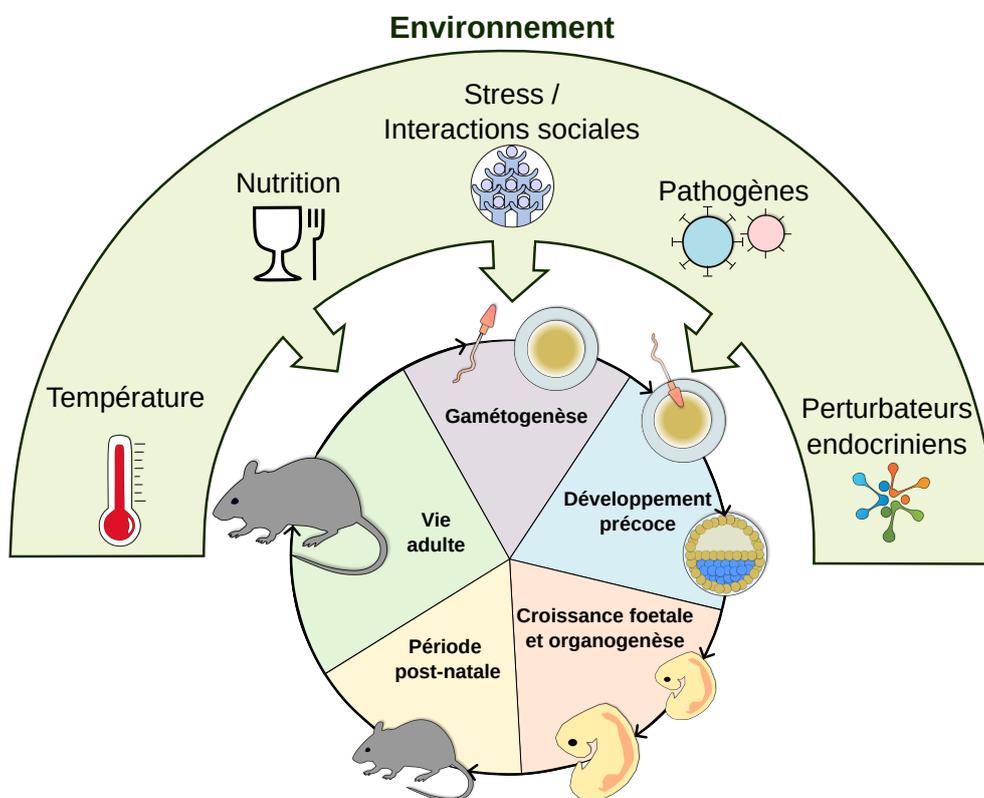


FIGURE 3.1 – Impact de l'environnement sur le développement.

### 3.1 Phases critiques du développement embryonnaire

Pendant la période embryonnaire, le développement d'un individu peut-être impacté sur le long terme en raison d'une grande sensibilité à son environnement (D. H. Ho et al., 2011 ; Reed & Clark, 2011). Il s'agit d'une période critique pendant laquelle un individu va subir des transformations majeures et rapides. Chez les mammifères, le développement est constitué de quatre phases critiques : la gamétogenèse, le développement pré-implantatoire précoce, la croissance fœtale et l'organogenèse ainsi que la période post-natale (Panchenko et al., 2015) (**Figure 3.2.A**), phases au cours desquelles l'épigénome évolue, notamment le profil de méthylation de l'ADN (**Figure 3.2.B**).

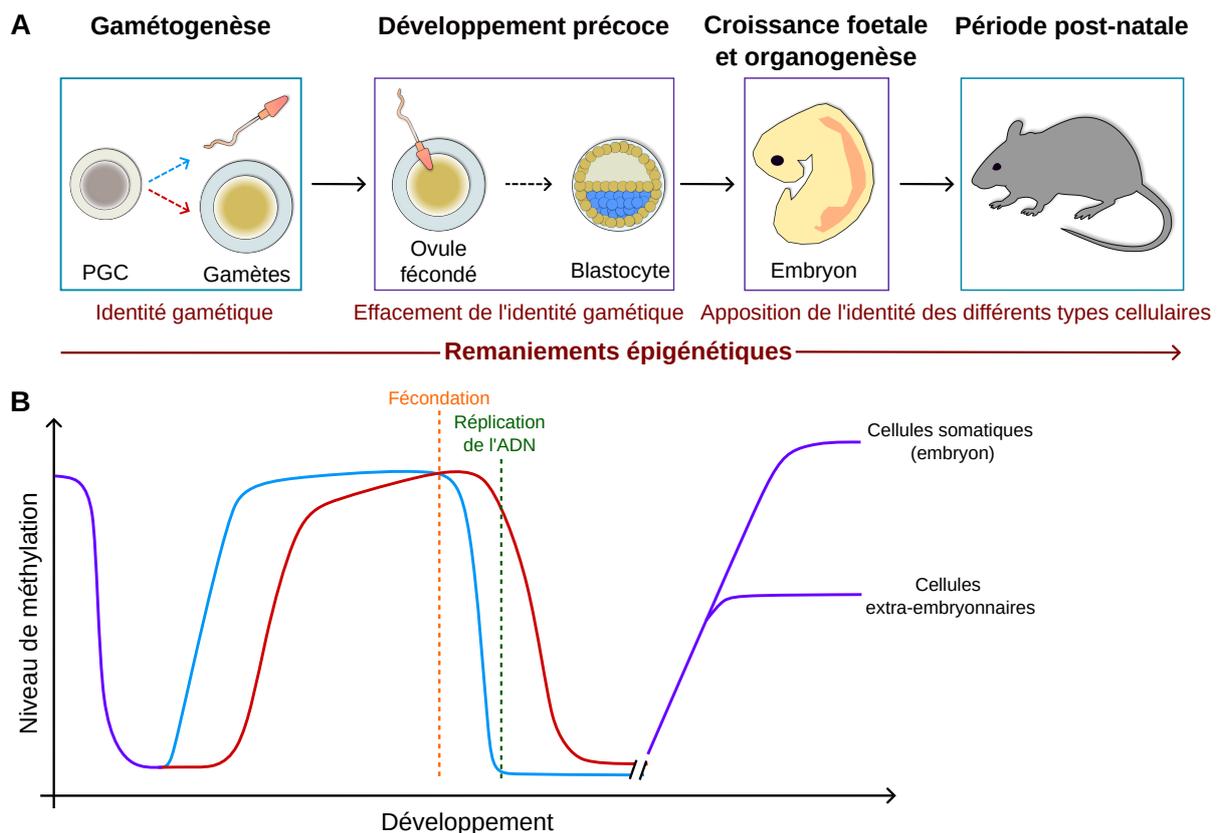


FIGURE 3.2 – Mémoire de l'environnement précoce chez la souris. (A) Les quatre phases critiques du développement : la gamétogenèse, le développement préimplantatoire précoce, la croissance fœtale et l'organogenèse, et la période postnatale. Les flèches discontinues bleue et rouge représentent respectivement la spermatogenèse et l'ovogenèse. PGC = cellules germinales primordiales. Les cadres bleus et violets représentent respectivement l'environnement maternel/paternel et maternel. (B) Evolution du profil de méthylation au cours du développement embryonnaire : une première déméthylation a lieu durant la gamétogenèse dans les PGC et touche le génome maternel et paternel (en violet), suivie d'une méthylation du génome paternel (en bleu) pendant la spermatogenèse et, plus tardivement, du génome maternel (en rouge) durant l'ovogenèse. Après la fécondation, une seconde vague consiste en une déméthylation rapide et active du génome paternel et passive du génome maternel, suivie d'une reméthylation (en violet) du génome embryonnaire dans les cellules somatiques et extra-embryonnaires. L'évolution du profil de méthylation peut varier en fonction des espèces, elle est bien décrite chez les mammifères mais est méconnue chez les oiseaux. *Figures adaptées de Reik et al. (2001), Panchenko et al. (2015) et Gabory (2008).*

La gamétogenèse, qui permet la formation de gamètes fonctionnels à partir des cellules germinales primordiales (PGC), garantit la diversité génétique ainsi que la transmission de l'information génétique à la descendance (**Figure 3.2.A**, gauche). Le développement pré-implantatoire précoce désigne l'étape initiale du développement embryonnaire qui se produit immédiatement après la fécondation d'un ovule par un spermatozoïde et avant l'implantation de l'embryon dans l'utérus. Il s'agit donc d'une étape critique et déterminante pour le bon développement de l'embryon (**Figure 3.2.A**, milieu). Pendant la croissance fœtale et l'organogenèse, les organes se forment et se développent à partir de cellules embryonnaires différenciées. La période post-natale désigne l'étape survenant après la naissance durant laquelle l'individu subit une croissance rapide et continue ainsi qu'un renforcement de ses capacités physiques, cognitives, émotionnelles et de son système immunitaire. L'alimentation, les interactions sociales et d'autres facteurs intervenant durant cette période peuvent influencer son développement. Cette période d'acquisition de compétences et d'habitudes est donc cruciale pour le développement ultérieur de l'individu (**Figure 3.2.A**, droite).

Les marques épigénétiques, telles que la méthylation de l'ADN, sont connues pour être dynamiques lors du développement embryonnaire (Zeng & Chen, 2019 ; Frésard et al., 2013) et peuvent, dans certains cas, être perturbées par l'environnement précoce, et garder des traces de ces perturbations à plus long terme. Ces marques sont sensibles à l'environnement et peuvent être transmises au cours des divisions cellulaires. Durant le développement, elles subissent un remodelage qui va définir l'identité cellulaire. Le profil de méthylation évolue et est parcouru par deux grandes phases de déméthylation et reméthylation, appelées phases de reprogrammation épigénétique (**Figure 3.2.B**). L'évolution de la méthylation a surtout été étudiée chez les mammifères (souris, Homme) (Monk, 2015), elle peut varier légèrement en fonction des espèces mais est moins connue chez les oiseaux. La première phase correspond à la reprogrammation de la méthylation des cellules germinales. Durant la gamétogenèse, les marques épigénétiques sont effacées, les cellules germinales primordiales (PGC) subissant une déméthylation. Un épigénome spécifique des gamètes est ensuite établi, se traduisant par une augmentation de la méthylation de l'ADN en premier lieu du génome paternel, lors de la spermatogenèse, puis du génome maternel, lors de l'ovogenèse (**Figure 3.2.B**, gauche). Lors de la spermatogenèse, les histones des cellules germinales mâles sont remplacées par des protamines, ce qui permet la condensation de l'ADN dans la tête des spermatozoïdes (Gaucher et al., 2010). En effet, ces dernières se lient fortement à l'ADN permettant la condensation de la chromatine spermatique ainsi que l'inactivation du génome pendant la spermatogenèse. Ce remplacement induit l'effacement des marques histones. La seconde phase de reprogrammation de la méthylation a lieu après la fécondation. Lors du développement précoce, l'identité gamétique est effacée, se traduisant par une déméthylation rapide et active du génome paternel, immédiatement après la fécondation, suivie d'une déméthylation passive du génome maternel, dépendant de la réplication de l'ADN. L'ensemble est ensuite reméthylé au moment de l'implantation de l'embryon, dans les cellules embryonnaires et extra-embryonnaires (**Figure 3.2.B**, droite). Lors de la fécondation, les protamines sont de nouveau remplacées par des histones, juste après l'entrée du spermatozoïde dans l'ovule, permettant la formation des pronoyaux. L'environnement précoce peut donc avoir des conséquences durables sur la santé, le développement

physique et cognitif des individus. Des changements dans l'environnement paternel et maternel pourraient induire des modifications de l'établissement de ces marques épigénétiques (**Figure 3.2.A**). Cela peut modifier la régulation de l'expression de gènes impliqués dans le développement et influencer sur le long terme le développement et la santé de l'individu (Attig et al., 2010 ; A. Gabory et al., 2011a).

## 3.2 Facteurs environnementaux

Des facteurs environnementaux tels que la fluctuation de la température, l'alimentation, les interactions sociales, le stress, la présence de pathogènes, les substances chimiques ou biologiquement actives, et d'autres éléments environnementaux peuvent façonner le développement et influencer les caractéristiques phénotypiques des individus (Feil & Fraga, 2012 ; Tammen et al., 2013).

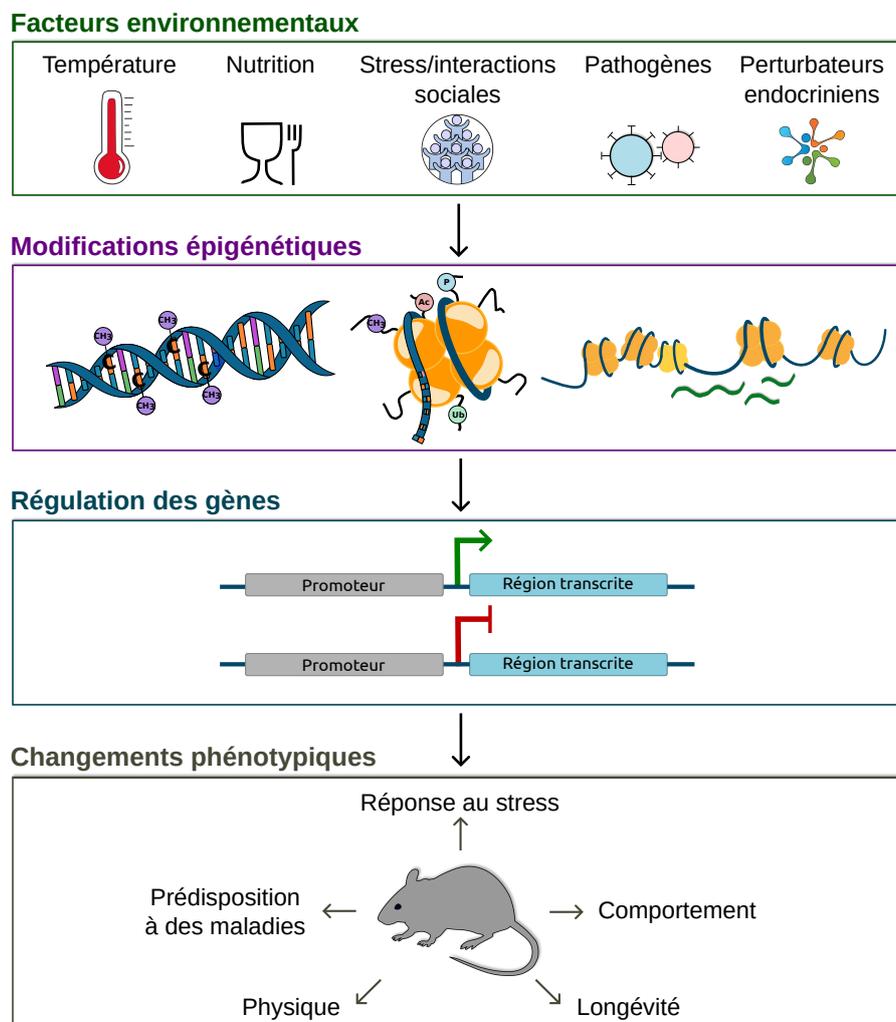


FIGURE 3.3 – **Effets des facteurs environnementaux sur les phénotypes.** Présentation des différents facteurs de l'environnement susceptibles d'engendrer des modifications épigénétiques. La régulation des gènes est en partie médiée par ces marques épigénétiques ce qui peut avoir des répercussions sur le développement d'un individu. *Figure adaptée de Tammen et al. (2013).*

### 3.2.1 Température

Chez de nombreuses espèces de reptiles (tortues et alligators) et de poissons, la fluctuation de la température, lors de la période embryonnaire et larvaire, peut avoir un effet sur le développement, notamment le déterminisme sexuel (Mizoguchi et al., 2022).

Par exemple, chez la tortue de Floride (*Trachemys scripta elegans*), en fonction de la température, un même œuf peut donner un mâle ou une femelle (**Figure 3.4**). Une température faible (26°C) va favoriser l'expression du gène *Kdm6b*, la lysine déméthylase 6B (KDM6B) ainsi produite déméthyle le promoteur de *Dmrt1*<sup>1</sup> (facteur de transcription 1 lié au double sexe et au mab-3), ce qui augmente son expression et favorise le développement d'un mâle (**Figure 3.4.A**). Si la température est plus élevée (32°C), l'expression de *Kdm6b* et de *Jarid2* est inhibée mais le gène *Jarid2* serait suffisamment exprimé pour permettre à PRC2<sup>2</sup> de triméthyle H3K27 au niveau du promoteur de *Dmrt1*. La répression du gène *Dmrt1* conduirait ainsi au développement d'une femelle (**Figure 3.4.B**).

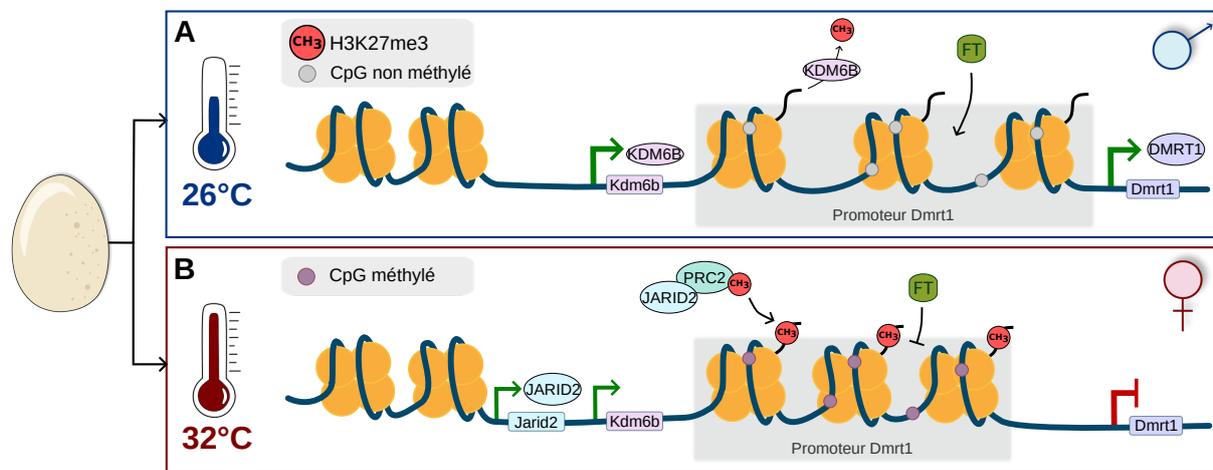
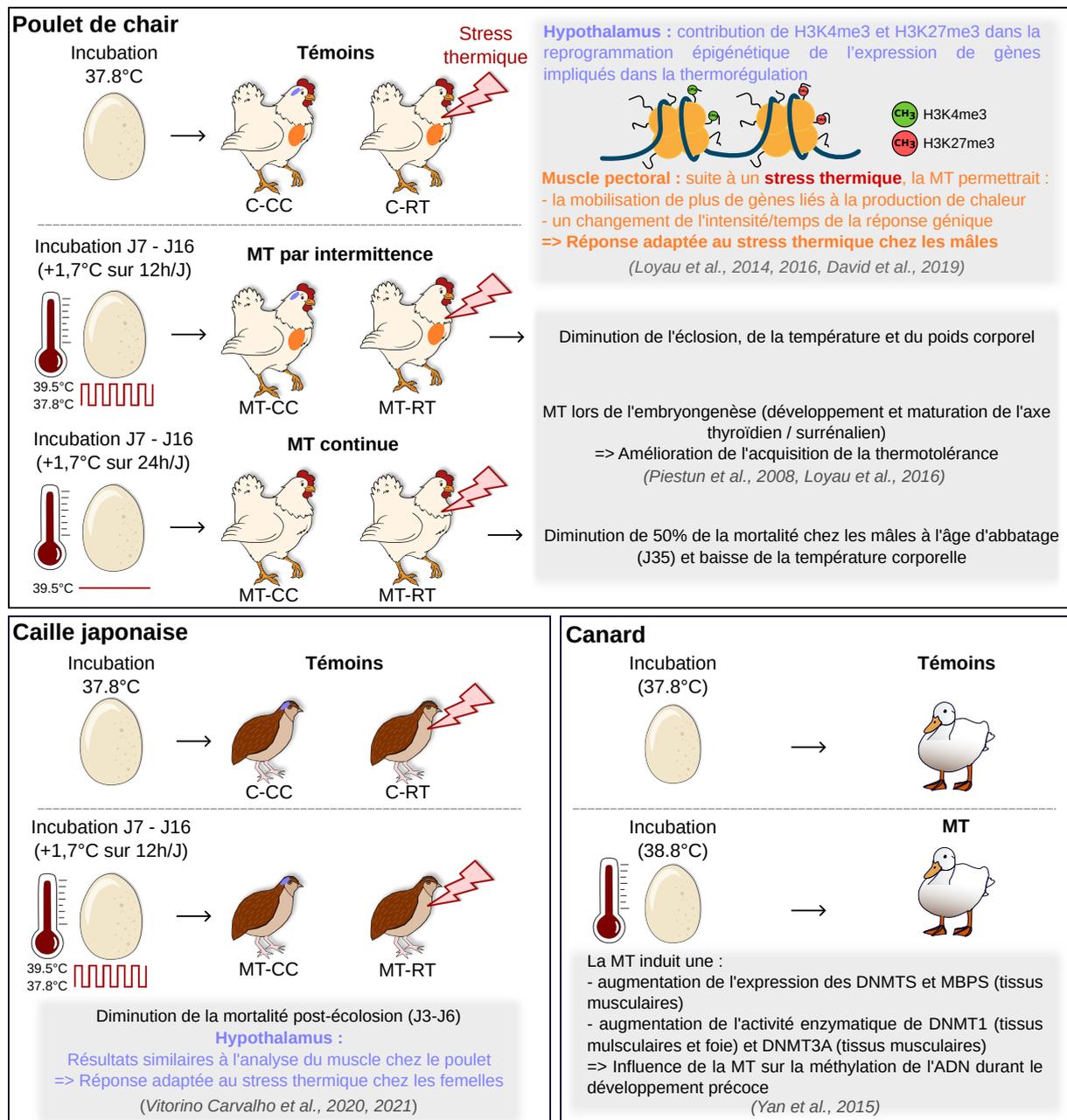


FIGURE 3.4 – Effet de la température sur le déterminisme sexuel chez la tortue (*Trachemys scripta elegans*). (A) Une température faible (26°C) produit un mâle (♂). Le gène *Kdm6b* est sur-exprimé et la protéine KDM6B produite va supprimer les marques H3K27me3 au niveau du promoteur de *Dmrt1*, conduisant à l'ouverture de la chromatine et laissant l'accès aux facteurs de transcription. (B) Une température plus élevée (32°C) conduit au développement d'une femelle (♀). Les gènes *Kdm6b* et *Jarid2* sont inhibés, mais le gène *Jarid2* serait suffisamment exprimé pour permettre au complexe PRC2, lié à JARID2, de méthyle la H3K27 (H3K27me3) au niveau du promoteur de *Dmrt1*, réprimant la transcription de ce dernier. Figure adaptée de Castelli et al. (2020) et Georges et Holleley (2018).

Dans le contexte actuel, l'un des enjeux majeurs des productions animales est de conserver la productivité des animaux tout en diminuant l'utilisation des ressources, l'impact environnemental et en préservant leur santé et leur bien-être. L'objectif est également d'avoir des animaux capables de s'adapter au mieux à un environnement en constante évolution, dans le but d'assurer une certaine pérennité des filières. Chez les espèces aviaires, notamment le poulet, leurs performances n'ont pas cessé de croître ces dernières années parfois au détriment de leur capacité à s'adapter à leur milieu. Concernant les poulets de chair par exemple, la masse corporelle des animaux étant disproportionnée par rapport à leur appareil cardio-respiratoire, leur thermorégulation n'est pas optimale, les rendant particulièrement sensibles à des fluctuations de température

1. Doublesex and mab-3 related transcription factor 1  
2. Polycomb Repressive Complex 2

(Settar et al., 1999 ; Deeb & Cahaner, 2002). Des chercheurs ont ainsi élaboré une stratégie de programmation thermique précoce dans le but d'améliorer la résistance à la chaleur des espèces aviaires, notamment le poulet de chair. Cette partie a été détaillée dans la revue de Productions Animales (Coustham et al., 2023, **Annexe 19**, partie 2.2. Programmation thermique), dont je suis co-auteur, portant sur la contribution de l'épigénétique dans la sélection de volailles plus performantes et les mieux adaptées à leur environnement. Certains éléments présentés dans cette partie sont synthétisés dans la figure ci-dessous (**Figure 3.5**).



**FIGURE 3.5 – Programmation thermique chez les volailles.** Exemple de traitement de manipulation thermique embryonnaire à la chaleur (MT) chez le poulet (en haut), la caille japonaise (en bas à gauche) et le canard (en bas à droite). RT et CC correspondent respectivement aux groupes ayant reçus ou non un coup de chaleur. *Figure adaptée de Coustham et al., 2023 (Annexe 19).*

### 3.2.2 Alimentation

L'alimentation à un stade précoce peut également impacter le développement des individus (Tiffon, 2018). Seuls quelques exemples illustratifs seront présentés dans cette partie.

Chez les abeilles, nous avons vu précédemment qu'un même œuf peut donner soit une reine, si la larve est nourrie avec de la gelée royale, soit une ouvrière, si la larve est nourrie à la bouillie larvaire (**Chapitre 1.2**). La méthylation de l'ADN jouerait un rôle important dans la différenciation entre reine et ouvrière (Kucharski et al., 2008a). L'administration précoce de gelée royale modifie le profil de méthylation de certains gènes, ces derniers étant hypométhylés. Ce faible taux de méthylation, notamment le gène *dynactin p62*, serait dû à une inhibition de l'expression de DNMT3 favorisant ainsi le développement d'une reine (**Figure 3.6.A**). Une alimentation principalement composée de bouillie larvaire, quant à elle, n'induirait pas de répression de l'expression de DNMT3 qui pourra méthyler certains gènes, ces derniers étant inhibés, favorisant ainsi le développement d'une ouvrière (**Figure 3.6.B**). Plus de 550 gènes seraient différenciellement méthylés entre reine et ouvrière (Lyko et al., 2011). Récemment, une étude a mis en évidence que le gène *Map3k15* serait impliqué dans le dimorphisme phénotypique entre reine et ouvrière. Cette différenciation serait également en partie expliquée par la présence de marques d'histones (Jin et al., 2023). Les reines auraient plus de modifications H3K27ac que les ouvrières (**Figure 3.6.A**), et ces dernières auraient plus de modifications H3K4me1 que les reines (**Figure 3.6.B**).

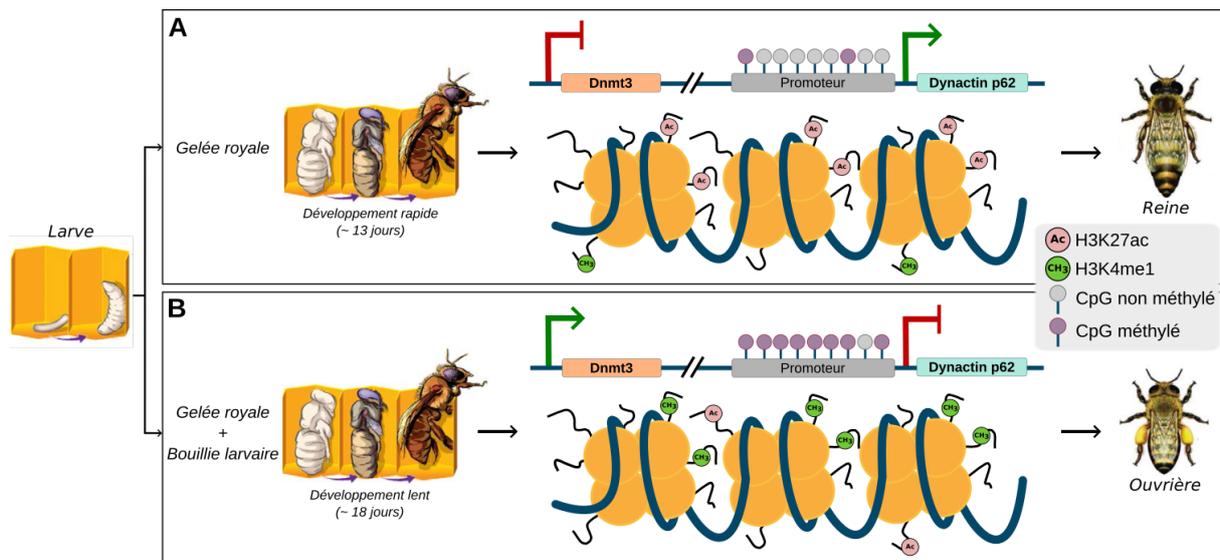


FIGURE 3.6 – Effet de l'alimentation sur le développement des abeilles (*Apis mellifera*, **Apidae**). (A) Une alimentation à base de gelée royale induit une inhibition de l'expression de DNMT3. Certains gènes, tels que *dynactin p62*, sont hypométhylés, favorisant leur expression et le développement d'une reine. (B) Une alimentation à base de bouillie larvaire induit la méthylation de certains gènes, inhibant leur expression et favorisant le développement d'une ouvrière. Figure adaptée de Shi et al. (2011), Jin et al. (2023) et Cridge et al. (2015).

Chez la souris, une étude a mis en évidence que l'exposition à un régime alimentaire obésogène (riche en matière grasse et en sucre) modifie la méthylation de l'ADN et l'expression de certains gènes dans l'hippocampe des souris mâles (Vander Velden & Osborne, 2022). Des changements significatifs dans l'expression de l'ARNm de gènes liés à la neurodégénérescence, codant notamment pour les protéines HDAC5 (Histone Deacetylase 5) et BIN1 (Bridging Integrator 1), ont été

observés. Les effets du régime sur la méthylation de l'hippocampe sont d'autant plus importants que la durée du régime est longue. Le vieillissement chez les souris témoins a modifié la méthylation de l'hippocampe, mais pas aussi fortement que chez les souris obèses. L'obésité interagit avec le vieillissement pour affecter de manière significative le méthylome de l'hippocampe, modifiant l'expression des gènes impliqués dans la neurodégénérescence et le métabolisme.

En productions animales, la supplémentation directe en nutriments dans l'œuf (alimentation *in ovo*) peut influencer le développement des oiseaux et améliorer leurs performances (robustesse, croissance, composition corporelle). Cette partie a également été détaillée dans la revue de production animale (Coustham et al., 2023, partie 2.1. Programmation nutritionnelle). Une partie des éléments concernant l'injection de nutriments *in ovo* présentés dans cette partie sont synthétisés dans la figure ci-dessous (Figure 3.7).

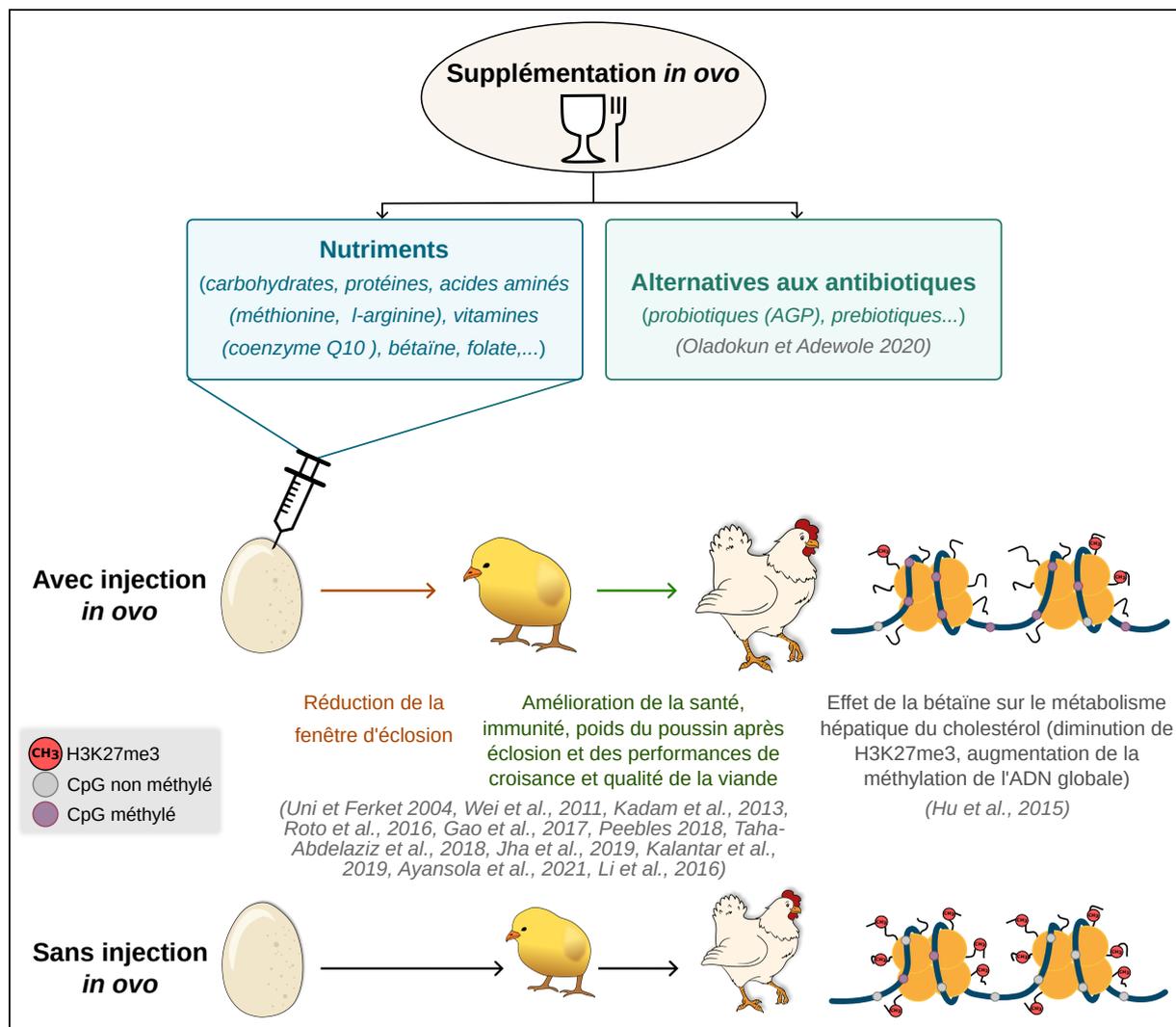


FIGURE 3.7 – Programmation nutritionnelle chez les volailles. Exemple d'injection *in ovo* de nutriments. Figure adaptée de Coustham et al., 2023 (Annexe 19).

### 3.2.3 Interactions sociales

La relation entre les parents et les descendants au début de la vie ainsi que les traumatismes vécus pendant l'enfance peuvent avoir des effets sur le comportement d'un individu, ces effets étant potentiellement en partie médiés par les marques épigénétiques.

Chez les rongeurs (rats), les soins apportés par les parents, en particulier la mère, sont très importants pour le bon développement de la progéniture. Des études ont mis en évidence que des rats n'ayant pas reçu d'affection de la part de leur mère, mesurée par la fréquence de léchage et de toilettage, développent à l'âge adulte une hypersensibilité au stress contrairement aux rats ayant reçu suffisamment d'affection (D. Liu et al., 1997; Weaver et al., 2004). En réponse à un stress, l'organisme produit du cortisol. Après liaison à son récepteur, ce dernier étant produit par le gène *nr3c1*, le cortisol va induire une réponse du système nerveux sympathique qui permet d'adapter la réaction de l'organisme au stress, contribuant ainsi à rétablir un état d'équilibre (**Figure 3.8.A**). Le dosage du cortisol libre (non lié au récepteur) dans l'hippocampe, a montré que, chez les individus en manque d'affection, cette hormone présente des niveaux anormalement élevés contrairement aux autres rats. Ce taux élevé de cortisol libre est expliqué par l'inactivation du gène *nr3c1* codant pour le récepteur au cortisol, en partie due à une méthylation plus élevée de son promoteur, mais le mécanisme par lequel le manque d'affection maternelle conduit à l'hyperméthylation de *nr3c1* est méconnu (Weaver et al., 2004). En l'absence de liaison au récepteur, un excès de cortisol libre peut perturber la transduction du signal hormonal. Cette perturbation peut entraîner une activation excessive du système nerveux sympathique, ce qui se traduit par une augmentation de la réponse au stress, provoquant ainsi une hypersensibilité au stress (**Figure 3.8.B**).

Le gène *nr3c1* a également été étudié chez l'Homme. Une étude menée sur trois groupes d'individus, un comportant des personnes décédées de mort naturelle et n'ayant pas subi d'abus durant leur enfance (contact sexuel, abus physique sévère, négligence sévère), un autre groupe composés de suicidés sans abus dans leur enfance, et un dernier groupe constitué de suicidés ayant subi des abus dans leur enfance, a mis en évidence que, chez ce dernier groupe, le gène *nr3c1* est également inactivé (McGowan et al., 2009).

Ces études mettent en évidence l'influence des soins parentaux sur la régulation épigénétique de l'expression du récepteur au cortisol. Ainsi, les soins maternels peuvent avoir des effets directs sur les réponses comportementales et neuroendocriniennes au stress chez leur progéniture, par l'intermédiaire de marques épigénétiques.

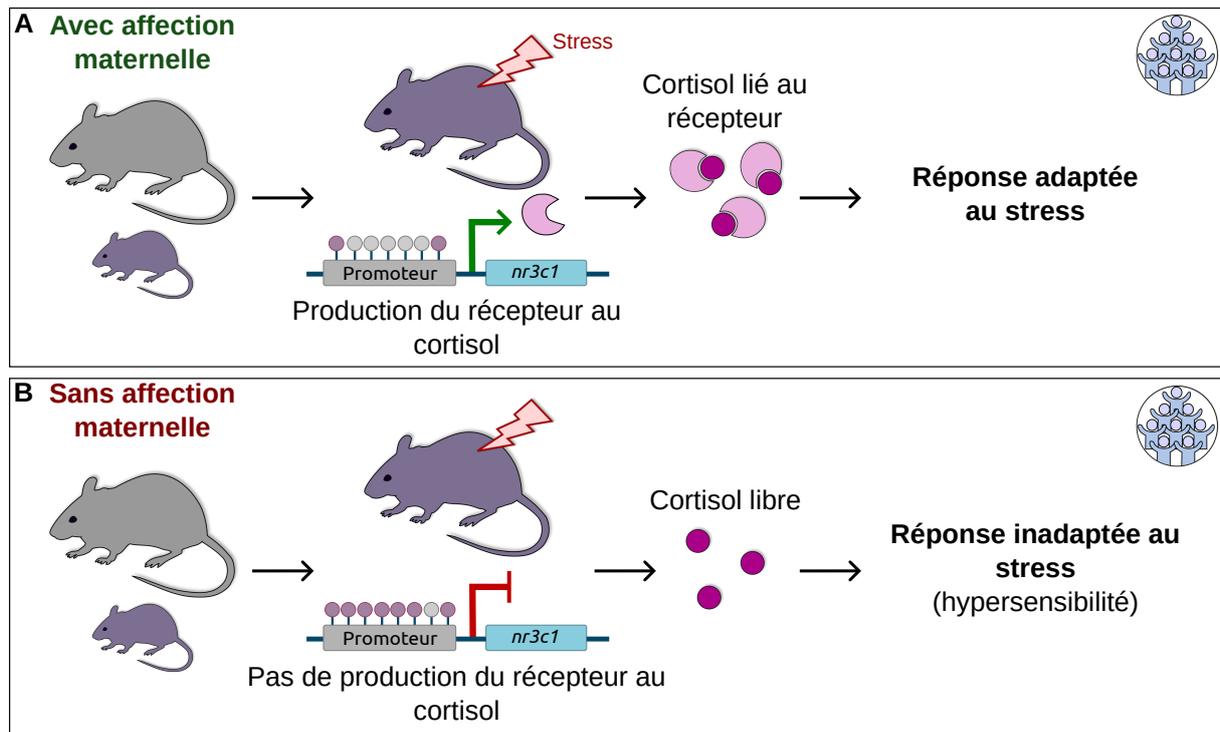


FIGURE 3.8 – Effet du comportement maternel sur la réponse au stress chez la descendance. (A) Chez des rats ayant reçu de l'affection (toiletage, léchage) de la part de leur mère durant l'enfance, à l'âge adulte, le gène *nr3c1*, n'étant pas méthylé, produit les récepteurs (en rose) au cortisol (en violet). Les complexes cortisol-promoteur influencent le système nerveux lymphatique permettant une réponse adaptée au stress. (B) Chez des rats n'ayant pas reçu d'affection de la part de leur mère durant l'enfance, à l'âge adulte, le gène *nr3c1*, étant méthylé, les récepteurs au cortisol sont produits en moins grande quantité, induisant une accumulation de cortisol libre et ainsi une réponse inadaptée au stress, voire excessive (hypersensibilité au stress).

### 3.2.4 Molécules biologiquement actives

L'exposition à des substances biologiquement actives pendant le développement embryonnaire peut affecter une grande variété de mécanismes impliqués dans la construction de l'individu. Parmi ces molécules actives, les perturbateurs endocriniens (PE) sont connus pour dérégler le système hormonal et leur nombre ne cesse de croître dans l'environnement (Santé-Publique-France, 2022). L'effet de ces molécules sur la santé est encore mal connu : elles sont suspectées d'être impliquées dans l'apparition de certaines maladies ou d'avoir une influence sur la croissance et la fertilité. En effet, certaines de ces molécules sont connues pour impacter le développement et la santé des individus exposés. Elles seraient impliquées dans des dysfonctionnements au niveau du système reproducteur, de la croissance des nourrissons et des enfants, du neurodéveloppement et des cancers (Yilmaz et al., 2020). Il existe plusieurs types de molécules pouvant affecter la construction de l'individu, mais cette partie se concentrera spécifiquement sur les PE, certains d'entre eux étant utilisés au cours de cette thèse.

### 3.2.4.1 Principaux perturbateurs endocriniens

Dans les années 90, les PE ont été pour la première fois nommés par Théo Colborn, un zoologiste et épidémiologiste (Colborn, 1992). Un PE désigne toute substance ou mélange extérieur à l'organisme pouvant être d'origine naturelle, la plupart étant des molécules xénobiotiques, c'est-à-dire étrangères à la vie (molécules de synthèse), interférant avec le système endocrinien. Les PE sont des agents environnementaux d'origine biologique ou artificielle dont la structure chimique est suffisamment similaire à celle des hormones pour qu'ils puissent se lier aux récepteurs des hormones et inhiber ou renforcer l'action des hormones endogènes. De très nombreux PE sont présents dans l'environnement (Casals-Casas & Desvergne, 2011 ; Albini et al., 2014). Certains de ces PE sont d'origine artificielle et peuvent être retrouvés dans l'alimentation. C'est notamment le cas des bisphénols, tels que le bisphénol A (BPA), un constituant du plastique (Qin et al., 2020). Les bisphénols sont principalement retrouvés dans les plastiques alimentaires tels que les revêtements de canettes et boîtes de conserve, certains contenants plastiques, ciments dentaires et papier thermique. Les pesticides sont également connus pour interférer avec le système endocrinien et regroupent les insecticides, herbicides et fongicides (ex : DDT - Dichlorodiphényltrichloroéthane). Les phtalates (ex : DEHP - phtalate de bis(2-éthylhexyle)), sont retrouvés dans les assouplissements des plastiques présents dans les emballages alimentaires, certains jouets, produits cosmétiques mais encore dans certains produits hospitaliers tels que les cathéters et poches à sang. Les parabènes (ex : butylparaben) sont présents dans certains produits cosmétiques et savons. D'autres PE sont retrouvés à l'état naturel, comme notamment les phytoœstrogènes et sont présents dans une grande variété de plantes. L'une des plus étudiées est la génistéine de soja, connue pour avoir un effet féminisant, cette hormone étant proche des œstrogènes. Les PE organiques sont présents dans de nombreuses plantes telles que le soja, les légumineuses et la luzerne.

### 3.2.4.2 Mode d'action des perturbateurs endocriniens

Les PE représentent toute molécule qui va perturber l'équilibre hormonal d'un individu et qui va affecter sa santé en interférant avec le système endocrinien. Les deux grands systèmes de communication dans l'organisme sont le système endocrinien et le système nerveux. Le système endocrinien regroupe plusieurs organes et glandes, notamment l'hypothalamus, l'hypophyse, la thyroïde, la parathyroïde, les surrénales, le pancréas, les ovaires et les testicules. Ces glandes et organes sécrètent des hormones qui passent dans le sang pour aller agir à distance sur un organe cible. L'hypothalamus (faisant partie du cerveau) est considéré comme le chef d'orchestre du système endocrinien. Il sécrète des hormones qui vont contrôler l'hypophyse qui elle-même contrôle les autres glandes endocrines. Le système endocrinien repose sur la sécrétion d'hormones dans le sang (signaux chimiques) et ce dernier a une action lente, continue, diffuse et généralisée contrairement au système endocrinien. Ces deux systèmes sont étroitement imbriqués, le tissu nerveux faisant partie des cibles des hormones sécrétées par le système endocrinien (Hiller-Sturmhöfel & Bartke, 1998). Ainsi, la plupart des PE sont donc aussi des perturbateurs nerveux, appelés perturbateurs endocriniens et nerveux (END - *Endocrine and Nervous Disruptor*) (Serlini & Jungers, 2021) (**Figure 3.9**).

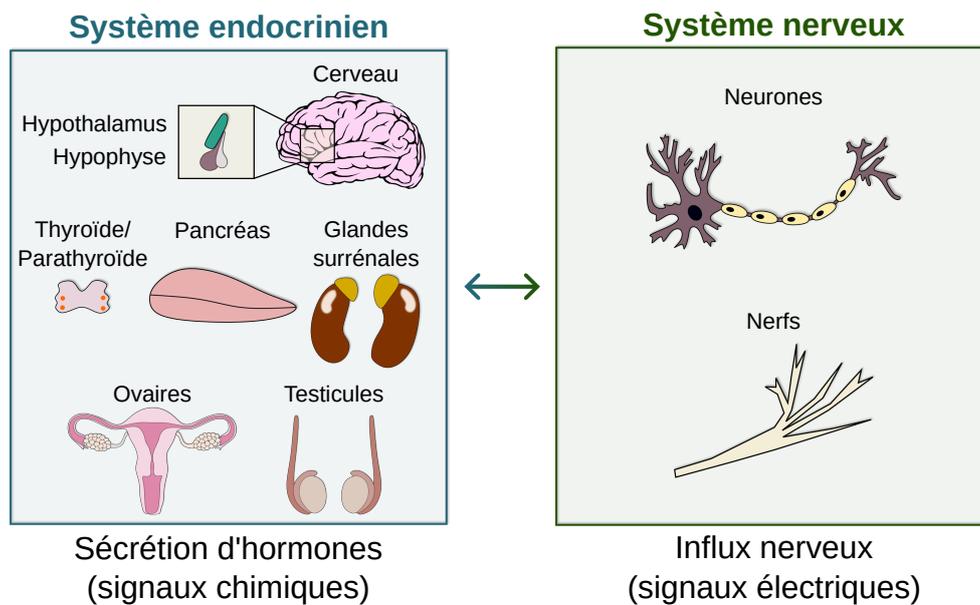


FIGURE 3.9 – Interaction entre le système endocrinien et le système nerveux. *Figure adaptée de Hiller-Sturmhöfel et Bartke (1998).*

En l'absence de PE, les hormones peuvent influencer la modulation de l'expression des gènes. Les hormones peuvent se fixer à des récepteurs membranaires ou intracellulaires induisant l'activation de gènes (Keller & Wahli, 1997). Il peut donc y avoir une modulation de l'expression de gènes en réponse à un signal extérieur hormonal. Certains PE peuvent agir par mimétisme, lorsqu'ils prennent la place de l'hormone naturelle en son absence, soit par inhibition, lorsqu'ils entrent en compétition avec l'hormone naturelle (Escher et al., 2014). Certains PE, ressemblant à l'hormone naturelle, vont remplacer l'hormone (effet mimétique) et activer le récepteur. Ces PE ont une action agoniste sur le récepteur, déclenchant la même réponse que l'hormone et peuvent induire une suractivation du récepteur et une augmentation des effets de l'hormone endogène. D'autres PE peuvent agir par inhibition, c'est-à-dire que si l'hormone et le PE sont tous les deux présents, ces derniers entrent en compétition, le PE l'emporte sur l'hormone naturelle et se fixe au récepteur à sa place sans l'activer. Le blocage du récepteur entraîne une inhibition des effets de l'hormone empêchant la signalisation de se faire normalement induisant une inhibition de l'expression des gènes. Ces PE ont une action antagoniste sur le récepteur. L'exposition à des PE peut avoir des répercussions sur la santé des individus et ils n'ont pas le même impact selon le moment de la vie où ils agissent. La période de la vie où un individu est le plus vulnérable en terme d'exposition correspond à la période embryonnaire et fœtale. Les PE sont le plus souvent des molécules très stables (bioaccumulables), ont une longue durée de vie et sont le plus souvent lipophiles (solubles dans les graisses) (Geyer et al., 2000 ; Elobeid & Allison, 2008).

Afin d'étudier l'effet de l'environnement sur le méthylome des cailles, ces dernières ont été exposées à des PE tels que le BPA et la génistéine. Une autre molécule, la 5-aza-2'deoxyctidine (aussi appelée decitabine ou 5Aza), qui n'est pas un PE mais un inhibiteur des DNMT (Christman, 2002) est souvent utilisée dans les études visant à démontrer l'effet d'un PE sur la méthylation de l'ADN (Majid et al., 2010 ; Weng et al., 2010). Contrairement aux PE, cette molécule n'est pas présente dans l'environnement. Elle est également utilisée comme traitement contre la leucémie (Rivard et al., 1981). Le BPA et la génistéine étant étudiées au cours de cette thèse, seul leur mode

d'action sera détaillé dans cette partie. Le BPA et la génistéine font partie des PE les plus étudiés et sont connus pour avoir un effet sur l'épigénome, notamment la méthylation de l'ADN. Le BPA est un xœstrogène synthétisé la première fois en 1891 et dont les propriétés œstrogéniques ont été découvertes en 1930 (Wolstenholme et al., 2011) (**Figure 3.10.A**). Depuis les années 1950, le BPA est utilisé comme constituant principal du plastique polycarbonate (Onn Wong et al., 2005) et est également retrouvé dans l'eau et certains aliments (Vom Saal et al., 2007). La génistéine est un phytoœstrogène naturel, une substance chimique œstrogénique, possédant une structure similaire aux œstrogènes. La génistéine fait partie des phytoœstrogènes les plus étudiés et son activité œstrogénique a été mise en évidence chez de nombreuses espèces. Cette molécule est présente dans l'alimentation, en particulier dans les produits à base de soja (Viggiani et al., 2019) (**Figure 3.10.B**). Le BPA et la génistéine peuvent être des agonistes ou des antagonistes des œstrogènes (hormones sexuelles féminines), tels que l'œstradiol (**Figure 3.10.C**), une hormone responsable du développement des organes sexuels et des caractères sexuels secondaires chez les femelles (Kurosawa et al., 2002 ; Setchell, 2001 ; Miodini et al., 1999). Ces deux molécules peuvent également être agonistes ou antagonistes des hormones androgènes, telles que la testostérone (**Figure 3.10.D**), une hormone essentielle à la fonction reproductive masculine qui intervient notamment dans la spermatogenèse (Bonfeld-Jorgensen et al., 2007 ; Pihlajamaa et al., 2011). Ainsi, en agissant comme agonistes et/ou antagonistes des récepteurs stéroïdiens, les PE peuvent interférer avec la régulation hormonale normale et le bon fonctionnement des systèmes endocrinien et neuro-endocrinien (Anway & Skinner, 2006 ; Gore, 2008).

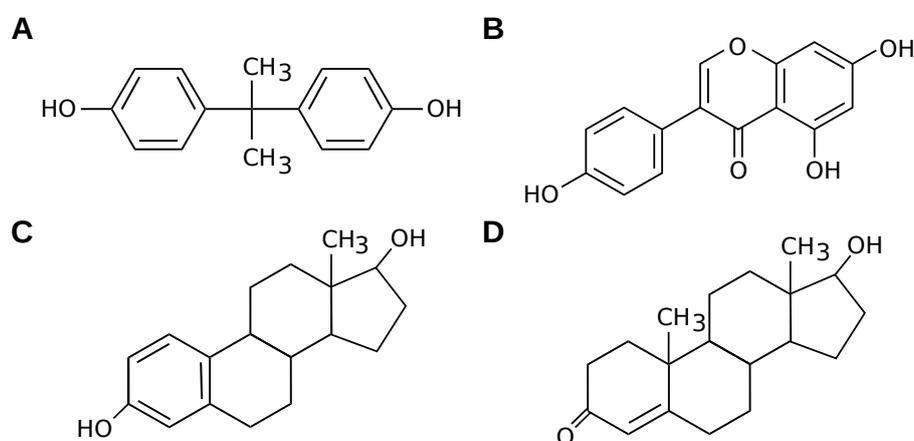


FIGURE 3.10 – **Structure des molécules.** Les structures chimiques (A) du bisphénol A, (B) de la génistéine, (C) de l'œstradiol et (D) de la testostérone sont présentées. La similarité structurale entre les quatre molécules permet au BPA et à la génistéine d'interagir avec les récepteurs des œstrogènes ( $\alpha$  et  $\beta$ ) et d'autres récepteurs semblables aux œstrogènes. *Figures issues de Wikipédia : BPA (Calvero), génistéine (Emeldir), œstradiol (NEUROtiker) et testostérone (NEUROtiker).*

Le BPA et la génistéine sont donc capables de se lier aux récepteurs nucléaires des œstrogènes (RE, tels que les RE  $\alpha$  et  $\beta$  (Kuiper et al., 1998). Le BPA est également un inhibiteur compétitif et antagoniste du récepteur des androgènes (RA), tels que la testostérone (Sohoni & Sumpster, 1998 ; Bonfeld-Jorgensen et al., 2007 ; Pihlajamaa et al., 2011), le RA étant le principal élément régulateur de la signalisation cellulaire des androgènes (**Figure 3.11**).

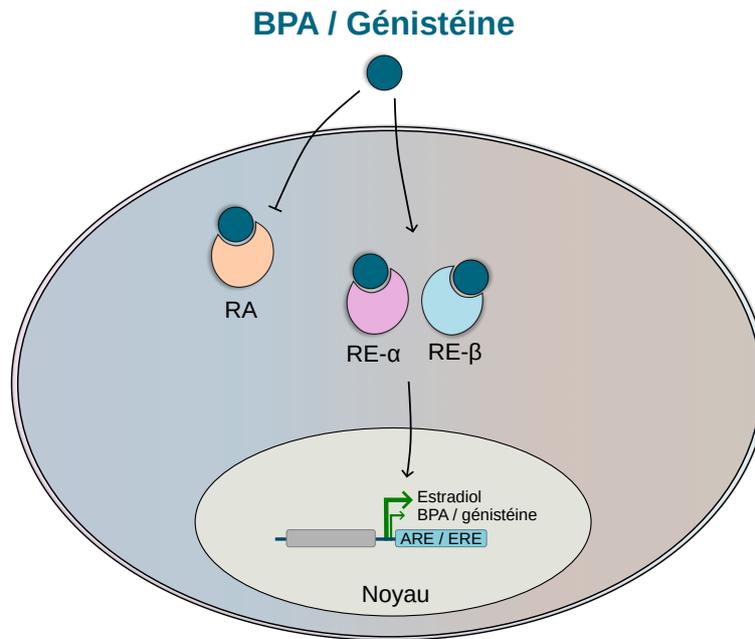


FIGURE 3.11 – **Un modèle des modes d'action potentiels du BPA et de la génistéine.** Le BPA et la génistéine peuvent activer plusieurs récepteurs d'œstrogènes (RE) et récepteurs liés aux œstrogènes tels que RE- $\alpha$  et RE- $\beta$  et/ou sont des antagonistes au récepteur des androgènes (AR). En activant ces récepteurs, ces PE peuvent modifier l'expression des gènes contenant des éléments de réponse aux œstrogènes (ERE) ou aux androgènes (ARE) par rapport à celle du ligand endogène. L'activation des gènes peut ainsi entraîner des modifications épigénétiques en aval, telles que la méthylation de l'ADN ou des modifications des histones, pouvant impacter la transcription.

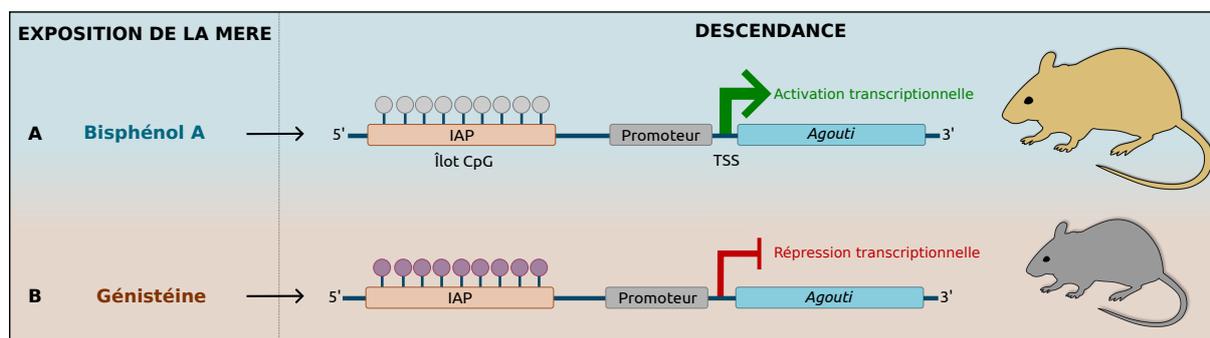
L'exposition au BPA pendant la période précoce du développement peut impacter le développement des individus. Une exposition à de faibles niveaux de BPA peut affecter la différenciation et la prolifération cellulaires, la croissance corporelle, la maturation sexuelle, la fonction et le développement reproductifs ainsi que le comportement chez les invertébrés et les vertébrés (Maffini et al., 2006 ; Vom Saal & Welshons, 2006). Le BPA induirait une baisse de la qualité du sperme chez les mâles et une maturation sexuelle précoce chez les femelles (Howdeshell et al., 1999) ainsi qu'une modification du développement des glandes mammaires chez les pubères (Muñoz-de Toro et al., 2005). Suite à l'exposition au BPA, une augmentation des cas de cancers à médiation hormonale, de cancers de la prostate et du sein, des troubles métaboliques (diabète résistant à l'insuline, obésité) mais encore des problèmes neuro-comportementaux (ex : autisme) ont été mis en évidence (Farabollini et al., 1999 ; Howdeshell et al., 1999 ; Nakamura et al., 2006 ; Timms et al., 2005 ; Vom Saal et al., 2007 ; Markey et al., 2005 ; Patisaul et al., 2006). Chez le rat, une exposition au cours du développement à de faibles doses de BPA induit une prédisposition aux lésions précancéreuses et à la carcinogenèse de la prostate (S.-M. Ho et al., 2006 ; Prins et al., 2008). Une étude a également mis en évidence l'effet du BPA sur le poids corporel des caillies japonaises (Halldin et al., 2001).

L'exposition à la génistéine au cours du développement a des effets néfastes sur la santé mais aurait aussi de potentiels effets bénéfiques (Silva et al., 2019), en fonction de la dose et du moment de l'exposition. La génistéine est connue pour impacter sur le long terme l'appareil reproducteur chez de nombreuses espèces, y compris chez l'Homme, les rongeurs, les poissons et les oiseaux (Wocławek-Potocka et al., 2013 ; Colborn et al., 1993). Chez l'Homme, le fœtus

peut être exposé à ces molécules durant le développement *in utero*, ainsi que pendant l'enfance et l'allaitement (Franke et al., 1998). Le nombre de cas d'hypospadias (malformation de la verge) chez les garçons nés de femmes ayant eu un régime végétarien pendant la grossesse est plus important (North & Golding, 2000). Une étude sur de jeunes adultes ayant consommé du soja étant nourrissons a révélé une augmentation de l'utilisation de médicaments contre les allergies, ainsi que des saignements menstruels plus longs au cours du cycle menstruel chez les femmes (Goldman et al., 2001 ; Strom et al., 2001), alors qu'une autre étude a révélé une amélioration des taux de synthèse du cholestérol (Cruz et al., 1994). Chez la souris, l'exposition néonatale à la génistéine favorise la formation d'un adénocarcinome utérin d'apparition plus tardive au cours de la vie du sujet (Newbold et al., 2001). Chez le médaka, une exposition à la génistéine pendant 100 jours a impacté les gonades, provoquant une atrésie des ovocytes préovulatoires et l'altération des caractéristiques sexuelles secondaires chez les femelles, ainsi que le développement de gonades intersexuées chez les mâles (Kiparissis et al., 2003). D'autres exemples, chez des guépards en captivité, nourris avec des aliments à base de soja, et des moutons, consommant du trèfle rouge, présentaient respectivement une diminution de la fertilité (Setchell et al., 1987) et une infertilité (Hearnshaw et al., 1972). Une alimentation composée de protéines de poulet et dépourvue de soja, a rétabli la fertilité des guépards. Chez la caille japonaise, des chercheurs se sont intéressés à l'impact de l'exposition embryonnaire aux œstrogènes chez les oiseaux adultes, et ces derniers altèrent le système neural et reproducteur des embryons chez l'oiseau (Halldin, 2005). La génistéine aurait également des effets protecteurs contre des maladies métaboliques ou des cancers (Silva et al., 2019). L'exposition prépubertaire à la génistéine préviendrait l'apparition de cancer de la glande mammaire chez les rats (Hilakivi-Clarke, Onojafe, et al., 1999 ; Lamartiniere et al., 1998), tandis qu'une exposition durant la vie prénatale induirait une augmentation des cas de cancer (Hilakivi-Clarke, Cho, et al., 1999). Chez le poulet, l'injection génistéine dans l'œuf a induit un gain de poids corporel significatif, impactant ainsi leur croissance (Lv, Fan, Zhang, Xing, & Guo, 2018). Des études antérieures avaient mis en évidence une augmentation du gain de poids corporel chez les poulets de chair de 21 jours après une faible dose de supplémentation en génistéine dans l'alimentation de leur mère, ce qui entraîne la présence de génistéine dans l'œuf (Lv, Fan, Zhang, Xing, & Guo, 2018), mais une diminution du poids de l'embryon après exposition à forte dose de génistéine par supplémentation alimentaire de la mère (Lv, Fan, Zhang, Ning, et al., 2018). Ainsi, les phytoœstrogènes sont présents dans notre environnement à des concentrations suffisamment élevées pour influencer sur le long terme la santé des individus.

Le BPA et la génistéine peuvent également modifier la méthylation de l'ADN. Chez l'exemple très connu de la souris Agouti jaune viable (A<sup>vy</sup>), le gène *Agouti* est modifié par l'insertion d'un rétro-transposon de particule intracisternales de type A (IAP) dans la région 5' du gène *Agouti*, constitué d'un promoteur sensible à la méthylation de l'ADN. L'expression de ce promoteur conduit à l'expression ectopique d'Agouti, conduisant au développement d'une fourrure jaune, de l'obésité, du diabète et à la formation de tumeurs chez les souris (Morgan et al., 1999). Le traitement au BPA (*in utero* ou ingéré par la mère) induit une diminution de la méthylation de l'ADN au niveau de l'élément IAP conduisant à l'activation du promoteur et la production de la protéine ASIP, favorisant une couleur de pelage jaune chez la descendance (Dolinoy et al.,

2006) (**Figure 3.12.A**). Une supplémentation alimentaire maternelle composée d'acide folique (donneur de méthyle) ou de génistéine induit également une modulation du phénotype *Agouti* (Dolinoy et al., 2006, 2007). La disponibilité des donneurs de méthyle contrôle la production de S-adenosylméthionine (SAM) qui agit comme le principal donneur de méthyle pour la méthylation de l'ADN. L'augmentation de la disponibilité de la méthionine dans le cerveau augmente la SAM et la méthylation de l'ADN (Tremolizzo et al., 2002 ; Guidotti et al., 2007). Les traitements à l'acide folique et à la génistéine augmentent la méthylation de l'ADN au niveau de l'élément IAP favorisant le phénotype *Agouti* (pelage gris) (**Figure 3.12.B**). La diminution du taux de méthylation induite par le BPA pourrait ainsi être contrebalancée par la génistéine. L'expression plus ou moins élevée du gène *Agouti*, induite par des méthylation différentielles des sites CpG spécifiques, induit une couleur du pelage variée, allant du jaune au gris (**Figure 2.8**).



**FIGURE 3.12 – Effet des perturbateurs endocriniens sur la méthylation de l'ADN chez la souris *Avy*.** Régulation du gène *Agouti* (*Avy*, *Agouti viable yellow*) chez la souris par la méthylation de l'ADN. L'allèle *Avy* résulte de l'insertion d'un rétrotransposon IAP (particule intracisternale de type A) en amont (100 kb) du site d'initiation de la transcription (TSS) du gène *Agouti*. Une mère exposée au (A) bisphénol A, dans alimentation ou *in utero*, engendre une hypométhylation des sites CpG du gène *Agouti*, favorisant la production de la protéine ASIP, un pelage jaune et l'obésité chez la descendance. Une mère exposée à la (B) génistéine, engendre une hyperméthylation des sites CpG du gène *Agouti*, inhibant la production de la protéine ASIP et l'obésité et favorisant un pelage gris chez la descendance.

D'autres effets de ces molécules sur la méthylation de l'ADN ont été rapportés. Chez les rats, l'exposition néonatale au BPA a entraîné une hyperméthylation de l'ADN dans les testicules, suggérant que les effets néfastes du BPA sur la spermatogenèse et la fertilité sont en partie médiés par la méthylation de l'ADN (Doshi et al., 2011). Une autre étude chez l'Homme a montré que l'exposition maternelle au BPA est associée à une hyperméthylation de gènes associés au développement neuro-comportemental (*CAPS2*), immunitaire (*TNFRSF25*) et physique (*HKR1*) (Song et al., 2020). La génistéine peut aussi inhiber la prolifération des cellules cancéreuses du rein via l'hypométhylation du gène suppresseur de tumeur *CDKN2a*, favorisant son expression (Ji et al., 2020).

Les exemples cités précédemment montrent l'influence de l'environnement maternel sur la descendance. Ainsi, les PE peuvent avoir des répercussions sur la santé des individus directement exposés mais aussi au niveau de la descendance via l'environnement maternel. Les effets de l'environnement maternel peuvent également être transmis à la descendance via les cellules germinales et sur plusieurs générations, notamment par l'intermédiaire de marques épigénétiques.

## 4. Transmission des effets environnementaux

Cette partie se concentre sur la transmission non génétique des effets de l'environnement et son effet potentiel sur les phénotypes. Depuis quelques années, des scientifiques ont émis l'hypothèse d'une transmission non génétique des effets de l'environnement à travers les générations (Miska & Ferguson-Smith, 2016a ; David et al., 2019a). La transmission des marques épigénétiques entre les générations pourrait influencer les phénotypes des adultes (Triantaphyllopoulos et al., 2016 ; Frésard et al., 2013). Les effets de l'environnement peuvent être transmis via l'environnement maternel, notamment lors du développement précoce (embryonnaire et foetal), et via les cellules germinales (ovocytes et spermatozoïdes), cette transmission pouvant être intergénérationnelle ou transgénérationnelle (**Figure 4.1,4.2**). Dans le cas où un mâle ou une femelle non gestante (G0) sont impactés par l'environnement, les ovocytes ou spermatozoïdes (G1) de ces derniers sont également considérés comme directement impactés. Dans ce cas, la transmission est dite intergénérationnelle entre la G0 et la G1 et transgénérationnelle à partir de la G2 et au-delà (**Figure 4.1.A**). Chez les mammifères, un changement dans l'environnement d'une mère gestante (G0) peut avoir des répercussions directes sur le développement de l'embryon (G1) ainsi qu'au niveau des cellules germinales (G2) portées par ce dernier (**Figure 4.1.B**). La transmission est dite intergénérationnelle lorsqu'elle se fait entre les générations ayant été directement exposées à un stress, c'est-à-dire lorsque des effets subis en G0 sont observables au niveau des générations G1 et G2. La transmission des effets de l'environnement entre la G0 et la G2 est également qualifiée de multigénérationnelle (David et al., 2019b). Ce n'est qu'à partir de la troisième génération (G3) que la transmission est dite transgénérationnelle, c'est-à-dire lorsque les effets subis en G0 sont observables chez les individus n'ayant pas été directement exposés au stress (G3 et au-delà).

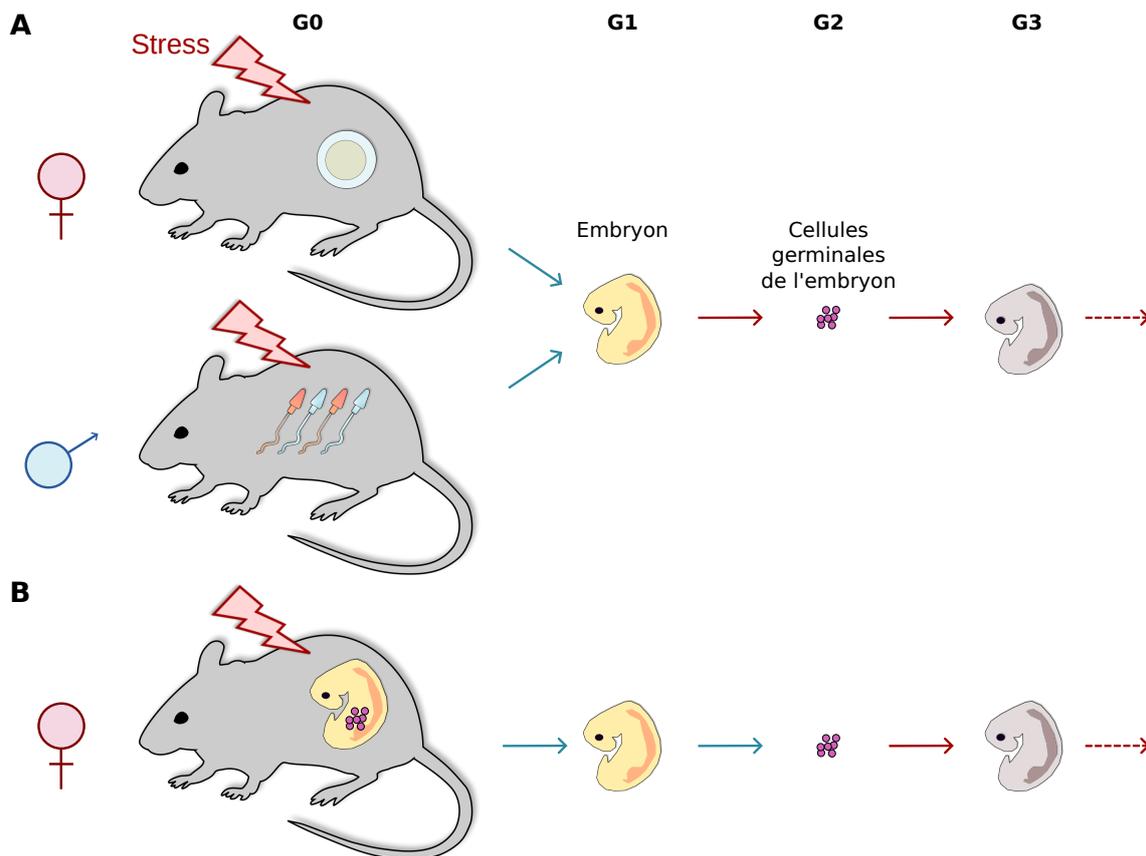


FIGURE 4.1 – **Transmission des effets de l'environnement chez les mammifères.** (A) Transmission des effets de l'environnement via les cellules germinales : un stress impactant une mère ou père porteurs de cellules germinales (ovocytes ou spermatozoïdes) (G1) peut être transmis de manière intergénérationnelle à le G1 et transgénérationnelle à la G2 et au-delà. (B) Transmission des effets via l'environnement maternel : un stress impactant une mère gestante (G0) peut être transmis de manière intergénérationnelle à l'embryon (G1) et aux cellules germinales (G2) portées par ce dernier ainsi que de manière transgénérationnelle aux autres générations (G3 et au-delà). Les flèches bleues et rouges représentent respectivement les transmissions intergénérationnelle et transgénérationnelle.

Chez les oiseaux, la transmission est également qualifiée de transgénérationnelle à partir de la G2 (**Figure 4.2**). La modification de l'environnement d'un individu en plein développement peut, comme nous l'avons vu précédemment, avoir des répercussions au niveau des marques épigénétiques, dont la méthylation de l'ADN, et engendrer une modification phénotypique induite par un changement de l'expression des gènes. Ces modifications des marques épigénétiques peuvent être transmises au cours du développement aux cellules germinales pouvant ainsi affecter les générations suivantes, notamment celles n'ayant pas été directement exposées aux effets de l'environnement. Ce phénomène est désigné sous le terme d'épigénétique transgénérationnelle (Heard & Martienssen, 2014 ; Szyf, 2015 ; Miska & Ferguson-Smith, 2016b).

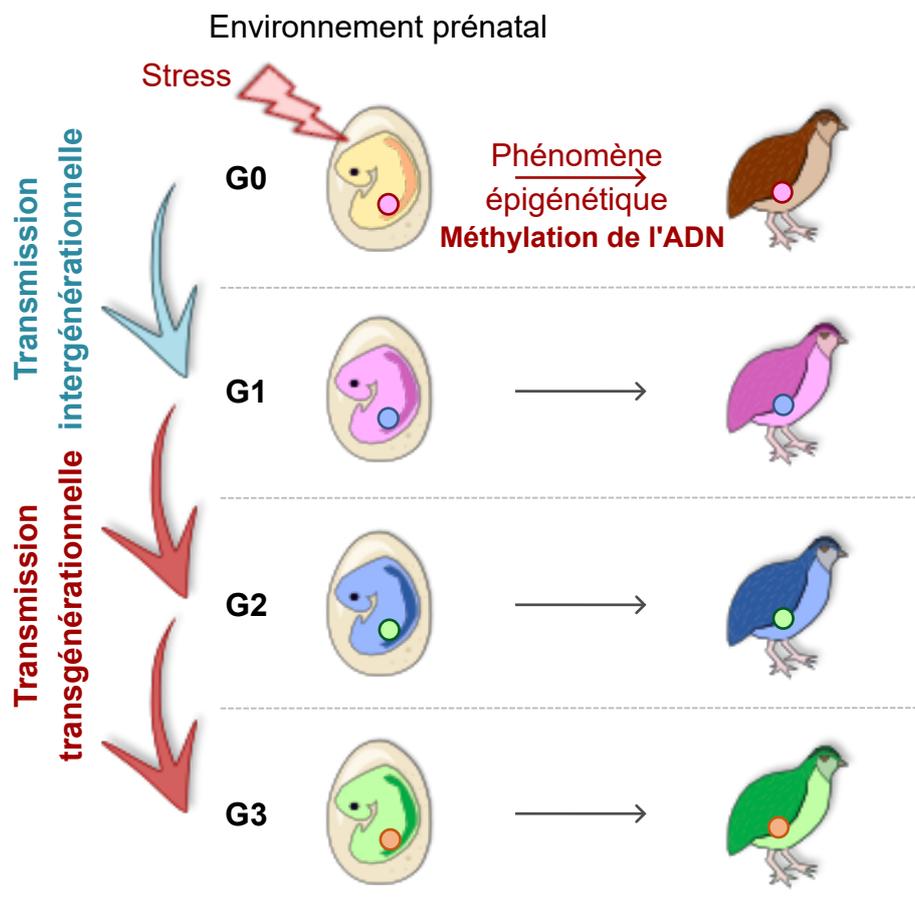


FIGURE 4.2 – **Transmission des effets de l'environnement.** La modification de l'environnement d'un embryon en plein développement (G0) peut impacter les marques épigénétiques, notamment la méthylation de l'ADN. Ces modifications peuvent être transmises aux cellules germinales (G1) au cours du développement et être transmises aux générations suivantes (G2 et G3). Les flèches bleues et rouges représentent respectivement les transmissions intergénérationnelle et transgénérationnelle.

#### 4.1 Transmission intergénérationnelle

La transmission intergénérationnelle désigne une transmission des effets de l'environnement parental (G0) à la descendance sur une (G1) ou deux (G2) générations.

Le régime alimentaire maternel peut impacter le développement de la descendance. Comme nous l'avons vu précédemment, l'exemple très connu de la souris Agouti jaune viable (Avy) est très illustratif de l'effet de la supplémentation en BPA ou génistéine dans l'alimentation de la mère gestante sur le développement de la descendance (Dolinoy et al., 2006, 2007 ; Tremolizzo et al., 2002 ; Guidotti et al., 2007) (**Figure 2.8**). De nombreuses études ont pu mettre en évidence un effet sur le long terme de l'environnement embryonnaire par le biais de l'alimentation maternelle, notamment chez les volailles. Chez la poule, une supplémentation de bêtaïne, un agent antioxydant et donneur de méthyle, dans l'alimentation maternelle a eu des répercussions sur la descendance. Les poussins femelles présentaient un poids corporel et un niveau d'hormones thyroïdiennes biologiquement actives significativement plus faible par rapport au groupe contrôle. Une diminution significative de l'expression du gène *Dio1* (iodothyronine déiodinase de type 1) ainsi qu'une hyperméthylation de son promoteur ont été mises en évidence. Chez les poussins mâles, seule l'expression de *Dio1* a été impactée. Ces résultats indiquent que l'administration

maternelle de bétaine affecte la croissance de la progéniture par le biais d'une modification de la méthylation de l'ADN et de l'expression du gène *Dio1* dans le foie (Hou et al., 2018; Hu et al., 2020) (**Figure 4.3.A**).

L'exposition d'une mère gestante à un stress intense peut également avoir des répercussions sur le développement de la descendance. Chez l'Homme, quelques exemples bien connus ont pu souligner la transmission des effets de l'environnement maternel à la descendance lors de périodes de grandes privations. Une étude en Suède menée sur une large cohorte de mères ayant subi un stress (deuil) autour de la grossesse (Class et al., 2014) mais encore les conséquences du ressenti de mères ayant subi la tempête de verglas de 1998 au Québec (Cao-Lei et al., 2015), ont pu mettre en évidence une transmission intergénérationnelle des phénotypes par l'intermédiaire des marques épigénétiques. Une autre population très étudiée est celle des descendants de mères ayant subi la grande famine en période de grossesse aux Pays-Bas pendant la seconde guerre mondiale (*Dutch Hunger Winter*) (Heijmans et al., 2008). Ces dernières ont donné naissance à des enfants plus susceptibles de développer des affections cardiaques et métaboliques par rapport à ceux nés l'année suivante. La région différentiellement méthylée (DMR) responsable du phénomène d'empreinte génomique du gène *IGF2* impliqué dans la croissance était hypométhylée chez les enfants dont les mères enceintes avaient été exposées par rapport aux autres enfants (Heijmans et al., 2008).

Une étude portant sur l'effet du stress maternel chez la caille (Charrier et al., 2022), tel que la diffusion de bruits soudains ou d'appels de prédateurs, des pulvérisations d'eau et d'air, la contention, des restrictions alimentaires, l'instabilité sociale ou la déambulation sur un chariot, a mis en évidence que le stress prénatal maternel (SPM, lors de la formation de l'œuf) augmente la réactivité émotionnelle de la descendance (G1). Le SPM induit une augmentation de H3K27me3 dans des régions du cerveau connues pour être impliquées dans le contrôle des réponses émotionnelles chez les vertébrés. L'effet du SPM sur la marque H3K27me3 pourrait être impliqué dans les effets comportementaux observés chez les cailles G1 (**Figure 4.3.B**).

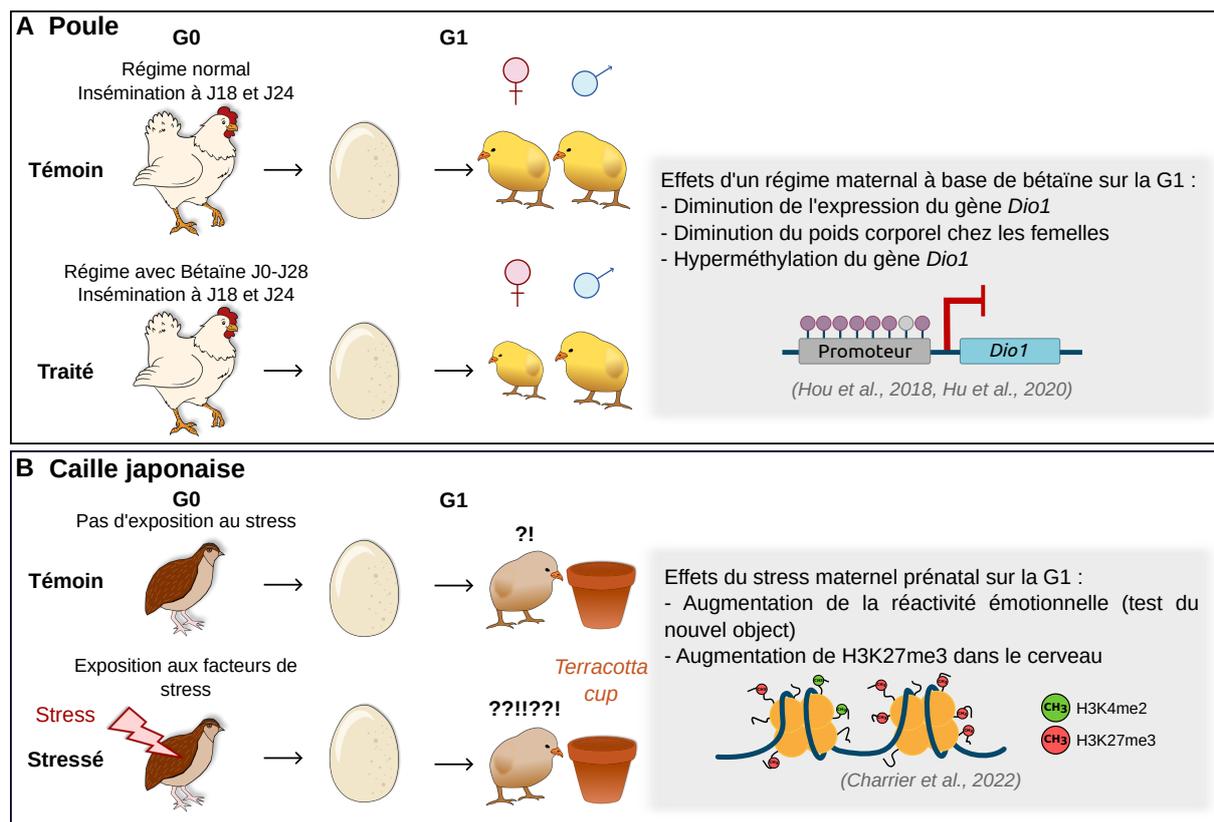


FIGURE 4.3 – **Transmission des effets de l'environnement maternel chez les volailles.** (A) Représentation des effets d'une supplémentation de bétaïne dans l'alimentation maternelle chez la poule (G0) sur la descendance (G1) (Figure adaptée de Hou et al.; Hu et al. (2018; 2020)). (B) Représentation des effets du stress maternel chez la caille japonaise (G0) sur la descendance (G1) (Figure adaptée de Charrier et al. (2022)).

Certaines études ont montré que les marques épigénétiques peuvent être transmises d'une génération à l'autre via les cellules germinales (Probst et al., 2009; Meaney & Szyf, 2022).

Chez l'Homme, une étude menée sur des victimes de l'Holocauste ainsi que sur leur descendance a mis en évidence une transmission intergénérationnelle de l'effet du traumatisme, subi avant la conception, à la descendance, se traduisant par une hyperméthylation de l'intron 7 du gène *FKBP5*, agissant comme régulateur du complexe du récepteur des glucocorticoïdes (ex : cortisol, hormone du stress), chez les parents exposés et une hypométhylation chez les descendants (Yehuda et al., 2016). Les effets sur la descendance sont donc liés à l'exposition maternelle à l'Holocauste pendant l'enfance.

Dans une autre étude, des souris mâles (G0) ont été soumises à un conditionnement à la peur d'une substance odorante (acétophénone couplée à un choc électrique) avant conception (Dias & Ressler, 2014). Les souris conditionnées à la peur ont développé une sensibilité comportementale ainsi qu'une modification de la méthylation de l'ADN, au niveau du cerveau et du sperme, se traduisant par une hypométhylation au niveau de plusieurs CpG du gène *Olfr151*, codant pour le récepteur M71, ce dernier étant impliqué dans la détection de l'acétophénone. De plus, la génération suivante (G1), issue de mâles G0 conditionnés et de femelles naïves (non conditionnées), présentait une sensibilité comportementale accrue à l'odeur de conditionnement des G0, mais pas à d'autres odeurs. Une hypométhylation de CpG dans le gène *Olfr151* a été identifiée dans les spermatozoïdes des mâles G0 conditionnés ainsi que des descendants G1 naïfs (Figure 4.4).

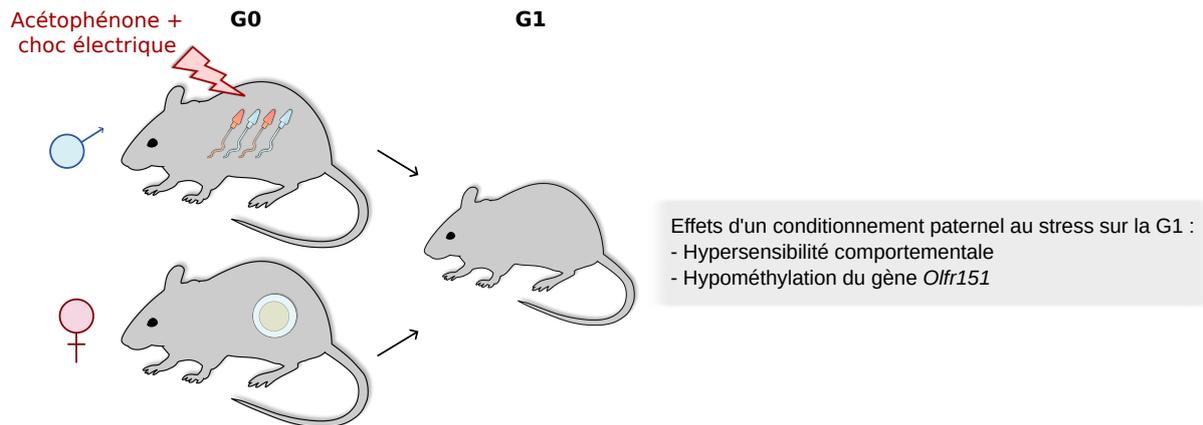


FIGURE 4.4 – **Transmission intergénérationnelle des effets de l’environnement paternel chez la souris.** Des souris mâles (G0) ont subi un stress odorant d’acétophénone couplé à un choc électrique. Après accouplement de ces mâles conditionnés avec des femelles naïves (non conditionnées), la descendance (G1) a développé une sensibilité comportementale à la même odeur ainsi qu’une hypométhylation du gène *Olf151*. *Figure adaptée de Dias et Ressler (2014).*

Les effets de l’environnement maternel peuvent aussi être transmis au-delà de la G1. Une étude, chez la souris *Agouti*, a pu mettre en évidence que la supplémentation dans l’alimentation de la mère (G0) en donneurs de méthyle (folate, choline, betaine et vitamine B12) engendrait une diminution du nombre de sujets obèses deux générations après (G2), cette observation étant potentiellement due à une hyperméthylation du gène *Agouti* (Cropley et al., 2006) (**Figure 4.5**). Toutes ces études montrent l’influence de l’environnement parental, qu’il soit maternel ou paternel, sur la descendance. Ces effets peuvent également être transmis sur plusieurs générations et concerner des individus n’ayant pas été directement impactés, cette transmission étant qualifiée de transgénérationnelle.

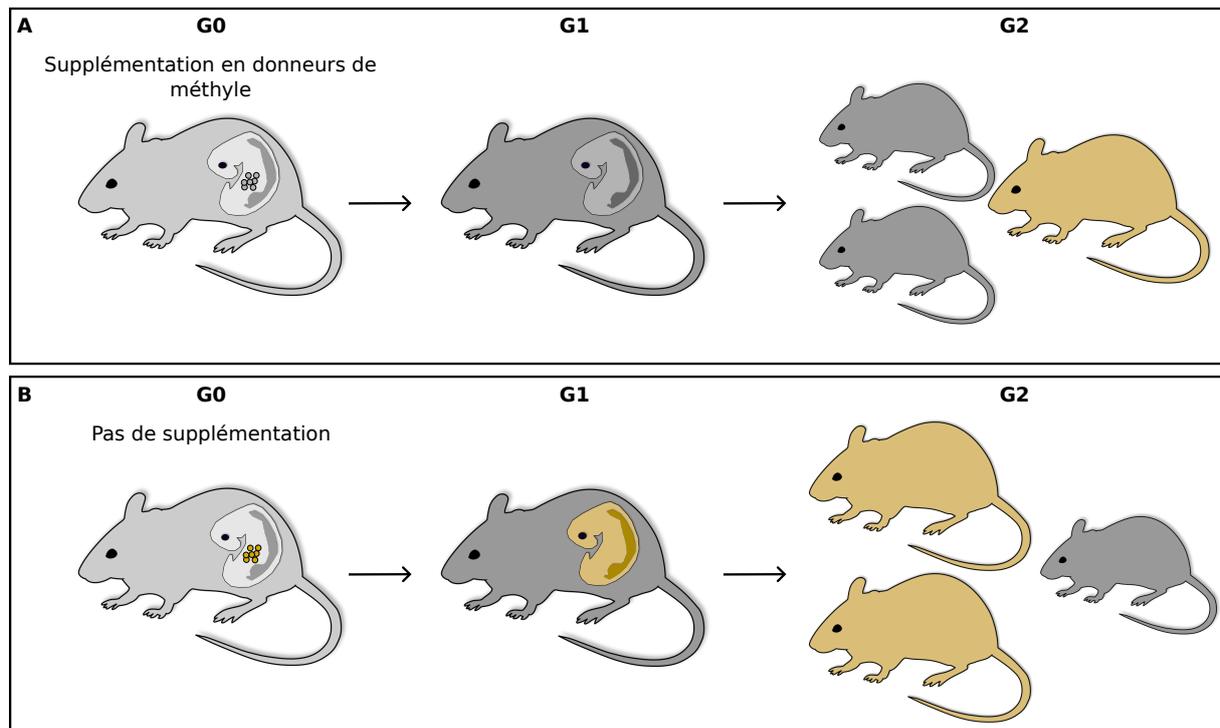


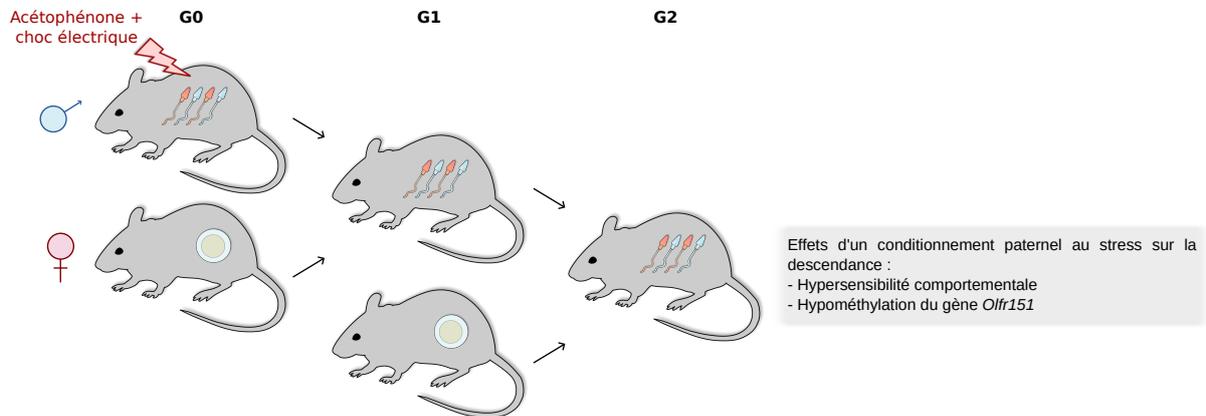
FIGURE 4.5 – **Transmission intergénérationnelle des effets de l'environnement maternel chez la souris.** Une mère gestante (G0), porteuse d'un embryon (G1) lui-même porteur de cellules germinales (G2), ayant reçu un régime alimentaire supplémenté en donneurs de méthyle (A) aura moins de descendants obèses en G2 par rapport à une mère ayant reçu un régime alimentaire normal (B). *Figure adaptée de Cropley et al. (2006).*

## 4.2 Transmission transgénérationnelle

Les effets de l'environnement peuvent également être transmis à travers les générations et au-delà des générations ayant été directement exposées. Dans le cas d'une hérédité transgénérationnelle des marques épigénétiques, la modification épigénétique peut être transmise de manière stable à la descendance, via les gamètes, et sur plusieurs générations. Les informations épigénétiques transgénérationnelles semblent être transmises soit via des mécanismes directs, où une marque épigénétique est maintenue au cours des générations, soit via des mécanismes dits coopératifs, lorsque les informations sont transmises par l'intermédiaire de plusieurs marques épigénétiques agissant en coopération (Fitz-James & Cavalli, 2022 ; Miska & Ferguson-Smith, 2016b). Cette hérédité transgénérationnelle a été mise en évidence dans de nombreuses espèces (Perez & Lehner, 2019 ; Feeney et al., 2014), notamment chez la souris (Dias & Ressler, 2014 ; Jung et al., 2022), le rat (Beck et al., 2021 ; Bruner-Tran & Osteen, 2011 ; Anway et al., 2006, 2005), l'Homme (Golding et al., 2022), le poisson zèbre (Knecht et al., 2017 ; Terrazas-Salgado et al., 2022), le medaka (R. K. Bhandari, 2016 ; R. Bhandari et al., 2015), le porc (Braunschweig et al., 2012) et la caille (Leroux et al., 2017 ; Vitorino Carvalho et al., 2023).

La transmission des effets de l'environnement peut s'effectuer via les gamètes paternels (spermatozoïdes). Elle est qualifiée de transgénérationnelle lorsque les effets sont visibles à partir de la G2 et au-delà. L'étude réalisée sur la souris (Dias & Ressler, 2014), présentée dans la partie précédente (**Figure 4.4**), a montré un héritage des effets via les cellules germinales parentales à la descendance. En effet, les générations de souris suivantes (G1 et G2), issues de l'accouplement

de mâles conditionnés à la peur de l'odeur et de femelles naïves (non conditionnées), présentaient une sensibilité comportementale accrue à l'odeur de conditionnement des G0. De plus, une hypométhylation des CpG dans le gène *Olf151* a été identifiée dans les spermatozoïdes des mâles G1 et G2 naïfs. Ainsi, les changements observés ont également persisté chez les descendants mâles G2, mettant en évidence l'existence d'une hérédité transgénérationnelle des effets via les gamètes paternels (**Figure 4.6**).



**FIGURE 4.6 – Transmission transgénérationnelle des effets de l'environnement paternel chez la souris.** Des souris mâles (G0) ont subi un stress odorant d'acétophénone couplé à un choc électrique. Après accouplement de ces mâles conditionnés avec des femelles naïves (non conditionnées), la descendance (G1 et G2) a développé une sensibilité comportementale à la même odeur ainsi qu'une hypométhylation du gène *Olf151*. *Figure adaptée de Dias et Ressler (2014)*.

Dans le cas d'une mère gestante (G0), la transmission est dite transgénérationnelle à partir de la troisième génération (G3). Une étude menée sur la souris a montré que l'exposition de souris gestantes au BPA induisait une augmentation significative du nombre de souris obèses dans la descendance, phénomène observé sur six générations, sans que ces individus n'aient été directement exposés à la molécule au cours de leur vie (Jung et al., 2022). Dans une autre étude menée chez le rat, des femelles gestantes ont été exposées à des PE, la vinclozoline, un fongicide viticole aux propriétés féminisantes, et au méthoxychlore, un composé œstrogénique utilisé comme insecticide. La descendance mâle des rates exposées à ces molécules présentait une diminution de la spermatogenèse, sans être stérile, ainsi qu'une augmentation de l'infertilité (Anway et al., 2005). Après la 2<sup>ème</sup> génération, sans aucune autre exposition, les mâles avaient le même phénotype que leur père, et au total, quatre générations ont été touchées (G1 à G4). Une corrélation entre les effets sur la reproduction et l'altération des schémas de méthylation de l'ADN dans la lignée germinale a été mise en évidence (**Figure 4.7**).

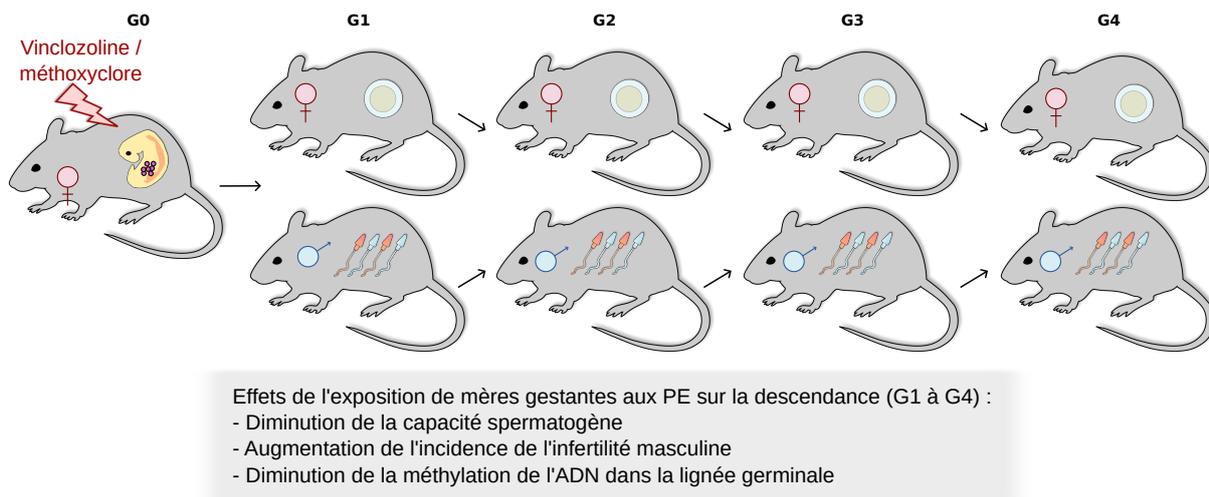


FIGURE 4.7 – **Transmission transgénérationnelle des effets de l'environnement maternel chez la souris.** Des mères gestantes (G0) ont été exposées à la vinclozoline et au méthoxychlor. Les effets de cette exposition, transmis via les cellules germinales, ont eu des répercussions sur quatre générations (G1 à G4). *Figure adaptée de Anway et al. (2005).*

Chez les oiseaux, très peu d'études ont démontré l'existence d'une transmission épigénétique transgénérationnelle des effets de l'environnement embryonnaire. Une étude pilote effectuée chez la caille japonaise a suggéré que l'environnement embryonnaire affecte le phénotype de la descendance trois générations après exposition à la génistéine (Leroux et al., 2017). Dans cette étude, deux lignées de cailles (épilignées), ont été croisées "en miroir", une épilignée ayant reçu directement l'injection de génistéine dans l'œuf (epi+) et une n'ayant pas reçu d'injection (contrôle, epi-). Le croisement "en miroir" visait à minimiser la variabilité génétique entre les lignées par le maintien des contributions d'ancêtres similaires à travers les générations dans chaque lignée. Cette étude a montré un effet de l'injection de génistéine sur la descendance adulte trois générations après l'injection (G3), notamment au niveau de leur poids corporel (diminué), la maturité sexuelle (retard de ponte) ainsi que leur comportement social, se traduisant notamment par moins de stress lors d'un isolement social (**Figure 4.8**). Il s'agit de la première étude, chez la caille, démontrant un effet transgénérationnel de l'environnement embryonnaire sur le phénotype de la descendance trois générations (G3) après le stress. Même si l'analyse de la méthylation de l'ADN globale n'a pas pu mettre en évidence une différence entre les deux épilignées, les modifications phénotypiques de la G3 peuvent être en partie expliquées par une transmission transgénérationnelle des marques épigénétiques, dont la méthylation de l'ADN.

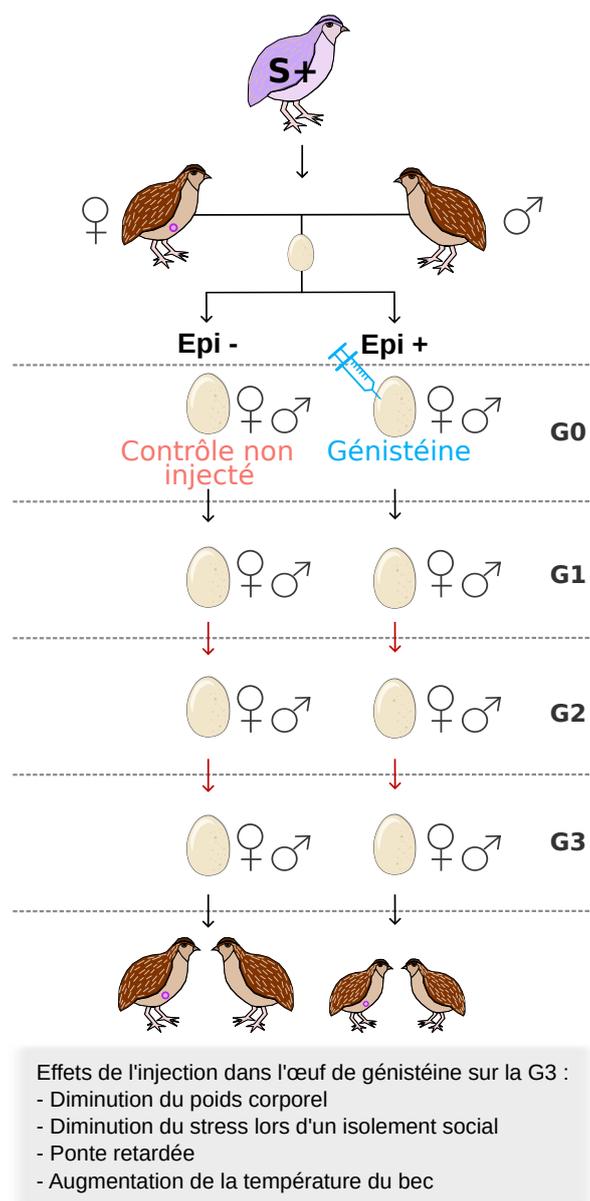


FIGURE 4.8 – **Transmission transgénérationnelle des effets de l'injection de génistéine dans l'œuf chez la caille japonaise.** Deux épilignées de cailles (issues de la lignée S+) ont été croisées "en miroir", une épilignée ayant reçu directement l'injection de génistéine dans l'œuf (epi+) et une n'ayant pas reçu d'injection (contrôle, epi-) (G0). L'injection dans l'œuf en G0 a eu des répercussions phénotypiques sur la descendance trois générations plus tard (G3), sans aucune autre injection. *Figure adaptée de Leroux et al. (2017).*

Une étude plus récente menée sur sept générations de cailles japonaises a également pu mettre en évidence une transmission transgénérationnelle des effets de la manipulation thermique (MT), cette dernière consistant à incuber les œufs à 39.5°C pendant 12h par jour (Vitorino Carvalho et al., 2023) (**Figure 4.9**). La répétition de MT sur plusieurs générations de cailles (G1 à G4) a induit une diminution du poids corporel des cailles ainsi que de leurs œufs, ces effets étant plus importants au cours des générations exposées. Ces effets sont également transmis de manière transgénérationnelle mais sont aussi en partie réversibles, selon l'âge et le sexe des individus. En effet, les effets sur la croissance s'estompent après cinq générations de cailles non exposées. Cette étude a ainsi pu mettre en évidence l'effet quantitatif, transgénérationnel et réversible de la manipulation thermique.

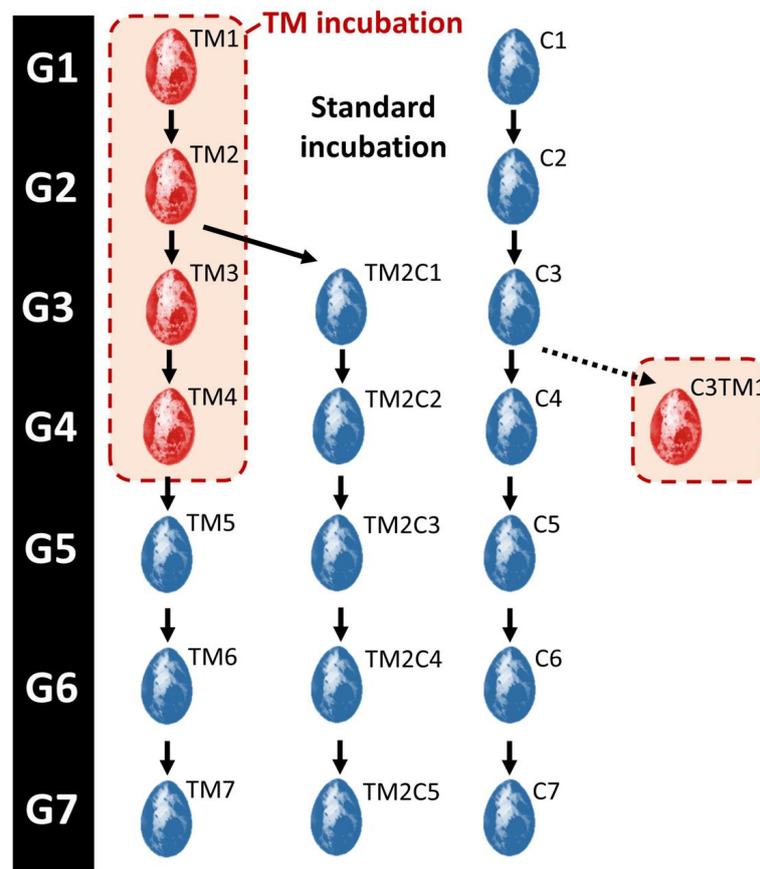


FIGURE 4.9 – **Transmission transgénérationnelle des effets de la manipulation thermique chez la caille japonaise.** Les œufs incubés dans des conditions de contrôle standard (37,8°C) sont représentés en bleu. Les œufs thermo-manipulés (TM; +1,7 °C 12h / jour des jours d'incubation 0-13) sont représentés en rouge. Le nom des traitements pour chaque génération est indiqué en haut à droite de l'œuf comme suit : Cn = n générations incubées dans des conditions de contrôle ; TMn = n générations consécutives de TM ; TM2Cn-2 = 2 générations consécutives de TM suivies de n-2 générations incubées dans des conditions de contrôle ; Gn = Génération n. *Figure issue de Vitorino Carvalho et al. (2023).*

Les études menées sur la caille japonaise ont pu mettre en évidence une transmission transgénérationnelle des effets de l'environnement et renforcent l'hypothèse d'une hérédité épigénétique transgénérationnelle, même si cette dernière n'a pas été démontrée.

Les deux dispositifs précédents ont été utilisés dans le cadre de cette thèse et seront présentés plus en détail dans la **Partie II**. L'ensemble de ces dispositifs s'inscrit dans l'étude des mécanismes d'action pouvant impliquer des régulations épigénétiques, notamment par le biais de modifications de la méthylation de l'ADN, et permet d'analyser les effets potentiellement hérités de manière non génétique. Pour cela, de nombreuses techniques de séquençage haut débit ont été développées, notamment certaines méthodes basées sur le traitement de l'ADN au bisulfite de sodium ou sur nanopores. Ainsi, une meilleure connaissance de la relation entre le génotype et le phénotype, en prenant en compte les mécanismes génétiques et épigénétiques, permettrait d'améliorer les systèmes de production animale.



## 5. Technologies de séquençage pour l'étude du méthylome

De nombreuses techniques de séquençage permettent d'étudier les marques épigénétiques telles que la méthylation de l'ADN (Kurdyukov & Bullock, 2016) (**Table 5.1**). Chez les vertébrés, la méthylation de l'ADN cible principalement les 5mC au niveau des CpG (Gardiner-Garden & Frommer, 1987). La méthode de référence pour étudier l'état de méthylation de l'ADN au niveau des 5mC est la conversion de l'ADN par le bisulfite de sodium. Certaines techniques de séquençage sont donc basées sur le traitement de l'ADN préalable au bisulfite, telles que le séquençage au bisulfite à représentation réduite (RRBS - *Reduced Representation Bisulfite Sequencing*) (Meissner et al., 2005a; Harris et al., 2010) et le séquençage au bisulfite tout génome (WGBS - *Whole Genome Bisulfite Sequencing*) (Urich et al., 2015). D'autres techniques d'étude de la 5mC, n'utilisant pas la conversion au bisulfite, peuvent également être utilisées, notamment :

- ONT (*Oxford Nanopore Technologies*) : technique de séquençage du génome entier ou ciblé basée sur l'utilisation de nanopores,
- MeDIP-seq (*Methylated DNA Immuno-Precipitation sequencing*) : immuno-précipitation de l'ADN méthylé utilisant des anticorps anti-mCpG,
- MeDP-seq (*Methylated DNA Precipitation sequencing*) ou MBD-seq (*Methyl-binding domain sequencing*) : techniques de séquençage utilisant une protéine recombinante reconnaissant l'ADN méthylé (domaine MBD),
- MRE-seq (*Methylation sensitive Restriction Enzyme sequencing*) : séquençage par enzyme de restriction sensible à la méthylation permettent également d'étudier l'état de méthylation au niveau des 5mC,
- SMRT (*Single Molecule Real Time*) de PacBio : séquençage en temps réel de molécules uniques, se focalisant sur les marques de méthylation m6A (Adénine méthylée sur l'azote 6) et m4C (Cytosine méthylée sur l'azote 4),
- EM-seq (*Enzymatic Methyl-seq*) : méthode récente méthyl-seq enzymatique permet de détecter les 5mC et les 5hmC.

Seules les techniques utilisées au cours de cette thèse, notamment le séquençage RRBS, WGBS et ONT, seront détaillées (**Figure 5.1**).

TABLE 5.1 – Principales technologies d'analyse de la méthylation de l'ADN au niveau des 5mC. La taille des lectures mentionnée est approximative, cette dernière pouvant varier en fonction de la plate-forme de séquençage et du protocole utilisés. *Table adaptée de Mersch et al. (2018) et Gowil et Keniry (2019).*

Technique	Préparation de l'ADN	Principe d'identification	Fraction d'ADN étudiée	Taille des lectures	Résolution
<i>Whole Genome Bisulfite Sequencing</i> (WGBS ou MEthylC-seq)	Extraction + conversion au bisulfite de sodium	Séquençage haut-débit	Génome entier	Courte (~ 50 - 300 pb)	À la base près
<i>Reduced Representation Bisulfite Sequencing</i> (RRBS)	Extraction + conversion au bisulfite de sodium + digestion avec enzyme de restriction + sélection de fragments par leur taille	Séquençage haut-débit	Sous-représentation du génome enrichie en dinucléotides CpG	Courte (~ 50 - 300 pb)	À la base près
<i>Oxford Nanopore Technologies sequencing</i> (ONT)	Extraction d'ADN préservant les longs fragments	Séquençage haut-débit	Génome entier/ciblé	Longue (~ 500 pb - 1 Mpb)	À la base près
<i>Enzymatic Methyl-seq</i> (EM-seq)	Extraction + conversion enzymatique	Séquençage haut-débit	Génome entier	Courte (~ 50 - 300 pb)	À la base près
<i>Methylated DNA ImmunoPrecipitation</i> (MeDIP-seq ou MBD-seq)	Extraction classique	Reconnaissance par anticorps de l'ADN méthylé + séquençage haut-débit	Génome entier	Courte (~ 50 - 300 pb)	Limitée à la taille des fragments immunoprécipités
<i>Methylation sensitive Restriction Enzyme sequencing</i> (MRE-seq)	Extraction + digestion avec enzyme de restriction sensible à la méthylation	Séquençage haut-débit	Génome entier	Courte (~ 50 - 300 pb)	Limitée aux sites de coupure enzymatique

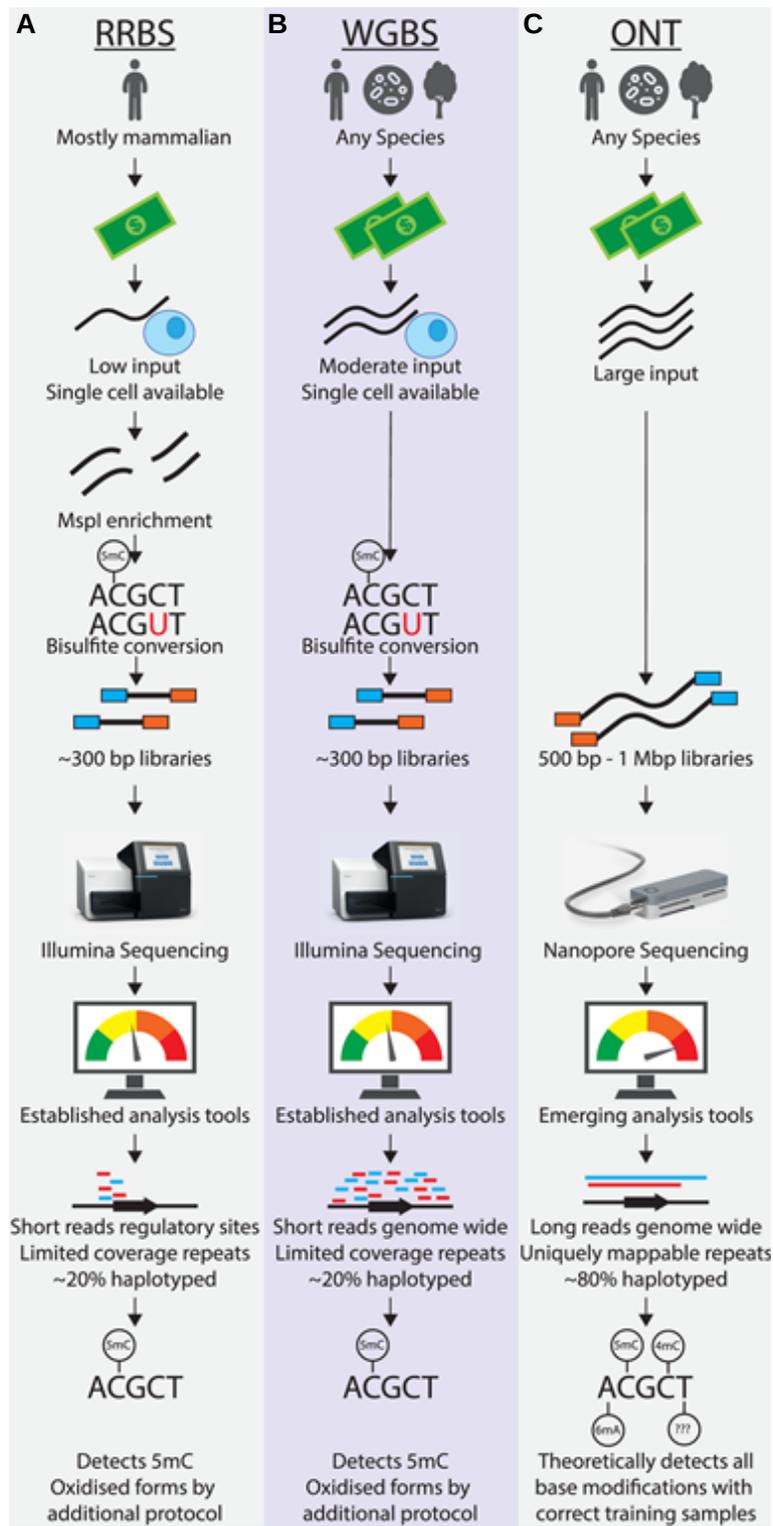


FIGURE 5.1 – **Détection des modifications de bases par séquençage bisulfite et à longue lecture.** (A) Séquençage au bisulfite à représentation réduite (RRBS) : lectures courtes (~ 300 pb). (B) Séquençage au bisulfite tout génome (WGBS) : lectures courtes (~ 300 pb). (C) Séquençage par nanopores d'Oxford Nanopore Technologies (ONT) : lectures longues (500 pb - 1 Mb). *Figure issue de Gouil et Keniry (2019).*

## 5.1 Traitement de l'ADN au bisulfite de sodium

Les techniques RRBS et WGBS sont basées sur le séquençage après traitement au bisulfite de sodium : elles reposent sur la désamination des cytosines non méthylées par le bisulfite de sodium, ce dernier n'affectant pas les cytosines méthylées (**Figure 5.2**) (Meissner et al., 2005b). Les méthodes RRBS et WGBS utilisées au cours de cette thèse reposent sur la technologie de séquençage Illumina.

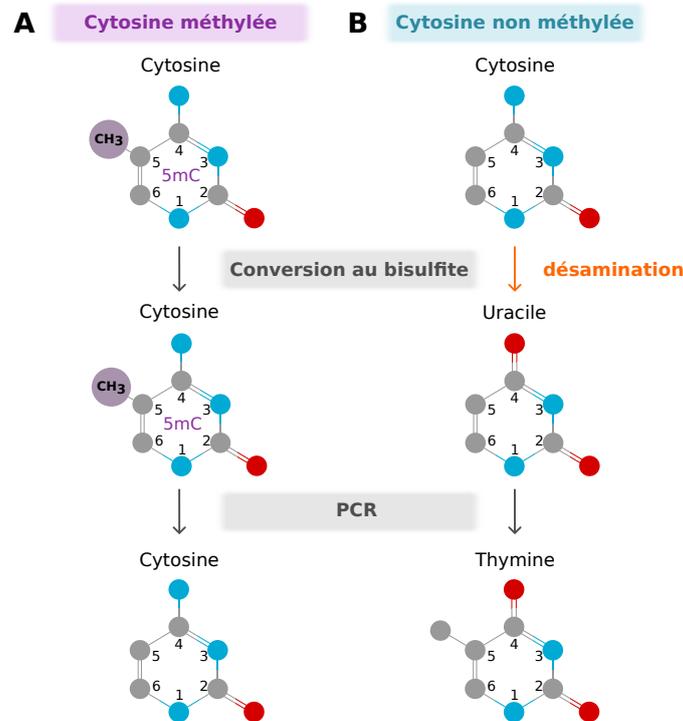
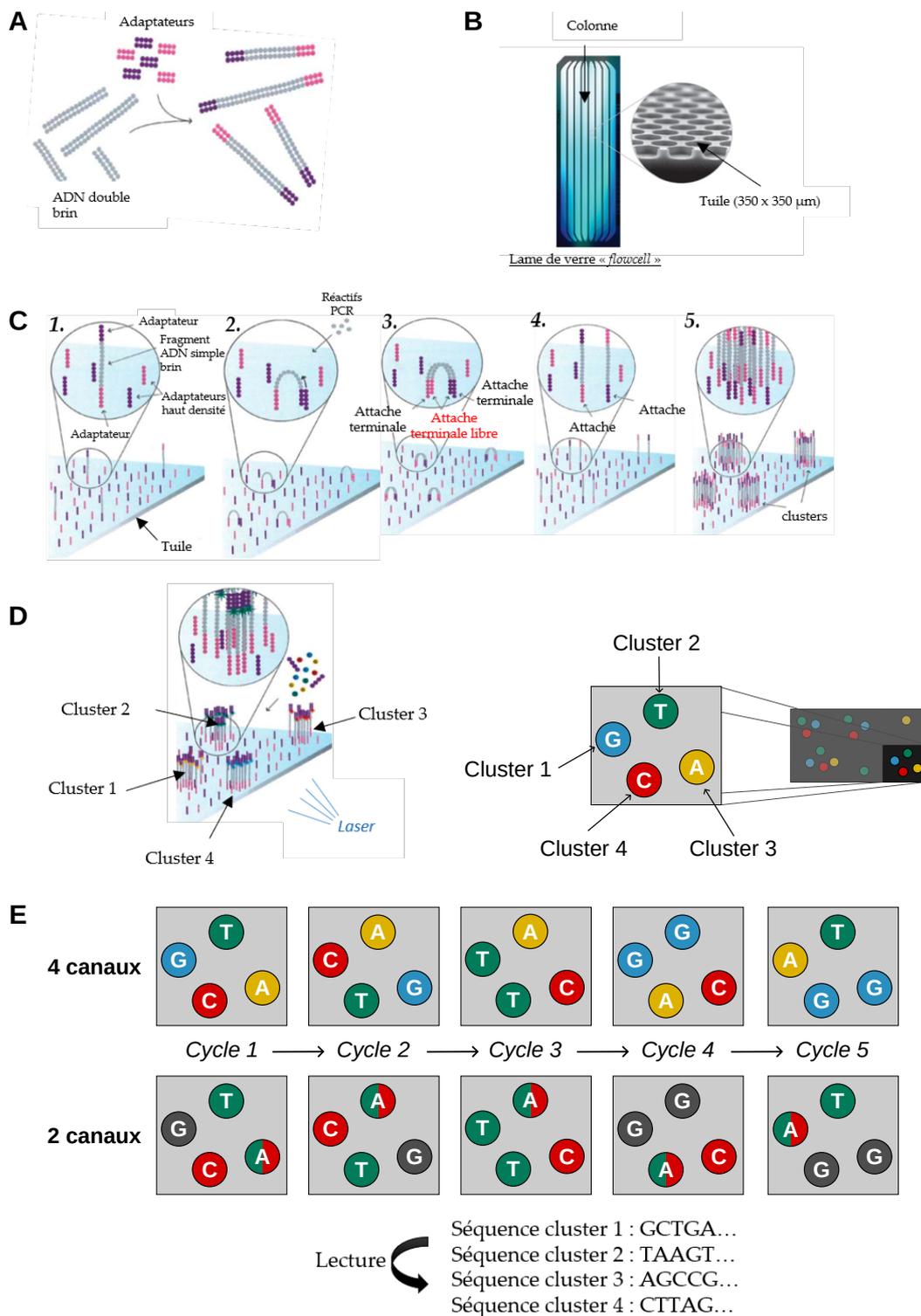


FIGURE 5.2 – **Conversion au Bisulfite de Sodium.** (A) Conversion d'une cytosine initialement méthylée. (B) Conversion d'une cytosine initialement non méthylée. Les groupements méthyles (CH<sub>3</sub>) sont représentés en violet. Les atomes d'oxygène (O), de carbone (C) et d'azote (N) sont représentés respectivement en rouge, gris et bleu.

### 5.1.1 Séquençage haut-débit de l'ADN par la technologie Illumina

Plusieurs techniques de séquençage haut-débit actuelles reposent sur le principe du séquençage Sanger ou séquençage par synthèse, ce dernier étant miniaturisé et parallélisé. La technique Illumina est résumée sur la **Figure 5.3**. Elle consiste globalement à amplifier des clusters de fragments similaires qui sont séquencés ensuite par incorporation de nucléotides fluorescents. En fonction de la technologie utilisée, le nombre de dNTP fluorescents peut varier. En effet, certaines technologies, telles que HiSeq, peuvent utiliser quatre dNTP fluorescents (rouge, vert, jaune et bleu) pour séquencer en parallèle les millions de clusters sur la surface du *flowcell*. Le fluorophore lié à chaque base des fragments va être excité par un laser et la fluorescence émise est lue et enregistrée (**Figure 5.3.E**, en haut). Afin d'accélérer la détection des 4 bases de l'ADN, certaines technologies, telles que NovaSeq, peuvent aussi utiliser seulement deux dNTP fluorescents (rouge et vert). Dans ce cas, le signal émis par chacun des deux fluorophores, la superposition des deux ainsi que l'absence de signal seront lus et enregistrés (**Figure 5.3.E**, en bas).



**FIGURE 5.3 – Séquençage Illumina.** L'ADN double brin est découpé en fragments puis deux adaptateurs double brin spécifiques sont ligués aux extrémités de l'ADN (A). Les oligonucléotides complémentaires aux adaptateurs de séquençage sont déposés sur une lame de verre (*flowcell*) (B). Chaque fragment d'ADN simple brin s'hybride sur les amorces complémentaires fixées sur la lame (C.1). L'hybridation de l'adaptateur libre sur son complémentaire proche permet la création d'un pont (C.2). L'ajout de réactifs pour faire la PCR permet l'amplification du brin complémentaire d'ADN (C.3). L'ADN double brin produit permet de générer deux molécules d'ADN simple brin attachées à la lame, qui sont ensuite séquencées par incorporation de désoxynucléotides triphosphates (dNTP) fluorescents (C.4). Les étapes C.1 à C.4 sont répétées pour générer des clusters de molécules identiques (C.5 et D). En fonction de la technologie utilisée, le nombre de dNTP fluorescents peut varier (quatre (rouge, vert, jaune et bleu) ou deux (rouge et vert) dNTP fluorescents), pour séquencer en parallèle les millions de clusters sur la surface du *flowcell*. Le fluorophore lié à chaque base des fragments va être excité par un laser et la fluorescence émise ou l'absence de signal (en fonction de la technologie) sont lus et enregistrés (E). *Image adaptée de Illumina (2022).*

### 5.1.2 Séquençage au bisulfite tout génome (WGBS)

Le séquençage WGBS permet d'analyser la totalité des sites CpG du génome d'un individu et est une technique assez onéreuse. Cette technique est basée sur une conversion des cytosines non méthylées (**Figure 5.2**). Lors du séquençage WGBS (**Figure 5.4**), après extraction de l'ADN des tissus et fragmentation de l'ADN, un A est ajouté en extrémité 3' de chaque côté du double brin pour permettre la liaison avec les adaptateurs méthylés. Après traitement de l'ADN au bisulfite de sodium, une C méthylée est inchangée alors qu'une C non-méthylée est désaminée en U. Une étape d'amplification par PCR (réaction en chaîne par polymérase) est ensuite effectuée afin d'amplifier *in vitro* les fragments d'ADN. Puis le séquençage Illumina est effectué, durant lequel les brins Original Top (OT), Original Bottom (OB) et les brins complémentaires (OTC et OBC), sont séquencés (**Figure 5.3**). Après les étapes de PCR et de séquençage, les séquences récupérées (appelées lectures), comportent des différences par rapport à la séquence d'origine. Une C méthylée de la séquence d'origine est maintenue en C et l'U est remplacée par une T. Ainsi, une T observée sur une lecture peut correspondre à une C non-méthylée sur la séquence d'ADN d'origine (Frommer et al., 1992 ; Mardis, 2008) (**Figure 5.2**). Pour chaque position, la proportion de cytosines non converties, c'est-à-dire le nombre de C observées, permettra de calculer le taux de méthylation à cette position.

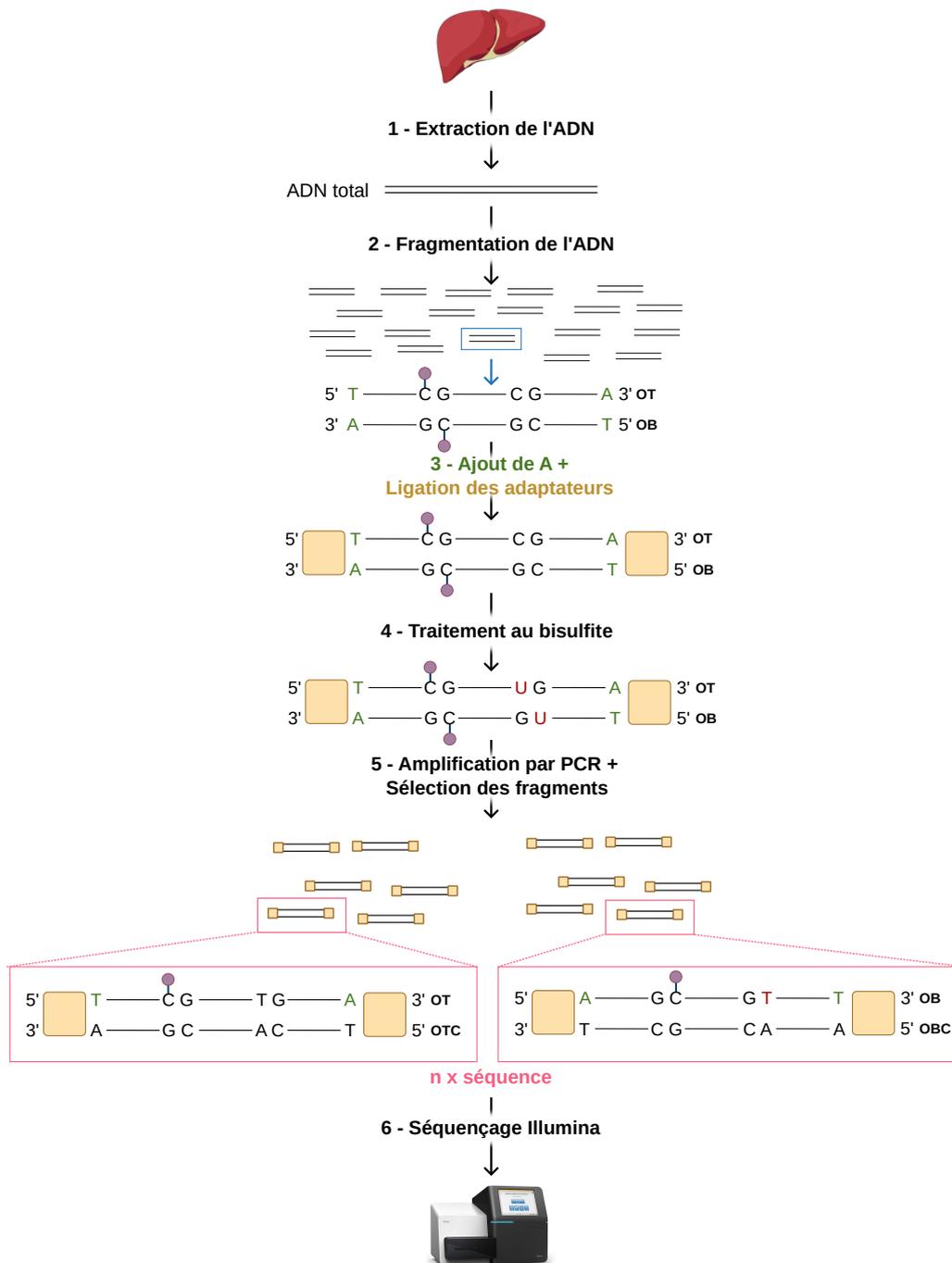


FIGURE 5.4 – **Séquençage WGBS.** (1) L'ADN est extrait des tissus (2) et est ensuite fragmenté. (3) Un A est ajouté et les adaptateurs sont ligués à chaque extrémité des fragments d'ADN. (4) L'ADN est traité au bisulfite de sodium convertissant les C non-méthylées en U. (5) Une étape d'amplification par PCR des fragments d'ADN est effectuée. (6) Les fragments récupérés sont ensuite utilisés pour la préparation des banques de séquençage Illumina. Les brins OT et OB sont séquençés et en général les brins OTC et OBC ne le sont pas : un séquençage directionnel (*directional*) est effectué. OT = *Original Top* ; OTC = *Original Top Complementary* ; OB = *Original Bottom* ; OBC = *Original Bottom Complementary*.

### 5.1.3 Séquençage au bisulfite à représentation réduite (RRBS)

Le séquençage RRBS, beaucoup moins coûteux que le WGBS, permet d'identifier une sous-sélection de sites CpG répartis sur l'ensemble du génome, enrichis au niveau des régions promotrices des gènes, lorsque l'enzyme *MspI* est utilisée (~2% du génome séquencé), ces dernières étant le plus souvent riches en îlots CpG. Environ 60% des promoteurs des gènes contiennent des îlots CpG chez l'Homme (Hannenhalli & Levy, 2001 ; Sarda & Hannenhalli, 2018). Cette technique permet de cibler les régions ayant un taux de méthylation variable, 70% à 80% des CpG étant méthylés chez les mammifères (Jabbari & Bernardi, 2004). Ainsi, le niveau de méthylation moyen observé après une analyse RRBS utilisant l'enzyme *MspI* est inférieur à la moyenne de tous les CpG du génome. Lors du séquençage RRBS (**Figure 5.5**), l'ADN est digéré par une enzyme de restriction, en général *MspI*, qui n'est pas sensible à la méthylation des sites CpG. Cette enzyme permet de cliver de l'ADN au niveau des sites C/CGG et génère des extrémités cohésives. Les régions riches en GC étant dotées de plus de sites CCGG, les sites CpG identifiés se retrouvent majoritairement dans ces régions. Dans le protocole utilisé classiquement, les nucléotides G et C sont ajoutés afin d'avoir un double brin d'ADN avec des extrémités franches. Comme pour le WGBS, un A est ajouté en extrémité 3' de chaque côté du double brin pour permettre la liaison avec les adaptateurs méthylés. Après traitement de l'ADN au bisulfite de sodium, une étape d'amplification par PCR suivie du séquençage Illumina (paired-end) est effectuée (**Figure 5.3**), durant laquelle seuls les brins Original Top (OT) et Original Bottom (OB), et non les brins complémentaires (OTC et OBC), sont séquencés, cette méthode de construction des bibliothèques étant la plus couramment utilisée en RRBS. Comme pour les WGBS, après les étapes de PCR et de séquençage, une C méthylée de la séquence d'origine est maintenue en C et l'U est remplacée par une T. Ainsi, une T observée sur une lecture peut correspondre à une C non-méthylée sur la séquence d'ADN d'origine.

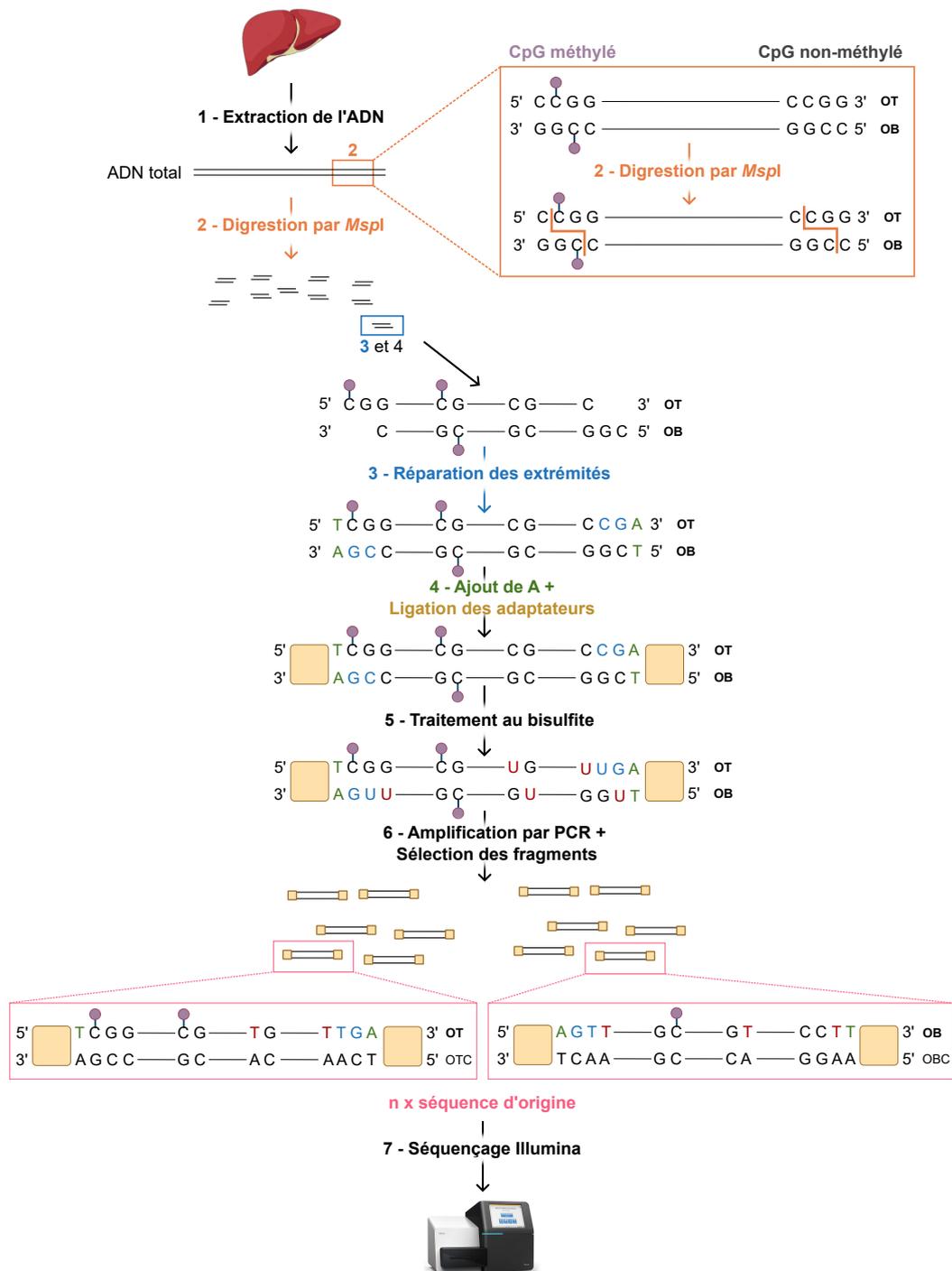


FIGURE 5.5 – **Séquençage RRBS.** (1) L'ADN est extrait des tissus (2) puis subit une digestion enzymatique (enzyme *MspI*) formant des extrémités cohésives. (3) Ces extrémités sont réparées (ajout de GC) et les fragments sont sélectionnés en fonction de leur taille. (4) Un A et des adaptateurs sont ajoutés. (5) L'ADN est traité au bisulfite de sodium convertissant les C non-méthylées en U. (6) Une étape d'amplification par PCR des fragments d'ADN est effectuée. Les fragments récupérés sont ensuite utilisés pour la préparation des banques de séquençage Illumina (7). Seuls les brins OT et OB sont séquencés. OT = *Original Top*; OTC = *Original Top Complementary*; OB = *Original Bottom*; OBC = *Original Bottom Complementary*.

### 5.1.4 Analyse du méthylome et biais d'analyse

L'ensemble des techniques utilisées permet de calculer le taux de méthylation de chaque site CpG. Le taux de méthylation est calculé à l'aide de l'équation suivante (équation 5.1), où C représente le nombre de cytosines et T le nombre de thymines :

$$\text{Taux de méthylation} = \frac{\text{Nombre de C}}{\text{Nombre de C} + \text{Nombre de T}} \quad (5.1)$$

Le taux de méthylation correspond donc au ratio du nombre de C sur la somme des C et T répertoriés à l'aide des lectures. La somme du nombre de C et de T pour chaque CpG correspond à la profondeur de lecture (ou couverture). Notons que le taux est souvent différent de 0 ou 100% puisqu'il représente un grand nombre de cellules qui ne sont pas forcément toutes 100% identiques : on peut avoir des populations cellulaires hétérogènes avec des taux de méthylation différents, et même en cas de population homogène, il arrive que certaines cellules présentent une méthylation spécifique. Les lectures s'alignant sur une même position, peuvent présenter un C, qui correspond à un C méthylé sur la séquence d'origine, et d'autres lectures peuvent avoir un T, qui représente initialement un C non-méthylé.

Afin de choisir le meilleur outil et interpréter correctement les résultats, il est important de connaître les biais de chaque technique (Bock, 2012 ; Mersch et al., 2018). Le RRBS, se focalisant principalement au niveau des régions promotrices des gènes, est une méthode moins coûteuse que le WGBS. Cependant, comme son nom l'indique, le RRBS ne permet d'étudier qu'une sous-partie du génome. Les sites CpG isolés sont le plus souvent exclus de l'analyse (Gu et al., 2011), ce qui peut être limité par l'utilisation d'une combinaison d'enzymes pour cibler une portion plus grande du génome (Mersch et al., 2018). Le niveau de méthylation moyen est en général plus élevé pour le WGBS que pour le RRBS et la couverture au niveau des îlots CpG est plus élevée pour le RRBS (Doherty & Couldrey, 2014 ; Beck et al., 2022). Les différences observées entre ces deux technologies sont en partie expliquées par l'utilisation d'une enzyme de restriction spécifique (*MspI*) pour le RRBS. Les sites de reconnaissance (motif "CCGG") de cette enzyme sont principalement situés dans des régions à fort taux de CG (îlots CpG) telles que les régions promotrices, connues pour être faiblement méthylées (A. Bird, 2002).

Le traitement de l'ADN au bisulfite de sodium est connu pour entraîner une dégradation de l'ADN, se traduisant par une dépyrimidination de l'ADN (rupture du lien entre un désoxyribose et une pyrimidine) induisant ainsi une diminution de la taille des fragments d'ADN (Olova et al., 2018 ; Tanaka & Okamoto, 2007). Le taux de mapping est souvent légèrement supérieur en WGBS qu'en RRBS, les séquences étant en général plus longues et en paired-ends, mais ce n'est pas systématique.

La résolution à la base près du séquençage au bisulfite permet d'étudier avec précision l'état de méthylation des cytosines. Cependant, elle ne permet pas de distinguer les 5mC des 5hmC ou de détecter d'autres formes oxydées (Globisch et al., 2010). L'ADN doit être de très bonne qualité pour l'obtention de fragments non dégradés de taille correcte (Ziller et al., 2015). Plus la taille des fragments séquencés est grande, plus la longueur des haplotypes déterminés au sein de chaque lecture sera importante.

L'haplotypage consiste à déterminer les loci (allèles) présents sur une région du génome. Si plusieurs SNP sont présents au sein de la même lecture, la reconstitution d'haplotypes parentaux sera facilitée par l'association directe des allèles pour chaque SNP (*Single Nucleotide Polymorphism*). Pour cela, les lectures plus longues sont donc souvent plus informatives. Bien que les SNP puissent être des éléments ajoutant une information intéressante, par exemple quand on s'intéresse à la méthylation allèle-spécifique, certains de ces polymorphismes peuvent induire des erreurs d'interprétation. En effet, des sites CpG peuvent être mutés, induisant la formation d'un SNP. Dans le cas d'un CpG non muté, le C de ce dernier pourra être méthylé par la machinerie de méthylation de l'ADN (**Figure 2.2.B et 5.6.A**). En revanche, si le C ou le G d'un site CpG est muté, alors la méthylation du C ne pourra pas s'effectuer (**Figure 5.6.B**). Les cytosines méthylées sont plus sensibles à la désamination, ce qui conduit à des taux de mutations de type  $C \rightarrow T$  supérieurs à la moyenne dans les sites CpG des vertébrés. Ainsi, si le génome de référence contient un dinucléotide CG lu en TG sur des données issues de traitement bisulfite, on calcule un taux de méthylation à zéro, alors qu'il s'agit d'une position qui ne doit pas être prise en compte en analyse de méthylation. Par exemple, dans le cas d'une comparaison entre deux échantillons, si pour une même position, l'un d'entre eux présente un site CpG intact alors que l'autre présente un dinucléotide CA, CT ou CC, le taux de méthylation du second groupe sera de 0 car le traitement au bisulfite convertit les C non-méthylés en T (**Figure 5.2**). Ainsi, si le taux du premier groupe est non nul, un différentiel de méthylation pourrait être identifié à tort. La présence de ces SNP dans les données est une source de biais importante. Il est donc essentiel de minimiser leur impact pour la suite des analyses.

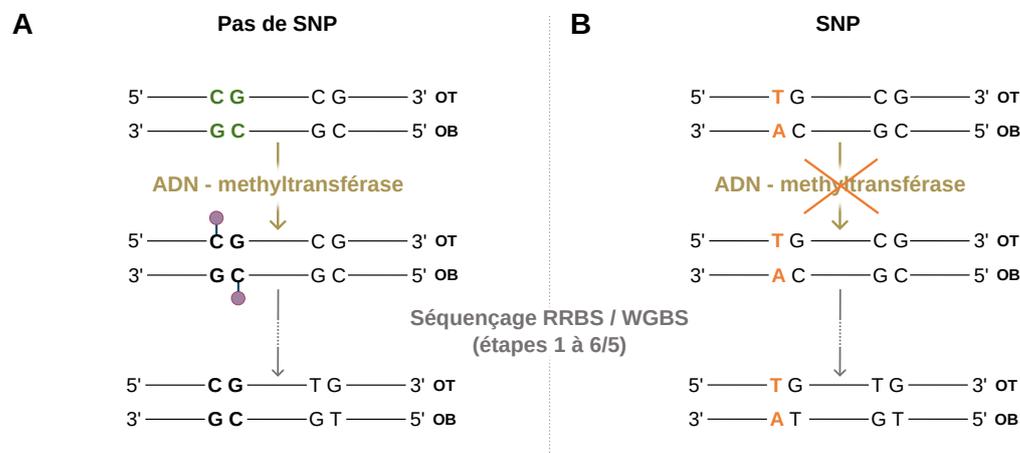


FIGURE 5.6 – **Conséquence de la présence d'un SNP sur les analyses.** Comparaison d'un site CpG localisé à la même position chez deux individus différents. (A) Un site CpG intact. (B) Le même site CpG muté générant un SNP.

Pour limiter le nombre de faux positifs, de nombreux outils de détection de SNP à partir de données issues de séquençage après traitement bisulfite peuvent être utilisés, tels que BisSNP (Y. Liu et al., 2012), biscuit (<https://github.com/zhou-lab/biscuit>), BS-SNPer (Gao et al., 2015), CGMAPTools (Guo et al., 2018), Epidiverse (Nunn et al., 2022), methylExtract (Barturen et al., 2013) et gemBS (Merkel, Fernández-Callejo, et al., 2019). Parmi ces outils, gemBS offre un bon compromis concernant le taux de vrais positifs et de faux négatifs détectés par rapport aux autres outils disponibles (Lindner et al., 2022 ; Merkel, Fernández-Callejo, et al., 2019). Le

pipeline *gemBS* aligne des lectures sur le génome de référence à l'aide de *GEM3* puis effectue le calling de variants à l'aide de *BScall*, ces deux outils étant développés spécifiquement pour les données issues de séquençage au bisulfite.

Ainsi, afin de limiter les biais d'analyse pour l'obtention de résultats plus fiables, d'autres technologies de séquençage, basées sur l'utilisation de lectures longues et ne nécessitant pas de traitement préalable de l'ADN au bisulfite de sodium, peuvent être utilisées, notamment le séquençage par nanopores (ONT).

## 5.2 Technologie Nanopore

Certaines technologies de séquençage à longues lectures peuvent différencier les cytosines méthylées et non méthylées sans traitement préalable de l'ADN au bisulfite de sodium. C'est notamment le cas de la technique de séquençage Nanopore (ONT) (<https://nanoporetech.com>) qui est utilisée pour étudier l'état de méthylation de l'ADN (Simpson et al., 2017).

### 5.2.1 Séquençage nanopore tout génome

Le séquençage par nanopores d'Oxford Nanopore Technologies (ONT) permet de lire simultanément le génome et le méthylome d'un même échantillon et ne nécessite pas de traitement de l'ADN au bisulfite de sodium ni d'une amplification en amont (Simpson et al., 2017; Tagu & Jaubert-Possamai, 2018). Ainsi, suite à l'extraction et à la fragmentation (en longs fragments) de l'ADN de haut poids moléculaire (**Figure 5.7.1-2**), les fragments d'ADN sont ligués au niveau des deux extrémités à des adaptateurs double brin, un adaptateur en forme d'épingle à cheveux ("*hairpin*") et un adaptateur contenant une séquence nucléotidique ("*tether*"). Ces adaptateurs contiennent des enzymes motrices qui permettent de guider le fragment d'ADN dans le pore protéique (nanopore) (**Figure 5.7.3**). Des centaines de nanopores sont intégrés dans une membrane en polymère synthétique fixée sur une puce, cette dernière étant incorporée dans une *flowcell*. Un champ électromagnétique est ensuite appliqué à travers la membrane générant un flux ionique et des charges opposées (positives et négatives) de chaque côté de la membrane (**Figure 5.7.4**, gauche). Les enzymes motrices permettent de guider le fragment d'ADN vers le nanopore. La séquence "*tether*" va ensuite se fixer à la membrane permettant l'immobilisation de l'ADN à proximité du nanopore. L'enzyme motrice induit l'ouverture de la double hélice d'ADN et permet le glissement du simple brin d'ADN dans le nanopore. Au fur et à mesure que le brin d'ADN passe dans le pore, chaque base de l'ADN crée des perturbations du courant électrique à travers la membrane. Ces perturbations sont ensuite interprétées et converties en bases à l'aide d'un algorithme basé sur un apprentissage de type *deep learning*, avec des modèles de plus en plus précis (**Figure 5.7.4**, droite). Cette technique permet notamment d'identifier directement des marques de méthylation au niveau des bases de l'ADN, ces marques engendrant des modifications spécifiques du profil électrique.

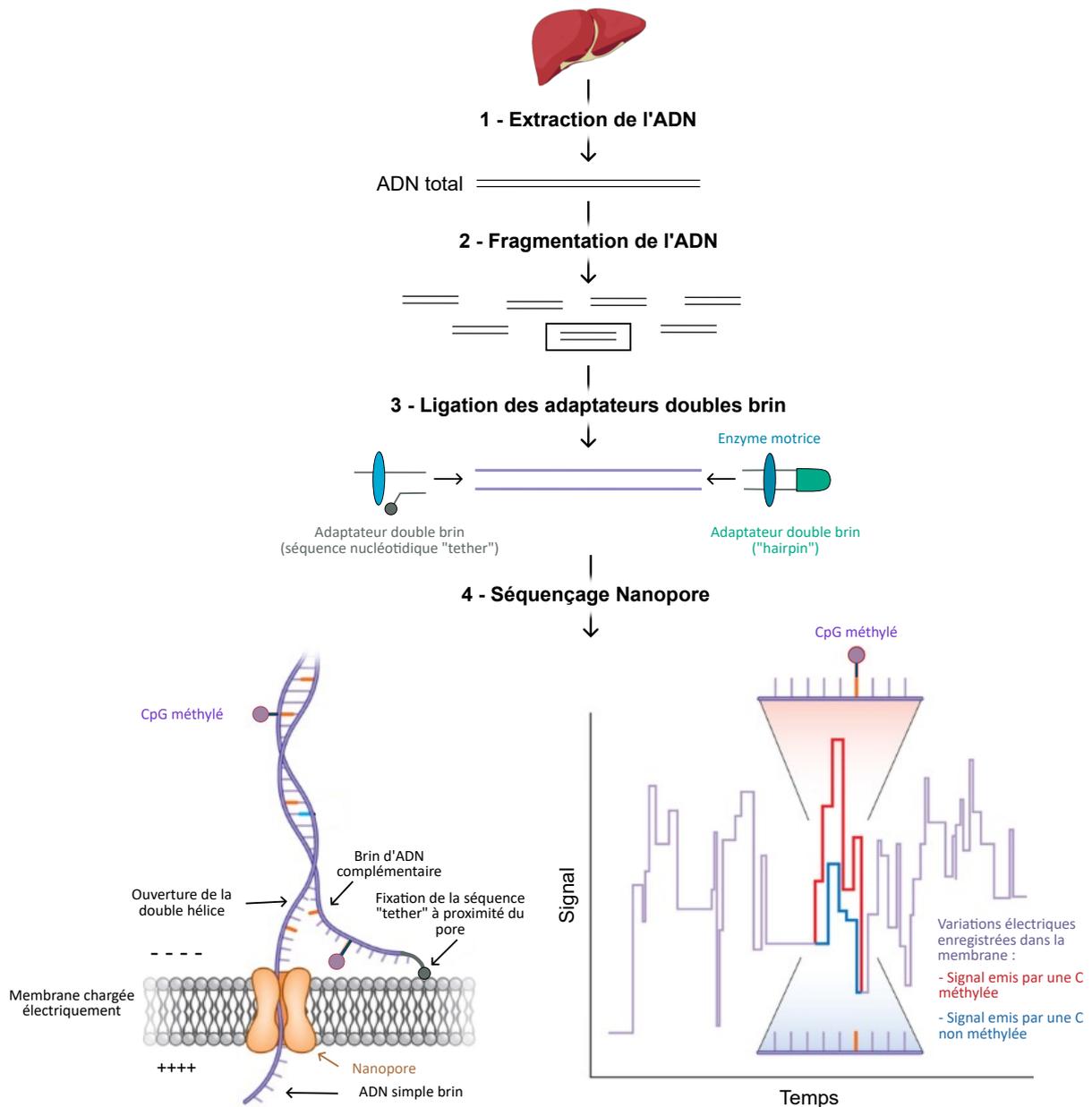


FIGURE 5.7 – **Séquençage ONT.** Après extraction (1) et fragmentation de l'ADN (2), des adaptateurs couplés à des enzymes motrices vont se liquer à l'ADN (adaptateur double brin contenant une séquence nucléotidique "tether" et un adaptateur double brin "hairpin") (3). L'ADN double brin est ensuite guidé vers le nanopore par l'intermédiaire des enzymes motrices et par la fixation de la séquence "tether" à proximité du nanopore (4, gauche). L'enzyme motrice induit l'ouverture de la double hélice d'ADN et permet la progression du simple brin d'ADN à travers le nanopore. Au fur et à mesure que le brin d'ADN passe dans le pore, chaque base de l'ADN crée des perturbations du courant électrique dans la membrane permettant ainsi de déterminer la base ainsi que les marques de méthylation à l'aide d'un algorithme (4, droite). *Figure adaptée de Schatz (2017) et Tagu et Jaubert-Possamai (2018).*

### 5.2.2 Analyse du méthylome et biais d'analyse

Le séquençage par nanopores d'Oxford Nanopore Technologies (ONT) à longues lectures (500 pb à plus de 2 Mb) permet d'étudier plusieurs marques de méthylation, telles que les 5mC, 5hmC (en fonction du *basecaller* utilisé), 4mC et 6mA (Payne et al., 2019). Le coût de séquençage reste assez élevé et est à peu près équivalent à celui du WGBS. Cette technologie comporte quelques limites, notamment un taux d'erreur global de séquençage qui reste supérieur au séquençage Illumina (< 0.1% d'erreur), malgré une amélioration certaine depuis les débuts : le taux d'erreur actuel en utilisation standard est de l'ordre de 1%, il descend à 0.1% avec la technologie *Duplex basecalling* (<https://nanoporetech.com/accuracy>). De plus, cette technologie nécessite de grandes quantités d'ADN de départ (méthode d'extraction d'ADN de haut poids moléculaire) pour la préparation des bibliothèques et une très bonne qualité d'échantillon (Van Dijk et al., 2018). La technologie ONT a également de nombreux avantages. Elle ne nécessite pas de traitement préalable de l'ADN au bisulfite de sodium, limitant ainsi la dégradation de l'ADN, ainsi qu'une amplification de l'ADN, limitant les biais d'amplification. L'utilisation de longues lectures améliore la cartographie et l'haplotypage (Gouil & Keniry, 2019). De plus, le séquençage ONT permet de lire simultanément le génome ainsi que le méthylome d'un même échantillon (Simpson et al., 2017). Cela facilite les analyses et permet donc de détecter directement l'état de méthylation de l'ADN ainsi que d'effectuer la détection directe des variants natifs à partir des mêmes données. Les analyses de type méthylation allèle-spécifique en sont grandement facilitées. Il existe de nombreux outils permettant la détection de SNP à partir de données ONT, tels que Human-SNP-wf, Clair3, Clair, NanoCaller Medaka et Longshot (Helal et al., 2022 ; Edge & Bansal, 2019). Après l'étape de séquençage, des outils bioinformatiques sont utilisés pour analyser les données de méthylation de l'ADN.

## 6. Approches pour la détection et l'analyse d'effets épigénétiques

Suite au séquençage, des outils bioinformatiques / statistiques sont utilisés pour analyser les données de méthylation de l'ADN afin d'identifier les marques de méthylation dans les échantillons ainsi que des différentiels de méthylation entre groupes d'échantillons. Dans cette partie, seuls les outils développés pour l'analyse de données issues des séquençages RRBS et WGBS seront présentés.

### 6.1 Détection des marques de méthylation

A partir des lectures issues du séquençage, le calcul du taux de méthylation de chaque CpG observable est effectué en deux étapes : dans un premier temps les lectures sont alignées sur un génome de référence, puis en chaque position CG du génome sont comptés le nombre de C et de T (données bisulfite) ou le nombre de lectures comportant un C méthylé (Nanopore). Pour effectuer l'identification des marques de méthylation, de nombreux outils bioinformatiques ont été développés et le choix de l'outil dépend du type de données à analyser (**Table 6.1**). En effet, concernant les données traitées au bisulfite, des outils développés spécifiquement pour ce type de données et prenant en considération la conversion des bases doivent être utilisés, en particulier pour tenir compte de la très faible teneur en bases C par rapport au génome de référence au moment de l'alignement. Pour identifier les sites CpG des données, des pipelines (enchaînement d'outils), tels que *nf-core/methylseq* (P. A. Ewels et al., 2020) ou *megalodon* (Oxford Nanopore Technologies, 2021) peuvent être utilisés pour analyser respectivement des données issues de séquençage bisulfite (T. Gong et al., 2022) ou nanopore (*Megalodon*) (Y. Liu et al., 2021). Les données issues de techniques de séquençage Illumina (NextSeq et NovaSeq) basées sur l'utilisation de deux dNTP fluorescents peuvent présenter des artefacts sous la forme de séquences poly-G. En effet, au fur et à mesure du séquençage, l'intensité du signal diminue et lorsque le signal est trop faible pour être détecté, il est reconnu comme un G lors de l'étape d'appel des bases induisant une erreur de séquençage. Cela se traduit par un ajout de G (poly-G) au niveau des extrémités des lectures. Pour retirer les séquences poly-G des lectures, l'outil de nettoyage des lectures *fastp* peut être utilisé en amont de la détection des sites CpG (S. Chen et al., 2018).

TABLE 6.1 – Exemples d'outils de détection d'informations de méthylation à partir de données traitées au bisulfite. Table adaptée de T. Gong et al. (2022) et Y. Liu et al. (2021).

Pipeline	Type de données	Outil d'alignement	Lien	Référence
<i>BAT</i>	RRBS/WGBS	Module <i>BAT_mapping</i>	<a href="https://github.com/helenebioinf/BAT">https://github.com/helenebioinf/BAT</a>	Kretzmer et al. (2017)
<i>Bicycle</i>	RRBS/WGBS	Bowtie1/2	<a href="http://www.sing-group.org/bicycle/">http://www.sing-group.org/bicycle/</a>	Graña et al. (2018)
<i>BSmooth</i>	RRBS/WGBS	Bowtie2, Merman	<a href="https://github.com/BenLangmead/bsmooth-align">https://github.com/BenLangmead/bsmooth-align</a>	Hansen et al. (2012)
<i>Clusterflow</i>	RRBS/WGBS	Bismark, Bowtie1/2 et autres	<a href="https://github.com/ewels/ClusterFlow">https://github.com/ewels/ClusterFlow</a>	P. Ewels, Krueger, et al. (2016)
<i>GemBS</i>	RRBS/WGBS	Module <i>GEM3</i>	<a href="http://statgen.cnag.cat/gemBS/">http://statgen.cnag.cat/gemBS/</a>	Merkel, Fernández-Callejo, et al. (2019)
<i>MOABS</i>	RRBS/WGBS	BSMAP	<a href="https://github.com/sunnyisgalaxy/moabs">https://github.com/sunnyisgalaxy/moabs</a>	Sun et al. (2014)
<i>Nf-core methylseq</i>	RRBS/WGBS	Bismark, Bwa-meth	<a href="https://github.com/nf-core/methylseq">https://github.com/nf-core/methylseq</a>	P. A. Ewels et al. (2020)
<i>PiGx bsseq</i>	RRBS/WGBS	Bismark	<a href="https://github.com/BIMSBbioinfo/pigx_bsseq">https://github.com/BIMSBbioinfo/pigx_bsseq</a>	Osberg et al. (2018)
<i>SnakePipes</i>	RRBS/WGBS	Bwa-meth	<a href="https://snakepipes.readthedocs.io/en/latest/content/workflows/WGBS.html">https://snakepipes.readthedocs.io/en/latest/content/workflows/WGBS.html</a>	Bhardwaj et al. (2019)
<i>Wg-blimp</i>	RRBS/WGBS	GemBS, Bwa-meth	<a href="https://github.com/MarWoes/wgblimp">https://github.com/MarWoes/wgblimp</a>	Wöste et al. (2020)
<i>DeepSignal</i>	ONT	Bidirectional RNN with LSTM + Inception structure	<a href="https://github.com/MarWoes/wgblimp">https://github.com/MarWoes/wgblimp</a>	Ni et al. (2019)
<i>Megalodon</i>	ONT	Minimap2	<a href="https://github.com/nanoporetech/megalodon">https://github.com/nanoporetech/megalodon</a>	-
<i>Nanopolish</i>	ONT	Bwa-mem	<a href="https://github.com/jts/nanopolish">https://github.com/jts/nanopolish</a>	Simpson et al. (2017)
<i>SignalAlign</i>	ONT	Bwa-mem	<a href="https://github.com/ArtRand/signalAlign">https://github.com/ArtRand/signalAlign</a>	Rand et al. (2017)

L'outil `nf-core/methylseq` v.1.5 (P. A. Ewels et al., 2020) est un pipeline d'analyse bioinformatique utilisé pour étudier des données de méthylation issues du séquençage après traitement de l'ADN au bisulfite de sodium. Le pipeline est construit à l'aide de `Nextflow`, un gestionnaire de workflow permettant d'exécuter des tâches en parallèle et pouvant tourner sur plusieurs infrastructures de calcul différentes. Ces données de séquençage sont sous la forme de fichiers au format `fastq` et contiennent la liste des lectures récupérées à la fin du séquençage RRBS ou WGBS. Ce pipeline se décompose en plusieurs étapes (**Figure 6.1**), dont seules les étapes clés seront détaillées :

1. Contrôle qualité des lectures avec l'outil `FastQC`
2. Nettoyage des séquences avec `TrimGalore`,
3. Alignement des lectures sur le génome de référence avec `Bismark`,
4. Extraction des taux de méthylation,
5. Génération d'un rapport `Bismark`,
6. Contrôle qualité de l'alignement avec `Qualimap BamQC`,
7. Génération d'un rapport html pour l'ensemble des échantillons avec `MultiQC`,
8. Génération d'un rapport d'exécution du pipeline.

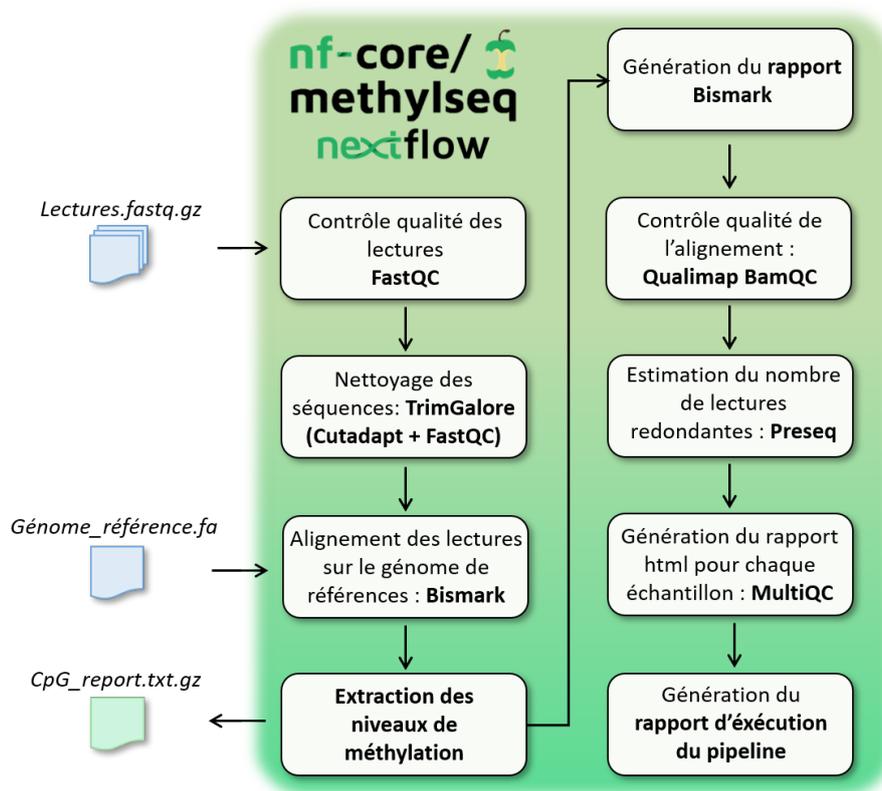


FIGURE 6.1 – Pipeline `nf-core/methylseq`.

Notons que l'outil `nf-core/methylseq` donne également la possibilité d'utiliser un autre logiciel d'alignement, `bwa-meth`, donnant des résultats comparables (W. Gong et al., 2022). Le pipeline d'analyses statistiques construit pour la suite des analyses étant basé sur les formats de sortie de `Bismark`, c'est ce dernier qui a été utilisé pour l'alignement.

### 6.1.1 Contrôle qualité des lectures avec FastQC

Avant d'effectuer l'analyse différentielle des données de méthylation, il faut s'assurer de la bonne qualité des données. Pour cela, l'outil FastQC (Andrews et al., 2010) est couramment utilisé et donne un certain nombre d'informations sur les lectures, notamment :

- la qualité moyenne des bases le long des lectures
- la qualité moyenne par lecture
- vérification de l'homogénéité du séquençage sur la *flowcell*
- le pourcentage de chaque base à chaque position de la lecture
- la distribution de la longueur des lectures
- la mesure du contenu en GC de chaque lecture et comparaison à la distribution attendue
- vérification que les adaptateurs utilisés lors du séquençage soient absents dans les lectures : dans le cas où leur nombre est important, un outil permettant de les retirer sera utilisé (ex : Cutadapt)

### 6.1.2 Nettoyage des séquences avec TrimGalore

Après le contrôle qualité des lectures, la deuxième étape du pipeline consiste à nettoyer les lectures de séquences non désirées. Il peut s'agir de morceaux de séquences ajoutés au cours de la fabrication des librairies, comme par exemple les séquences des adaptateurs et des adénosines de fixation. Pour effectuer cette étape de nettoyage, l'outil TrimGalore (Krueger et al., 2021) est utilisé (utilisant Cutadapt (Martin, 2011) associé à FastQC). Cet outil a été initialement développé pour l'analyse des données RRBS avec une digestion par l'enzyme *MspI*. Les extrémités de séquences non désirées, comme par exemple les adaptateurs, sont éliminées des lectures.

### 6.1.3 Alignement des lectures sur le génome de référence avec Bismark

Suite au nettoyage des lectures, ces dernières sont alignées sur le génome de référence à l'aide de l'outil Bismark (Krueger & Andrews, 2011). Les séquences étant modifiées avant le séquençage par le traitement au bisulfite de sodium, seules les bases A, T et G de l'ADN issues du séquençage, ainsi que les C méthylées, peuvent être alignées avec les bases du génome de référence. Les C non méthylées ont été converties en T et ne s'alignent pas en tant que C. Bismark est adapté pour aligner ce type de données et est conçu pour l'analyse de séquences de petites tailles. Il donne également des informations sur l'état de méthylation des C.

Concernant l'analyse des données WGBS, une étape de *déduplication* est effectuée. Cette étape consiste à supprimer les lectures dupliquées s'alignant au même endroit sur le génome. En effet, lorsque plusieurs lectures identiques s'alignent au même endroit, il est probable qu'elles proviennent du même fragment d'ADN et qu'elles aient été dupliquées durant l'amplification PCR. Ces duplications techniques ne représentent pas de nouvelles informations biologiques, mais représentent un biais potentiel pour la suite des analyses. Ainsi, la suppression de ces séquences dupliquées permet de s'assurer de ne conserver qu'une lecture par fragment d'ADN d'origine et d'améliorer la qualité des données. En RRBS, comme par construction les séquences commencent toutes à un site de restriction *MspI*, cette étape de *déduplication* est en général omise, pour ne pas éliminer de séquences identiques provenant de fragments d'ADN différents.

Dans les deux cas ne sont conservées que les lectures s'alignant de manière unique sur le génome : les régions répétées sont éliminées pour ne pas générer de biais d'analyse.

#### 6.1.4 Extraction des niveaux de méthylation et mise en forme des résultats

Après l'alignement des lectures sur le génome de référence, les informations, telles que les niveaux de méthylation et la couverture au niveau des sites CpG, sont extraites à partir des brins OT et OB par `Bismark`. Ce dernier extrait les sites CpG dans un fichier pour chaque échantillon en considérant toutes les cytosines sur les deux brins et rapporte leur position, brin, contexte trinuécléotidique et état de méthylation. Des rapports `Bismark` sont ensuite générés et un contrôle qualité de l'alignement (statistiques sur la taille et la couverture des lectures) est réalisé par l'outil `Qualimap BamQC` (García-Alcalde et al., 2012). L'estimation du nombre de lectures redondantes est effectuée par `Preseq` (Daley & Smith, 2013). Enfin, un rapport html est généré pour chaque échantillon de l'analyse à l'aide de `MultiQC` (P. Ewels, Magnusson, et al., 2016). L'ensemble des informations d'exécution du pipeline sont également disponibles en fin d'exécution. Après la détection et la récupération des informations de méthylation au niveau des sites CpG de chaque échantillon, la détection d'une différence du taux de méthylation entre des groupes de plusieurs échantillons peut être effectuée.

## 6.2 Outils d'analyse de méthylation différentielle

Pour identifier l'impact de l'environnement sur le méthylome des individus, de nombreux outils bioinformatiques, permettant de comparer les taux de méthylation des CpG entre plusieurs conditions expérimentales, ont été développés (Piao et al., 2021) (**Table 6.3**).

TABLE 6.3 – **Exemples d'outils d'analyse de méthylation différentielle.** Table adaptée de Feng et al. (2014a), Y. Chen et al. (2017) et Piao et al. (2021). Test du rapport de vraisemblance = *Likelihood ratio test*.

Outil	Modèle	Test statistique
Fisher's	-	Test exact de Fisher
BSmooth	Modèle binomial	T-test modifié
methylKit	Modèle logistique	Test de régression logistique
methylSig	Modèle bêta-binomial	Test du rapport de vraisemblance
DSS	Modèle hiérarchique bayésien (comparaison entre 2 groupes)	Test de Wald
DSS	Modèle hiérarchique bayésien (plus de 2 groupes comparés)	Test-F
metilene	Méthode non-paramétrique	Test de Kolmogorov-Smirnov bidimensionnel
RADMeth	Modèle bêta-binomial	Test du rapport du log de vraisemblance
Biseq	Modèle bêta-binomial	Test de Wald
edgeR	Modèle linéaire généralisé binomial négatif	Test du rapport de vraisemblance

Parmi l'ensemble de ces outils, DSS (*Dispersion Shrinkage for Sequencing data*) (Feng et al., 2014a) et edgeR (Robinson et al., 2010) ont été utilisés au cours de cette thèse.

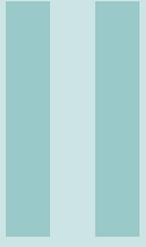
L'outil DSS a été développé pour améliorer les résultats d'analyse d'expression différentielle de gènes, et adapté pour l'analyse de méthylation différentielle à partir de données traitées au bisulfite. La méthylation différentielle est détectée à l'aide de tests statistiques sur chaque site CpG. Les tests sont effectués sur chaque site CpG sous l'hypothèse nulle que les moyennes des deux groupes sont égales. L'estimation du paramètre de dispersion est effectuée par un estimateur de rétrécissement basé sur un modèle hiérarchique bayésien utilisant un modèle bêta-binomial (Feng et al., 2014b). Cela permet de tenir compte en partie des variations biologiques et/ou techniques au sein des groupes d'échantillons. En effet, la distribution binomiale permet de tenir compte de la variation résultant de l'échantillonnage aléatoire des fragments d'ADN lors du séquençage, les lectures pouvant être méthylées ou non méthylées. La distribution bêta modélise la variation du taux de méthylation, ce dernier pouvant varier entre 0 et 100%. Un lissage (*smoothing*) des données peut être effectué ce qui permet de considérer les sites CpG voisins comme des pseudo-réplicats et d'estimer avec une précision raisonnable la dispersion. De plus, pour des données avec une couverture des CpG clairsemée, comme le RRBS, le lissage n'a pas énormément d'impact sur les résultats mais il est tout de même recommandé. Les sites CpG sont susceptibles d'être regroupés localement dans de petites régions génomiques, le lissage pourrait

potentiellement aider à l'estimation de la méthylation. Concernant les données WGBS, le lissage est recommandé pour que les informations des sites CpG proches puissent être combinées et ainsi améliorer l'estimation des niveaux de méthylation (Lister et al., 2009a). Les données issues d'un design expérimental complexe, c'est-à-dire comprenant plusieurs paramètres, sont modélisées par un modèle linéaire généralisé. Les données sont normalisées à l'aide de la fonction *arcsine* au lieu de la fonction *logit* car celle-ci est connue pour mieux gérer les données extrêmes à 0 et 1. L'ajustement du modèle linéaire est effectué par les moindres carrés sur les niveaux de méthylation transformés. Cette méthode consiste à trouver la droite qui minimise la somme des carrés des distances entre chaque point et la droite. Les matrices de variance/covariance des estimations sont dérivées en tenant compte de la distribution et de la transformation des données de comptage. Après la normalisation des données, le modèle linéaire incluant tous les paramètres à tester est ajusté et chacun des paramètres du modèle est testé. Si les termes à tester sont des variables catégorielles à plus de deux niveaux, plusieurs paramètres sont testés en même temps. Il s'agit donc d'un test d'hypothèse composé et un test F est effectué (où la statistique de test suit la loi de Fisher sous l'hypothèse nulle). Si le terme à tester n'a que deux niveaux, un test de Wald est effectué. L'outil DSS permet aussi de comparer les niveaux de méthylation pour chaque site CpG entre deux groupes seulement. Pour chaque site CpG, les niveaux moyens de méthylation sont estimés. Afin d'obtenir une meilleure estimation de la méthylation moyenne, le lissage des données est effectué. Les dispersions de chaque site CpG sont estimées et le test de *Wald* est effectué pour chaque site CpG.

Comme pour DSS, l'outil *edgeR* a initialement été développé pour l'analyse d'expression différentielle de gènes à partir de données RNA-seq (Robinson et al., 2010; Y. Chen et al., 2016). Cet outil peut également être utilisé pour détecter des différentiels de méthylation à partir de données issues de séquençage bisulfite. Cette méthode est basée sur un modèle log-linéaire généralisé binomial négatif appliqué au nombre de lectures pour chaque site CpG. Les modèles de dispersion négative binomiale tiennent compte de la surdispersion des données, c'est-à-dire de la variation supplémentaire par rapport à ce qui serait attendu d'une simple distribution binomiale, permettant ainsi de prendre en compte la variabilité biologique et technique entre les échantillons. *edgeR* est différent des autres outils par sa façon de prendre en compte le nombre de lectures méthylées et non méthylées. En effet, au lieu de les analyser ensemble, il les analyse séparément, sous la forme d'une matrice de comptage binaire, où chaque site CpG est considéré comme soit méthylé (1), soit non méthylé (0) pour chaque échantillon et ne prend donc pas directement en compte la proportion de méthylation dans chaque échantillon. Les modèles de dispersion négative binomiale sont ensuite appliqués aux données binarisées permettant de modéliser la variation entre les échantillons et tester les différences de méthylation entre les groupes. Après avoir examiné la manière dont les données de méthylation traitées au bisulfite varient en fonction de leur moyenne, les chercheurs n'ont pas trouvé de relation entre les deux (Feng et al., 2014a). Par conséquent, les auteurs d'*edgeR* ont choisi de ne pas prendre en compte de tendance particulière entre la variation et la moyenne des données de méthylation bisulfitées. Ainsi, aucune relation spécifique entre la moyenne et la variation des données n'est modélisée ou ajustée pour les données bisulfitées, et le modèle de variation est estimé sans tenir compte de cette relation. L'utilisation d'un modèle linéaire

généralisé est plus flexible pour l'analyse de designs expérimentaux plus complexes et permet de prendre en compte l'hétérogénéité des niveaux de méthylation au sein des échantillons (McCarthy et al., 2012). Un test du rapport de vraisemblance est ensuite appliqué pour un ou plusieurs coefficients du modèle linéaire afin de détecter des différentiels de méthylation. L'hypothèse nulle de ce test suggère que tous les coefficients du modèle à tester sont égaux à 0, c'est-à-dire que les groupes comparés ont la même distribution de comptages et la variation observée dans les niveaux de méthylation entre les groupes étudiés n'est pas statistiquement significative. Cela suppose que toute variation observée dans les données est due au hasard ou à des fluctuations aléatoires. A l'issue de ce test, des p-valeurs significatives permettent de rejeter l'hypothèse nulle et suggèrent que la méthylation diffère de manière significative entre les conditions étudiées.

Les p-valeurs issues des différents tests sont ensuite ajustées, notamment par la méthode Benjamini-Hochberg (BH) (Y. Benjamini, 1995) avec un seuil  $\alpha$  de 5%. Cette méthode d'ajustement est connue pour être plus puissante et permet de contrôler le taux de fausses découvertes (FDR ou taux d'erreur global), c'est-à-dire le taux de faux positifs acceptés à tort parmi les DMC. Les p-valeurs sont augmentées en fonction de leur nombre et du taux de p-valeurs significatives. Etant peu stringente, elle est bien adaptée aux situations où l'on cherche à sélectionner un grand nombre de caractères potentiellement intéressants, notamment dans le cas d'une analyse différentielle. Seuls les CpG ayant un différentiel de méthylation significatif (p-valeur  $< 0.05$  après ajustement) sont considérés comme des DMC.



# Situation du sujet et objectifs de la thèse



Le changement climatique, les contraintes économiques et les préoccupations sociétales pour une agriculture durable vont fortement influencer les systèmes de production animale. De nouveaux modes de production potentiellement en rupture avec les pratiques utilisées jusqu'à présent nécessitent une caractérisation plus fine de la capacité d'adaptation du génome des animaux à des fluctuations environnementales. Cela permettrait notamment une meilleure résistance des animaux à d'éventuelles vagues de chaleur, une utilisation plus efficace d'aliments moins riches mais aussi le maintien d'un niveau de production acceptable en allongeant la carrière des individus. Un individu sera d'autant plus sensible aux effets de l'environnement durant la période embryonnaire, où de nombreux mécanismes nécessaires pour son développement se mettent en place. Des conditions d'élevage inappropriées en début de vie, comme la présence de contaminants chimiques pendant le développement embryonnaire, peuvent donc impacter la santé et le bien-être des animaux. L'influence de l'environnement embryonnaire sur la construction des phénotypes d'un individu passe en partie par des phénomènes épigénétiques, ces derniers étant impliqués dans la régulation de l'expression des gènes et pouvant être transmis à la descendance. La méthylation de l'ADN, induisant l'ajout d'un groupement méthyle au niveau des bases de l'ADN, fait partie des marques épigénétiques les plus étudiées. L'analyse de ces marques épigénétiques est rendue possible grâce au développement de technologies de séquençage haut-débit innovantes ainsi que d'outils bioinformatiques. Cependant, l'influence de ces marques épigénétiques sur l'expression des gènes ainsi que leur transmission à la descendance sur plusieurs générations n'est pas très étudiée, en particulier chez les oiseaux. Une compréhension globale du contrôle génétique et épigénétique de l'adaptation des animaux aux changements environnementaux est donc particulièrement importante pour contribuer à la durabilité de ces filières, notamment chez les volailles. Il est donc important de mieux comprendre les mécanismes épigénétiques résultant de ces effets précoces de l'environnement, ces derniers pouvant être pris en compte dans les schémas de sélection.

La question scientifique majeure de ce projet de thèse est l'étude chez l'oiseau de la transmission entre les générations des effets de l'environnement, notamment chez la caille japonaise. La caille japonaise (*Coturnix japonica*) est une espèce de volaille domestique populaire de plus en plus utilisée comme espèce modèle notamment pour étudier le développement, le comportement et les maladies aviaires. Cette espèce possède de nombreux avantages, tels qu'un temps de génération court (inférieur à 6 mois) par rapport à d'autres espèces, et une petite taille, facilitant son élevage, tous ces avantages étant associés à un faible coût de production. De plus, la bonne qualité du génome de référence de la caille facilite les analyses (Morris et al., 2020). Un modèle permettant de modifier l'environnement de développement sans affecter la santé des embryons a été développé chez la caille, par injection de modificateurs de méthylation dans l'œuf avant l'incubation (Bisphénol A, Décitabine (5-aza-2'-deoxycytidine), Génistéine) (**Figure 6.2**, encadré bleu). Ces traitements ont été effectués sur 4 fonds génétiques différents. Pour l'un des fonds génétiques et l'un des produits (la génistéine, phyto-œstrogène présent notamment dans le soja), deux 'épilignées' de caille (Epi+ et Epi-) ont été produites en utilisant des œuf fécondés provenant de la même population fondatrice. Après 3 générations parallèles de reproduction intra-épilignée, sans autre injection, plusieurs caractères ont été affectés par le traitement des ancêtres, incluant à la fois la reproduction (ex : l'âge au premier œuf retardé de 8 jours) et des

caractères de comportement (**Figure 6.2**, encadrés orange). Même si l'analyse de la méthylation de l'ADN globale n'a pas pu mettre en évidence une différence entre les deux épilignées, les modifications phénotypiques de la G3 peuvent être en partie expliquées par une transmission transgénérationnelle des marques épigénétiques, dont la méthylation de l'ADN. La variabilité génétique entre épilignées ayant été minimisée grâce à un dispositif de croisements 'en miroir', les différences observées sont probablement dues, au moins en partie, à une transmission épigénétique transgénérationnelle. Cette étude pilote montrerait pour la première fois chez l'oiseau - sous réserve d'écarter l'implication de modifications génétiques 'épilignée-spécifiques' - l'existence d'une transmission transgénérationnelle des effets de l'environnement (Leroux et al., 2017). L'analyse de la méthylation de l'ADN globale ne permettant d'étudier que l'information du taux de méthylation global sur l'ensemble du génome, d'autres méthodes permettant d'étudier les informations sur les changements spécifiques au niveau des sites CpG individuels peuvent être utilisées.

L'objectif principal de cette thèse est donc de mieux comprendre la transmission transgénérationnelle des effets de l'environnement pendant le développement embryonnaire chez la caille japonaise. Pour répondre à cette question, un pipeline bioinformatique permettant d'effectuer une analyse différentielle du taux de méthylation des sites CpG entre plusieurs conditions expérimentales a été développé. En plus d'une partie consacrée au développement du pipeline (**Partie III**), les résultats de cette thèse seront présentés en deux grandes parties (**Parties IV et V**). Dans un premier temps, nous avons caractérisé la réponse de l'organisme aux variations de l'environnement embryonnaire par l'analyse de données phénotypiques (caractères de ponte et de croissance) et de méthylation de l'ADN (RRBS et WGBS) issues de cailles ayant reçu directement une injection de molécules dans l'œuf (**Figure 6.2**, encadré bleu, **Partie IV**). Plusieurs tissus ont été étudiés : le foie, le cerveau et le sang. Le profil de méthylation de l'ADN étant spécifique à chaque tissu (Lokk et al., 2014), cela suggère des modes distincts de régulation de l'expression des gènes. De plus, contrairement au sang des mammifères, le sang aviaire contient des globules rouges nucléés (Scanes & Dridi, 2021), ce qui rend ce tissu relativement homogène chez les oiseaux (plus de 90% de l'ADN isolé provient des érythrocytes, (Verhulst et al., 2016)). Le foie, organe clé de l'homéostasie énergétique, et le cerveau, principal lieu de mise en place des comportements, ont été sélectionnés car des études antérieures ont suggéré un impact de l'exposition chimique *in-ovo* sur le poids et les comportements (Leroux et al., 2017). L'ensemble des données de la G0 a été généré dans le cadre du projet Epiques (INRAE, métaprogramme SelGen). Dans une deuxième partie, nous avons également caractérisé la réponse de l'épigénome aux effets environnementaux trois générations après injection par l'analyse de données de méthylation de l'ADN (WGBS et ONT) issues de cailles dont les ancêtres ont reçu une injection de génistéine dans l'œuf (**Figure 6.2**, encadrés orange, **Partie V**). Les données WGBS et ONT de la G3 ont été générées respectivement dans le cadre des projets Caramel (INRAE, département de Génétique Animale) et SeqOccin (Région Occitanie, FEDER). De plus, d'autres données de méthylation de l'ADN (WGBS) issues du projet QuailHeatE (ANR), provenant de cailles japonaises ayant reçu une manipulation thermique (MT) en période embryonnaire, ont également été analysées, afin de mettre en évidence une potentielle transmission des effets de la MT sur la descendance des cailles exposées (Vitorino Carvalho et al., 2023) (**Figure 4.9**, **Partie V**).

L'ensemble de ces analyses a permis de caractériser la réponse épigénétique aux variations de l'environnement embryonnaire et de mettre en évidence les régions du génome affectées par la transmission entre les générations de ces effets afin de décortiquer les mécanismes moléculaires en jeu. Le pipeline bioinformatique *SeqOccin* utilisé pour l'analyse de méthylation différentielle sera présenté (**Partie III**). L'ensemble des résultats et les différentes perspectives relatifs à ce projet de thèse seront ensuite discutés (**Partie VI**).

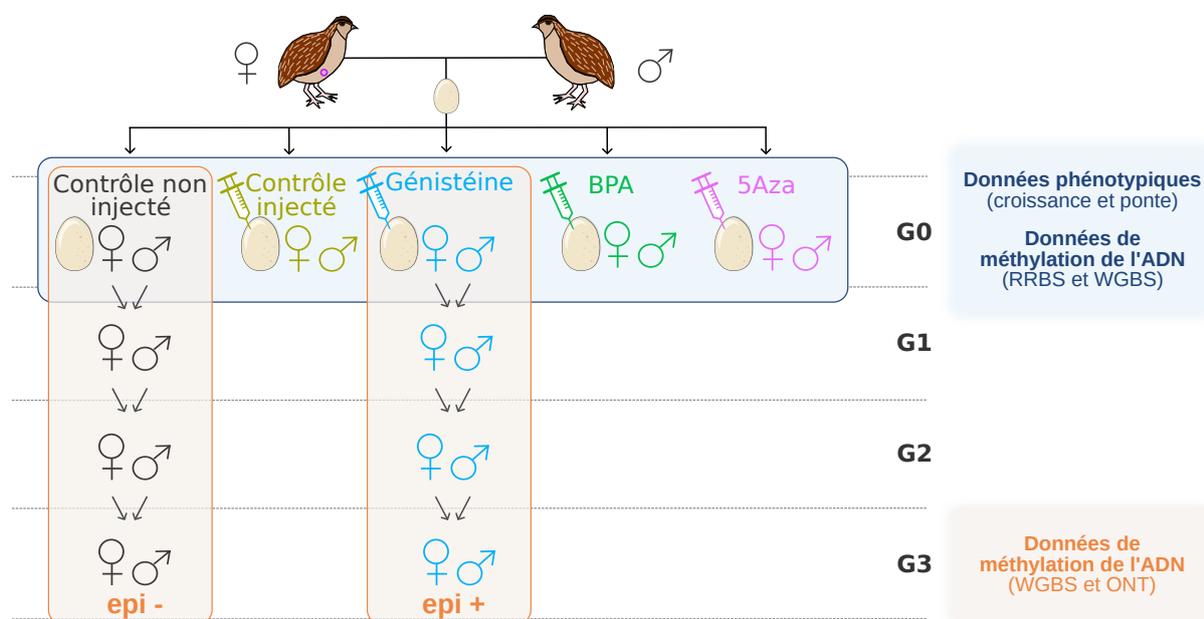


FIGURE 6.2 – **Design expérimental ayant permis d'obtenir les données analysées dans cette thèse.** Les encadrés bleu et orange représentent respectivement les données issues de la génération injectée (G0) et de la troisième génération (G3). BPA = Bisphénol A ; 5Aza = 5-Aza-2'-deoxycytidine.

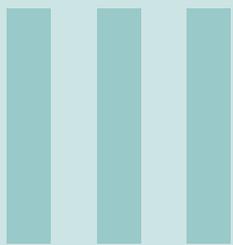
L'ensemble des résultats présentés repose sur un article publié et deux articles rédigés :

**Article 1** (publié) - Cerutti et al. (2022) - *Animal* - **Short communication : Effects of *in-ovo* injection of endocrine disruptors and methyltransferase inhibitor on quail growth and egg-laying performances.** doi : 10.1016/j.animal.2022.100464.

**Article 2** (rédigé pour *Epigenetics*) - Cerutti et al. (2024a) - **Effects of *in-ovo* injection of endocrine disruptors and a methyltransferase inhibitor on the DNA methylation of adult quails.**

**Article 3** (rédigé) - Cerutti et al. (2024b) - **Transgenerational effects of *in-ovo* injection of genistein on the DNA methylation of offspring quails.**





# Développement bioinformatique : pipeline *SeqOccin*

<b>7</b>	<b>Pré-traitement des données</b> .....	<b>113</b>
<b>8</b>	<b>Analyse de méthylation différentielle</b> .....	<b>115</b>
8.1	DSS	
8.2	edgeR	



Le pipeline utilisé pour la détection de cytosines différemment méthylées (DMC) entre groupes expérimentaux est inscrit dans le projet INRAE **SeqOccin**. Ce pipeline est issu du travail de plusieurs développeurs successifs. Je suis la dernière intervenante à avoir amélioré le pipeline, en collaboration avec Paul Terzian (MIAT). Initialement l'outil permettait d'effectuer l'analyse différentielle entre deux conditions seulement. J'ai par la suite activement contribué à l'implémentation du pipeline SeqOccin, ma contribution majeure portant sur l'ajout de l'analyse multifactorielle de DSS ainsi que l'ajout de fonctionnalités d'**edgeR** (**Figure 6.3**). Ce pipeline (codé en R) analyse les sorties de l'outil **Bismark**, implémenté dans le pipeline **nf-core/methylseq**, contenant les informations de méthylation de chaque site CpG et est constitué de deux sous-parties : une étape de pré-traitement (**Chapitre 7**) suivie de l'analyse différentielle (**Chapitre 8**). Cet outil a dans un premier temps été basé sur l'outil R **DSS** (*Dispersion Shrinkage for Sequencing data*, **Chapitre 8.1**) et dans un deuxième temps sur l'outil R **edgeR** (**Chapitre 8.2**). En parallèle du pipeline, des scripts R complémentaires ont été développés pour la visualisation des résultats. Chaque étape du pipeline (pré-traitement et analyse différentielle) peut être lancée séparément à l'aide de scripts *Bash* adaptés pour le cluster de calcul Genotoul (INRAE, *bioinfo.genotoul*, Bioinfo (2018)). L'utilisation de ce cluster de calcul, basé sur le gestionnaire de ressources SLURM (*Simple Linux Utility for Resource Management*, genologin), permet d'avoir des ressources suffisantes pour l'analyse de jeux de données conséquents. La gestion des données a été effectuée à l'aide du package R *data.table* développé pour diminuer le temps de programmation et de calcul. L'ensemble des scripts et le détail des options et commandes d'utilisation du pipeline SeqOccin sont disponibles via le dépôt Github suivant : <https://forgemia.inra.fr/chloe.cerutti/bsseqmethdiffanalysis>.

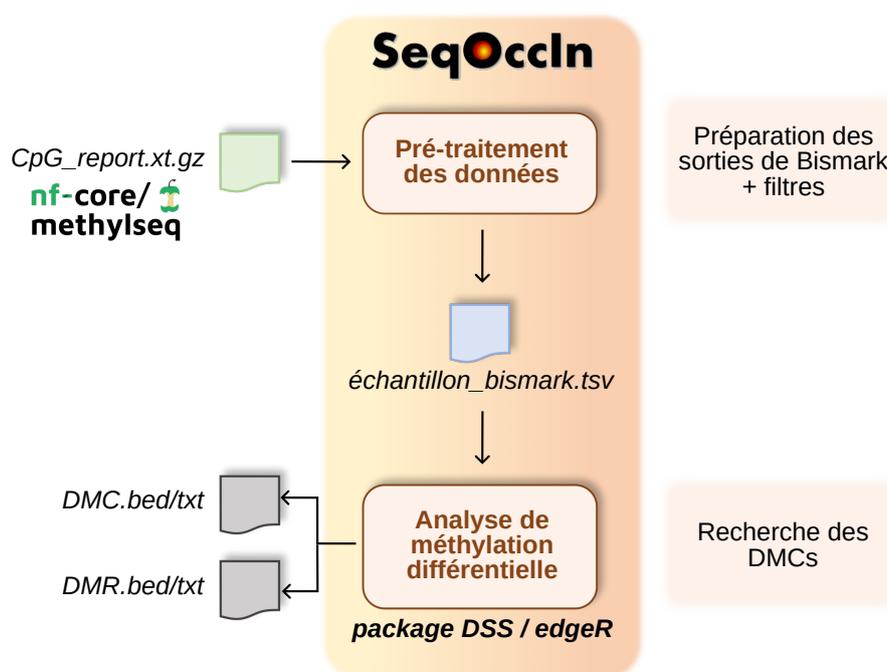


FIGURE 6.3 – Pipeline **SeqOccin**. Les fichiers en sortie de l'outil **Bismark**, contenant les informations de méthylation des sites CpG, subissent un pré-traitement (formatage des données et application de filtres) pour pouvoir être ensuite utilisées pour l'analyse de méthylation différentielle, basée sur les packages **DSS** (*Dispersion Shrinkage for Sequencing data*) et **edgeR**.



## 7. Pré-traitement des données

L'alignement des lectures issues du séquençage sur le génome de référence effectué par le pipeline `nf-core/methylseq` génère des fichiers contenant la liste des sites CpG (*CpG\_report.txt.gz*). Ces fichiers regroupent certaines informations, telles que le chromosome, la position du site CpG, le nombre total de lectures et le nombre de lectures méthylées s'alignant à la position ainsi que le brin (+/-). La première étape du pipeline consiste à mettre en forme et à filtrer les sorties de `Bismark` (*CpG\_report.txt.gz*), contenant la liste des sites CpG, notamment à l'aide d'un fichier VCF contenant la liste de SNP (fichier facultatif). Une autre option facultative peut être utilisée afin de filtrer les CpG dont la moyenne par condition est égale à 0, ces derniers pouvant être considérés comme des SNP potentiels. Avant d'effectuer la détection des DMC, les fichiers de sortie de `Bismark`, contenant l'information de méthylation à chaque site CpG sont utilisés mais le format de ces derniers doit être adapté pour être exploité par `DSS` (Table 7.1) ou `edgeR` (Table 7.2). Cette étape de pré-traitement des données est dotée de trois paramètres de filtration choisis par l'utilisateur : (i) le minimum et (ii) le maximum de couverture acceptée par site CpG ainsi que (iii) la fraction minimale d'échantillons couvrant chaque position (%).

Chromosome	Position	N	X	Strand
1	175489384	39	39	+
1	175491407	14	0	+
1	175513474	65	52	+
1	175513480	65	52	+

TABLE 7.1 – **Extrait d'un fichier en sortie de Bismark mis en forme pour DSS.** Extrait d'un fichier *CpG\_report.txt.gz*. Chaque ligne du fichier correspond à un site CpG et fournit les informations suivantes : le chromosome, la position génomique, le nombre total de lectures s'alignant sur la position (N, nombre de C + nombre de T) et le nombre de lectures montrant une méthylation à cette position (X, nombre de C) pour chaque CpG et le brin.

Chromosome	Start	End	FreqC	Nb <sub>C</sub>	Nb <sub>T</sub>
1	175489384	175489384	100	39	0
1	175491407	175491407	0	0	14
1	175513474	175513474	80	52	13
1	175513480	175513480	80	52	13

TABLE 7.2 – **Extrait d'un fichier en sortie de Bismark mis en forme pour edgeR.** Extrait d'un fichier *CpG\_report.txt.gz*. Chaque ligne du fichier correspond à un site CpG et fournit les informations suivantes : le chromosome, la position start et end, le pourcentage de méthylation ( $\text{freqC} = \frac{\text{Nb}_C}{\text{Nb}_C + \text{Nb}_T} \times 100$ ), le nombre de lectures méthylées ( $\text{nb}_C$ ) et non méthylées ( $\text{nb}_T$ ) pour chaque CpG.



## 8. Analyse de méthylation différentielle

La deuxième étape du pipeline consiste à effectuer l'analyse de méthylation différentielle. Pour cela, les fonctionnalités de l'outil DSS de R ont été implémentées (Feng et al., 2014b) dans le pipeline `Seq0ccin`. Les résultats des analyses avec ce dernier sur les données ONT n'étant pas concluants (**Partie VI**), les analyses ont finalement été effectuées avec l'outil `edgeR` de R (Robinson et al., 2010; Y. Chen et al., 2016) et ce dernier a donc été également implémenté dans le pipeline `Seq0ccin` (**Chapitre 6.2**). L'analyse différentielle des niveaux de méthylation consiste à rechercher des cytosines (C) différentiellement méthylées (DMC) entre les groupes comparés.

### 8.1 DSS

#### 8.1.1 Ajustement des données

Avec l'outil DSS, la méthylation différentielle est détectée à l'aide de tests statistiques sur chaque site CpG sous l'hypothèse nulle que les moyennes des deux groupes sont égales. L'estimation du paramètre de dispersion est effectuée par un estimateur de rétrécissement basé sur un modèle hiérarchique bayésien (Feng et al., 2014b). La méthode d'ajustement des p-valeurs du test (`--correct` : Benjamini-Hochberg ou BH ou fdr, Holm, Hochberg, Hommel, Bonferroni, Benjamini-Yekutieli ou BY) ainsi que le seuil  $\alpha$  (`--alpha`) sont choisis par l'utilisateur. Seuls les CpG ayant un différentiel de méthylation significatif (p-valeur  $< \alpha$  après ajustement) sont considérés comme des DMC. Une option de lissage (`--smoothing`) peut être utilisée, cette dernière permettant aux sites CpG voisins d'être considérés comme des pseudo-réplicats et d'estimer la dispersion avec une précision raisonnable. Les régions différentiellement méthylées peuvent aussi être récupérées (DMR, fonction `callDMR` de DSS). Ces régions correspondent au regroupement de plusieurs DMC, le nombre minimum de DMC (`--dmr.numC`) et la proportion de DMC minimale dans chaque région (`--propDMC`) peuvent être également renseignés par l'utilisateur.

#### 8.1.2 Comparaison entre deux groupes

Le pipeline permet de comparer les niveaux de méthylation pour chaque site CpG entre deux groupes (`--comp`). Cette fonctionnalité a été initialement implémentée par Marjorie Mersch dans le cadre de sa thèse (2018). Pour chaque site CpG, les niveaux moyens de méthylation sont estimés. Afin d'obtenir une meilleure estimation de la méthylation moyenne, le lissage des données peut être effectué (`--smoothing`). Pour chaque site CpG, les dispersions sont estimées et le test de *Wald* est effectué entre les deux groupes. Après ajustement des p-valeurs issues du test précédent par la méthode choisie (seuil  $< \alpha$ ), la liste des DMC est récupérée.

### 8.1.3 Détection de DMC à partir d'un design expérimental complexe

Les données issues d'un design expérimental complexe sont modélisées par un modèle linéaire (`--formula`). Les données sont normalisées à l'aide de la fonction *arcsine* car celle-ci permet de mieux gérer les données extrêmes à 0 et 1. L'ajustement du modèle linéaire est effectué par les moindres carrés sur les niveaux de méthylation transformés. Cette méthode consiste à trouver la droite qui minimise la somme des carrés des distances entre chaque point et la droite. Les matrices de variance/covariance des estimations sont dérivées en tenant compte de la distribution et de la transformation des données de comptage. Après la normalisation des données, le modèle linéaire est ajusté et chacun des paramètres du modèle mentionnés par l'utilisateur (`--term`) est testé. Dans le cas où les termes à tester sont des variables catégorielles à plus de deux niveaux, plusieurs paramètres sont testés en même temps et un test F est effectué (où la statistique de test suit la loi de Fisher sous l'hypothèse nulle). L'utilisateur peut également choisir le groupe de référence pour chacun des termes à tester (`--ref`). En fonction du design expérimental et du modèle linéaire renseigné (**Table 8.1**), une matrice du design expérimental est créée servant de base au modèle pour tester l'effet des différents paramètres (**Table 8.2**). Après ajustement des p-valeurs à l'aide de la méthode choisie, la liste des DMC est récupérée.

Sp	tissu	treat
1	Foie	Contrôle
2	Sang	BPA
3	Sang	Contrôle
4	Sang	Genisteine

TABLE 8.1 – Exemple de design expérimental. Sp = échantillon, treat = traitement, BPA = Bisphénol A.

(Intercept)	tissuSang	treatGenisteine	treatBPA
1	0	0	0
1	1	0	1
1	1	0	0
1	1	1	0

TABLE 8.2 – Exemple de matrice du design expérimental de DSS. Modèle linéaire correspondant :  $\sim$  tissu + treat. Chaque ligne correspond à un échantillon. Les groupes tissuFoie, treatContrôle sont situés dans l'Intercept et correspondent aux groupes de référence pour les analyses de méthylation différentielle.

## 8.2 edgeR

### 8.2.1 Ajustement des données

La méthode de détection du différentiel de méthylation de l'outil **edgeR** est basée sur un modèle log-linéaire généralisé binomial négatif appliqué au nombre de lectures pour chaque site CpG. Comme pour **DSS**, l'utilisateur peut choisir la méthode d'ajustement des p-valeurs du test (`--correct` : Benjamini-Hochberg ou BH ou fdr, Holm, Hochberg, Hommel, Bonferroni, Benjamini-Yekutieli ou BY) ainsi que le seuil  $\alpha$  (`--alpha`). Seuls les CpG ayant un différentiel de méthylation significatif (p-valeur  $< \alpha$  après ajustement) sont considérés comme des DMC.

### 8.2.2 Détection de DMC à partir d'un design expérimental

Contrairement à DSS, edgeR est basé sur l'utilisation de modèle linéaire généralisé (`--formula`). Ce type de modèle est plus flexible pour l'analyse de design expérimentaux plus complexes et permet de prendre en compte l'hétérogénéité des niveaux de méthylation au sein des échantillons (McCarthy et al., 2012). Un test du rapport de vraisemblance est ensuite appliqué pour les termes renseignés du modèle linéaire (`--term`) afin de détecter des différentiels de méthylation. Comme pour DSS, le groupe de référence pour chacun des termes à tester peut être renseigné (`--ref`). En fonction du design expérimental et du modèle linéaire renseigné (**Table 8.1**), une matrice du design expérimental est créée servant de base au modèle pour tester l'effet des différents paramètres (**Table 8.3**). L'hypothèse nulle de ce test suggère que tous les coefficients du modèle à tester sont égaux à 0, les groupes comparés ont donc la même distribution de comptages et la variation observée dans les niveaux de méthylation entre les groupes n'est pas statistiquement significative. Les p-valeurs de ce test sont ajustées avec la méthode BH. A l'issue de ce test, les CpG dont les p-valeurs ajustées sont significatives (`--alpha`, seuil  $< \alpha$ ) correspondent à des DMC. Comme edgeR ne permet pas de récupérer la liste des DMR, ces dernières sont récupérées à l'aide de la fonction `callDMR` (DSS) et les mêmes options ont été implémentées (`--dmr.numC`, `--dmr.propDMC`).

Sp1	Sp2	Sp3	Sp4	(Intercept)	tissuSang	treatGenisteine	treatBPA
1	0	0	0	1	0	0	0
1	0	0	0	0	0	0	0
0	1	0	0	1	1	0	1
0	1	0	0	0	1	0	1
0	0	1	0	1	1	0	0
0	0	1	0	0	1	0	0
0	0	0	1	1	1	1	0
0	0	0	1	0	1	1	0

TABLE 8.3 – **Exemple de matrice du design expérimental d'edgeR.** Modèle linéaire correspondant :  $\sim$  tissu + treat. Les lignes 2 à 2 correspondent à un échantillon. Les groupes tissuFoie et treatContrôle correspondant aux groupes de référence pour les analyses de méthylation différentielle ainsi que l'information méthylée (1) ou non méthylée (0) sont situés dans l'Intercept.



# IV Caractérisation de la réponse de l'organisme aux variations de l'environnement embryonnaire

<b>9</b>	<b>Impacts sur les phénotypes de croissance et de ponte</b> . . . . .	<b>123</b>
9.1	Dispositif expérimental	
9.2	Article 1 : Effets de l'injection <i>in-ovo</i> de perturbateurs endocriniens et d'un inhibiteur de l'ADN méthyltransférase sur la croissance et les performances de ponte chez la caille japonaise.	
<b>10</b>	<b>Impacts sur la méthylation de l'ADN</b> . . . . .	<b>139</b>
10.1	Dispositif expérimental	
10.2	Article 2 : Effets de l'injection <i>in-ovo</i> de perturbateurs endocriniens et d'un inhibiteur de l'ADN méthyltransférase sur la méthylation de l'ADN chez la caille adulte.	



Nous savons que l'environnement prénatal influence le phénotype adulte chez plusieurs espèces, notamment par l'intermédiaire de mécanismes épigénétiques. L'objectif de cette première partie est de caractériser la réponse de l'organisme aux variations de l'environnement embryonnaire. Pour cela, nous avons étudié les effets de l'injection de perturbateurs endocriniens et d'un inhibiteur de l'enzyme ADN méthyltransférase dans l'œuf au cours le développement embryonnaire sur des caractères phénotypiques et sur la méthylation de l'ADN de cailles japonaises adultes (G0).

Mon travail de thèse s'inscrit dans la continuité du projet Epiques (INRAE, métaprogramme SelGen, 2016) (**Partie II**). Dans cette partie, j'ai réalisé les analyses statistiques afin d'identifier les effets potentiels de différentes molécules sur les phénotypes (**Chapitre 9**). J'ai également effectué l'alignement des lectures issues du séquençage bisulfite et récupéré les informations de méthylation de chaque site CpG avec le pipeline `nf-core/methylseq`. J'ai ensuite effectué les analyses de méthylation différentielle et leur interprétation (**Chapitre 10**) à l'aide du pipeline `Seq0ccin` que j'ai contribué à développer (**Partie III**).

Les dispositifs expérimentaux sont détaillés dans les chapitres ci-dessous. L'ensemble des résultats décrits dans cette partie repose sur les articles suivants :

**Article 1** (publié) - Cerutti et al. (2022) - *Animal* - **Short communication : Effects of *in-ovo* injection of endocrine disruptors and methyltransferase inhibitor on quail growth and egg-laying performances.** doi : 10.1016/j.animal.2022.100464.

**Article 2** (rédigé pour *Epigenetics*) - Cerutti et al. (2024a) - **Effects of *in-ovo* injection of endocrine disruptors and a methyltransferase inhibitor on the DNA methylation of adult quails.**



## 9. Impacts sur les phénotypes de croissance et de ponte

### 9.1 Dispositif expérimental

A travers cette étude, nous avons étudié l'effet de l'ajout de perturbateurs endocriniens (Génistéine et Bisphénol A (BPA)) ou d'un inhibiteur de l'enzyme ADN méthyltransférase (5-Aza-2'-déoxycytidine (5Aza)) au cours du développement embryonnaire de cailles japonaises (*Coturnix japonica*) issues de quatre fonds génétiques différents sur leur croissance et leurs performances de ponte. Nous avons utilisé deux lignées de cailles expérimentales et deux lignées de cailles commerciales : la lignée à forte motivation sociale obtenue par sélection divergente sur ce caractère (Mills & Faure, 1991), appelée lignée S+, la lignée DD qui a été sélectionnée pour la production précoce d'œufs (M. J. C. J. F. A. M. Y. Minvielle B.F., 1999) et deux lignées industrielles A et B. Cinquante-quatre couples parentaux par lignée ont été utilisés et les œufs fécondés de chaque couple ont été répartis aléatoirement en cinq groupes : un groupe témoin sans aucune injection, un groupe témoin traité par injection dans l'œuf d'huile de sésame (diluant des 3 molécules actives), et trois groupes traités par injection de génistéine, de BPA ou de 5Aza. Toutes les cailles ont été pesées individuellement à 8, 21, 36 et 78 jours. L'âge de la première ponte et le nombre d'œufs pondus ont été enregistrés (**Figure 9.1**).

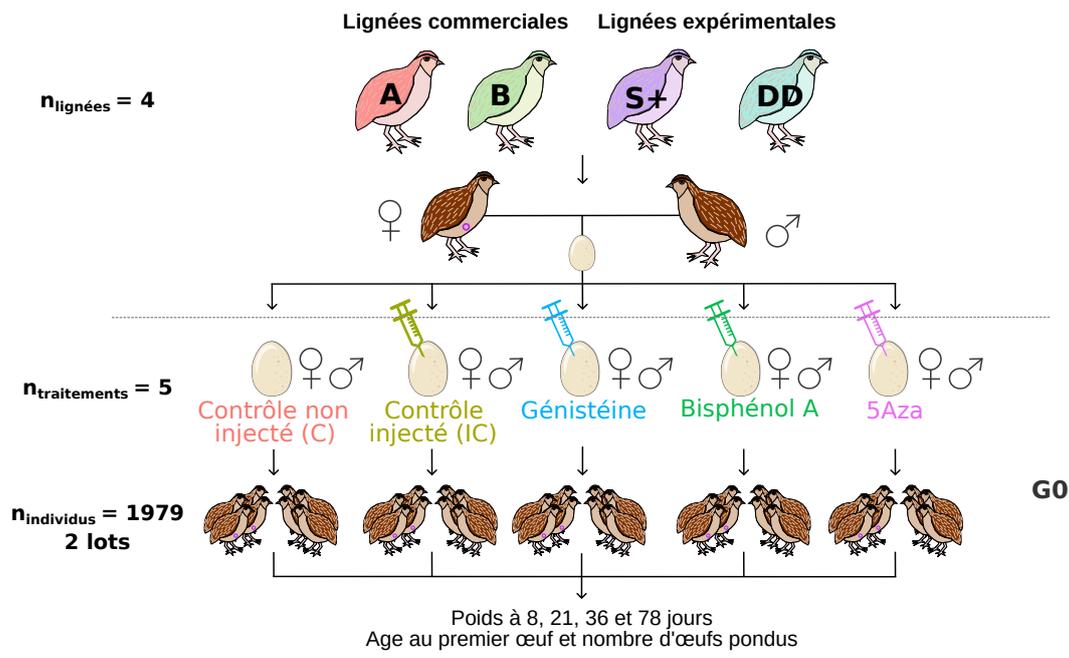


FIGURE 9.1 – Design expérimental de la première génération (données phénotypiques). Données issues de cailles (mâles ♂ et femelles ♀) séparées en 5 groupes : 2 groupes contrôles (non injecté et injecté avec de l'huile de sésame), 3 groupes traités avec du bisphénol A, génistéine et 5Aza. Les cailles étudiées sont issues de 4 lignées : 2 commerciales (A et B) et 2 expérimentales (S+ et DD). 1979 cailles séparées en 2 lots ont été produites.

## 9.2 Article 1 : Effets de l'injection *in-ovo* de perturbateurs endocriniens et d'un inhibiteur de l'ADN méthyltransférase sur la croissance et les performances de ponte chez la caille japonaise.

Afin de caractériser les effets des molécules étudiées sur les phénotypes de ponte et de croissance de la G0 (données quantitatives), nous avons utilisé un modèle linéaire mixte. Ce type de modèle permet d'établir une relation linéaire entre une variable expliquée et plusieurs variables explicatives en incluant des effets aléatoires (Figure 9.2). A l'aide de ce modèle, l'effet des différents facteurs sur le poids à un âge donné (8, 21, 36 et 78 jours) et sur les caractéristiques de ponte des cailles a été estimé (âge au premier œuf et nombre d'œufs pondus). Les variables expliquées correspondent donc aux poids et caractères de ponte et les variables explicatives correspondent au traitement injecté, lignée, lot, sexe et âge. L'impact des différents facteurs sur l'évolution du poids au cours du temps a été évalué à l'aide d'un modèle de régression aléatoire appliqué aux mesures longitudinales du poids. Pour tous les modèles, nous avons inclus l'information généalogique (famille) comme effet aléatoire pour prendre en compte le lien de parenté entre les individus. En raison de la non-linéarité de l'évolution du poids par rapport à l'âge, la partie fixe du modèle de régression aléatoire a pris en compte un polynôme du second ordre (âge<sup>2</sup> ou âge quadratique) dans le modèle du poids global. Pour tenir compte des mesures de poids répétées pour chaque individu, un effet aléatoire identifiant (ID) inclus dans la famille et lié à l'âge, a été spécifié pour l'analyse du poids global. Chaque modèle comprenait des effets fixes tels que la lignée, le lot, le traitement injecté, le sexe et l'âge le cas échéant, ainsi que leur interaction. Des modèles mixtes linéaires ont également été réalisés pour le poids à chaque âge, similaires à ceux utilisés pour l'analyse des caractères

de ponte. L'injection seule d'huile de sésame et/ou d'éthanol pouvant avoir un effet par rapport au contrôle non injecté, la comparaison ANOVA entre chaque groupe a été effectuée en utilisant le contrôle injecté (IC) comme groupe de référence (IC vs contrôle et IC vs traitements). Une sélection de modèles par étapes selon le critère d'information d'Akaike (AIC) a été réalisée et le modèle associé à la valeur AIC la plus faible a été sélectionné. Le test de Tukey de la différence franchement significative (HSD - *Honestly Significant Difference*) a été utilisé pour effectuer une comparaison par paire des groupes de traitement et tester les différences significatives entre leurs moyennes en utilisant un seuil de significativité pour les p-valeurs ajustées de  $\alpha = 5\%$ .

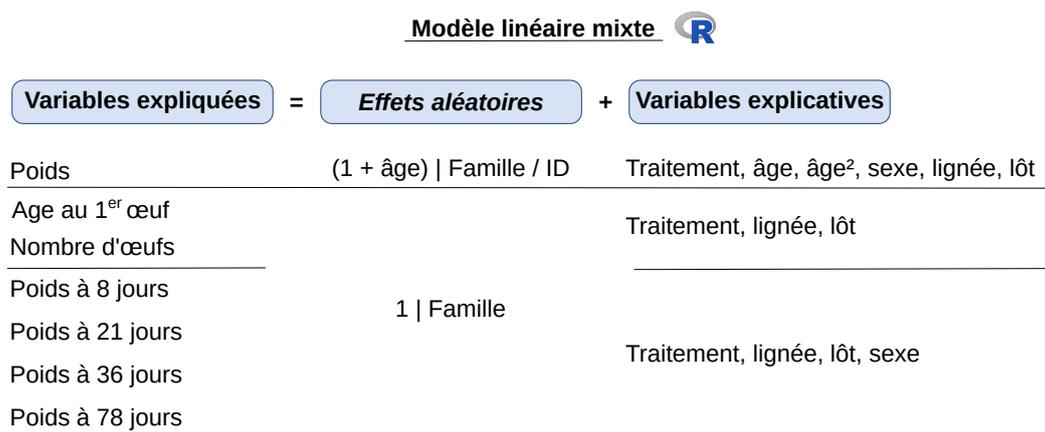


FIGURE 9.2 – Modèles linéaires utilisés pour estimer l'effet des molécules sur les phénotypes.

Cette étude a permis de mettre en évidence que l'exposition à des molécules pendant le développement peut affecter la construction du phénotype. Les analyses ont pu mettre en évidence un effet significatif du traitement sur le poids global des cailles. Les effets de la génistéine et du 5Aza sont significatifs sur le poids corporel jusqu'à 36 jours d'âge, et significatifs sur la croissance globale. Le BPA affecte significativement le poids à 8 jours d'âge. La différence observée entre les IC et C peut démontrer une influence d'une intervention sur l'œuf à un stade précoce du développement. Les molécules injectées ont un effet sur le poids des cailles, les individus injectés ayant un poids plus faible que les individus non injectés. Ces perturbateurs endocriniens et l'inhibiteur de l'ADN méthyltransférase n'affectent pas les caractéristiques de la ponte. Toutefois, les effets détectés ne semblent pas dépendre du fond génétique testé. Ces composés peuvent se trouver dans l'alimentation des animaux (génistéine) ou comme contaminants chimiques (BPA) et leur effet sur les phénotypes, tant des animaux exposés que de leur descendance, doit être analysé plus en détail. Pour la suite des analyses, nous nous intéressons à l'effet de l'injection de ces traitements sur la méthylation de l'ADN de l'embryon (G0).

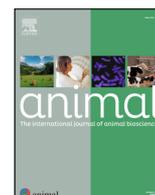
Le script utilisé pour les analyses est disponible via le lien suivant : [https://github.com/ccerutti88/Quail\\_Phenotypes.git](https://github.com/ccerutti88/Quail_Phenotypes.git).



Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

# Animal

## The international journal of animal biosciences



### Short communication: Effects of *in-ovo* injection of endocrine disruptors and methyltransferase inhibitor on quail growth and egg-laying performances



C. Cerutti<sup>a</sup>, S. Leroux<sup>a</sup>, D. Gourichon<sup>b</sup>, Y. Labrune<sup>a</sup>, I. David<sup>a</sup>, T. Zerjal<sup>c</sup>, V. Coustham<sup>d,e</sup>, G. Devailly<sup>a</sup>, F. Pitel<sup>a,\*</sup>

<sup>a</sup>GenPhySE, Université de Toulouse, INRAE, ENVT, 31320 Castanet Tolosan, France

<sup>b</sup>INRAE, PEAT, 37380 Nouzilly, France

<sup>c</sup>AgroParisTech, INRAE, Université Paris-Saclay, GABI, 78350 Jouy-en-Josas, France

<sup>d</sup>INRAE, Université de Tours, BOA, 37380 Nouzilly, France

<sup>e</sup>Université de Pau et des Pays de l'Adour, E2S UPPA, INRAE, NUMEA, Saint-Pée-sur-Nivelle 64310, France

#### ARTICLE INFO

##### Article history:

Received 27 July 2021

Revised 11 January 2022

Accepted 13 January 2022

Available online 15 February 2022

##### Keywords:

5-Aza-2'-deoxycytidine

Bisphenol A

Coturnix Japonica

Epigenetics

Genistein

#### ABSTRACT

Early experiences, including prenatal environment, are known to influence a wide variety of mechanisms involved in the phenotype elaboration. We investigated the effect of the addition of endocrine disruptors or of a methyltransferase inhibitor during the embryonic development of quails from different genetic backgrounds (four different quail lines) on their growth and egg-laying performances. Fifty-four pairs of parents per line were used and fertilised eggs from each pair were randomly divided into five groups: a control group without any injection, an injected control group treated by injection into the egg of sesame oil, and three groups treated by injection of Genistein, Bisphenol A or 5-Aza-2'-deoxycytidine. All quails were individually weighed at 8, 21, 36 and 78 days. The age at first egg laid and the number of eggs laid were recorded. These analyses revealed a significant impact of the treatment on growth but no influence on the egg-laying traits. All three molecules significantly affected at least one of the analysed growth traits. In conclusion, we showed that the injection of endocrine disruptors or DNA methyltransferase inhibitor into the egg had significant effects on quail development; these effects were specific to each treatment, but no interaction between line and treatment was observed.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of The Animal Consortium. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

#### Implications

Early experiences, including exposure to biologically active substances during embryonic development, can affect a wide variety of mechanisms involved in the construction of the individual. Inappropriate early-life rearing conditions may lead to the development of unfavourable consequences on animal health and welfare. This study demonstrates the significant effects on adult phenotypes of endocrine disruptors and methyltransferase inhibitors during embryonic development in quail. It paves the way to study the extent to which the individual early environment can influence future generations, notably in the case of contamination by endocrine disruptors into the maternal diet and accumulation into the egg.

#### Introduction

The prenatal environment is known to influence the adult phenotype in several species, in part through epigenetic mechanisms. Epigenetic phenomena are “mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence” (Riggs et al., 1996). These mechanisms take part in the regulation of gene expression and can induce modifications of phenotype. There are different types of epigenetic phenomena including histone modifications, changes in chromatin structure, effects of non-coding RNAs, and DNA methylation. Modifications in the embryo environment, such as the presence of chemical contaminants, can induce epigenetic changes during the development of somatic cells and eventually of germ cells affecting the offspring of the next generations (Skinner, 2011). Here, we propose to analyse in quail the impact of *in-ovo* exposure to endocrine disruptors, Genistein and Bisphenol A (BPA), and of a DNA methyltransferase

\* Corresponding author.

E-mail address: [frederique.pitel@inrae.fr](mailto:frederique.pitel@inrae.fr) (F. Pitel).

inhibitor, 5-Aza-2'-deoxycytidine (**5Aza** or Decitabine) on transmissible traits, in different genetic backgrounds.

Several molecules have been shown to induce epigenetic modifications in many vertebrate species such as humans, mice, rats and zebrafish (Silva et al., 2019; Qin et al., 2021). Genistein is a phytoestrogen notably found in soy, known to have protective effects against metabolic diseases or cancers that has been shown to induce epigenetic changes (Silva et al., 2019). Bisphenol A is a widespread environmental contaminant that has notably been demonstrated to promote epigenetic modifications (Qin et al., 2021). The DNA methyltransferase inhibitor 5Aza is a drug that has long been known to be beneficial in the treatment of leukaemia (Mompalmer et al., 1997). For this study, four different quail lines were used to analyse the putative interaction between the injected treatment and the genetic background. This study hypothesised that changing the embryo environment by injecting molecules into the egg may trigger phenotypic alterations later in life, putatively depending on the molecule or the genetic background.

## Material and methods

### Experimental design

For this study, we used two experimental and two commercial quail lines: the high social reinstatement quail line obtained by divergent selection on social motivation (Mills and Faure, 1991), referred to as the "S+" line, the DD line that was selected for early egg production (Minvielle et al., 1999) and the line A and the line B, two parental lines kindly provided by commercial breeders. Fertilised eggs produced from 54 pairs of parents per line were randomly divided into five experimental groups: four received an injection before incubation and one received no injection and served as control. The injection was performed the day the freshly laid eggs were incubated, directly into the egg yolk. Of the four groups that received an injection, one was treated with 50 µl of delivery vehicle (10% ethanol/90% sesame oil, Acros Organics), and three others were treated with an injection of 50 µl of one of the tested compounds dissolved in 100% ethanol, and then diluted ten-fold in sesame oil. The dose used per egg was of 500 µg of genistein (Molekula), or of 200 µg of BPA (Sigma-Aldrich), or of 34 µg (50 µM) of 5Aza (Sigma-Aldrich). Doses were established according to previous studies in quail (Halldin et al., 2001; Leroux et al., 2017), or rodents (Zhang et al., 2013) in order to maximise the effects while keeping the hatching rate as high as possible. Untreated eggs and eggs injected with carrier fluid only were used as controls to analyse the putative impact of the *in-ovo* injection on the phenotypes measured. After hatching, we obtained 1 979 live chicks issued from two distinct batches, representing approximately 50 individuals per sex, line and treatment. All birds were weighted at 8, 21, 36 and 78 days of age. The age at first egg laid and the total number of eggs produced over 30 days of experiment were also recorded (Supplementary Material S1). The record of the egg number is the same period for each line: 30 days from the date of the first egg laying. Growth and egg-laying traits were measured to determine a potential phenotypic effect of the chemical agents.

### Statistical analysis

All the analyses were carried out using R (v.4.0.2) (Supplementary Material S2). Effect of the different factors on BW at a given age (8, 21, 36 and 78 days) and on laying traits of quails was estimated using linear mixed models. In addition, impact of the different factors on the trajectory of BW over time was evaluated using a random regression (RR) model applied to longitudinal BW mea-

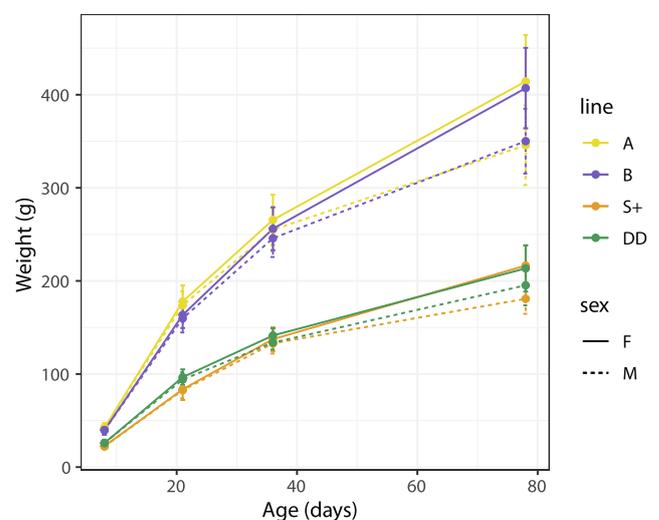
surements. For all models, we included the pedigree information (family) as a random effect to take into account the kinship between individuals. According to the non-linearity of the weight evolution relative to the age (Fig. 1), the fixed part of the RR model considered a second-order polynomial (age<sup>2</sup>) and the random part consisted in first-order polynomials for the family and animal random effects. Each model included fixed effects such as line, batch, injected treatment, sex and age when appropriate, and their interaction. The *lme* function from R MASS package (v.3.1-149) was used to apply the mixed models (Supplementary Table S1).

The variance was analysed for each trait using the ANOVA function from R car package (v.3.0-10) according to the given linear model. As the injection of sesame oil alone may have an effect compared to the non-injected control, the ANOVA comparison between each group was performed using the injected control (IC) as the reference group (IC vs control and IC vs treatments). The *stepAIC* function from R MASS package (v.7.3-54) was used to perform stepwise model selection by Akaike Information Criteria (AIC), and the model associated with lower AIC value was selected (Supplementary Table S1). The posthoc Tukey's Honestly Significant Difference statistical test (*emmeans* function from R *emmeans* package v.1.6.1) was used to perform pairwise comparison of the treatment groups and test for significant differences between their means using a significance threshold for adjusted *P*-values of  $\alpha = 5\%$ .

## Results

### Line and sex effects

Analyses of growth and egg-laying phenotyping data of quails highlighted a weight difference between lines and sexes of individuals, notably at 78 days of age (Fig. 1, Supplementary Figure S1, Supplementary Table S2). Line had a significant effect on all measured traits (Table 1). The two experimental lines S+ and DD had a lower weight than the two commercial lines A and B at all measured ages (Fig. 1, Supplementary Figure S1.a, Supplementary Table S2). The S+ line laid the first egg significantly later than the other three lines, and its total egg number was significantly lower than the other lines for the same time period (Supplementary Figure S1.a, Supplementary Table S2). Sex had also a significant effect



**Fig. 1.** Plot indicating the weight of quails according to the age, line and sex. The yellow, violet, orange and green straight lines show respectively the quails from A, B, S+, DD lines. The solid lines and dotted lines refer to female (F) and male (M), respectively.

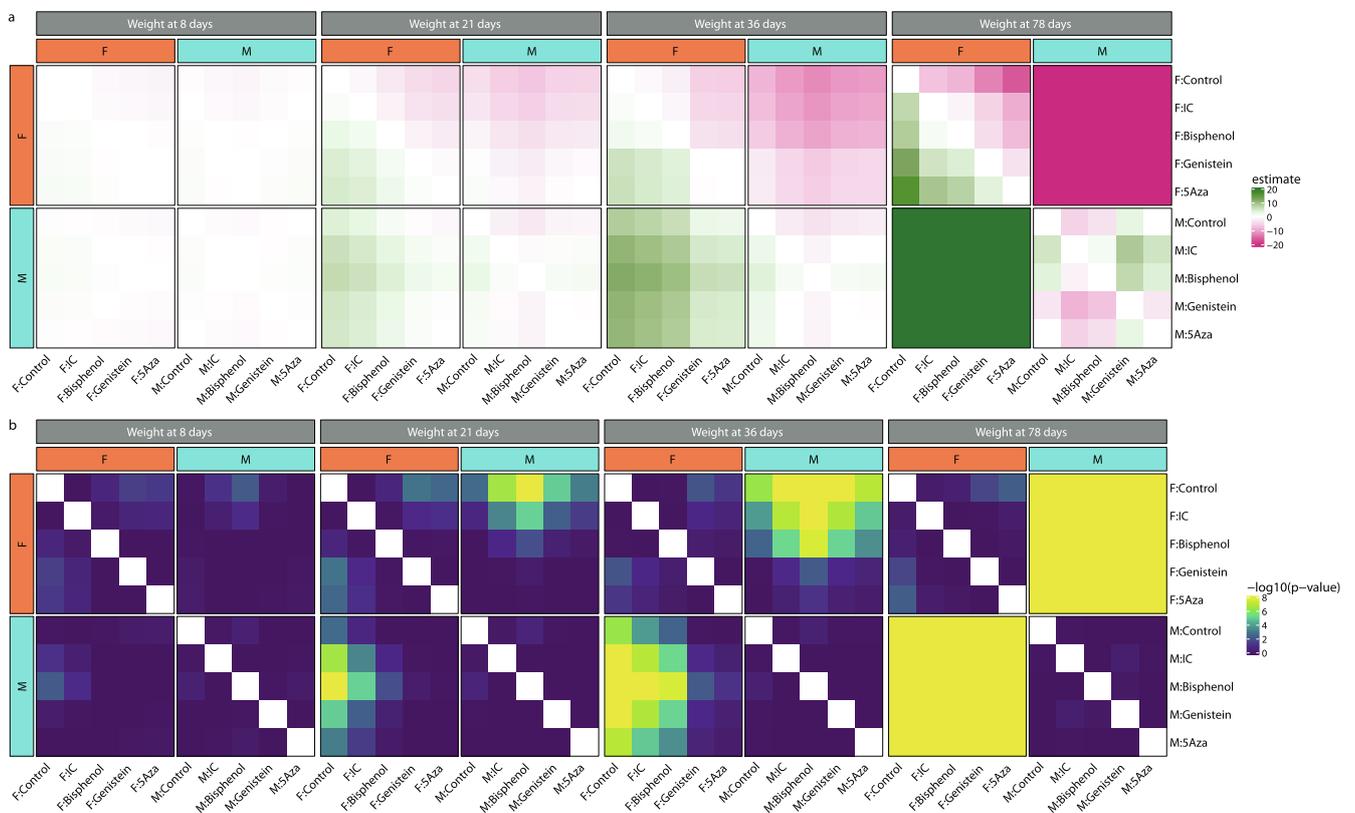
**Table 1**

Table of ANOVA *P*-values for growth and egg-laying traits of quails (a, top) and comparison of treatments impact on the weight (b, bottom). Results were obtained by combining the following methods: *aov* + *stepAIC*. Only values less than 0.05 are considered as significant.

	Growth and egg-laying traits			
	Weight	Age 1st egg	Egg number	
line	<2.2e-16*	<2e-16*	<2.2e-16*	
batch	1.23e-04*	1.05e-02*	4.78e-10*	
sex	1.06e-01	NA	NA	
age	<2.2e-16*	NA	NA	
age <sup>2</sup>	<2.2e-16*	NA	NA	
treat	3.48e-03*	6.39e-02	NR	
age:sex	4.72e-02*	NA	NA	
age:line	<2.2e-16*	NA	NA	
age:batch	1.15e-02*	NA	NA	
treat:age <sup>2</sup>	1.64e-02*	NA	NA	
sex:age <sup>2</sup>	6.21e-13*	NA	NA	
line:age <sup>2</sup>	<2.2e-16*	NA	NA	
batch:age <sup>2</sup>	8.36e-04*	NA	NA	
line:batch	NR	5.91e-02	NR	
age:sex:line	1.89e-03*	NA	NA	
treat:sex:age <sup>2</sup>	4.74e-02*	NA	NA	
sex:line:age <sup>2</sup>	<2.2e-16*	NA	NA	
sex:batch:age <sup>2</sup>	4.12e-02*	NA	NA	
	Treatment impact on the weight (vs IC)			
	Con	BPA	Gen	5Aza
treat	6.06e-01	2.02e-01	1.08e-02*	3.34e-02*
treat:age <sup>2</sup>	2.58e-01	7.22e-01	3.35e-01	2.19e-02*
treat:age <sup>2</sup> :sex	8.76e-01	5.34e-01	7.69e-02	1.13e-02*

Abbreviations: *aov* = R function for analysis of variance; *stepAIC* = R function performing stepwise model selection by Akaike Information Criterion; treat = treatment; NA = untested effect; NR = effect not retained by *stepAIC*; IC = injected control (sesame oil); Con = non-injected control; BPA = Bisphenol A; Gen = Genistein; 5Aza = 5-Aza-2'-deoxycytidine.

\* Value significantly different at  $P < 0.05$ .



**Fig. 2.** Heatmaps representing tables of Tukey HSD contrast estimate values (a) and  $-\log_{10}(P\text{-value})$  (b) for each pairwise treatment comparison according to the sex at each measurement age of quails. The estimate value refers to the column condition minus the line condition. The colour gradient (blue to green) is proportional to the significance of Tukey test: blue and green for respectively low and high Tukey test *P*-value. The diverging colour gradient refers to a loss (pink) or a gain (green) of weight. Abbreviations: HSD = honestly significant difference; F = female; M = male; IC = injected control (sesame oil); 5Aza = 5-Aza-2'-deoxycytidine.

on BW at all ages (Supplementary Table S3, Supplementary Figure S1.b), females being generally heavier than males (Fig. 1, Supplementary Figure S1.b). As illustrated by the evolution of the estimate values between males and females in Tukey Honestly Significant Difference (HSD) tests (Fig. 2.a), females were heavier than males in all lines, the difference between sexes increasing along life.

#### Treatment effect

Treatment had a significant effect on BW and a significant interaction between treatment, sex and age<sup>2</sup> effect was observed (Table 1.a), reflecting different weight gain dynamics. The analysis according to the age revealed a significant treatment effect at 8, 21 and 36 days (Supplementary Table S3). We observed a significant interaction between sex and treatment for BW at 8 and 21 days. No treatment effect was detected for egg-laying traits. We did not observe any line × treatment interaction (Table 1.a, Supplementary Table S3).

#### Individual treatment comparison

The groups' comparison showed that the injection of sesame oil had no significant effect on overall growth (Table 1.b), but sesame oil seemed to decrease the weight at 78 days (Supplementary Table S3, Fig. 2.a). A treatment effect on weights at 8, 21 and 36 days was observed for Genistein and 5Aza when compared to the IC groups, whereas BPA had an effect only at 8 days but no influence was identified on the growth (Table 1.b, Supplementary Table S3).

According to the multiple comparison analyses (Tukey HSD), individuals injected with a chemical agent had a lower weight compared to the non-injected controls (Fig. 2.a, Supplementary Table S4). Genistein and 5Aza groups had a significant lower weight compared to non-injected control group for females at 8, 21, 36, 78 days and a trend is observed compared to the IC group only at 21 days whereas no mean difference was observed for the BPA group (Fig. 2.b, Supplementary Table S5). We observed no significant effect of the treatment on the male growth.

#### Discussion

In this study, we compared in different quail lines the effects on growth and egg-laying related phenotypes of three compounds expected to modify the embryonic environment after injection into the egg. These lines were of different origins and as expected, the commercial lines (A, B) were heavier than the experimental lines (S+, DD). The observed increase in the weight difference between sexes and between S+/DD and A/B lines over time can explain the line by sex effect observed for BW at 36 and 78 days (Supplementary Table S3). The observed difference between the injected and non-injected controls may demonstrate an influence of an intervention on the egg at an early stage of development.

The effects of treatments were highly significant on BW until 36 days of age, and significant on overall growth. The less significant effect on BW observed after 36 days of age is in accordance with previous studies that used BPA in quail (Hallidin et al., 2001) or genistein in chicken, for which the effect on BW gain was significant at 21 days of age but not at 42 days of age (Lv et al., 2018b). More specifically, only 5Aza and Genistein impacted growth compared to the injected control. Previous studies have shown an increased BW gain in 21-day-old broiler chickens after a low-dose genistein supplementation into the diet of their mother, which results in genistein presence into the egg (Lv et al., 2018b), but a decreased embryo weight after high-dose genistein

exposure by feed supplementation of the mother (Lv et al., 2018a). No treatment effect on egg-laying traits was observed.

#### Conclusion

Our study showed significant effects of injection of endocrine disruptors or methylation modifiers into the egg on adult phenotypes. These effects do not seem to depend on the genetic background tested. These compounds may be found in animal diet (genistein) or as chemical contaminants (BPA) and their effect on phenotypes, both from the exposed animals and their progeny, need to be further analysed. Mechanisms of action may involve epigenetic regulations, notably through DNA methylation alterations, and need to be further analysed, including the analysis of putative non-genetic inheritance effects. Improved knowledge of genome-to-phenome relationship, accounting for genetic and non-genetic mechanisms, should ultimately improve livestock production systems.

#### Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.animal.2022.100464>.

#### Ethics approval

Animals were bred at the INRAE Poultry Experimental Facility (PEAT) in Nouzilly, France (<https://doi.org/10.15454/1.5572326250887292E12>) in accordance with European Union Guidelines for animal care with an approval by the local ethical committee in animal experimentation (Val de Loire) and the French Ministry of Higher Education and Scientific Research (authorization n°4609-2016032115196879). Animals were maintained under standard breeding conditions and were subjected to minimal disturbance.

#### Data and model availability statement

All data and scripts used during this study are public. They are available on Github ([https://github.com/ccerutti88/Quail\\_Phenotypes.git](https://github.com/ccerutti88/Quail_Phenotypes.git)).

#### Author ORCIDs

**C. Cerutti:** <https://orcid.org/0000-0002-3207-3461>  
**S. Leroux:** <https://orcid.org/0000-0003-2840-5109>  
**D. Gourichon:** <https://orcid.org/0000-0002-4631-0149>  
**Y. Labrune:** <https://orcid.org/0000-0001-5099-7580>  
**I. David:** <https://orcid.org/0000-0002-2514-6693>  
**T. Zerjal:** <https://orcid.org/0000-0002-0666-5420>  
**V. Coustham:** <https://orcid.org/0000-0002-5399-2723>  
**G. Devailly:** <https://orcid.org/0000-0001-8878-9357>  
**F. Pitel:** <https://orcid.org/0000-0002-1477-7633>

#### Author contributions

**Chloé Cerutti:** Methodology, Formal analysis, Data Curation, Visualization, Writing - Original Draft, Writing - Review & Editing; **Sophie Leroux:** Methodology, Validation, Investigation, Writing - Review & Editing; **David Gourichon:** Methodology, Resources, Investigation, Funding acquisition, Writing - Review & Editing; **Yann Labrune:** Methodology, Software, Writing - Review & Editing; **Ingrid David:** Methodology, Software, Writing - Review & Editing; **Tatiana Zerjal:** Conceptualization, Resources, Funding acquisition, Writing - Review & Editing; **Vincent Coustham:** Con-

C. Cerutti, S. Leroux, D. Gourichon et al.

Animal 16 (2022) 100464

ceptualization, Methodology, Resources, Investigation, Funding acquisition, Writing - Review & Editing; **Guillaume Devailly**: Methodology, Visualization, Supervision, Writing - Original Draft, Writing - Review & Editing; **Frédérique Pitel**: Conceptualization, Methodology, Investigation, Supervision, Funding acquisition, Writing - Original Draft, Writing - Review & Editing.

#### Declaration of interest

None.

#### Acknowledgements

The authors want to acknowledge PEAT staff for their conscientious animal care and phenotyping, and Caillor (<https://caillor-genetics.com/home>) and Caille Robin (<https://www.cailles-robin.fr/>) breeders for providing animals. We are grateful to C. Leterrier for providing the S+ experimental line and helpful discussions.

#### Financial support statement

This work was supported by the EPIQUES grant (INRAE, SelGen Metaprogramme). C. Cerutti is supported by a PhD scholarship from Région Occitanie and INRAE Animal Genetics Division.

#### References

- Haldin, K., Berg, C., Bergman, A., Brandt, I., Brunstrom, B., 2001. Distribution of bisphenol A and tetrabromobisphenol A in quail eggs, embryos and laying birds and studies on reproduction variables in adults following in ovo exposure. *Archives of Toxicology* 75, 597–603. <https://doi.org/10.1007/s002040100277>.
- Leroux, S., Gourichon, D., Leterrier, C., Labrune, Y., Coustham, V., Riviere, S., Zerjal, T., Coville, J.L., Morisson, M., Minvielle, F., Pitel, F., 2017. Embryonic environment and transgenerational effects in quail. *Genetics, Selection, Evolution* 49, 14. <https://doi.org/10.1186/s12711-017-0292-7>.
- Lv, Z., Fan, H., Zhang, B., Ning, C., Xing, K., Guo, Y., 2018a. Dietary genistein supplementation in laying broiler breeder hens alters the development and metabolism of offspring embryos as revealed by hepatic transcriptome analysis. *FASEB Journal* 32, 4214–4228. <https://doi.org/10.1096/fj.201701457R>.
- Lv, Z., Fan, H., Zhang, B., Xing, K., Guo, Y., 2018b. Dietary genistein supplementation for breeders and their offspring improves the growth performance and immune function of broilers. *Scientific Reports* 8, 5161. <https://doi.org/10.1038/s41598-018-23530-z>.
- Mills, A.D., Faure, J.M., 1991. Divergent selection for duration of tonic immobility and social reinstatement behavior in Japanese quail (*Coturnix coturnix japonica*) chicks. *Journal of Comparative Psychology* 105, 25–38. <https://doi.org/10.1037/0735-7036.105.1.25>.
- Minvielle, B.F., Monvoisin, J.L., Costa, J., Frenot, A., Maeda, Y., 1999. Changes in heterosis under within-line selection or reciprocal recurrent selection: an experiment on early egg production in Japanese quail. *Journal of Animal Breeding and Genetics* 116, 363–377. <https://doi.org/10.1046/j.1439-0388.1999.00218.x>.
- Momparler, R.L., Cote, S., Eliopoulos, N., 1997. Pharmacological approach for optimization of the dose schedule of 5-Aza-2'-deoxycytidine (Decitabine) for the therapy of leukemia. *Leukemia* 11, 175–180. <https://doi.org/10.1038/sj.leu.2400550>.
- Qin, T., Zhang, X., Guo, T., Yang, T., Gao, Y., Hao, W., Xiao, X., 2021. Epigenetic Alteration Shaped by the Environmental Chemical Bisphenol A. *Frontiers in Genetics* 11. <https://doi.org/10.3389/fgene.2020.618966> 618966.
- Riggs, A.D., Martienssen, R.A., Russo, V.E.A., 1996. Introduction. In: Russo, E., Martienssen, R., Riggs, A.D. (Eds.), *Epigenetic mechanisms of gene regulation*. Cold Spring Harbor Laboratory Press, New York, NY, USA, pp. 1–4. 10.1101/0.1–4.
- Silva, L., Pinheiro-Castro, N., Novaes, G.M., Pascoal, G.F.L., Ong, T.P., 2019. Bioactive food compounds, epigenetics and chronic disease prevention: Focus on early-life interventions with polyphenols. *Food Research International* 125. <https://doi.org/10.1016/j.foodres.2019.108646> 108646.
- Skinner, M.K., 2011. Environmental epigenetic transgenerational inheritance and somatic epigenetic mitotic stability. *Epigenetics* 6, 838–842. <https://doi.org/10.4161/epi.6.7.16537>.
- Zhang, Y., Sun, J., Gao, Y., Jin, L., Xu, Y., Lian, H., Sun, Y., Sun, Y., Liu, J., Fan, R., Zhang, T., He, Z., 2013. A carrier-mediated prodrug approach to improve the oral absorption of antileukemic drug decitabine. *Molecular Pharmaceutics* 10, 3195–3202. <https://doi.org/10.1021/mp400233x>.

## Short communication: Effects of *in-ovo* injection of endocrine disruptors and methyltransferase inhibitor on quail growth and egg-laying performances

C. Cerutti<sup>a</sup>, S. Leroux<sup>a</sup>, D. Gourichon<sup>b</sup>, Y. Labrune<sup>a</sup>, I. David<sup>a</sup>, T. Zerjal<sup>c</sup>, V. Coustham<sup>d,e</sup>, G. Devailly<sup>a</sup>, F. Pitel<sup>a</sup>

*animal, The International Journal of Animal Biosciences*

### Supplementary Material S1: Validation and Quality Insurance.

All measurements (weight, age at first egg, number of eggs) are routinely performed on the experimental farm, and ensure a quality already demonstrated on thousands of animals in peer-reviewed articles. The weight measurements were performed by the same audited balance, which allows to optimize their robustness. The weights measured as well as the egg-laying performances are in concordance with the expert knowledge on these quail lines, ensuring the validity of the data. The statistical methods used and their implementation in R are standard and recognized by the literature as being reliable.

### Supplementary Material S2. Description of critical methods.

All the data (input files, R script) are available on the following github repository: [https://github.com/ccerutti88/Quail\\_Phenotypes.git](https://github.com/ccerutti88/Quail_Phenotypes.git)

## Mathematical model equations used for the analyses

### Global weight

Random regression model used for studying longitudinal records of body weight is the following:

$$Weight_{ijkl} = \beta_{0,k} + \beta_{1,k} age_j + \beta_{2,k} age_j^2 + u_{1,i} + u_{2,i} age_j + v_{1,l} + v_{2,l} age_j + \varepsilon_{ijkl}$$

where  $Weight_{ijkl}$  is the weight of animal  $i$  from the family  $l$  at age  $j$  in situation  $k$ ,  $u_{1,i}$  and  $u_{2,i}$  are the first and second random regression coefficients for animal  $i$  with

$[u_1 u_2] \sim N\left(0, \begin{bmatrix} \sigma_{u1}^2 & \sigma_{u12} & \sigma_{u12} & \sigma_{u2}^2 \end{bmatrix} \otimes I_n\right)$ ,  $v_{1,l}$  and  $v_{2,l}$  are the first and second random regression coefficients for family  $l$  with  $[v_1 v_2] \sim N\left(0, \begin{bmatrix} \sigma_{v1}^2 & \sigma_{v12} & \sigma_{v12} & \sigma_{v2}^2 \end{bmatrix} \otimes I_L\right)$ ,  $\varepsilon$  is the

vector of residuals with  $\varepsilon \sim N\left(0, I_N \sigma_\varepsilon^2\right)$ ,  $n$ ,  $L$  and  $N$  are the number of animals, families and records, respectively.  $\beta_{l,k}$  is the  $(l+1)^{th}$  fixed regression coefficients in situation  $k$ .

The situation  $k$  is defined, in the saturated model, by the combination of treatment, line, batch and sex (80 levels).

### Weight for each age and laying traits

Model for the weight at a given age and laying traits (egg number and age at first egg) is:

$$y_{ilk} = \mu_k + v_l + \varepsilon_{ilk}$$

where  $y_{ilk}$  is the weight at a given age or the laying trait of animal  $i$  from family  $l$  in situation  $k$ ,  $v_l$  is the random effect for family  $l$  with  $v \sim N(0, I_L \sigma_u^2)$ ,  $\varepsilon$  is the vector of residuals with  $\varepsilon \sim N(0, I_N \sigma_\varepsilon^2)$ .  $N$  and  $L$  correspond to the number of observations and families,  $\mu_k$  is the fixed effect in situation  $k$ . The situation  $k$  is defined, in the saturated model, by the combination of treatment, line, batch and sex (80 levels) for the weight and the combination of treatment, line and batch for laying traits.

## Extract of the R script used for all the analyses

### Global weight

```
# Fitting of a linear mixed-effects model allowing for nested random effects:
lmemod_weight = lme(weight ~ 1 + (treat * age * sex * line * batch) + (l(age^2) *
treat * sex * line * batch), data=df[,c("ID", "sex", "line", "batch", "treat", "family",
"weight", "age")], random=~1|ID, na.action=na.exclude, method="ML",
control=lmeControl(opt='optim'))
# Stepwise model selection by AIC with stepAIC:
mod_weight = stepAIC(lmemod_weight)
# Fitting of a linear mixed-effects model with model selected by stepAIC excluding
interactions between more than 3 effects:
lmemod_weight_selected_tmp = lme(weight ~ treat + age + sex + line + batch +
l(age^2) + treat:sex + age:sex + treat:line + age:line + sex:line + treat:batch +
age:batch + sex:batch + line:batch + treat:l(age^2) + sex:l(age^2) + line:l(age^2) +
batch:l(age^2) + treat:sex:line + age:sex:line + treat:sex:batch + treat:line:batch +
sex:line:batch + treat:sex:l(age^2) + treat:line:l(age^2) + sex:line:l(age^2) +
treat:batch:l(age^2) + sex:batch:l(age^2) + line:batch:l(age^2), data=df[,c("ID",
"sex", "line", "batch", "treat", "family", "weight", "age")], random=~(1+age)|family/ID,
na.action=na.exclude, method="ML", control=lmeControl(opt='optim'))
# Identify significant interactions with the anova:
Anova(lmemod_weight_selected_tmp, type=3)
# Fitting of a linear mixed-effects model with relevant interactions according to the
anova:
lmemod_weight_selected = lme(weight ~ treat + age + sex + line + batch +
l(age^2) + treat:sex + age:sex + age:line + sex:line + age:batch + sex:batch +
treat:l(age^2) + sex:l(age^2) + line:l(age^2) + batch:l(age^2) + age:sex:line +
treat:sex:l(age^2) + sex:line:l(age^2) + sex:batch:l(age^2), data=df[,c("ID", "sex",
"line", "batch", "treat", "family", "weight", "age")], random=~(1+age)|family/ID,
na.action=na.exclude, method="ML", control=lmeControl(opt='optim'))
```

### **Weight for each age**

```
# Weight at 8 days:
## Fitting of a linear mixed-effects model with a family random effect:
lmemod_weight_d8 = lme(weight ~ treat * line * batch * sex, random=~1|family,
data=df_8, na.action=na.exclude, method="ML", control=lmeControl(opt='optim'))
## Stepwise model selection by AIC with stepAIC:
stepAIC(lmemod_weight_d8)

# Weight at 21 days:
lmemod_weight_d21 = lme(weight ~ treat * line * batch * sex, random=~1|family,
data=df_21, na.action=na.exclude, method="ML", control=lmeControl(opt='optim'))
stepAIC(lmemod_weight_d21)

# Weight at 36 days:
lmemod_weight_d36 = lme(weight ~ treat * line * batch * sex, random=~1|family,
data=df_36, na.action=na.exclude, method="ML", control=lmeControl(opt='optim'))
stepAIC(lmemod_weight_d36)

# Weight at 78 days:
lmemod_weight_d78 = lme(weight ~ treat * line * batch * sex,
random=~1|family,data=df_78, na.action=na.exclude, method="ML",
control=lmeControl(opt='optim'))
stepAIC(lmemod_weight_d78)
```

### **Age at first egg**

```
lmemod_age1stegg = lme(age1stegg ~ treat * line * batch, random=~1|family,
data=df_egg[,c("ID", "line", "batch", "treat", "family", "age1stegg")],
na.action=na.exclude, method="ML")
stepAIC(lmemod_age1stegg)
```

### **Egg number**

```
lmemod_eggnb = lme(eggNumber ~ treat * line * batch, random=~1|family,
data=df_egg[,c("ID", "line", "batch", "treat", "family", "eggNumber")],
na.action=na.exclude, method="ML")
stepAIC(lmemod_eggnb)
```

**Supplementary Table S1.** Models before and after selection by *stepAIC* for each trait.

Trait	Random effect	Models before selection by <i>stepAIC</i>	Models after selection by <i>stepAIC</i>
weight	(1+age)   family / ID	weight ~ treat * sex * line * batch + (treat * sex * line * batch) * age + (treat * sex * line * batch) * age <sup>2</sup>	weight ~ treat + sex + line + batch + treat:sex + sex:line + sex:batch + age + age:sex + age:line + age:batch + age:sex:line + age <sup>2</sup> + treat:age <sup>2</sup> + sex:age <sup>2</sup> + line:age <sup>2</sup> + batch:age <sup>2</sup> + treat:sex:age <sup>2</sup> + sex:line:age <sup>2</sup> + sex:batch:age <sup>2</sup>
d8 weight	1   family	d8 weight ~ treat * line * batch * sex	d8 weight ~ treat + line + batch + sex + treat:sex + batch:sex
d21 weight	1   family	d21 weight ~ treat * line * batch * sex	d21 weight ~ treat + line + batch + sex + treat:sex
d36 weight	1   family	d36 weight ~ treat * line * batch * sex	d36 weight ~ treat + line + sex + treat:sex + line:sex
d78 weight	1   family	d78 weight ~ treat * line * batch * sex	d78 weight ~ treat + line + batch + sex + treat:batch + treat:sex + line:sex + batch:sex + treat:batch:sex
age 1st egg	1   family	age 1st egg ~ treat * line * batch	age 1st egg ~ line + batch + treat + line:batch
egg number	1   family	egg number ~ treat * line * batch	egg number ~ line + batch

Abbreviations: *stepAIC* = R function performing stepwise model selection by Akaike Information Criterion; ID = identifier; d = days; treat = treatment.

**Supplementary Table S2.** Mean and standard deviation for each egg-laying (female) trait and for the weight at 78 days, grouped by line of quails. P-values for the effects of line on age at first egg, egg number and weight at 78 days are respectively  $<2e-16$ ,  $<2.2e-16$  and  $<2.2e-16$ . S+ is the reference line for the anova. Only values less than 0.05 are considered as significant.

line	age 1st egg			egg number			d78 weight		
	mean	sd	P-value	mean	sd	P-value	mean	sd	P-value
S+	60.72	±3.85		13.33	±5.69		199.79	±26.27	
DD	57.48	±4.16	1.24e-3*	16.42	±8.37	6.77e-4*	203.72	±24.79	2.86e-1*
A	55.56	±3.96	2.43e-10*	21.75	±6.5	$<2.2e-16$ *	378.78	±57.66	$<2.2e-16$ *
B	54.5	±3.13	2.29e-14*	22.09	±5.47	$<2.2e-16$ *	378.23	±48.45	$<2.2e-16$ *

Abbreviations: d = days.

\* Value significantly different at  $P < 0.05$ .

**Supplementary Table S3.** Table of anova p-values for the weight of quails at a given age. Results obtained by combining the following methods: *aov* + *stepAIC*. Only values less than 0.05 are considered as significant.

	d8 weight	d21 weight	d36 weight	d78 weight
line	$< 2.2e-16$ *	$< 2.2e-16$ *	$< 2.2e-16$ *	$< 2.2e-16$ *
batch	3.62e-12*	4.96e-14*	NR	7.19e-02
sex	1.86e-03*	9.39e-06*	3.21e-03*	2.79e-08*
treat	4.15e-04*	1.05e-05*	3.41e-04*	2.75e-02*
(vs IC)				
Con	6.14e-01	3.41e-01	5.6e-01	1.32e-02*
BPA	4.83e-02*	1.2e-01	4.75e-01	6.79e-02
Gen	9.38e-03*	3.02e-03*	3.85e-03*	3.46e-01
5Aza	7.88e-03*	2.35e-03*	9.68e-03*	5.55e-01
batch:sex	2.2e-02*	NR	NR	1.61e-02*
sex:treat	1.94e-02*	2.31e-02*	8.45e-02	6.91e-02
line:sex	NR	NR	2.12e-02*	$< 2.2e-16$ *
batch:treat	NR	NR	NR	1.79e-02*
batch:sex:treat	NR	NR	NR	1.37e-02*

Abbreviations: *aov* = R function for analysis of variance; *stepAIC* = R function performing stepwise model selection by Akaike Information Criterion; d = days; treat = treatment; IC = Injected control (sesame oil); Con = Non-injected control; BPA = Bisphenol A; Gen = Genistein; 5Aza = 5-Aza-2'-deoxycytidine; NR = effect not retained by *stepAIC*.

\* Value significantly different at  $P < 0.05$ .

**Supplementary Table S4.** Tables of Tukey HSD contrast estimate values for pairwise treatment comparison between quail females at each measurement age. Results obtained by combining the following methods: *aov* + *stepAIC* + *emmeans*. The contrast estimate values reflect the weight loss or gain (in g) between the compared groups.

Female	d8 weight	d21 weight	d36 weight	d78 weight
Con - BPA	+0.84	+2.67	+1.98	+8.04
Con - Gen	+1.08	+4.28	+5.56	+12.15
Con - 5Aza	+1.26	+4.93	+5.8	+15.59
IC - Con	-0.16	-0.96	-0.85	-6.78
IC - BPA	+0.68	+1.70	+1.13	+1.25
IC - Gen	+0.92	+3.31	+4.71	+5.36
IC - 5Aza	+1.1	+3.96	+4.94	+8.8

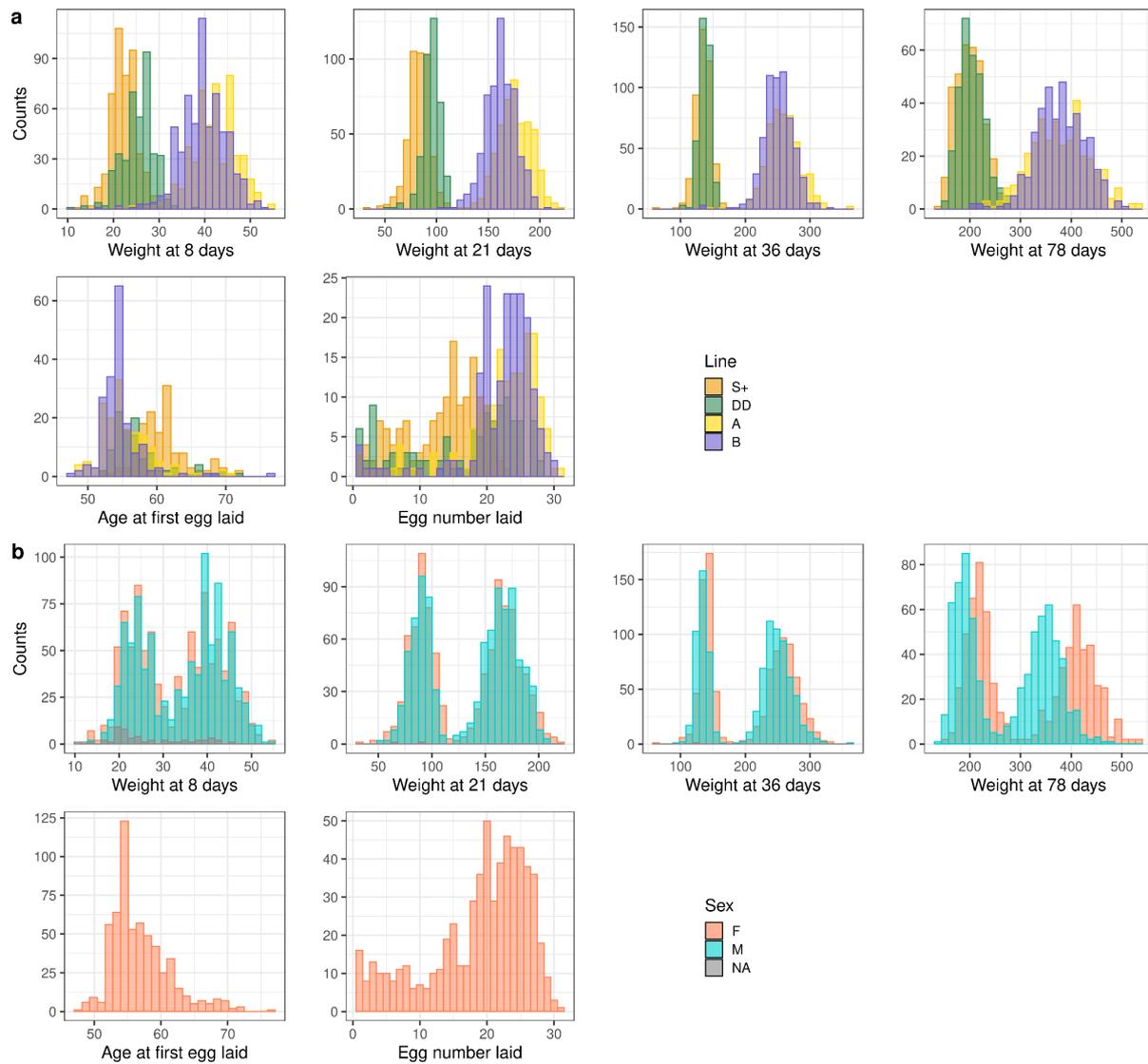
Abbreviations: HSD = honestly significant difference; *aov* = R function for analysis of variance; *stepAIC* = R function performing stepwise model selection by Akaike Information Criterion; *emmeans* = R function to compute estimated marginal means; d = days; IC = injected control (sesame oil); Con = non-injected control; BPA = Bisphenol A; Gen = Genistein; 5Aza = 5-Aza-2'-deoxycytidine.

**Supplementary Table S5.** Tables of  $-\log_{10}(\text{p-values})$  from Tukey HSD for pairwise treatment comparison between quail females at each measurement age. Results obtained by combining the following methods: *aov* + *stepAIC* + *emmeans*. Only values less than 0.05 are considered as significant.

Female	d8 weight	d21 weight	d36 weight	d78 weight
Con - BPA	1.86e-01	1.87e-01	9.36e-01	3.56e-01
Con - Gen	3.07e-02*	1.26e-03*	7.81e-03*	1.92e-02*
Con - 5Aza	4.25e-02*	2.63e-03*	4.53e-02*	4.65e-03*
IC - Con	1	9.95e-01	1	6.04e-01
IC - BPA	6.16e-01	8.68e-01	1	1
IC - Gen	2.18e-01	8.82e-02	1.08e-01	9e-01
IC - 5Aza	1.91e-01	7.13e-02	2.23e-01	4.94e-01

Abbreviations: HSD = honestly significant difference; *aov* = R function for analysis of variance; *stepAIC* = R function performing stepwise model selection by Akaike Information Criterion; *emmeans* = R function to compute estimated marginal means; d = days; IC = Injected control (sesame oil); Con = Non-injected control; BPA = Bisphenol A; Gen = Genistein; 5Aza = 5-Aza-2'-deoxycytidine.

\* Value significantly different at  $P < 0.05$ .



**Supplementary Figure S1.** Distribution of quail phenotyping data (weight at 8, 21, 36, and 78 days, age at first egg and egg number laid) according to the line (a) and sex (b). (a) The orange, green, yellow and violet bars show respectively the proportion of quails from S+, DD, A and B lines. (b) The blue and red bars show respectively the proportion of male and female quails.

Abbreviations: F = female; M = male; NA = untested effect.



# 10. Impacts sur la méthylation de l'ADN

## 10.1 Dispositif expérimental

Dans cette étude, nous avons étudié l'effet de l'injection de perturbateurs endocriniens (Génistéine et Bisphénol A (BPA)) ou d'un inhibiteur de l'ADN méthyltransférase (5-Aza-2'-déoxycytidine (5Aza)) au cours du développement embryonnaire sur la méthylation de l'ADN de cailles japonaises adultes. Afin d'identifier les régions génomiques affectées par les modifications épigénétiques induites par ces molécules, des données de méthylation de l'ADN en pool issues de la génération G0 ont été analysées. Ces données ont été obtenues par séquençage haut-débit après traitement de l'ADN au bisulfite de sodium, par séquençage RRBS et WGBS. Les données RRBS proviennent des individus dont les phénotypes ont été étudiés dans la première étude de la thèse (**Chapitre 9, Figure 10.1**, Cerutti et al. (2022)), abattus à 79 jours. Les données RRBS proviennent d'échantillons de foie, cerveau et sang de cailles issues de quatre lignées différentes (A, B, S+ et DD). Ces individus sont séparés en 5 groupes : un groupe témoin sans aucune injection, un groupe témoin traité par injection dans l'œuf d'huile de sésame et trois groupes traités par injection de génistéine, BPA ou 5Aza. Pour ces données, les échantillons présentant les mêmes caractéristiques, c'est-à-dire provenant du même type de tissu, d'individus ayant reçu le même traitement et de la même lignée, ont été regroupés avant le séquençage. Au total, 60 échantillons ont été analysés, chaque échantillon étant un pool de 8 individus (4 mâles et 4 femelles) (**Figure 10.1.A**). Cette étude étant effectuée sur des données RRBS, seulement environ 2% du génome de la caille est étudié, les informations de méthylation au niveau de la majorité du génome ne sont donc pas disponibles. Afin de pouvoir étudier l'intégralité du génome, des données de séquençage WGBS issues de la G0, une technique également basée sur le traitement au bisulfite de l'ADN, ont également été analysées. Ces données ont été obtenues à partir d'échantillons de sang prélevés sur un sous-ensemble du dispositif, c'est-à-dire de cailles de la lignée S+, traitées au BPA, à la génistéine, au 5Aza ou non injectées (C), nous permettant d'obtenir des données de méthylation sur l'ensemble du génome. Au total, 8 échantillons ont été analysés, chaque échantillon étant un pool de 5 individus (**Figure 10.1.B**).

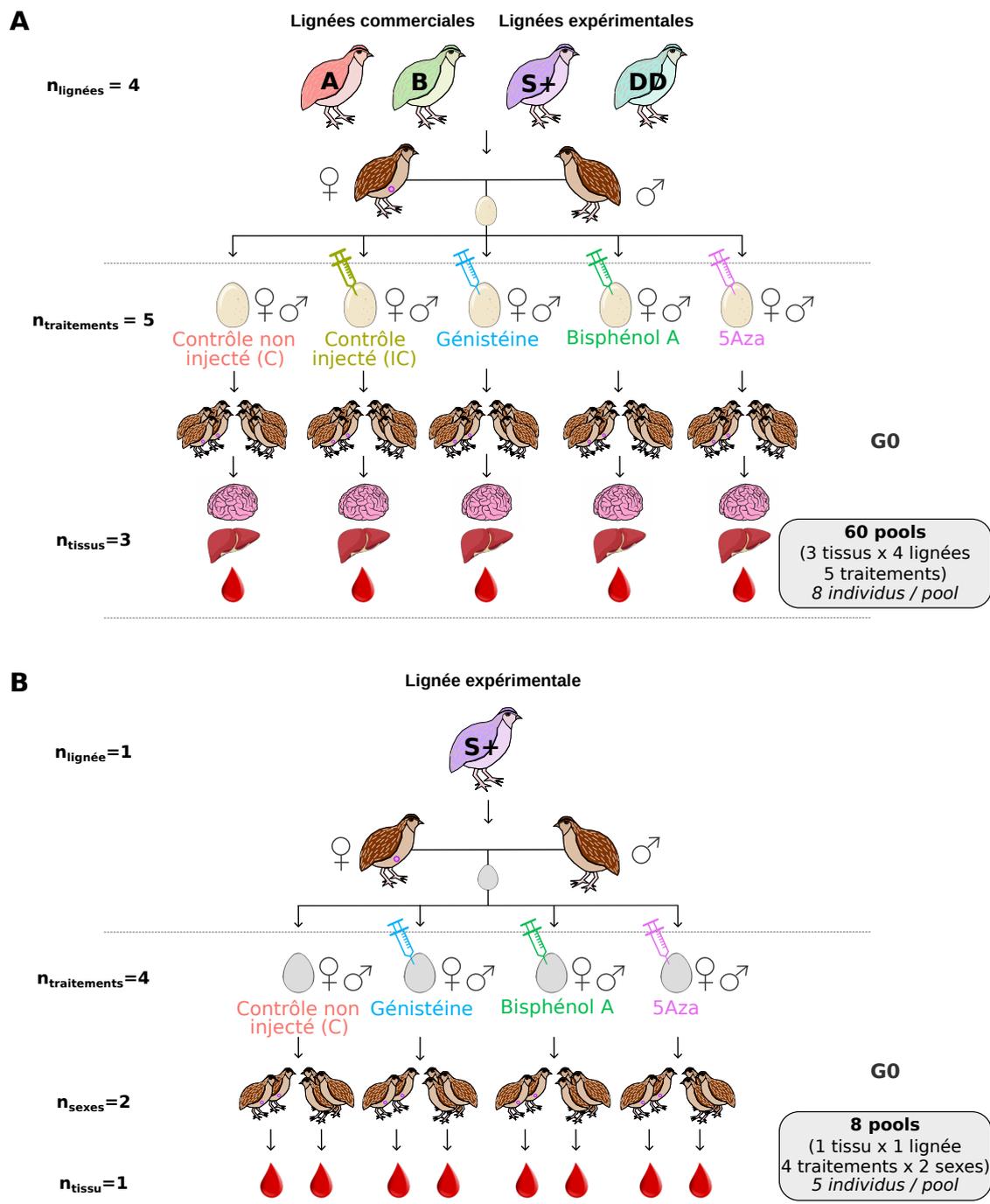


FIGURE 10.1 – Dispositif expérimental de la G0. (A) RRBS : des échantillons de cerveau, de sang et de foie ont été prélevés sur des cailles (mâles et femelles) de 5 groupes : 2 groupes de contrôle (non injectés (C) et injectés avec de l'huile de sésame (IC)), 3 groupes injectés avec du bisphénol A, de la génistéine ou du 5Aza. Les cailles étudiées provenaient de quatre lignées : deux commerciales (A et B) et deux expérimentales (S+ et DD). Des échantillons de tissus ont été prélevés sur 1979 cailles. Les échantillons présentant les mêmes caractéristiques ont été regroupés pour un total de 60 échantillons. (B) WGBS : des échantillons de sang ont été prélevés sur des cailles S+ (mâles et femelles) de 4 groupes : un groupe C et 3 groupes ayant reçu des injections de bisphénol A, de génistéine et de 5Aza. Les échantillons présentant les mêmes caractéristiques ont été regroupés pour un total de 8 échantillons.

## 10.2 Article 2 : Effets de l'injection *in-ovo* de perturbateurs endocriniens et d'un inhibiteur de l'ADN méthyltransférase sur la méthylation de l'ADN chez la caille adulte.

Afin de caractériser les effets des molécules injectées sur le méthylome de la G0, des données de méthylation de l'ADN RRBS et WGBS ont été analysées. Dans un premier temps, le traitement des lectures issues des séquençage RRBS et WGBS est réalisé par l'outil `nf-core/methylseq` v1.5 (P. A. Ewels et al., 2020 ; P. Ewels et al., 2023) (**Article 2 - Supplementary Figure S2.I**). Les lectures sont découpées par `Trim Galore` (Krueger & Andrews, 2011) afin d'éliminer les adaptateurs et les lectures de faible qualité. Puis les lectures découpées sont alignées sur le génome de référence de la caille par `Bismark` (Krueger & Andrews, 2011). Ce dernier extrait les sites CpG dans un fichier pour chaque échantillon en considérant toutes les cytosines sur les deux brins et rapporte leur position, brin, contexte trinuécléotidique et état de méthylation. Pour détecter les DMC entre les groupes, le pipeline `Seq0ccin` basé sur le package `edgeR` v3.30.3 de R (Robinson et al., 2010) a été utilisé (**Article 2 - Supplementary Figure S2.II.A**). Ce pipeline analyse les sorties de `nf-core/methylseq` et est composé de deux étapes : une étape de prétraitement suivie de l'analyse différentielle. Le prétraitement consiste tout d'abord à fusionner les informations de méthylation et de couverture des deux brins. Un filtre est ensuite appliqué pour supprimer les SNP qui pourraient être identifiés à tort comme des DMC. L'outil `gemBS` v4.0.2 (Merkel, Fernández-Callejo, et al., 2019) est utilisé pour détecter les SNP présents directement dans les échantillons (**Article 2 - Supplementary Figure S2.II.B**). `gemBS` effectue un alignement des lectures (`GEM3`) et un appel de variants (`BScall`) qui extrait les SNP de chaque échantillon. Les listes de SNP identifiées sont ensuite fusionnées avec le fichier VCF de la caille de référence (Bardou, 2022) par `bcftools` v1.16 (Danecek et al., 2021). Le fichier VCF combiné a été utilisé pour filtrer les SNP dans les listes de CpG issues de `Bismark`. Les sites CpG ayant un taux de méthylation nul dans tous les tissus, étant considérés comme des SNP potentiels, ont également été filtrés. D'autres filtres sont appliqués durant le prétraitement des données : les sites CpG avec une couverture minimale de 5 et maximale de 1000 lectures dans 80% des échantillons sont conservés pour la suite des analyses. Pour la détection des DMC, des modèles linéaires ont été utilisés. Pour le RRBS, ces derniers incluent les effets de lignée et de traitement, et pour le WGBS, ils incluent les effets de sexe et de traitement (**Table 10.1**). Le contrôle injecté (IC) a été utilisé comme groupe de référence dans l'analyse RRBS. Les p-valeurs ont été ajustées par la méthode Benjamini-Hochberg (BH), avec un seuil de significativité  $\alpha$  de 5%. L'appel des DMR a été réalisé avec la fonction `callDMR` du package `DSS` v2.38.0 à partir des sorties de `edgeR`. Un nombre minimum de 3 CpG est requis pour constituer une DMR et le pourcentage de sites CpG avec des p-valeurs significatives (p-valeurs  $< 0,05$ ) doit être supérieur à 50%.

TABLE 10.1 – **Modèles linéaires utilisés pour l'analyse différentielle.** Pour les données RRBS, le contrôle injecté (IC) et la lignée A ont été utilisés comme groupes de référence des tests. Pour les données WGBS, le contrôle non injecté (C) et le groupe des femelles (F) ont été utilisés comme groupes de référence des tests.

Données	Tissu	Modèle linéaire
RRBS	Sang	~ traitement + lignée
RRBS	Foie	~ traitement + lignée
RRBS	Cerveau	~ traitement + lignée
WGBS	Sang	~ traitement + sexe

L'annotation et l'analyse de la distribution des DMC ont été réalisées (package R `GenomeFeatures` v1.3) en utilisant le fichier GTF d'annotation d'Ensembl (v103). Après l'étape d'annotation, chaque DMC a été associée à des régions génomiques : promoteur (3 Kbp en amont et 500 bp en aval du site de début de transcription - TSS), 5'UTR, 3'UTR, exons, introns, aval (à 1 Kbp du site de fin de transcription - TES), séquence codante (CDS) et région intergénique. Une analyse de surreprésentation des termes de la Gene Ontology (GO) a ensuite été effectuée (package R `ViSEAGO` v1.4.0, (Brionne et al., 2019)) avec l'ensemble des gènes du génome de la caille comme référence ("*background*") et avec minimum 5 gènes associés à chaque terme GO. Pour chaque tissu et technique de séquençage, un test d'enrichissement a été réalisé à l'aide du test exact de Fisher sur chaque liste de gènes associés aux DMC (quatre listes de gènes au total). Les termes GO enrichis significatifs ( $p\text{-valeur} \leq 0,01$ ) ont été rassemblés en groupes fonctionnels à l'aide d'un clustering hiérarchique basé sur la distance de Wang entre les termes GO et le critère de Ward. Ces groupes fonctionnels ont ensuite été regroupés à l'aide d'un clustering hiérarchique basé sur la distance BMA entre les ensembles de termes GO et le critère de Ward ( $-\log_{10}(p\text{-valeurs ajustées})$ ).

Cette étude a pu mettre en évidence que l'exposition à des molécules lors du développement embryonnaire impacte la méthylation de l'ADN des individus. L'analyse de méthylation différentielle a permis d'identifier des régions différentiellement méthylées entre les conditions expérimentales. L'analyse de surreprésentation des termes GO a mis en évidence un effet des molécules sur des gènes liés aux voies métaboliques (croissance), au comportement (développement du système nerveux central, transport des lipides et réponse au cholestérol) et à la reproduction (maturité sexuelle). Nous savons que les effets de l'environnement peuvent être transmis aux générations suivantes, en partie par l'intermédiaire de marques épigénétiques. Afin d'étudier une potentielle transmission épigénétique transgénérationnelle des effets de l'environnement, les analyses suivantes consisteront à étudier l'influence de l'environnement précoce des individus sur les générations futures, notamment dans le cas d'une contamination dans l'environnement par des perturbateurs endocriniens et de leur accumulation dans l'œuf.

L'ensemble des scripts R utilisés pour les analyses sont disponibles via le lien suivant : <https://gitlab.com/Okami8/epiques>.

## **DNA methylation changes in adult quails following *in-ovo* injection of genistein, bisphenol A, or 5-Azacytidine.**

C. Cerutti<sup>1</sup>, S. Leroux<sup>1</sup>, P. Terzian<sup>2</sup>, D. Gourichon<sup>4</sup>, V. Coustham<sup>5,6</sup>, T. Zerjal<sup>7</sup>, J. Demars<sup>1</sup>, F. Pitel<sup>1</sup>, G. Devailly<sup>1</sup>

<sup>1</sup> *GenPhySE, Université de Toulouse, INRAE, ENVT, 31320, Castanet Tolosan, France*

<sup>2</sup> *MIAT, UR875 INRAE, 31326, Castanet-Tolosan, France*

<sup>3</sup> *GeT-PlaGe - Génome et Transcriptome - Plateforme Génomique, US1426 GET - INRAE, 31326, Castanet-Tolosan, France*

<sup>4</sup> *INRAE, PEAT, 37380, Nouzilly, France*

<sup>5</sup> *INRAE, Université de Tours, BOA, 37380, Nouzilly, France*

<sup>6</sup> *Université de Pau et des Pays de l'Adour, E2S UPPA, INRAE, NUMEA, Saint-Pée-sur-Nivelle, 64310, France*

<sup>7</sup> *GABI, AgroParisTech, INRAE, Université Paris-Saclay, 78350 Jouy-en-Josas, France*

Corresponding author: Guillaume Devailly

Email: [guillaume.devailly@inrae.fr](mailto:guillaume.devailly@inrae.fr)

### **Author ORCIDs :**

C. Cerutti : <https://orcid.org/0000-0002-3207-3461>

S. Leroux : <https://orcid.org/0000-0003-2840-5109>

P. Terzian : <https://orcid.org/0000-0002-7092-3906>

D. Gourichon : <https://orcid.org/0000-0002-4631-0149>

V. Coustham : <https://orcid.org/0000-0002-5399-2723>

T. Zerjal : <https://orcid.org/0000-0002-0666-5420>

J. Demars : <https://orcid.org/0000-0003-2451-0059>

F. Pitel : <https://orcid.org/0000-0002-1477-7633>

G. Devailly : <https://orcid.org/0000-0001-8878-9357>

### **Abstract**

Exposure to biologically active substances during embryonic development might affect a wide variety of mechanisms involved in the construction of the individual, and have consequences on the animal's health and welfare throughout its lifetime. Such effects can be mediated by epigenetic mechanisms, including DNA methylation patterns. The objective of this study is to characterise the DNA methylation response of the organism to variations in its embryonic environment. To this end, this study analysed the effects on the DNA methylation of adult quails (blood, brain, liver) of the injection of chemicals into the egg during embryonic development: endocrine disruptors (bisphenol A, genistein) and an inhibitor of the enzyme DNA methyltransferase (5Aza), using a combination of RRBS and WGBS. Embryonic exposure to the tested chemicals leave detectable traces on the adults DNA methylation profiles in all three tissues examined. The Gene Ontology (GO)

enrichment analysis highlights genes implicated in metabolic, behaviour and reproduction pathways. Therefore, *in-ovo* exposure of bioactive substances, notably bisphenol A and genistein, can have long lasting impacts on DNA methylation profiles in multiple bird tissues.

### Keywords

DNA methylation, avian species, *Coturnix japonica*, environment, endocrine disruptors, bisulfite, genistein, bisphenol A, 5-aza-deoxycytidine, differentially methylated cytosine

### Introduction

DNA methylation is one of the most studied epigenetic marks (Wang et al., 2022), notably thanks to the availability of numerous sequencing technologies that allow its analysis (Sun and Zhu, 2022). DNA methylation has been shown to partially reflect the organism's adaptation to its environment, and may have a major role in the regulation of gene expression (Feil and Fraga, 2012).

Early life seems to be an important window of sensitivity to the environment, and the prenatal environment is known to influence adult development in many species (Sinclair et al., 2016). Previous studies have shown that *in utero* exposure to stressors, such as malnutrition through the maternal diet, can alter the development of the fetus and increase the risk of developing metabolic diseases and some cancer types in adulthood (Silva et al., 2019).

Exposure to biologically active substances during embryonic development can affect a wide variety of mechanisms involved in the construction of the individual. Among these active molecules, endocrine disruptors, which deregulate the hormonal system, are a public health issue. The effect of these molecules on health is becoming more and more understood: they are suspected of being involved in the appearance of certain diseases such as cancers, or of having an influence on growth, neurodevelopment and fertility (Yilmaz et al., 2020). Some of these endocrine disruptors can be found in food, such as bisphenol A (BPA), a plastic constituent (Qin et al., 2021). Others are found in nature, such as phytoestrogens, and are present in a wide variety of plants. One of the best known phytoestrogens is soybean genistein, which is thought to have protective effects against metabolic diseases and cancers (Silva et al., 2019, Rasheed et al., 2022).

In livestock production, inappropriate rearing conditions in early life, such as the presence of chemical contaminants during embryonic development, can impact the health and well-being of animals. In a previous study, we investigated the impact of embryonic exposure to chemicals in adult birds (Japanese quail) and we revealed the impact of endocrine disruptors on quail growth. Quails injected with BPA or genistein in the egg had a significantly lower weight than those not exposed (Cerutti et al., 2022). The objective of this study is to characterise molecular mechanisms underlying this organism's response to variations in its embryonic environment. To this end, we analysed the effects of endocrine disruptors and an inhibitor of the

enzyme DNA methyltransferase injected into the egg during embryonic development on the DNA methylation profiles of adult quails. The Gene Ontology analysis performed on differentially methylated positions highlights genes related to metabolic (developmental growth), behaviour (central nervous system development, lipid transport and the response to cholesterol) and reproduction pathways (sexual maturity).

## **Materials and methods**

### ***Experimental designs***

The quail experimentation has been previously described (Cerutti et al, 2022) (**Supplementary Figure S1**).

Briefly, animals were reared at the Tours Avian Experimental Facility (PEAT), an INRAE experimental unit. They were from four different lines: two commercial lines (A and B) and two experimental lines (S+, Mills and Faure, 1991, and DD, Minvielle et al., 1999). Fifty-four parent pairs per line (216 in total) were used and fertilised eggs from each pair were divided into five groups in a randomised manner: two control groups, one non-injected (C) and one injected (IC) with 50µl of a mixture of 10% ethanol and 90% sesame oil (Acros Organics); two groups treated with 50 µl of endocrine disruptors, one group injected with 500 µg of Genistein (Molekula) and another injected with 200 µg of BPA (Sigma-Aldrich); a last group injected with 34 µg (50 µM) of a DNA methyltransferase enzyme inhibitor, 5-Aza-2'-deoxycytidine (5Aza, Sigma-Aldrich). All molecules were dissolved in 100% ethanol and then diluted 10 times in sesame oil prior to injection. The injection was performed on the day of incubation of freshly laid eggs, directly into the yolk. At the end of this procedure, 1979 quails were obtained. For each individual, blood, liver and whole brain samples were harvested at 79 days of age, and the tissues were flash frozen in liquid nitrogen.

### ***DNA extraction and library preparation***

Genomic DNA was extracted from blood and brain according to a rapid extraction protocol as in Roussot et al, 2003. To extract DNA from livers, we used the high-salt method described in Demars et al, 2022. To minimise sequencing costs, we performed pool sequencing.

To generate the RRBS data (**Supplementary data Table 1.A**), samples with the same characteristics, i.e. from the same tissue type, from individuals injected with the same treatment and from the same line, were pooled before sequencing. A total of 60 samples were analysed, each sample being a pool of 8 individuals (4 males and 4 females).

The WGBS data (**Supplementary data Table 1.B**) were obtained from blood samples taken from a subset of the animal design: quails from the S+ line, treated with BPA, genistein, 5Aza or non-injected (C), allowing us to obtain methylation data on the whole genome. A total of 8 samples were analysed, each sample being a pool of 5 individuals.

RRBS libraries were prepared from 100 ng of genomic DNA per sample, using the Premium Reduced Representation Bisulfite Sequencing (RRBS) Kit (Diagenode, #C02030033), according to the manufacturer's instructions. Briefly, the protocol consists in the digestion of DNA by the MspI enzyme, followed by fragment end repair, and ligation of adaptors. The size selection was done with AMPure XP Beads (Beckman Coulter). Next, samples were quantified by qPCR and the Ct values were used to pool samples by equimolarity. Then the bisulfite conversion was done on the pool and the final libraries were amplified using MethylTaq Plus Master Mix (Diagenode kit). After a clean-up with AMPure XP Beads, the quality of the RRBS library pools was evaluated by measuring their DNA concentration with the Qubit® dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham MA, USA), and the profile of the pools was verified using the High Sensitivity DNA chip for 2100 Bioanalyzer (Agilent) or DNF-474 NGS fragment kit on a Fragment Analyzer (Agilent).

### ***DNA methylation sequencing***

RRBS DNA methylation data from 60 samples including 20 blood, 20 liver and 20 brain samples of Japanese quail were obtained by sequencing (50 bp single end reads, Illumina HiSeq 4000) on the GenomEast platform.

WGBS DNA methylation data was obtained for 8 blood samples of Japanese quail. The library preparation and sequencing (2 x 150 bp) for the WGBS analyses were performed by Novogene (UK) (Illumina NovaSeq 6000).

### ***Differential methylation analyses***

The Ensembl reference genome of the Japanese quail (*Coturnix japonica* v.2.0, GCA\_001577835.1, Ensembl release 103) was used for the analyses ([https://www.ensembl.org/Coturnix\\_japonica/Info/Index](https://www.ensembl.org/Coturnix_japonica/Info/Index)). It includes 8,974,565 CpG sites.

The nf-core/methylseq v1.5 tool (Ewels et al., 2020, 2022) was used to process RRBS (single-end) and WGBS (paired-end) reads (**Supplementary Figure S2**). Briefly, reads were trimmed with Trim Galore v0.6.4 (Krueger and Andrews, 2011) to remove adapters and low-quality read (`--rrbs` and `--clip_r1 3` specified for RRBS) and trimmed reads were mapped to the quail reference genome by Bismark v0.22.3 (Krueger and Andrews, 2011). Bismark extracts methylation sites and generates a cytosine report for each sample (`--cytosine_report`) considering all cytosines on both strands and reports their position, strand, trinucleotide context and methylation state. To detect differentially methylated cytosines between experimental groups, we developed an EdgeR-based (Robinson et al., 2010) R pipeline, available at <https://forgemia.inra.fr/chloe.cerutti/bsseqmethdiffanalysis/-/tree/main>.

This pipeline analyses the outputs of nf-core/methylseq, and consists of two subparts: a pre-processing step followed by the differential analysis.

The pre-processing consists firstly of the merging of methylation and coverage information from both strands. A filter was then applied to remove SNPs that could be misidentified as DMCs. The gemBS v4.0.2 tool (Merkel et al., 2019) was used to

detect SNPs from the sequence data. The retrieved SNP list was merged with the reference quail VCF file (Bardou, 2022) by *bcftools* v1.14 (Danecek et al., 2021). The combined VCF file was used to filter putative SNP position from the Bismark outputs. CpG sites with a zero methylation rate in all tissues were considered as potential SNPs and were also filtered. This data pre-processing step has an additional filtering step: CpG sites needed to have a minimum of 5 and maximum of 1000 read coverage in 80% of samples.

To perform the differentially methylated cytosines (DMC) detection, the *edgeR* package (v3.30.3) (Robinson et al., 2010) was used. The differentially methylated regions (DMR) calling was performed with the *callDMR* function from the *DSS* package (v2.38.0) from the *edgeR* outputs. A minimum number of 3 CpGs was required to constitute a DMR and the percentage of CpG sites with significant adjusted p-values (less than 0.05) was set to more than 50%.

The linear models included the line and treatment effects for RRBS, and the sex and treatment effects for WGBS.

The injected control (IC) was used as the reference group in the RRBS analysis. The p-values were corrected with the Benjamini-Hochberg (BH) method (Benjamini and Hochberg, 1995), with a threshold of 5% defining significance after multiple testing correction.

For the principal component analysis (PCA), missing values were imputed with the mean methylation rate of all other samples for this position.

For the visual representations, we used *ggplot2*, *upSetR*, *pheatmap*, *qqman* and *dmrseq* R packages (**Supplementary Table S4**, see the package version on the code repository).

### ***DMC annotation and GO over-representation analysis***

The next analysis steps were performed as described previously (David et al., 2019). The annotation and distribution analysis of DMCs were performed by using the R package *GenomeFeatures* v1.3 (<https://forgemia.inra.fr/aurelien.brionne/GenomeFeatures>) using the annotation file (v103, *Coturnix\_japonica.Coturnix\_japonica\_2.0.103.gtf*) from Ensembl. Each DMC is associated to genomic features: promoter (3 Kbp upstream and 500 bp downstream of Transcriptional Start Site - TSS), 5'UTR, 3'UTR, exons, introns, downstream (1 Kbp from Transcriptional End Site - TES), coding sequence (CDS) and intergenic region.

A Gene Ontology (GO) term over-representation analysis was performed using the R package *ViSEAGO* v1.4.0 (Brionne et al., 2019) with quail genome genes associated with CpG as background and a minimum number of genes for each GO term of 5 (*create\_topGOdata* function). For each list of genes associated with DMCs identified by tissue and technology (four lists of genes), an over-representation test was performed using the exact Fisher's test (*runTest* function from the package *topGO* v2.42.0). Significant enriched GO terms (p-value  $\leq 0.01$ ) were grouped into functional

clusters using hierarchical clustering based on Wang distance between GO terms and Ward's criterion (*compute\_SS\_distances*, *GOterms\_heatmap* functions). These functional clusters were grouped using hierarchical clustering based on BMA distance between sets of GO terms and Ward's criterion (*compute\_SS\_distances*, *GOclusters\_heatmap* functions). The  $-\log_{10}$ (adjusted  $p$ -values) were used.

### **Data and script availability**

Sequencing data are available at <https://www.ebi.ac.uk/ena/browser/home> (study **PRJEB61896**). The differential methylation pipeline can be found here: <https://forgemia.inra.fr/chloe.cerutti/bsseqmethdiffanalysis/-/tree/main>. Scripts used for this analysis can be found in the following git repository: <https://gitlab.com/Okami8/epiques>

## **Results**

### **1 - Detection of the effects of in-ovo injection on adult DNA methylation by RRBS**

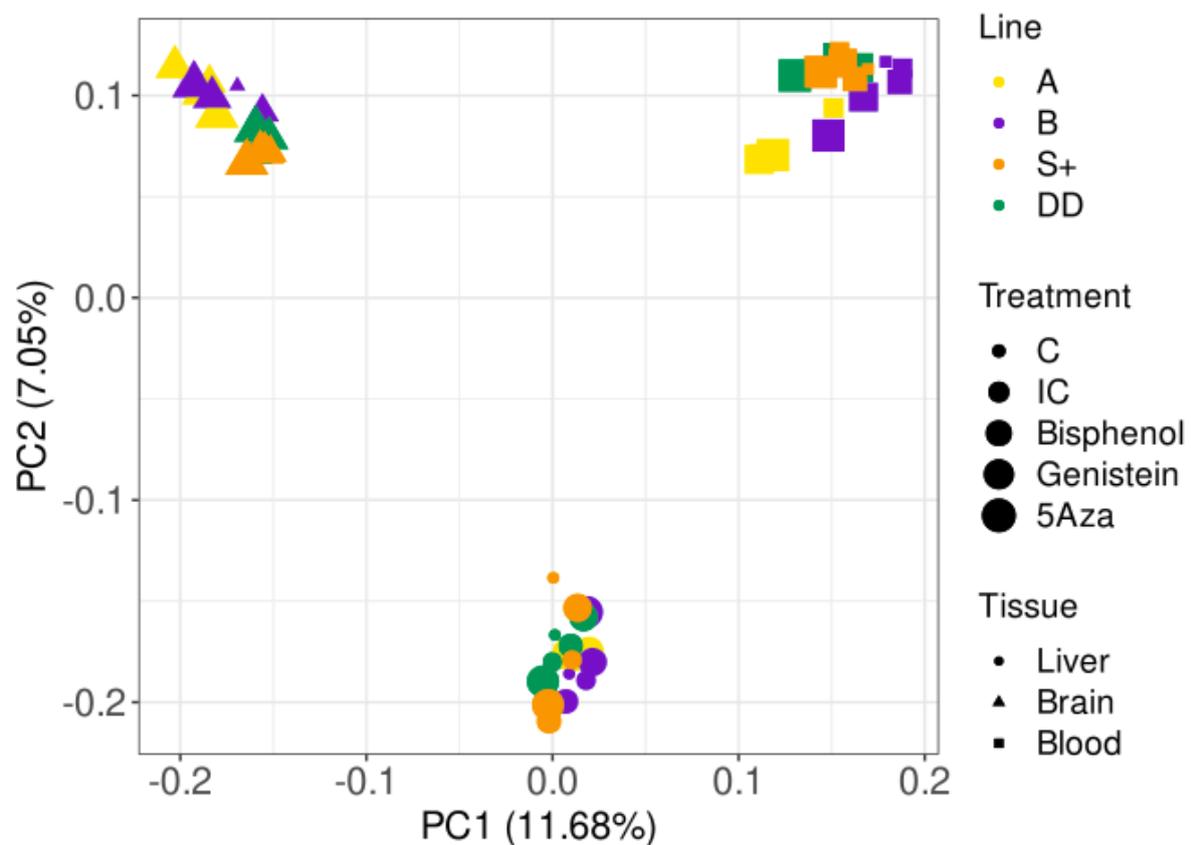
#### **DNA methylation profiling**

To study the DNA methylation profile, RRBS data from different tissues (blood, liver and brain), lines (A, B, S+ and DD) and treatment groups (non-injected control, injected control, genistein, bisphenol, 5Aza) were analysed.

After nf-core processing of the RRBS data, the sequence data showed an average of  $27.6 \pm 17.9$  M reads per sample (**Supplementary Table S2.A**). The quite low percentage of aligned reads ( $37.6 \pm 2\%$ ) is due to the presence of fragments only containing adapter sequences, that we chose to filter out thanks to their non-alignment instead of removing them prior to the alignment. After a first data pre-processing step applying several filters (**Supplementary Figure S3.A**), three samples were removed (FRPL299, FRPL310, FRPL313) from the 60 sample pools analysed for RRBS data, because their number of analysable CpGs was much lower than the others due to low depth or poor quality libraries.

At the end of the data preprocessing, a total number of 135,105 unique CpG sites were retained for further analysis, corresponding to 1.4% of the quail genome CpGs. Each CpG site has an average coverage of  $28.7 \pm 17.5$  and methylation level of  $28.1 \pm 2.2\%$ .

We performed a principal component analysis (PCA) on the methylation levels of retained CpG sites from the 57 RRBS pools (**Figure 1**). Principal component 1 (PC1, 11.68% of variance explained) clearly separates the liver, brain and blood samples. PC2 (7.05% of variance explained) separates the liver from brain and blood samples. A separation of the different quail lines can be distinguished for each tissue, while a possible treatment effect is not apparent in these first two components.



**Figure 1. Principal Component Analysis (PCA) of the methylation profiles of CpG sites from RRBS data.** The shapes differentiate the tissues: circles, triangles and squares represent brain, liver and blood samples, respectively. The point size represents the different treatments, from largest to smallest: non-injected controls (C), injected controls (IC), bisphenol A (BPA), genistein and 5Aza. The colours differentiate the quail lines: yellow and purple for the commercial lines A and B, respectively, orange and green for the experimental lines S+ and DD, respectively. Percentages indicate the proportion of variance explained by each PCA axis. Total number of CpGs sites: n = 135,105.

### ***In-ovo treatments and in-ovo injections both affect the adult DNA methylome***

As the methylation profiles were tissue specific, the DMC detection was conducted within each tissue. Multifactor analysis was used to identify a potential treatment effect.

From the RRBS data, 135,105 CpG sites were tested. A linear model including line and treatment effects was applied for each CpG site by tissue. DMCs associated with the treatment and line effects were identified (**Table 1**).

This analysis allowed to identify 76 treatment-associated DMCs in blood (**Figure 2.A**), 8 treatment-associated DMCs in liver (**Figure 2.B**) and 4 treatment-associated DMCs in brain samples (**Figure 2.C**). Among all these DMCs, 36 (24 hypomethylated, 12 hypermethylated as compared to the control), 2 (hypomethylated) and 4 (3 hypomethylated, 1 hypermethylated) could be attributed to

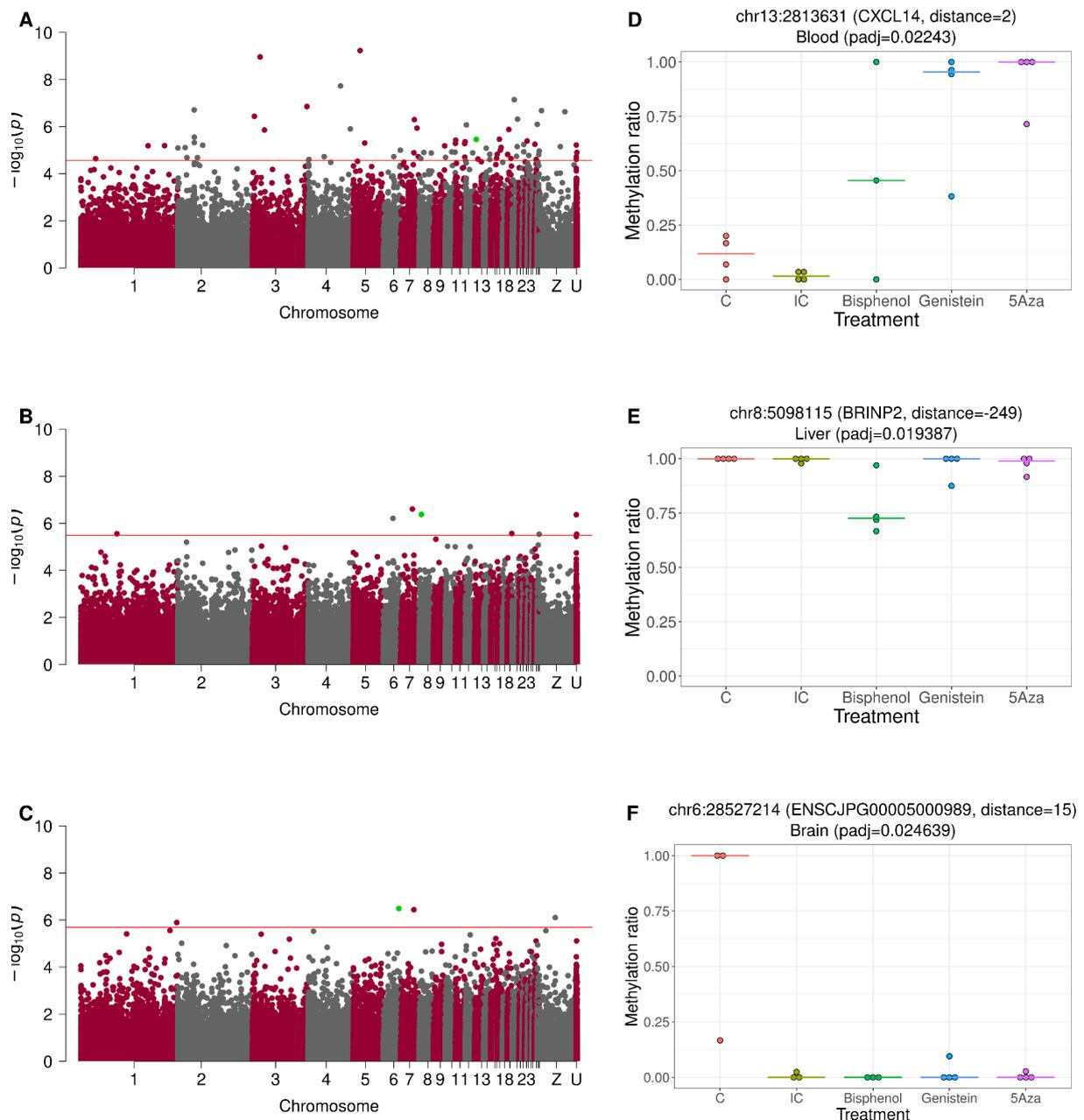
the effect of the egg injection, in blood (**Figure 3.A**), liver (**Figure 3.B**) and brain (**Figure 3.C**) samples, respectively.

Some DMCs are related to the injection of one or multiple molecules. One DMC identified in blood samples (chromosome 13, locus 2,813,631, adjusted p-value: 0.02) is hypermethylated in animals injected with genistein and 5Aza in the egg (**Figure 2.D**). Another DMC identified in liver samples (chromosome 8, locus 5,098,115, adjusted p-value: 0.02) is hypomethylated only in animals injected with bisphenol in the eggs (**Figure 2.E**). The 4 DMCs identified in the brain samples seem to be caused by the injection, and not, or not only, by the molecule themselves (**Figure 2.F**).

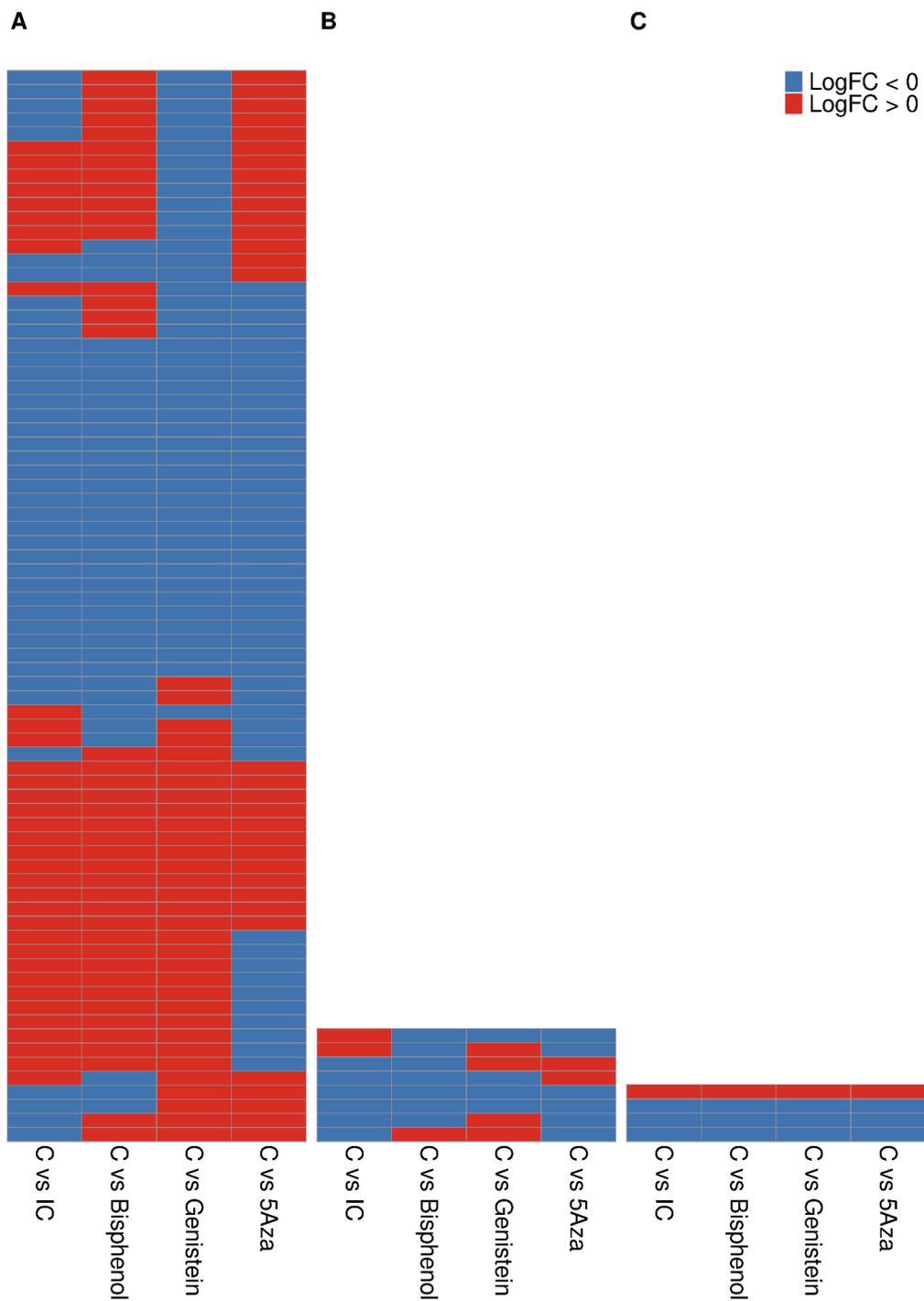
Thousands of line-associated DMCs were detected (**Table 1, Supplementary Figure S4**). When analysing the methylation profile of these DMCs, clusters separating lines are observed (**Supplementary Figure S4.A-C**). The intersection between line-associated DMCs shows 305 common DMCs between blood and liver, 230 between brain and liver, 63 between blood and brain and 137 between the three tissues. Most of the line-associated DMCs (3,300, 1,636 and 1,181 DMCs in respectively liver, blood and brain) are tissue-specific (**Supplementary Figure S4.D**).

Linear model	Tissue	Treatment DMCs	Line DMCs	Treatment DMRs	Line DMRs	Tested CpG
~ treatment + line	Blood	76	2,141	0	7	135,105
	Liver	8	3,972	0	47	
	Brain	4	1,611	0	4	

**Table 1.** DMC number identified with multifactor analysis by tissue on RRBS data. Injected control group is the reference treatment for the test. A likelihood ratio test is performed for the treatment and line effects.



**Figure 2. Treatment-associated DMCs identified from RRBS data.** (A-C) Manhattan plots showing the distribution along the genome of (A) 76 DMCs identified in blood samples, (B) 8 DMCs identified in liver samples, (C) 4 DMCs identified in brain samples. Each point refers to a CpG site ( $n=135,105$ ). The horizontal red bar represents the significance threshold for a False Discovery Rate (FDR) of 5%. X-axis: chromosome names (1-28, W, Z, U) with “W” and “Z” for sex chromosomes and “U” for unassigned contigs. Y-axis: the p-values for differential methylation levels by treatments, without multiple testing correction. (D-F) Boxplots of methylation levels for illustration DMC identified from (D) blood (here, a DMC hypermethylated in each treatment), (E) liver (hypomethylated in the Bisphenol treated group) and (F) brain (hypomethylation following any injection). For each DMC, the chromosome, position and BH (Benjamini-Hochberg) adjusted p-value are indicated in the title. The methylation rate is the ratio:  $C/(C+T)$ .



**Figure 3. Specificities and similarities of the effects of each experimental condition on the DNA methylation levels at treatment-associated DMCs.** Changes in DNA methylation between treatment groups and the non-injected control (T). 76, 8 and 4 DMCs detected in (A) blood, (B) liver and (C) brain, respectively. The injection effect can be detected by full blue or red lines : 36 in blood, 2 in liver and 4 in brain. Blue and red colours refer to hypo-methylated ( $\logFC < 0$ ) and hyper-methylated ( $\logFC > 0$ ) cytosines, respectively. IC refers to the injected control.

### ***Genes and biological functions associated with adult DMC after in-ovo treatments***

Each DMC was annotated by identifying the genomic features it falls into, its closest gene, and the distance between each DMC and its associated gene (**Supplementary Figure S5, Table 2**). The analysed CpGs sites were mostly found in promoters, exons and introns while the blood DMCs were found mostly in exons, introns and intergenic regions. A total of 23.9% of analysed CpG are located in promoter, 67.3% in gene and 8.8% in intergenic regions. The majority of CpGs falling into a gene are located in introns (33.8% of total analysed CpG). Analysed CpG could also fall into other types of genomic regions: 5'UTR (1.03%), 3'UTR (0.436%) and exons (18.3%) (**Supplementary Figure S5**). In total, the treatment-associated blood DMCs were associated with 34 different genes (**Table 2**).

An over-representation analysis of gene ontology (GO) terms was performed. We particularly focused on blood DMCs for further analysis, as the number of DMCs identified in the liver and brain was too low. After the over-representation test of functional GO terms associated with biological processes (BP), 42 enriched terms from the blood gene list were identified. These significant BP GO terms were grouped into 11 clusters (**Supplementary Figure S8.A**). The impacted BP GO terms can be classified as follows: response to stimulus and molecule such as lipid (cl1,2), developmental processes and cell differentiation (cl3-5,7,8,10,11), macromolecule modification (cl9), and finally cellular and pattern specification processes (cl6). These genes are also involved in different cellular component (CC) terms such as contractile fibre (cl1) and protein-containing complex (cl2). About the MF terms, these genes are enriched in nuclear hormone receptor, transcription factor and molecule binding (cl2-5,7,8) and lipid transporter (cl6) (**Supplementary Figure S8.A**).

Tissue	Chr	Locus	Gene name	Tissue	Chr	Locus	Gene name	
Blood	2	13289752	SVIL	Blood	17	7223123	PTGS1	
	3	14412959	DHX57		19	399749	MTMR4	
	4	1190567	P2RY4		19	6468713	BCAS3	
	4	76841699	CTBP1		20	9152614	RSPO4	
	5	12279082	PARVA		22	2754597	GFRA2	
	5	20729975	C11orf49		23	4629351	YARS1	
	6	19280666	NKX2-3		24	2483952	GRAMD1B	
	6	31255277	INPP5A		27	881656	FMNL1, MAP3K14	
	7	24573939	DES		28	318982	MARCHF2, KANK3	
	10	586547	ZNF609		28	3744876	PTPRS	
	11	417985	GNAO1		KQ966734.1	7145	PLD3	
	11	16731701	VPS9D1, ZNF276		LGE22C19W 28_E50C23	1084992	DNAJC22	
	13	2813631	CXCL14		Z	38013432	PTCH1	
	14	5212352	EPN2		Liver	6	17841436	CYP26A1
	14	11471879	SEPTIN12			8	5098115	BRINP2
	15	10958862	TBX3			19	8133550	MNT
	17	292283	EXD3		Brain	7	24496371	ABCB6
17	4134753	PTPA						

**Table 2. Genes associated with a treatment DMCs.**

## 2 - Whole genome analysis of the adult blood DNA methylome following in-ovo treatments

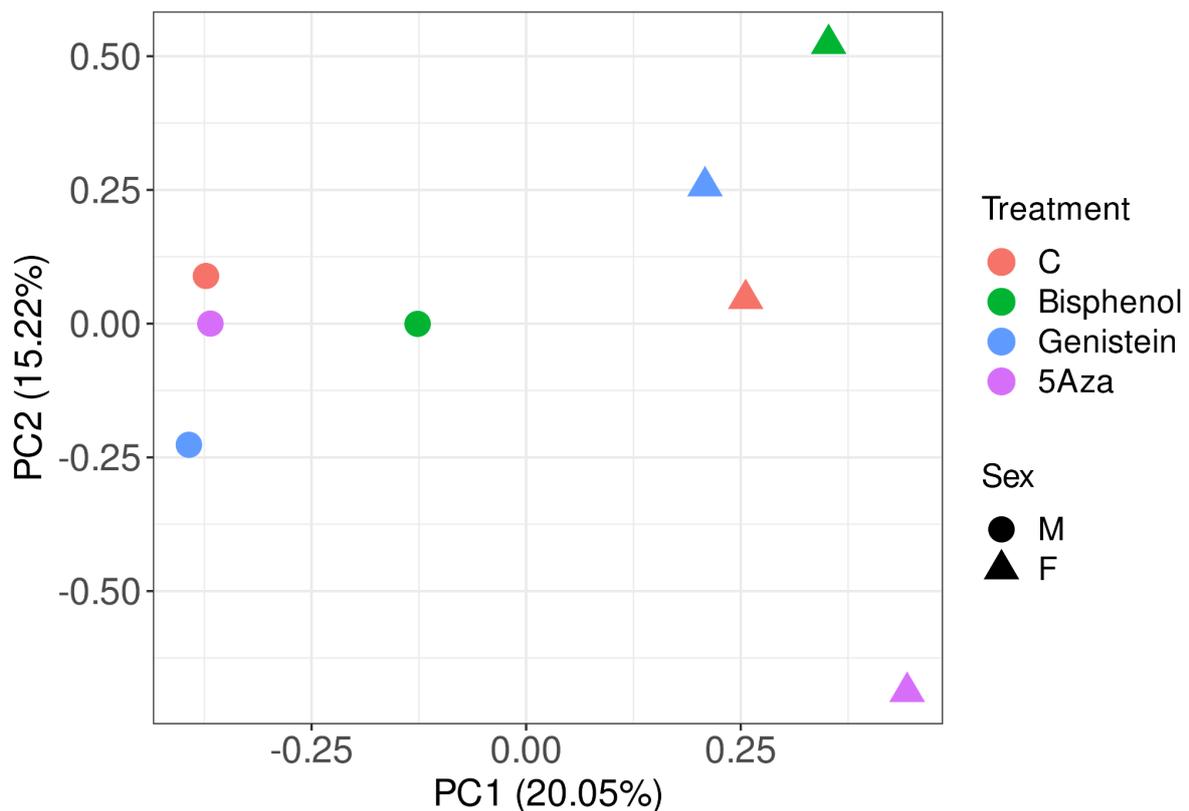
### DNA methylation profiling by WGBS

In order to detect the genome-wide effect of *in-ovo* treatments, WGBS sequencing data, also based on bisulfite treatment of DNA, were produced. We performed WGBS analysis on blood samples. Blood samples of S+ quails studied in the RRBS experiment, pooled per sex and treatment (non-injected control, bisphenol, genistein and 5Aza) were analysed.

After nf-core/methylseq processing of the WGBS data, the sequence data showed an average of  $143.1 \pm 18.5$  M reads per sample (**Supplementary Table S2.B**). The initial average number of stranded CpGs was  $15,024,654 \pm 612,121$  (83.7% of all stranded CpG sites in the whole genome). A high number of putative SNPs from the VCF file and from the gemBS analysis were removed with an average of  $6,790,227 \pm 138,668$  CpGs (75.7%) and  $5,263,066 \pm 22,980$  (58.6%), respectively. At the end of the data preprocessing, a total of 4,588,907 CpG sites were selected for further

analysis, corresponding to 51.1% of quail genome CpGs. Each CpG site has an average coverage of  $143 \pm 19.1$ , a methylation level about  $62 \pm 2.2\%$  and an average of  $64.5 \pm 1.1\%$  of the reads were aligned.

We performed a PCA on the methylation levels of retained CpG sites from the 8 WGBS pools (**Figure 3**). PC1 (explained variance of 20.33%) clearly separates the females and males samples. A treatment separation is not visible in these first two components.



**Figure 3. PCA of CpG methylation from WGBS data.** The shapes differentiate the sexes: triangles and circles for females and males, respectively. The colours refer to the treatments: red, green, blue and pink for non-injected controls (C), bisphenol A (BPA), genistein and 5Aza, respectively. Percentages indicate the proportion of variance explained by each axis. Number of strictly common CpGs between all samples:  $n = 4,588,907$ .

### Differential methylation analyses

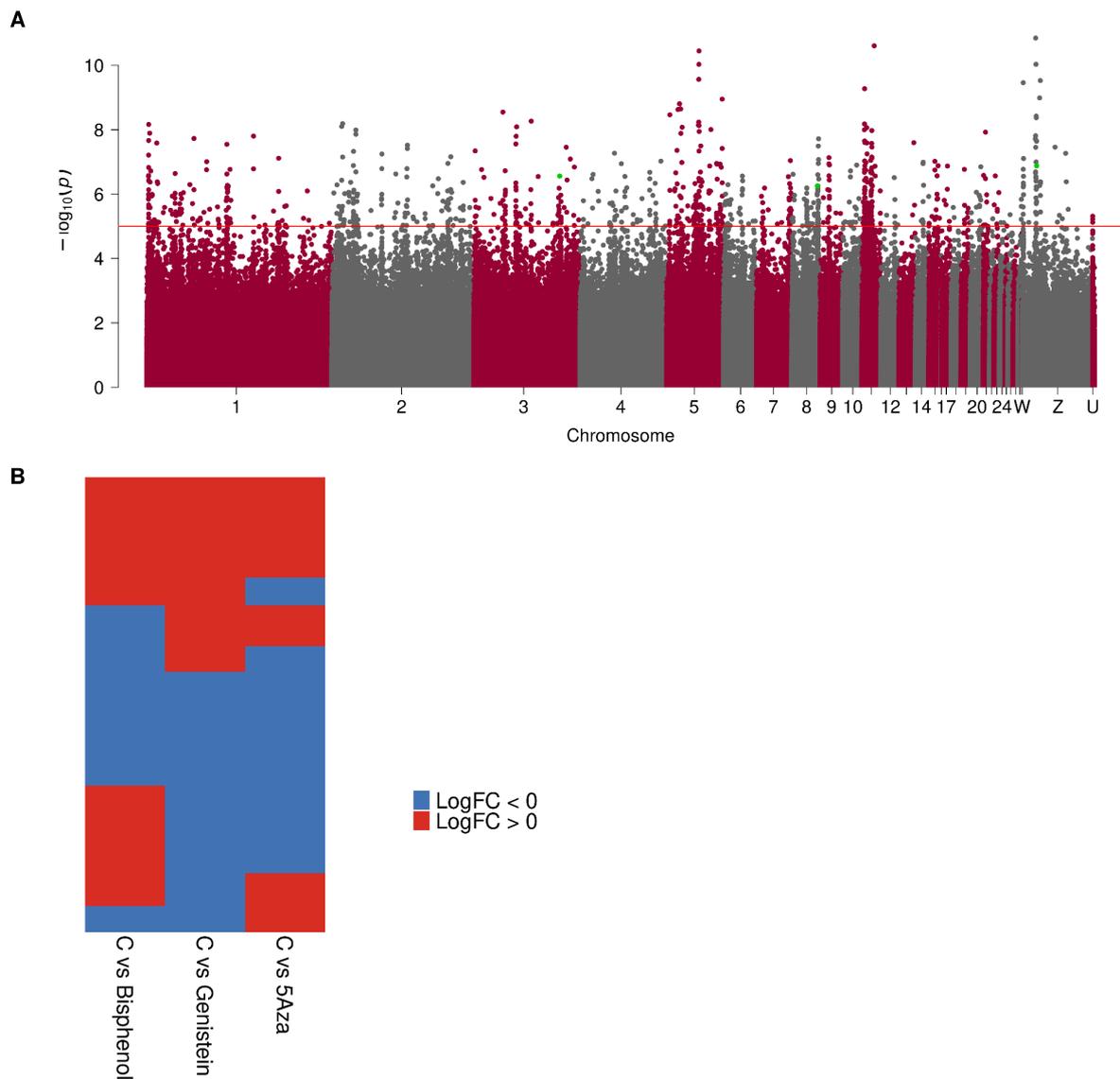
A total of 4,588,907 CpG sites were tested for differential methylation analysis. A linear model including sex and treatment effects was performed for each CpG site. This analysis identified 918 DMCs treatment-associated in blood (**Table 3**). Significant DMCs are observed all along the genome (**Figure 4.A**). The methylation profile analysis of these 918 DMCs highlights clusters with differential methylation levels between treatment groups (**Figure 4.B**), with DMCs hypomethylated or hypermethylated by one or multiple molecules (**Figure 4.B**). Among all these DMCs,

486 are hypomethylated or hypermethylated by one (**Figure 5.A**) or two molecules (**Figure 5.B**). For the other DMCs, 230 are hypomethylated and 202 are hypermethylated by the three molecules (**Figure 5.C**), which could be due to an *in-ovo* injection effect, independent of the nature of the injected molecule. The intersection of treatment-associated DMC identified in the three tissues from RRBS data and from WGBS data does not reveal common DMC (**Supplementary Figure S9.B**).

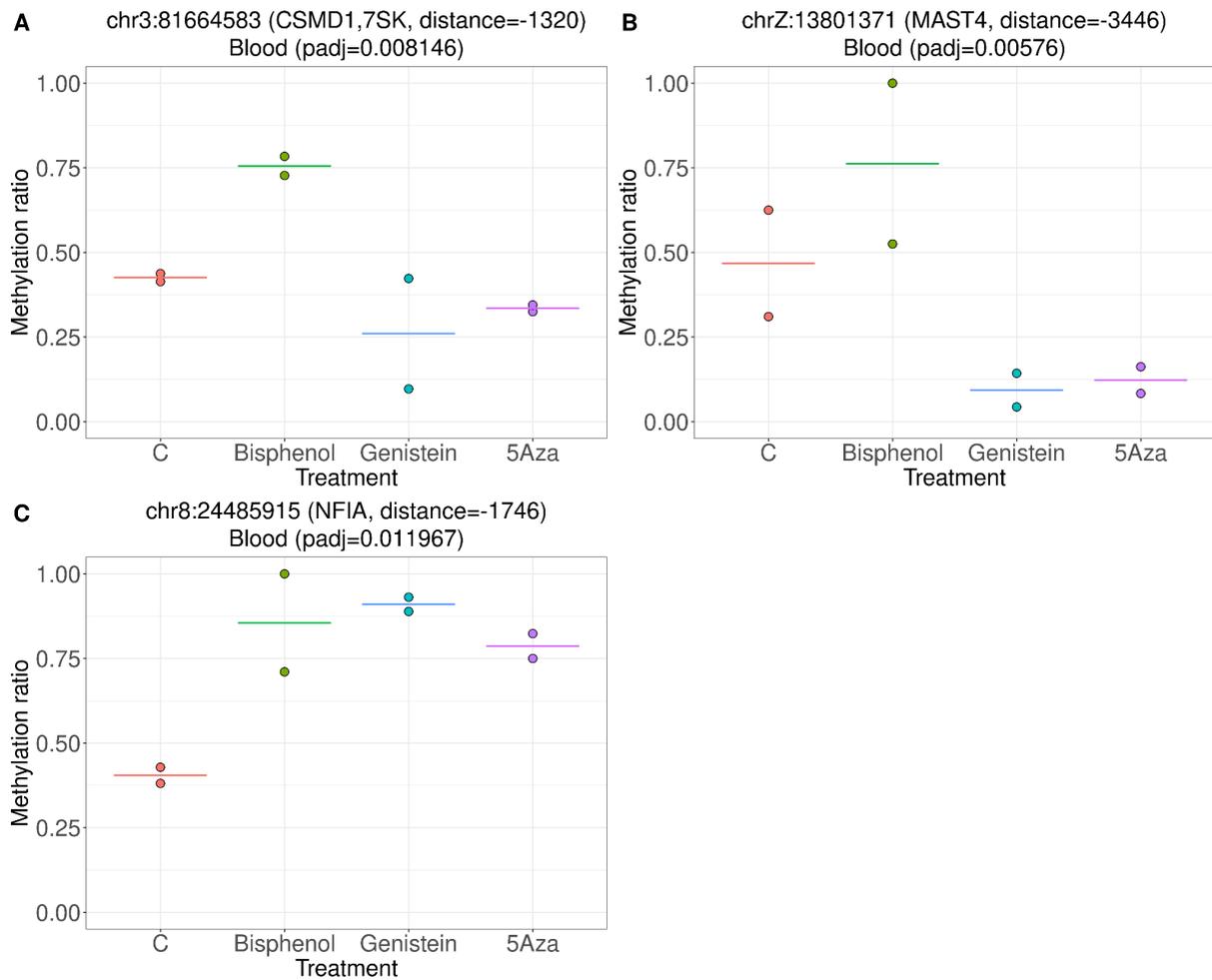
A total of 10,515 sex-associated DMCs and 113 differentially methylated regions (DMRs) between sexes were detected in WGBS data (**Table 3, Supplementary Figure S6**). The sex DMCs distribution along the genome shows that DMCs are distributed along the whole genome with a strong bias towards the sexual chromosome Z (**Supplementary Figure S6.A**). Clustering of sex DMCs led to the identification of male-specific and female-specific methylated CpG (**Supplementary Figure S6.B**).

Linear model	Tissue	Treatment DMCs	Sex DMCs	Treatment DMRs	Sex DMRs	Tested CpGs
~ treatment + sex	Blood	918	10,515	0	113	4,588,907

**Table 3.** DMC number identified with multifactor analysis by tissue on WGBS data. Non-injected control group is the reference treatment for the test. A likelihood ratio test and a Wald test are performed for the treatment and sex effects, respectively.



**Figure 4. Methylation profile of the 918 treatment-associated DMCs identified from WGBS data.** (A) Manhattan plot showing the DMCs distribution along the genome. Each point refers to a CpG site ( $n=4,588,907$ ). The horizontal red bar represents the significance threshold for a FDR of 5%. On the x-axis, the chromosome names (1-28, W, Z, U) with “U” refers to unassigned contigs. On the ordinate, the p-values (without multiple testing correction). The alternating red and grey bands represent the separation between chromosomes. (B) Heatmap of logFC comparison between treatment groups of the DMCs. LogFC results from the comparison of the methylation level between treatment groups and the non-injected control (T). Blue and red colours refer to hypo-methylated ( $\text{logFC} < 0$ ) and hyper-methylated cytosines ( $\text{logFC} > 0$ ), respectively. T refers to the non-injected control.



**Figure 5. Examples of DMCs identified in blood from WGBS data.** Example of a (A) hypermethylated DMC associated with the bisphenol (chr3:81,664,583, CSMD1 gene), (B) hypomethylated DMC associated with genistein and 5Aza (chrZ:13,801,371, MASTA gene) and (C) hypermethylated DMC associated with the three molecules (chr8:24,485,915, NFIA gene). For each DMC, the chromosome, the position and the p-value adjusted by the (Benjamini-Hochberg) method are provided. The methylation rate is the ratio:  $C/(C+T)$ .

### **Gene Ontology enrichment analysis**

After the annotation of treatment-associated DMCs and CpGs from WGBS data, their repartition by genomic feature was analysed (**Supplementary Figure S8**). A total of 5.7% of analysed CpGs are located in promoter, 79.6% in gene and 14.7% in intergenic regions. The majority of CpGs falling into a gene are located in introns (59.9% of total analysed CpG). Analysed CpG could also fall into other types of genomic regions: 5'UTR (0.142%), 3'UTR (0.329%) and exons (9.378%). The DMC repartition is broadly similar to the CpG repartition with a higher count in introns and intergenic regions. The DMCs were found in 388 different genes.

After the enrichment test of GO terms associated with biological processes (BP), 323 enriched terms from the gene list were identified. These enriched BP GO terms were grouped into 44 clusters (**Supplementary Figure S8.C**) and highlighted an important role of processes which can be classified as follow: developmental processes (cl1,3,5,7,24,39), developmental process involved in reproduction (cl6,31), nervous system development (cl8,9,18), response to stimulus (cl26,27,29,33), cellular organisation (cl11,15), double-strand break repair (cl22), cell cycle (cl30), regulation processes (cl2,4,10,13,14,16,19-21,23,25,28,32,34-38,41), interaction between organisms (cl40), and macromolecule transport (cl12,17,42-44). Some of the enriched GO terms associated with the lipid transport (cl42,44), tissue development (cl1,3,5-7,9,24,39) and cellular response to stimulus (cl26,27), are also highlighted in RRBS data from the blood samples. As for the analysis of blood samples from RRBS data, the cellular component (CC) assembly analysis allows to identify genes involved in protein-containing complexes (cl13). Regarding the molecular function (MF) terms, these genes are enriched in molecule binding (cl1-7,13,15) (**Supplementary Figure S8.A**) as for the RRBS analysis.

### **Discussion**

In a previous study, significant effects of injection of endocrine disruptors or a methylation modifier into the egg on adult phenotypes, notably on the growth of quails, were highlighted (Cerutti et al., 2022). In this subsequent study, we analysed in the same samples, from 4 different quail lines, the effects on the DNA methylation of blood, liver and brain of the three molecules expected to modify the embryonic environment after injection into the egg. It is indeed important to better understand the mechanisms linking early exposure to active molecules to the occurrence of developmental modifications or pathologies. In contrast to mammalian blood, the avian blood contains nucleated red blood cells (Scanes, 2015), making this tissue relatively homogeneous in birds (more than 90% of the DNA isolated originates from erythrocytes, Verhulst et al., 2016). The liver, key organ of energy homeostasis, and brain, the main place where behaviours are set up, were selected as previous studies suggested impact of *in-ovo* chemical exposure on weight and behaviours (Leroux et al., 2017, Cerutti et al., 2022).

Bisulfite treatment converts unmethylated cytosines to uracils that are sequenced as thymines. For this reason, a SNP in a CpG site can be inaccurately interpreted as a fully unmethylated site. We applied several filtration steps to minimise this risk: during the preprocessing step, the most prevalent SNPs from public datasets and those detected using the gemBS tool (Merkel et al., 2019) from the samples were removed from the list of CpGs' positions. Furthermore, a systematic zero methylation level at a CpG site in several tissues might correspond to a potential SNP absent from our filtering set, therefore we also filtered sites where all samples from a quail line were at zero methylation level. Although the gemBS tool was developed to search for variants from sodium bisulfite treated genomic sequences, as the genomic sequences are modified by the treatment, this makes the detection of SNPs less reliable than if this detection had been made after conventional sequencing of untreated DNA. There are other sequencing methods for analysing DNA methylation that do not require treatment of DNA with sodium bisulfite, such as Oxford Nanopore Technology (ONT) sequencing (Lu et al., 2016). Detection of variants from this type of data would potentially be more reliable but this sequencing technique is more expensive than RRBS or WGBS, and its use in agronomy is at an early stage.

The aligned reads rate is consistent with the alignment success from sequences initially treated with sodium bisulfite. As expected, the number of CpG sites retained after preprocessing and the average methylation level were higher for WGBS than RRBS (Doherty and Couldrey, 2014). A higher fraction of analysed CpGs are located in promoter regions in RRBS than in WGBS. The differences observed between these two technologies is expected, given the use of a specific restriction enzyme (Msp1) for RRBS. The recognition sites ("CCGG" motif) of this enzyme are mostly located in high-CG regions (CpG islands) such as promoter regions (Meissner et al., 2005), known to have low methylation levels (Bird, 2002). The analyses of RRBS data revealed a tissue-specific DNA methylation profile, which was expected (Lokk et al., 2014). The tissue-specific profile suggests distinct modes of gene expression regulation.

The differential methylation analyses revealed an *in-ovo* injection effect on DNA methylation in all tissues for several CpGs, independently of the nature of the injected molecule, as it could be observed in the sesame-oil carrier injection controls. This observation highlights the influence of a benign intervention on the egg at an early stage of development on the adult methylomes. In a previous study (Cerutti et al., 2022), we showed a difference of the weight between the injected and non-injected controls, resulting in a weight decrease of injected quails at 78 days. The observed effect could be induced by the injection in the eggs of sesame oil and/or ethanol. Indeed, Oppold et al. (2015) have previously observed an influence of the carrier control treatment by itself on adult DNA methylation profiles in Asian tiger mosquito *Aedes albopictus*. Other studies have also shown a long-term effect of early embryonic exposure to ethanol (Legault et al., 2018, Wallén et al., 2021).

We also detected an effect of each treatment molecule on DNA methylation profiles. Some treatment-associated DMCs were identified in all tissues from RRBS and WGBS data, but most effects were tissue specific. A higher number of treatment-associated DMCs were identified from blood compared to brain and liver. This could be due to the fact that blood is a more homogenous tissue (in birds and other species with nucleated red blood cells) than the brain and the liver, resulting in higher statistical power in the blood as compared to the liver and the brain, even if an hypothesis of tissue-specific response to the chemicals cannot be ruled out. A higher number of treatment-associated DMCs were identified from WGBS than from RRBS, in accordance with the higher number of analysed CpGs.

No common treatment-associated DMC was found between RRBS and WGBS analysis (**Supplementary Figure S9.B**). This can be partly explained because there are only 30,923 common CpGs between the two technologies after preprocessing (**Supplementary Figure S9.A**), while 104,182 CpG sites could be measured by RRBS only but not by WGBS. RRBS is enriched in GC rich regions such as CpG islands, as it analyses DNA fragments that are bordered by two CCGG sites. WGBS sequencing on Illumina suffers from a drop of sequencing depth on CpG-rich regions. Therefore many CpG sites in CpG rich regions did not pass our minimal coverage threshold for WGBS (at least 5 reads in 80% of the samples, **Supplementary Figure S9.C**), and could only be measured by RRBS. Higher sequencing depths or a lower GC bias for WGBS could have allowed a WGBS - RRBS comparison, but for these datasets they largely measured two distinct sets of CpG, and therefore constitute complementary analyses. Nevertheless, we can see a good correlation (0.85) between the 30,923 common CpGs from RRBS and WGBS data (**Supplementary Figure S9.D**).

The treatment-associated DMCs identified in blood were associated with 34 and 388 different genes for RRBS and WGBS, respectively. Genes were grouped by mechanism of interest for all tissues. Results revealed that *PTCH1* (Patched 1) and *PTPRS* (Protein Tyrosine Phosphatase Receptor Type S) genes, involved in the negative regulation of developmental growth, were impacted, which might explain some phenotypic differences we previously reported on growth (Cerutti et al. 2022). Some identified genes are involved in the developmental process of tissues such as muscles and glands such as *SVIL* (Supervillin), *CXCL14* (C-X-C Motif Chemokine Ligand 14) and *TBX3* (T-Box Transcription Factor 3). Other genes identified are involved in the response to transforming growth factor beta such as *ZFH3* (Zinc Finger Homeobox 3), *WNT5A* (Wnt Family Member 5A), and *NLK* (Nemo Like Kinase) and in developmental processes (*MYOM2*-Myomesin 2, *CTBP1*-C-Terminal Binding Protein 1, *EXOC4*-Exocyst Complex Component 4). *EXOC4* has the highest count of associated DMCs (n=18). This gene encodes a protein which is a component of the exocyst complex, essential for the biogenesis of epithelial cell surface polarity. *CTBP1* (C-Terminal Binding Protein 1) was identified from RRBS and WGBS data. This gene encodes a protein implicated in regulation of gene expression

during development. The inactivation of this gene would lead to a decrease of the body size in mice (Hildebrand and Soriano, 2002). These genes, involved in developmental processes, may be involved in the effect, previously highlighted, of the *in-ovo* injection of endocrine disruptors on weight and growth of quails directly injected (Cerutti et al., 2022) or quails whose ancestors, three generations before, received an injection of genistein (Leroux et al., 2017).

The *BRSK2* (BR Serine/Threonine Kinase 2) gene and the *WNT5A* (Wnt Family Member 5A), both implicated in cellular response to stimulus, were impacted. Some DMCs were associated with *MNT* (MAX Network Transcriptional Repressor), *CYP26A1* (Cytochrome P450 Family 26 Subfamily A Member 1) and *BRINP2* (BMP/Retinoic Acid Inducible Neural Specific 2) genes, all also associated with the cellular response to stimulus. Some genes, such as *PTCH1*, *PTPRS* and *RBPJ* (Recombination Signal Binding Protein For Immunoglobulin, *NFIA*-Nuclear Factor I A), are involved in central nervous system development. Other impacted genes such as *TBC1D23* (TBC1 Domain Family Member 23), *ANO4* (Anoctamin 4), *ABCC1* (ATP Binding Cassette Subfamily C Member 1) and *ATP9B* (ATPase Phospholipid Transporting 9B) were implicated in the lipid transport and the response to cholesterol. This could affect the transport of cholesterol and then impact the cortisol secretion, a stress hormone secreted by adrenal glands. A deregulation of the life cycle of this kind of hormone could lead to behavioural changes. This hypothesis is supported by a previous work showing that the injection of genistein in quail ancestors can affect the behaviour of the quail offspring, these quails being less stressed during a social isolation (Leroux et al., 2017).

Some genes are involved in protein-containing complexes, more precisely the transcription repressor complex (*MTA1*-Metastasis Associated 1) and ATPase complex (*PTPA*-Protein Phosphatase 2 Phosphatase Activator). Genes are enriched in transcription factor binding, more precisely the RNA polymerase II-specific DNA-binding transcription factor binding (*CTBP1*, *MTA1*, *FOXP1*, *TBX3*).

Some genes associated with DMC are involved in reproduction developmental processes, more precisely antral ovarian follicle growth and sex differentiation and development of primary sexual characteristics (*LHFPL2-LHFPL Tetraspan Subfamily Member 2*, *SCAPER-S-Phase Cyclin A Associated Protein In The ER*, *KIT-KIT Proto-Oncogene*, *Receptor Tyrosine Kinase*, *WNT5A-Wnt Family Member 5A*, *KIF18A-Kinesin Family Member 18A*, *BRIP1-BRCA1 Interacting Helicase 1*, *SALL1-Spalt Like Transcription Factor 1*, *CSMD1-CUB And Sushi Multiple Domains 1*). Even if no impact of *in-ovo* injection was identified on the egg laying of quails directly injected, the sexual maturity of the females whose ancestors were injected with genistein were impacted, the females starting to lay eggs later (Leroux et al., 2017).

Line-associated DMCs were found in all tissues and some of them are tissue-specific (**Supplementary Figure S4.D**). Some of these DMCs are hypomethylated or

hypermethylated in all tissues for a specific line suggesting that they may be under genetic control (**Supplementary Figure S4**). Therefore, methylation differentials were probably related to DNA polymorphisms involved in the regulation of methylation, as previously observed (Bell et al., 2011; Do et al., 2017). As no injected control group is available for WGBS data, it is likely that some of the DMCs identified from the WGBS data are induced by the effect of the injection alone and not directly related to the effect of a molecule.

In conclusion, this study shows that the *in-ovo* injection of chemicals during embryonic development in quail involves lasting epigenetic changes in the blood, liver and brain that may impact the growth and neural development of the animals. Several CpGs may serve as potential epigenetic biomarkers of exposure with the aim of identifying markers of past exposure to toxicants. This study paves the way to analyse the influence of the early environment on future generations, particularly in the case of contamination by endocrine disruptors in the maternal diet and their accumulation in the egg.

### Acknowledgments

The authors want to acknowledge PEAT staff for their conscientious animal care, and Caillor (<https://caillor-genetics.com/home>) and Caille Robin (<https://www.cailles-robin.fr/>) breeders for providing animals. We are grateful to the entire staff of the PEAT experimental unit for their excellent animal care. We thank C. Leterrier for providing the S+ experimental line and helpful discussions. RRBS Sequencing was performed by the GenomEast platform, a member of the "France Génomique" consortium (ANR-10-INBS-0009). We are grateful to the genotoul bioinformatics platform Toulouse Occitanie (GenoToul Bioinfo, 2018) for providing help, computing, and storage resources.

### Declaration of interest statement

This work was supported by the EPIQUES grant (INRAE, SelGen Metaprogramme). C. Cerutti is supported by a PhD scholarship from Région Occitanie and INRAE Animal Genetics Division.

### References

- Bardou, P., 2022. Coturnix japonica - VCF reference dataset. <https://doi.org/10.15454/TKMI53>
- Bell, J.T., Pai, A.A., Pickrell, J.K., Gaffney, D.J., Pique-Regi, R., Degner, J.F., Gilad, Y., Pritchard, J.K., 2011. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol* 12, R10. <https://doi.org/10.1186/gb-2011-12-1-r10>
- Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society:*

- Series B (Methodological) 57, 289–300.  
<https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bird, A., 2002. DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21. <https://doi.org/10.1101/gad.947102>
- Brionne, A., Juanchich, A., Hennequet-Antier, C., 2019. ViSEAGO: a Bioconductor package for clustering biological functions using Gene Ontology and semantic similarity. *BioData Mining* 12, 16. <https://doi.org/10.1186/s13040-019-0204-1>
- Cerutti, C., Leroux, S., Gourichon, D., Labrune, Y., David, I., Zerjal, T., Coustham, V., Devailly, G., Pitel, F., 2022. Short communication: Effects of in-ovo injection of endocrine disruptors and methyltransferase inhibitor on quail growth and egg-laying performances. *animal* 16, 100464. <https://doi.org/10.1016/j.animal.2022.100464>
- Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008. <https://doi.org/10.1093/gigascience/giab008>
- David, S.-A., Vitorino Carvalho, A., Gimonnet, C., Brionne, A., Hennequet-Antier, C., Piégu, B., Crochet, S., Couroussé, N., Bordeau, T., Bigot, Y., Collin, A., Coustham, V., 2019. Thermal Manipulation During Embryogenesis Impacts H3K4me3 and H3K27me3 Histone Marks in Chicken Hypothalamus. *Front. Genet.* 10, 1207. <https://doi.org/10.3389/fgene.2019.01207>
- Demars J, Labrune Y, Iannuccelli N, Deshayes A, Leroux S, Gilbert H, Aymard P, Benitez F, Riquet J. A genome-wide epistatic network underlies the molecular architecture of continuous color variation of body extremities. *Genomics.* 2022 May;114(3):110361. doi: 10.1016/j.ygeno.2022.110361.
- Do, C., Shearer, A., Suzuki, M., Terry, M.B., Gelernter, J., Grealley, J.M., Tycko, B., 2017. Genetic–epigenetic interactions in cis: a major focus in the post-GWAS era. *Genome Biol* 18, 120. <https://doi.org/10.1186/s13059-017-1250-y>
- Doherty, R., Couldrey, C., 2014. Exploring genome wide bisulfite sequencing for DNA methylation analysis in livestock: a technical assessment. *Front. Genet.* 5. <https://doi.org/10.3389/fgene.2014.00126>
- Ewels, P., Hüther, P., Spix, N., Miller, E., Nf-Core Bot, Hörtenhuber, M., Peltzer, A., Sven, F., Alneberg, J., Di Tommaso, P., Garcia, M.U., Davenport, C., Vikesh Ajith, Krueger, F., Harshil Patel, Alessia, Devailly, G., Syme, R., VIJAY, Menden, K., Noirod Céline, 2022. nf-core/methylseq: nf-core/methylseq version 2.3.0. <https://doi.org/10.5281/ZENODO.1343417>
- Ewels, P.A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M.U., Di Tommaso, P., Nahnsen, S., 2020. The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol* 38, 276–278. <https://doi.org/10.1038/s41587-020-0439-x>

Feil, R., Fraga, M.F., 2012. Epigenetics and the environment: emerging patterns and implications. *Nat Rev Genet* 13, 97–109. <https://doi.org/10.1038/nrg3142>

GenoToul Bioinfo, 2018. GenoToul Bioinformatics Facility. <https://doi.org/10.15454/1.5572369328961167E12>

Hildebrand, J.D., Soriano, P., 2002. Overlapping and Unique Roles for C-Terminal Binding Protein 1 (CtBP1) and CtBP2 during Mouse Development. *Molecular and Cellular Biology* 22, 5296–5307. <https://doi.org/10.1128/MCB.22.15.5296-5307.2002>

Krueger, F., Andrews, S.R., 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27, 1571–1572. <https://doi.org/10.1093/bioinformatics/btr167>

Krueger, F., James, F., Ewels, P., Afyounian, E., Schuster-Boeckler, B., 2021. FelixKrueger/TrimGalore: v0. 6.7-DOI via Zenodo. Zenodo <https://zenodo.org/record/5127899> (Accessed February 23, 2022).

Laine, V.N., Gossmann, T.I., Schachtschneider, K.M., Garroway, C.J., Madsen, O., Verhoeven, K.J.F., de Jager, V., Megens, H.-J., Warren, W.C., Minx, P., Crooijmans, R.P.M.A., Corcoran, P., The Great Tit HapMap Consortium, Adriaensen, F., Belda, E., Bushuev, A., Cichon, M., Charmantier, A., Dingemanse, N., Doligez, B., Eeva, T., Erikstad, K.E., Fedorov, S., Hau, M., Hille, S., Hinde, C., Kempnaers, B., Kerimov, A., Krist, M., Mand, R., Matthysen, E., Nager, R., Norte, C., Orell, M., Richner, H., Slagsvold, T., Tilgar, V., Tinbergen, J., Torok, J., Tschirren, B., Yuta, T., Sheldon, B.C., Slate, J., Zeng, K., van Oers, K., Visser, M.E., Groenen, M.A.M., 2016. Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nat Commun* 7, 10474. <https://doi.org/10.1038/ncomms10474>

Legault, L. M., Bertrand-Lehouillier, V., & McGraw, S. , 2018. Pre-implantation alcohol exposure and developmental programming of FASD: an epigenetic perspective. *Biochemistry and Cell Biology*, 96(2), 117-130. <https://doi.org/10.1139/bcb-2017-0141>

Leroux, S., Gourichon, D., Leterrier, C., Labrune, Y., Coustham, V., Rivière, S., Zerjal, T., Coville, J.-L., Morisson, M., Minvielle, F., Pitel, F., 2017. Embryonic environment and transgenerational effects in quail. *Genet Sel Evol* 49, 14. <https://doi.org/10.1186/s12711-017-0292-7>

Lokk, K., Modhukur, V., Rajashekar, B., Märtens, K., Mägi, R., Kolde, R., Koltšina, M., Nilsson, T.K., Vilo, J., Salumets, A., Tõnisson, N., 2014. DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. *Genome Biol* 15, 3248. <https://doi.org/10.1186/gb-2014-15-4-r54>

Lu, H., Giordano, F., Ning, Z., 2016. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics & Bioinformatics* 14, 265–279. <https://doi.org/10.1016/j.gpb.2016.05.004>

Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., Jaenisch, R., 2005. Reduced representation bisulfite sequencing for comparative high-resolution

DNA methylation analysis. *Nucleic Acids Research* 33, 5868–5877. <https://doi.org/10.1093/nar/gki901>

Merkel, A., Fernández-Callejo, M., Casals, E., Marco-Sola, S., Schuyler, R., Gut, I.G., Heath, S.C., 2019. gemBS: high throughput processing for DNA methylation data from bisulfite sequencing. *Bioinformatics* 35, 737–742. <https://doi.org/10.1093/bioinformatics/bty690>

Mills, A.D., Faure, J.-M., 1991. Divergent selection for duration of tonic immobility and social reinstatement behavior in Japanese quail (*Coturnix coturnix japonica*) chicks. *Journal of Comparative Psychology* 105, 25–38. <https://doi.org/10.1037/0735-7036.105.1.25>

Minvielle, B.F., Monvoisin, J.-L., Costa, J., Frenot, A., Maeda, Y., 1999. Changes in heterosis under within-line selection or reciprocal recurrent selection: an experiment on early egg production in Japanese quail. *J Anim Breed Genet* 116, 363–377. <https://doi.org/10.1046/j.1439-0388.1999.00218.x>

Oppold, A., Kreß, A., Vanden Bussche, J., Diogo, J.B., Kuch, U., Oehlmann, J., Vandegehuchte, M.B., Müller, R., 2015. Epigenetic alterations and decreasing insecticide sensitivity of the Asian tiger mosquito *Aedes albopictus*. *Ecotoxicology and Environmental Safety* 122, 45–53. <https://doi.org/10.1016/j.ecoenv.2015.06.036>

Qin, T., Zhang, X., Guo, T., Yang, T., Gao, Y., Hao, W., Xiao, X., 2021. Epigenetic Alteration Shaped by the Environmental Chemical Bisphenol A. *Front. Genet.* 11, 618966. <https://doi.org/10.3389/fgene.2020.618966>

Rasheed, S., Rehman, K., Shahid, M., Suhail, S., Akash, M.S.H., 2022. Therapeutic potentials of genistein: New insights and perspectives. *Journal of Food Biochemistry* 46. <https://doi.org/10.1111/jfbc.14228>

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. <https://doi.org/10.1093/bioinformatics/btp616>

Roussot O, Feve K, Plisson-Petit F, Pitel F, Faure JM, Beaumont C, et al. AFLP linkage map of the Japanese quail *Coturnix japonica*. *Genet Sel Evol.* 2003;35:559–72.

Scanes, C.G., 2015. Protein Metabolism, in: *Sturkie's Avian Physiology*. Elsevier, pp. 455–467. <https://doi.org/10.1016/B978-0-12-407160-5.00020-8>

Silva, L.B.A.R., Pinheiro-Castro, N., Novaes, G.M., Pascoal, G. de F.L., Ong, T.P., 2019. Bioactive food compounds, epigenetics and chronic disease prevention: Focus on early-life interventions with polyphenols. *Food Research International* 125, 108646. <https://doi.org/10.1016/j.foodres.2019.108646>

Sinclair, K.D., Rutherford, K.M.D., Wallace, J.M., Brameld, J.M., Stöger, R., Alberio, R., Sweetman, D., Gardner, D.S., Perry, V.E.A., Adam, C.L., Ashworth, C.J., Robinson, J.E., Dwyer, C.M., 2016. Epigenetics and developmental programming of

welfare and production traits in farm animals. *Reprod. Fertil. Dev.* 28, 1443. <https://doi.org/10.1071/RD16102>

Sun, R., Zhu, P., 2022. Advances in measuring DNA methylation. *Blood Science* 4, 8–15. <https://doi.org/10.1097/BS9.0000000000000098>

Verhulst, E.C., Mateman, A.C., Zwier, M.V., Caro, S.P., Verhoeven, K.J.F., Oers, K., 2016. Evidence from pyrosequencing indicates that natural variation in animal personality is associated with DRD 4 DNA methylation. *Mol Ecol* 25, 1801–1811. <https://doi.org/10.1111/mec.13519>

Wallén, E., Auvinen, P., & Kaminen-Ahola, N., 2021. The effects of early prenatal alcohol exposure on epigenome and embryonic development. *Genes*, 12(7), 1095. <https://doi.org/10.3390/genes12071095>

Wang, N., Sun, J., Pang, T., Zheng, H., Liang, F., He, X., Tang, D., Yu, T., Xiong, J., Chang, S., 2022. DNA Methylation Markers and Prediction Model for Depression and Their Contribution for Breast Cancer Risk. *Frontiers in Molecular Neuroscience* 15.

Yilmaz, B., Terekeci, H., Sandal, S., Kelestimur, F., 2020. Endocrine disrupting chemicals: exposure, effects on human health, mechanism of action, models for testing and strategies for prevention. *Rev Endocr Metab Disord* 21, 127–147. <https://doi.org/10.1007/s11154-019-09521-z>

## Effects of *in-ovo* injection of endocrine disruptors and a methyltransferase inhibitor on the DNA methylation of adult quails.

**Supplementary Table S1 - Metadata table for RRBS and WGBS data.** (A) RRBS data: Samples shown in bold were removed from the analysis due to low CpG counts. (B) WGBS data. 5Aza, IC and C refers to 5-Aza-2'-deoxycytidine, injected-control and non-injected control, respectively. M and F refers to males and females, respectively.

### A

ID	Tissue	Line	Treat	ID	Tissue	Line	Treat	ID	Tissue	Line	Treat
FRPL271	Liver	A	5Aza	FRPL291	Brain	A	5Aza	FRPL311	Blood	A	5Aza
FRPL272	Liver	A	Genistein	FRPL292	Brain	A	Genistein	FRPL312	Blood	A	Genistein
FRPL273	Liver	A	Bisphenol	FRPL293	Brain	A	Bisphenol	<b>FRPL313</b>	Blood	A	Bisphenol
FRPL274	Liver	A	IC	FRPL294	Brain	A	IC	FRPL314	Blood	A	IC
FRPL275	Liver	A	C	FRPL295	Brain	A	C	FRPL315	Blood	A	C
FRPL276	Liver	B	5Aza	FRPL296	Brain	B	5Aza	FRPL316	Blood	B	5Aza
FRPL277	Liver	B	Genistein	FRPL297	Brain	B	Genistein	FRPL317	Blood	B	Genistein
FRPL278	Liver	B	Bisphenol	FRPL298	Brain	B	Bisphenol	FRPL318	Blood	B	Bisphenol
FRPL279	Liver	B	IC	<b>FRPL299</b>	Brain	B	IC	FRPL319	Blood	B	IC
FRPL280	Liver	B	C	FRPL300	Brain	B	C	FRPL320	Blood	B	C
FRPL281	Liver	DD	5Aza	FRPL301	Brain	DD	5Aza	FRPL321	Blood	DD	5Aza
FRPL282	Liver	DD	Genistein	FRPL302	Brain	DD	Genistein	FRPL322	Blood	DD	Genistein
FRPL283	Liver	DD	Bisphenol	FRPL303	Brain	DD	Bisphenol	FRPL323	Blood	DD	Bisphenol
FRPL284	Liver	DD	IC	FRPL304	Brain	DD	IC	FRPL324	Blood	DD	IC
FRPL285	Liver	DD	C	FRPL305	Brain	DD	C	FRPL325	Blood	DD	C
FRPL286	Liver	S+	5Aza	FRPL306	Brain	S+	5Aza	FRPL326	Blood	S+	5Aza
FRPL287	Liver	S+	Genistein	FRPL307	Brain	S+	Genistein	FRPL327	Blood	S+	Genistein
FRPL288	Liver	S+	Bisphenol	FRPL308	Brain	S+	Bisphenol	FRPL328	Blood	S+	Bisphenol
FRPL289	Liver	S+	IC	FRPL309	Brain	S+	IC	FRPL329	Blood	S+	IC
FRPL290	Liver	S+	C	<b>FRPL310</b>	Brain	S+	C	FRPL330	Blood	S+	C

### B

ID	Tissue	Sex	Treat
1	Blood	M	5Aza
2	Blood	F	5Aza
3	Blood	M	Bisphenol
4	Blood	F	Bisphenol
5	Blood	M	Genistein
6	Blood	F	Genistein
7	Blood	M	C
8	Blood	F	C

**Supplementary Table S2. Table from the MultiQC report generated by the nf-core/methylseq pipeline on RRBS data (A) and WGBS data (B).** ID: pool identification, Reads (M): total number of sequenced reads (in millions), Aligned (%): percentage of aligned reads, CpG (K): total number after preprocessing of cytosines analysed (in thousands), mCpG (%): percentage of methylated cytosines in CpG context after preprocessing.

**A**

ID	Read (M)	Aligned (%)	CpG (K)	mCpG (%)
FRPL271	22.5	39.0	135.1	27.4
FRPL272	16.2	40.8	135.1	28.7
FRPL273	23.2	38.4	135.1	28.6
FRPL274	41.6	39.2	135.1	28.1
FRPL275	54.3	39.3	135.1	28
FRPL276	36.9	39.5	135.1	27.7
FRPL277	16.5	39.4	135	28.9
FRPL278	33.6	39.1	135.1	28.5
FRPL279	16.6	37.6	135.1	29
FRPL280	13.1	38.6	135	28.3
FRPL281	11.5	35.4	134.4	28.3
FRPL282	26.3	37.3	135.1	28.6
FRPL283	23.3	38.0	135.1	28.4
FRPL284	28.0	38.0	135.1	28
FRPL285	97.6	35.7	135.1	27.8
FRPL286	45.4	37.2	135.1	28.1
FRPL287	6.5	38.5	132.3	29.5
FRPL288	15.7	36.3	134.9	28.8
FRPL289	10.1	37.4	134.1	30.2
FRPL290	6.0	35.9	129.5	31.1

ID	Read (M)	Aligned (%)	CpG (K)	mCpG (%)
FRPL291	23.9	36.7	124.1	27
FRPL292	35.4	37.9	131.2	23.4
FRPL293	23.8	35.5	124.1	25.4
FRPL294	66.5	36.7	134.7	24.9
FRPL295	17.1	36.8	120.9	25.2
FRPL296	25.7	37.3	127.6	23.5
FRPL297	50.2	37.0	134.2	26.6
FRPL298	13.5	36.3	111.4	25.2
FRPL300	10.8	40.8	110.9	20.7
FRPL301	21.9	33.8	134.1	27.1
FRPL302	24.1	33.2	134.3	28.2
FRPL303	27.9	36.8	134.8	27.6
FRPL304	30.6	39.5	135	27.9
FRPL305	13.0	38.6	133.2	27.8
FRPL306	43.3	39.9	135.1	28
FRPL307	41.1	38.5	135	28.2
FRPL308	11.8	37.7	132.2	28.3
FRPL309	82.2	36.6	135.1	26.7

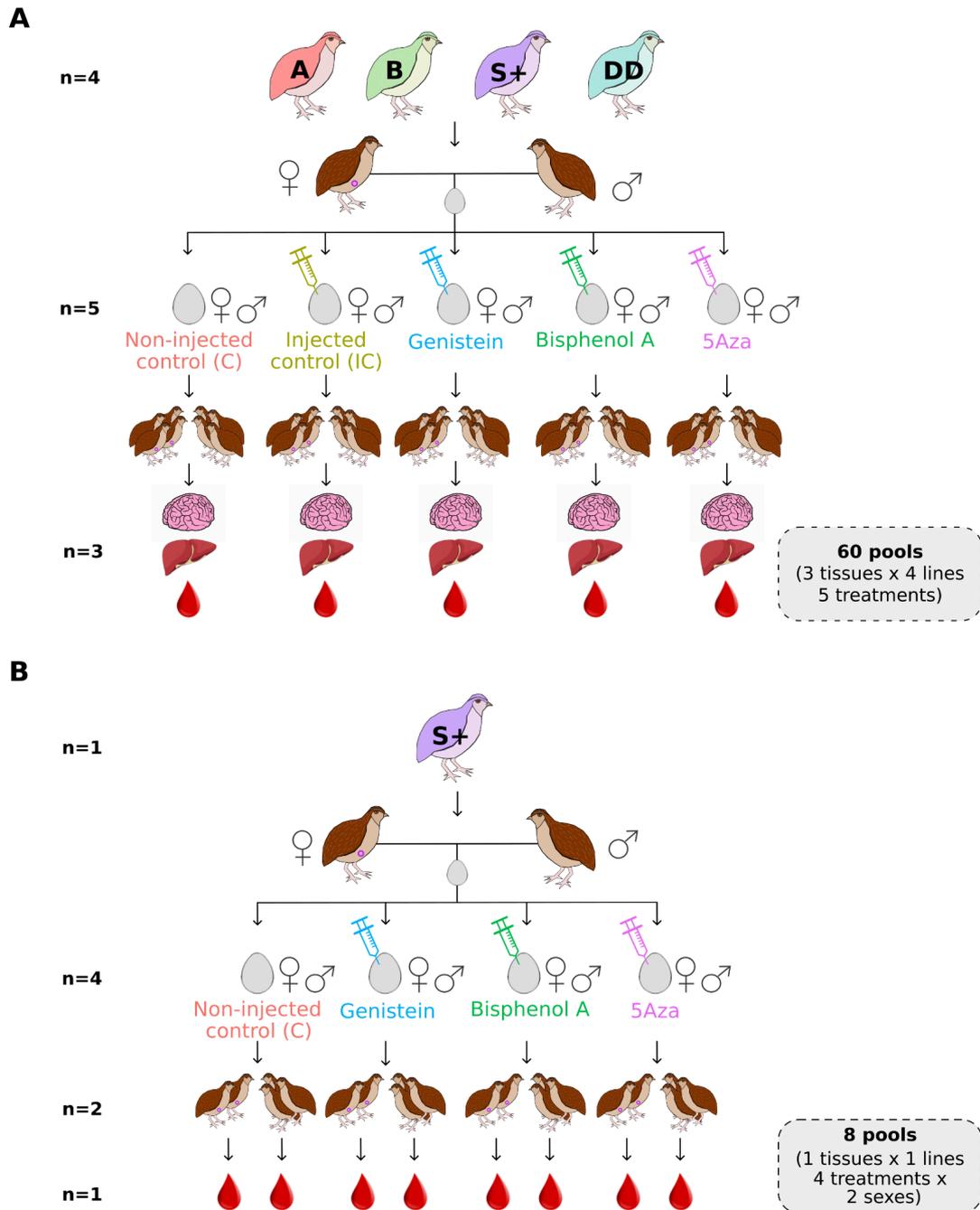
ID	Read (M)	Aligned (%)	CpG (K)	mCpG (%)
FRPL311	18.9	35.0	86.8	32.3
FRPL312	19.7	34.1	83.3	31.8
FRPL314	23.1	39.7	10.4	31.5
FRPL315	17.8	44.4	85.1	31.2
FRPL316	23.6	39.4	98.5	31.8
FRPL317	29.6	38.1	112.6	30.8
FRPL318	58.6	41.0	130.3	30.4
FRPL319	39.1	36.9	123.1	30.9
FRPL320	30.3	37.7	114.7	31.2
FRPL321	18.0	40.0	129.2	26.5
FRPL322	39.8	35.4	134.7	29.4
FRPL323	31.9	36.8	133.1	28.3
FRPL324	34.6	37.8	133.8	27.5
FRPL325	18.0	37.1	130.6	26.9
FRPL326	17.0	38.3	130.6	25.7
FRPL327	23.4	32.8	129.7	28.3
FRPL328	28.7	37.4	134.2	28.6
FRPL329	20.3	37.3	131.7	27
FRPL330	38.4	36.1	134.7	29.3

**B**

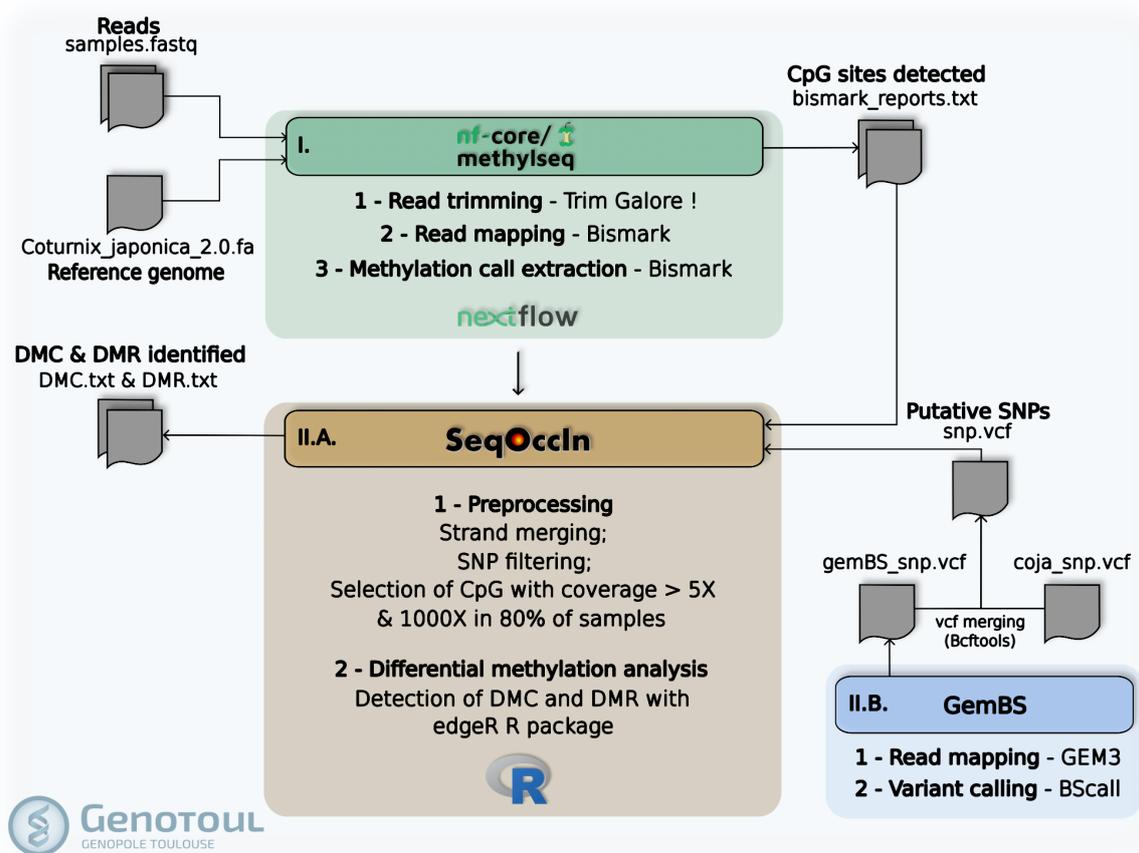
ID	Reads (M)	Aligned (%)	CpG (M)	mCpG (%)
1	157.2	65.4	4.6	58.8
2	122.4	64.1	4.6	64.4
3	180.2	65.5	4.6	61.5
4	128.2	62.9	4.6	63.9
5	132.2	65.3	4.6	60.1
6	143.5	62.8	4.6	63.1
7	130.4	65.4	4.6	60.2
8	150.3	64.2	4.6	64.4

**Supplementary Table S4. R packages used for the visual representations.**

Plot types	R packages	Package version	References
Histogram / Boxplot / Scatterplot / Dotplot / Barplot / PCA	<i>ggplot2</i>	3.4.2	H. Wickham. <i>ggplot2: Elegant Graphics for Data Analysis</i> . Springer-Verlag New York, 2016
Upset	<i>UpSetR</i>	1.4.0	Nils Gehlenborg (2019). <i>UpSetR: A More Scalable Alternative to Venn and Euler Diagrams for Visualizing Intersecting Sets</i> . R package version 1.4.0. <a href="https://CRAN.R-project.org/package=UpSetR">https://CRAN.R-project.org/package=UpSetR</a>
Heatmap	<i>pheatmap</i>	1.0.12	Raivo Kolde (2019). <i>pheatmap: Pretty Heatmaps</i> . R package version 1.0.12. <a href="https://CRAN.R-project.org/package=pheatmap">https://CRAN.R-project.org/package=pheatmap</a>
Manhattan	<i>qqman</i>	0.1.8	Stephen Turner (2017). <i>qqman: Q-Q and Manhattan Plots for GWAS Data</i> . R package version 0.1.4. <a href="https://CRAN.R-project.org/package=qqman">https://CRAN.R-project.org/package=qqman</a>
DMR line plot	<i>dmrseq</i>	1.10.1	Korthauer, K., Chakraborty, S., Benjamini, Y., and Irizarry, R. A. Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing. <i>Biostatistics</i> kxy007, <a href="https://doi.org/10.1093/biostatistics/kxy007">https://doi.org/10.1093/biostatistics/kxy007</a>

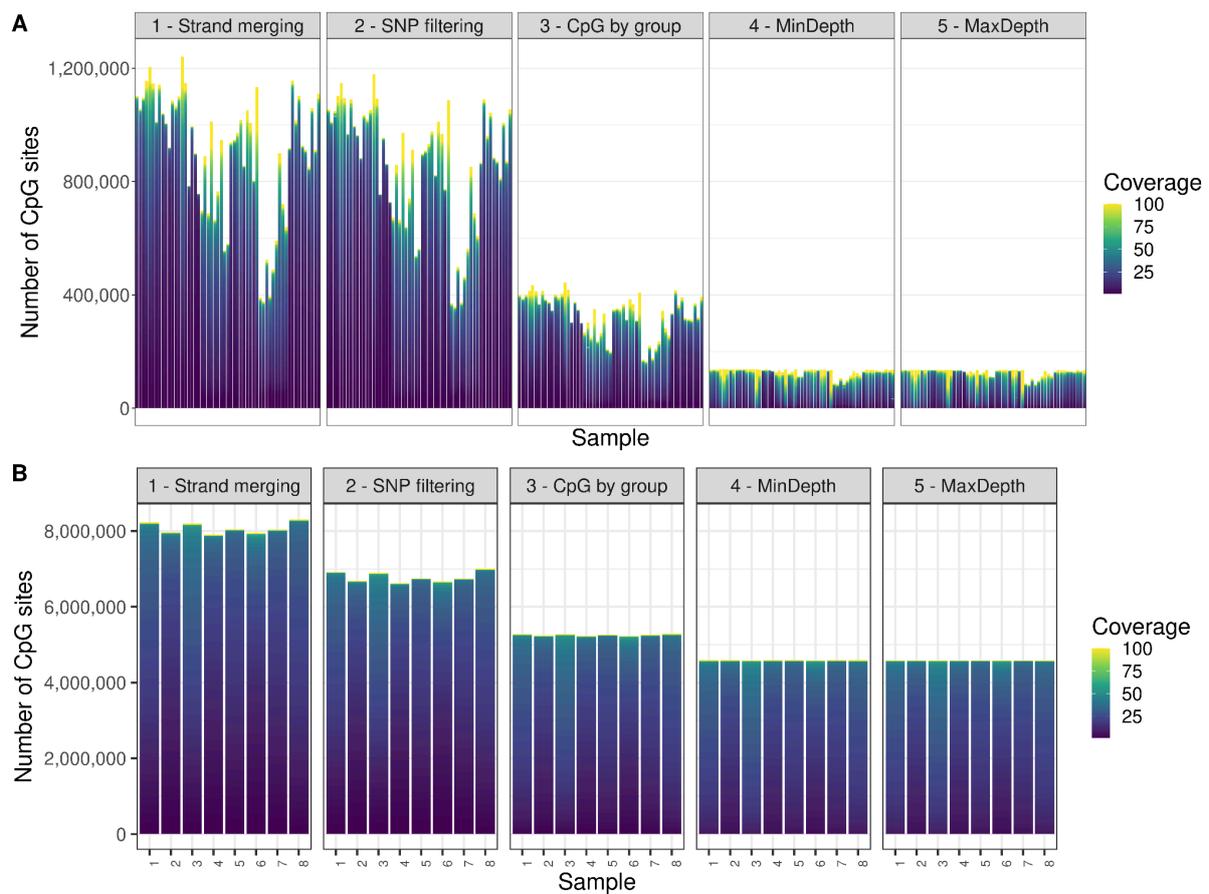


**Supplementary Figure S1. Experimental design from (A) RRBS and (B) WGBS.** (A) RRBS: brain, blood and liver samples were collected from quails (males and females) from five groups: two control groups (non-injected (C) and injected with sesame oil (IC)), two groups treated with endocrine disruptors (BPA and genistein) and one group treated with a methyltransferase inhibitor (5Aza). The quails studied were from four lines: two commercial (A and B) and two experimental (S+ and DD). 1979 quails separated into 2 batches were recovered. Tissues samples were taken from each quail. Samples with the same characteristics were pooled for a total of 60 samples. (B) Blood samples were collected from S+ quails (males and females) from four groups: one C and three injected with BPA, genistein and 5Aza. Samples with the same characteristics were pooled for a total of 8 samples.

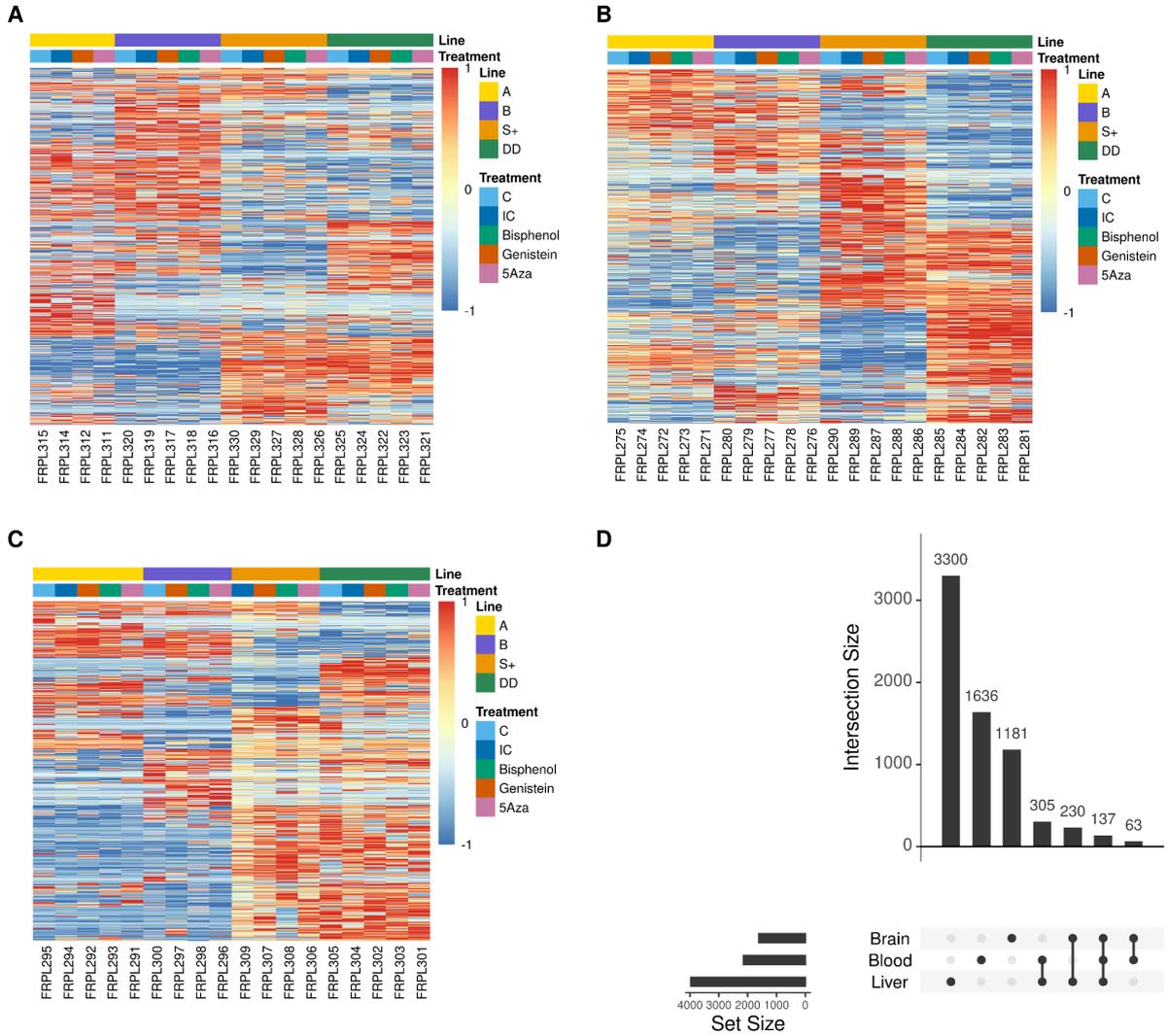


**Supplementary Figure S2. Methodology for bioinformatics analyses.** RRBS and WGBS reads were processed with the (I) *nf-core/methylseq* tool. (I.1) Reads were trimmed with Trim Galore and trimmed reads were (I.2) mapped to the quail reference genome by Bismark. (I.3) Bismark extracts methylation sites and generates a cytosine report for each sample. The outputs of *nf-core/methylseq* are analysed by an (II.A) in-house EdgeR-based R pipeline (*SeqOccin*) to detect differentially methylated cytosines (DMC) between experimental groups : (II.A.1) a pre-processing step and (II.A.2) a differential methylation analysis are performed. The pre-processing step consists of merging methylation and coverage information from both strands. A filter was then applied to remove SNPs and three additional filtering were applied : CpG sites needed to have a minimum of 5 and maximum of 1000 read coverage in 80% of samples. (II.B) The *gemBS* tool was used to detect SNPs present directly in samples : (II.B.1) GEM3 performs the short reads alignment and (II.B.2) BScall performs the variant calling extracting SNPs from BCF files. The retrieved SNP list was merged with the reference quail VCF file by *bcftools*. (II.A.1) The combined VCF file was used to filter putative SNP position from the Bismark outputs and the CpG sites with a zero methylation rate in all tissues considered as potential SNP were also filtered. (II.A.2) DMC were called by the *edgeR* package and differentially methylated regions (DMR) were called by *callDMR* (*DSS* package) from the *edgeR* outputs. The p-values were corrected with the Benjamini-Hochberg (BH) method, with a threshold of 5% defining significance after multiple testing corrections.

**Supplementary Figure S3. Evolution of the number of CpG sites during filtering steps for RRBS data (A) and WGBS data (B).** Samples are arranged along the x-axis according to their mean CpG coverage. Y-axis : Number of CpG sites (1) at the output of the nf-core/methylseq pipeline (after deduplication of CpG sites on the minus strand), (2) after SNP filtering, (3) after filtering of CpG sites with a zero methylation rate in all tissues, (3) after filtering out CpG sites with a coverage lower than 5 in at least 20% of the samples, (4) after filtering out CpG sites with a coverage higher than 1,000. (A) RRBS: initial number: 403,479 - 1,569,876 CpGs; after (1) and (2): 359,217 - 1,181,304 CpGs. (3): 166,826 - 446,983 CpGs. At the end of preprocessing: an average of 127,720 CpGs were obtained per sample (1.4% of the quail genome CpGs). (B) WGBS: initial number: 14,338,332 - 15,964,287 CpGs; after (1): 7,907,447 - 8,314,125 CpGs; (2): 6,623,713 - 7,008,383 CpGs; (3): 5,233,433 - 5,289,995 CpGs; at the end of preprocessing: an average of 4,588,415 CpGs were obtained per sample (51.1% of quail genome CpGs).

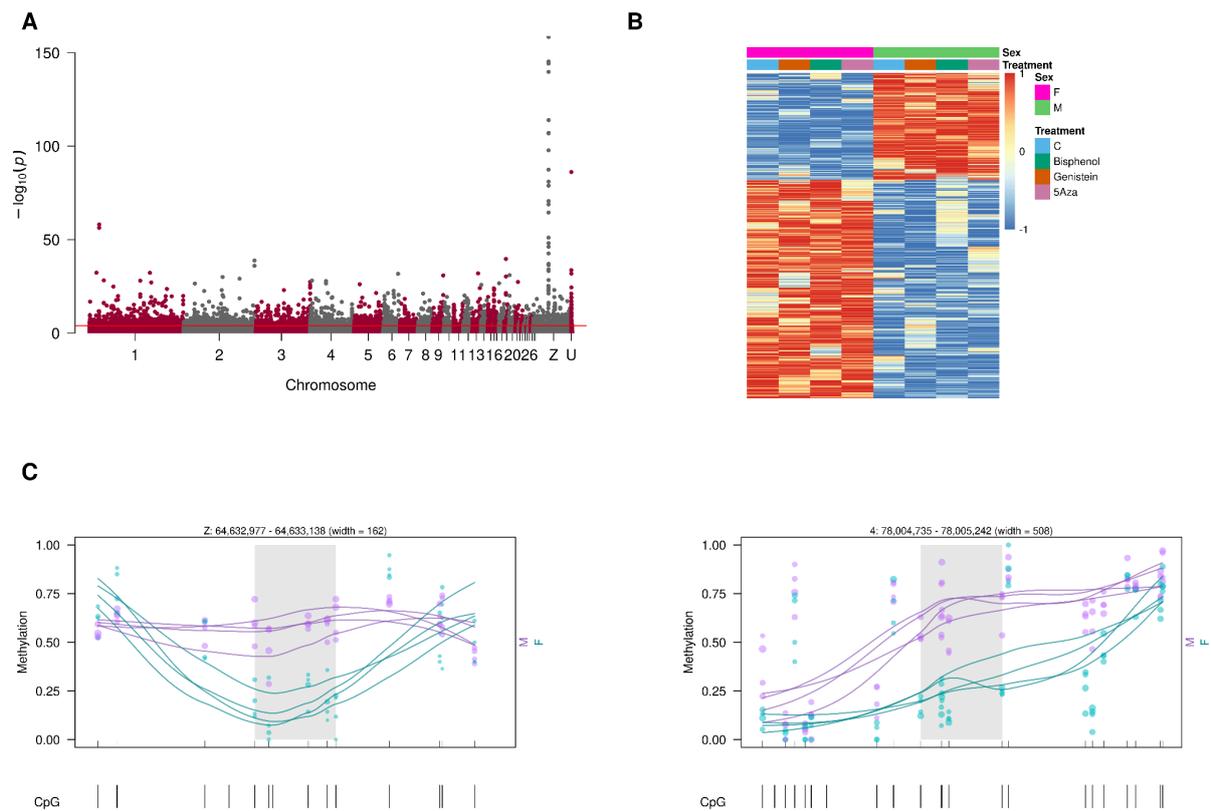


**Supplementary Figure S4. Methylation profile of line-associated DMCs identified from RRBS data.** (A-C) Heatmap of z-score methylation rate of (A) 2,141, (B) 3,972, (C) 1,611 DMCs identified in blood, liver and brain, respectively. The colour gradient (blue to red) is proportional to the z-score methylation rate: blue and red for respectively hypo-methylated and hyper-methylated cytosines. (D) Upset of DMC detected in blood, liver and brain.

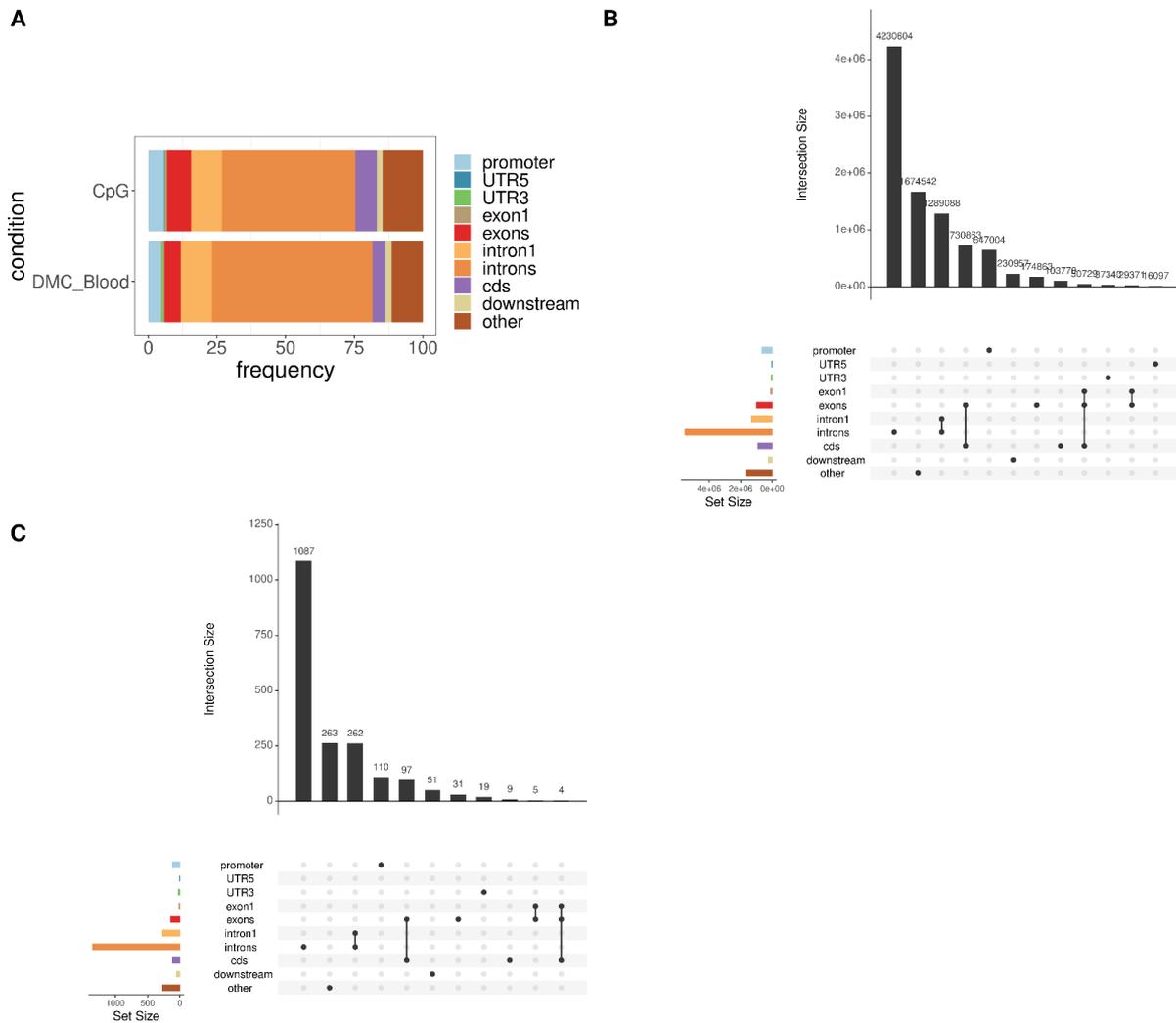




**Supplementary Figure S6. Methylation profile of sex-associated DMCs identified from WGBS data.** (A) Manhattan plot showing the DMCs distribution along the genome (generated by the R *qqman* package). Each point refers to a CpG site ( $n=4,588,907$ ). The horizontal red bar represents the significance threshold after adjustment of the p-values with the BH method (adjusted p-value < 0.05). On the abscissa, the chromosome names (1-28, W, Z, U) with "U" refers to unassigned contigs. On the ordinate, the p-values before multiple testing correction. The alternating red and grey bands represent the separation between chromosomes. The width of the bands is proportional to the size of the chromosomes. (B) Heatmap of the z-score methylation profile of the 10,515 DMCs. The colour gradient (blue to red) is proportional to the methylation rate: blue and red for respectively hypo-methylated and hyper-methylated cytosines. (C) Two examples of differentially methylated regions (DMRs) associated with the sex effect. DMRs identified by the *callDMR* function from the R *DSS* package. Plots were generated by the R *dmrseq* package.

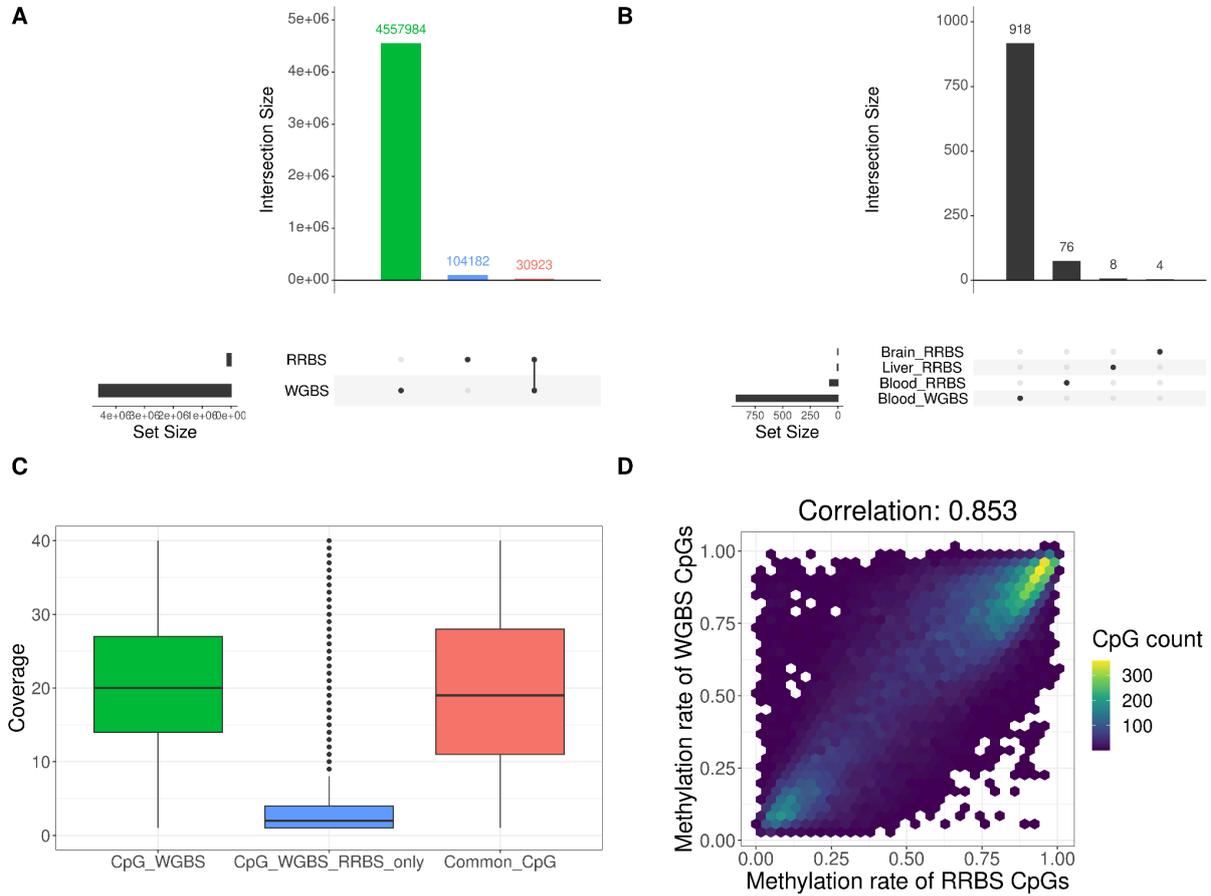


**Supplementary Figure S7. Annotation CpG and DMC associated with the treatment effect identified in WGBS data.** (A) Percentage of CpGs and identified DMCs across genomic features in blood using GenomeFeatures. Barplots showing the annotations count by feature of (B) CpG sites and (C) DMCs in blood samples. CpGs: promoter: 5.692%, UTR5: 0.142%, UTR3: 0.329%, exon1: 0.705%, exons: 8.673%, intron1: 11.342%, introns: 48.562%, cds: 7.79%, downstream: 2.032%, other: 14.733%; Blood DMCs: promoter: 4.762%, UTR3: 0.823%, exon1: 0.39%, exons: 5.931%, intron1: 11.342%, introns: 58.398%, cds: 4.762%, downstream: 2.208%, other: 11.385%.





**Supplementary Figure S9. Comparison between RRBS and WGBS data.** (A) Upset between tested CpG sites from RRBS and WGBS data. (B) Upset between treatment-associated DMCs identified from RRBS and WGBS data. (C) WGBS coverages at three sets of CpG sites. Red: common CpG between RRBS and WGBS data after preprocessing (n=30,923); green: CpG WGBS retained after preprocessing (n=4,588,907); blue: CpG only retained from RRBS data after preprocessing reported on WGBS data before preprocessing (n=99,299). (D) Correlation of the methylation rate of the common CpGs (n = 30,923) between RRBS and WGBS.





# V Caractérisation de la réponse de l'épigénome de la descendance aux effets de l'environnement embryonnaire

## **11 Des effets visibles au niveau du méthylome de la descendance 185**

11.1 Dispositif expérimental

11.2 Article 3 : Effets transgénérationnels de l'injection *in-ovo* de génistéine sur la méthylation de l'ADN de la descendance chez la caille

## **12 Programmation thermique transgénérationnelle ..... 219**

12.1 Présentation du projet QuailHeatE



---

Nous savons que les effets de l'environnement embryonnaire peuvent être transmis à la descendance par le biais de marques épigénétiques, dont la méthylation de l'ADN. L'objectif de cette deuxième partie est de caractériser la réponse de l'épigénome aux variations de l'environnement embryonnaire trois générations après exposition à un stress environnemental. Pour cela, nous avons étudié les effets de l'injection de la génistéine dans l'œuf pendant le développement embryonnaire sur la méthylation de l'ADN de cailles japonaises trois générations après l'injection (G3). Mon travail de thèse s'inscrit dans la continuité du projet Caramel (Département de génétique animale INRAE, 2018) (**Partie II**). Dans cette partie, j'ai réalisé l'alignement des lectures issues du séquençage bisulfite (WGBS) et récupéré les informations de méthylation de chaque site CpG avec le pipeline `nf-core/methylseq`. J'ai ensuite effectué les analyses de méthylation différentielle (**Chapitre 10**) à l'aide du pipeline `SeqOccin` (**Partie III**) à partir des informations sur les sites CpG récupérées des données WGBS mais aussi de données ONT (projet SeqOccin) issues du même dispositif. J'ai également participé à l'analyse de données WGBS issues de cailles ayant reçu une manipulation thermique lors du développement embryonnaire (projet QuailHeatE). Pour cette partie, j'ai également traité les données avec les pipelines `nf-core/methylseq` et le pipeline `SeqOccin`. Les dispositifs expérimentaux sont détaillés dans les chapitres ci-dessous. Les résultats décrits dans le **Chapitre 11** reposent sur un article et ont été présentés sous la forme d'un poster à l'ECCB et d'une communication orale à la journée Genotoul Bioinfo/biostat (**Annexe 19**) :

**Article 3** (1er jet rédigé, discussion à compléter après échanges avec les co-auteurs) - Cerutti et al. (2024b) - **Transgenerational effects of in-ovo injection of genistein on the DNA methylation of offspring quails.**

**Poster / communication orale** - ECCB / Genotoul Bioinfo/biostat - Cerutti et al. (2022) - **Transgenerational epigenetics in quail : whole genome DNA methylation analysis.**  
<https://hal.inrae.fr/hal-03935584v1>



# 11. Des effets visibles au niveau du méthylome de la descendance

## 11.1 Dispositif expérimental

Dans cette étude, nous avons étudié l'effet de l'injection de génistéine au cours du développement embryonnaire sur la méthylation de l'ADN de la descendance chez la caille japonaise. Afin de mettre en évidence une potentielle transmission transgénérationnelle des effets de l'environnement, les données de méthylation de l'ADN obtenues par séquençage haut-débit à partir d'individus de la troisième génération (G3) après injection ont été analysées. Afin d'identifier les régions génomiques affectées par les modifications épigénétiques induites par l'injection de génistéine dans l'œuf des ancêtres, des données de méthylation de l'ADN issues de la génération G3 ont été analysées. L'ensemble de ces données proviennent de 23 individus (lignée S+). Nous disposons de données WGBS provenant d'échantillons de sang prélevés sur la descendance de cailles séparées en 2 groupes : un groupe témoin non injecté (epi-, n = 11) et un groupe issus d'ancêtres ayant reçu une injection dans l'œuf de génistéine (epi+, n = 12). Nous avons également des données ONT sur les mêmes individus (**Figure 11.1**). Ces données sont issues du séquençage Nanopore qui permet de détecter l'état de méthylation des cytosines. Cette technologie permet de séquencer de grands fragments et produit à la fois la séquence du génome et du méthylome, sans traitement au bisulfite.

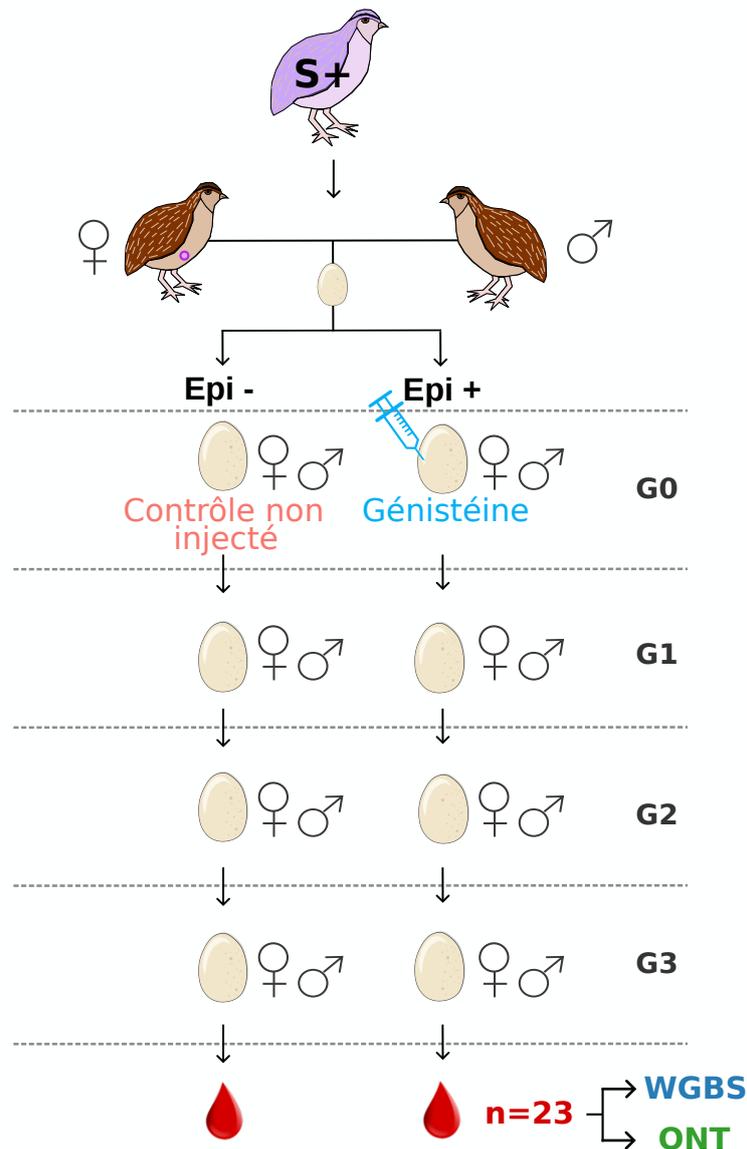


FIGURE 11.1 – Dispositif expérimental de la G3. Des échantillons de sang ont été prélevés sur 23 cailles (mâles et femelles) de 2 groupes : un groupe contrôle non-injecté (epi-) et un groupe injecté avec de la génistéine. Les cailles étudiées provenaient de la lignée expérimentale S+.

## 11.2 Article 3 : Effets transgénérationnels de l'injection *in-ovo* de génistéine sur la méthylation de l'ADN de la descendance chez la caille

Afin de caractériser les effets des molécules injectées sur le méthylome de la G0, des données de méthylation de l'ADN WGBS et ONT ont été analysées. Le traitement des lectures issues des séquençage WGBS est réalisé par l'outil *nf-core/methylseq* v1.5 (P. A. Ewels et al., 2020 ; P. Ewels et al., 2023) (**Article 3 - Supplementary Figure S2.I.A**). Les différentes étapes effectuées par ce pipeline ont été décrites dans la partie précédente (**Chapitre 10.2**). Les données WGBS étant issues du séquençage NovaSeq, l'outil *fastp* v0.21 (S. Chen et al., 2018) a été préalablement utilisé pour retirer les séquences polyG présentes dans les lectures. Les sites CpG sont extraits dans un fichier pour chaque échantillon. Concernant les données ONT, les sites CpG ont été extraits à l'aide de l'outil *Megalodon* v2.5.0 (*Oxford Nanopore Technologies Ltd*, **Article**

**3 - Supplementary Figure S2.I.B).** Pour détecter les DMC entre les groupes, le pipeline `Seq0ccin` (**Partie III**) basé sur le package `edgeR` v3.30.3 de R (Robinson et al., 2010) a été utilisé pour chaque donnée (**Article 3 - Supplementary Figure S2.II**). Ce pipeline analyse les sorties de `nf-core/methylseq` et de `Megalodon` et est composé de deux étapes : une étape de prétraitement suivie de l'analyse différentielle. Ces étapes ont été également détaillées dans la partie précédente (**Chapitre 10.2**). Les données n'étant pas traitées au bisulfite, la présence de SNP identifiés à tort comme DMC en lien avec la conversion des bases est exclue. Les SNP ont donc été directement détectés à partir des données ONT à l'aide de l'outil `Longshot` v0.4.1 (Edge & Bansal, 2019) (**Article 3 - Supplementary Figure S2.I.C**). Les listes de SNP identifiées sont ensuite fusionnées par `bcftools` v1.16 (Danecek et al., 2021). Le fichier VCF combiné a été utilisé pour filtrer les SNP dans les listes de CpG issues de `Bismark` et de `Megalodon`. Pour la détection des DMC, des modèles linéaires ont été utilisés. Ces derniers incluent les effets du traitement et du sexe (modèle linéaire :  $\sim$  traitement + sexe), le contrôle (C) étant utilisé comme groupe de référence dans l'analyse. Les p-valeurs ont été ajustées par la méthode Benjamini-Hochberg (BH), avec un seuil de significativité  $\alpha$  de 5%. L'appel des DMR a été réalisé avec la fonction `callDMR` du package `DSS` v2.38.0 à partir des sorties de `edgeR`. Un nombre minimum de 3 CpG est requis pour constituer une DMR et le pourcentage de sites CpG avec des p-valeurs significatives (p-valeurs  $<$  à 0,05) doit être supérieur à 50%.

L'annotation et l'analyse de la distribution des DMC ont également été détaillées dans la partie précédente (**Chapitre 10.2**). Une analyse complémentaire des SNP a été réalisée. La comparaison de la fréquence allélique entre les deux épilignées (epi- et epi+) a été effectuée à l'aide de l'outil `plink` et le package `gaston` de R.

Cette étude a pu mettre en évidence que l'exposition à la génistéine lors du développement embryonnaire impacte la méthylation de l'ADN de la descendance trois générations après injection. L'analyse de méthylation différentielle a permis d'identifier des régions différentiellement méthylées entre les deux épilignées (epi- et epi+). L'analyse de surreprésentation des termes GO a mis en évidence un effet du traitement sur des gènes liés aux voies métaboliques, au développement du cerveau et à la reproduction. Nous savons que les effets de l'environnement peuvent être transmis aux générations suivantes, en partie par l'intermédiaire de marques épigénétiques. Même si les effets génétiques ont été minimisés, une cause génétique à certains différentiels de méthylation observés n'a pas pu être écartée. Cependant, des analyses plus poussées nous ont permis d'identifier certaines régions qui semblent bien être liées à une transmission épigénétique transgénérationnelle des effets de l'environnement.

L'ensemble des scripts R utilisés pour les analyses sont disponibles via le lien suivant : <https://gitlab.com/Okami8/caramel>.

**Transgenerational changes of DNA methylation following *in-ovo* injection of genistein are mostly linked to genetic differences despite balancing genetic contribution of parental quails.**

C. Cerutti<sup>1</sup>, S. Leroux<sup>1</sup>, Y. Labrune, P. Terzian<sup>2</sup>, J. Lledo<sup>3</sup>, D. Gourichon<sup>4</sup>, V. Coustham<sup>5,6</sup>, R. F. Serre<sup>3</sup>, C. Vandecasteele<sup>3</sup>, C. Klopp<sup>2</sup>, C. Iampietro<sup>3</sup>, C. Gaspin<sup>2</sup>, D. Milan<sup>1</sup>, C. Donnadieu<sup>3</sup>, J. Demars<sup>1</sup>, F. Pitel<sup>1</sup>, G. Devailly<sup>1</sup>

<sup>1</sup> *GenPhySE, Université de Toulouse, INRAE, ENVT, 31320, Castanet Tolosan, France*

<sup>2</sup> *MIAT, UR875 INRAE, 31326, Castanet-Tolosan, France*

<sup>3</sup> *GeT-PlaGe - Génome et Transcriptome - Plateforme Génomique, US1426 GET - INRAE, 31326, Castanet-Tolosan, France*

<sup>4</sup> *INRAE, PEAT, 37380, Nouzilly, France*

<sup>5</sup> *INRAE, Université de Tours, BOA, 37380, Nouzilly, France*

<sup>6</sup> *Université de Pau et des Pays de l'Adour, E2S UPPA, INRAE, NUMEA, Saint-Pée-sur-Nivelle, 64310, France*

Corresponding author: Guillaume Devailly

Email: [guillaume.devailly@inrae.fr](mailto:guillaume.devailly@inrae.fr)

**Author ORCIDs :**

C. Cerutti: <https://orcid.org/0000-0002-3207-3461>

S. Leroux : <https://orcid.org/0000-0003-2840-5109>

Y. Labrune : <https://orcid.org/0000-0001-5099-7580>

P. Terzian : <https://orcid.org/0000-0002-7092-3906>

D. Gourichon : <https://orcid.org/0000-0002-4631-0149>

V. Coustham : <https://orcid.org/0000-0002-5399-2723>

J. Demars : <https://orcid.org/0000-0003-2451-0059>

F. Pitel : <https://orcid.org/0000-0002-1477-7633>

G. Devailly : <https://orcid.org/0000-0001-8878-9357>

**Abstract**

The influence of the prenatal environment on the adult phenotype development is partially mediated by epigenetic phenomena. Recently, an increasing number of studies highlighted the transmission of epigenetic marks between generations following an environmental exposure. However, there is much debate about their acquired transmission beyond the exposed individuals. Recent studies revealed that non-genetic inheritance was probably present in avian species. In one of them, performed in Japanese quails, fertilised eggs were injected with an endocrine disruptor (genistein) and after three generations without any other injection, several traits were impacted by the ancestor treatment such as the reproduction and the behaviour. The objective of this study is to compare the blood DNA methylation of the third generation between control and treated lines. Comparable results between ONT and WGBS data were obtained with a better detection of differential methylation from

ONT data. Thousands of DMCs were detected between epilines that highlights potential transgenerational effects of genistein on DNA methylation level. Most of the methylation level variability seems to be correlated with genetic variability. However, several candidate regions seem to be related to epigenetics and will be used for further analysis.

### **Keywords**

DNA methylation, *Coturnix japonica*, endocrine disruptors, genistein, bisulfite, nanopore, differentially methylated cytosine, transgenerational epigenetics

### **Introduction**

An increasing number of studies highlighted the transmission of epigenetics marks between generations following an environmental exposure. However, there is much debate about their transmission beyond the exposed individuals. Recent studies revealed that non-genetics inheritance was probably present in avian species. In one of them (Leroux et al., 2017), fertilised eggs were divided into two groups: one group injected with an endocrine disruptor, Genistein, and a non-injected control group. To minimise the genetic variability between epilines, a “mirrored” crossing was performed maintaining similar ancestor contributions across generations in each epiline. After three generations without any other injection, several traits were impacted by the ancestor treatment including development, reproduction and behaviour. This pilot study highlighted the potential existence of transgenerational transmission of environmental effects in quails.

A genetic cause for some of the observed differential methylation cannot be excluded. To better understand the transgenerational transmission of these environmental effects, we generated DNA methylation data from blood samples of the third generation using WGBS (Whole Genome Bisulfite Sequencing) and ONT (Oxford Nanopore Technologies) native DNA sequencing. We developed a bioinformatics pipeline in order to detect differentially methylated cytosines (DMCs) between the two epilines. We detected thousands of DMCs between both groups. In addition to these analyses, these data allowed us to perform a comparative analysis between the WGBS and ONT sequencing technologies. The aim of this study is to better understand the transgenerational transmission to the offspring of the effects of the early exposure to an active molecule. A Gene Ontology analysis performed on differentially methylated positions highlights genes involved in developmental processes, brain development and reproduction pathways. Using native DNA sequencing provided by ONT, we also compare allelic frequencies between the epilines, and found differences of allelic frequencies at close proximity of most differentially methylated regions despite the balancing cross used to generate the animals. This suggests that most of the DNA methylation differences between the two epilines are due to genetic differences, and not to transgenerational epigenetic inheritance.

## Materials and methods

### *Experimental design*

The quail experimentation has been previously described (Leroux et al, 2017) (**Supplementary Figure S1**). Briefly, animals were reared at tours avian experimental facility (PEAT), an INRAE experimental unit. They were derived from the experimental line S+ (Mills and Faure, 1991), a line developed by divergent selection on social motivation. A total of 10 founding parent pairs were used and fertilised eggs from each pair were divided into two groups (“epilines”) in a randomised manner: a non-injected control group (epi-) and a group treated (epi+) with 500 µg of genistein (Sigma-Aldrich) before incubation. Genistein was dissolved in 100% ethanol and then diluted 10 times in sesame oil prior to injection. The injection was performed on the day of incubation of freshly laid eggs, directly into the yolk. The quails hatched from these eggs constituted the G0 generation. For each line, two generations were produced by mirrored single-pair matings at each generation to minimise the genetic differences between the two epilines. This should have maximised the probability that the differences observed between third-generation epilines were due to the transgenerational transmission of epigenetic marks. At the end of this procedure, blood samples from 24 quails of the third generation (G3) were collected (12 controls and 12 issued from treated ancestors, **Supplementary Table S1**). Because the DNA extraction for one sample failed, only 23 samples were analysed.

### *DNA extraction, library preparation and sequencing*

High molecular weight genomic DNA was extracted from blood according to a high-salt extraction protocol as in Roussot et al, 2003. Genomic DNA concentrations were measured using a Qubit fluorimetry system (Thermo Fisher). Fragment size distributions were assessed using the Femto pulse Genomic DNA 165 kb Kit (Agilent).

The library preparation and sequencing (2 x 150 bp) for the WGBS analyses were performed by Novogene (UK) Company limited (Illumina NovaSeq 6000). ONT libraries were prepared according to the manufacturer’s instructions (SQK-LSK109). The libraries were loaded onto FLO-PRO002 R9.4.1 flow cells and sequenced on PromethION instrument at 20 fmol within 72H.

A total of 23 samples were analysed (**Supplementary data Table 1.B**).

### *DNA methylation analyses*

The Ensembl reference genome of the Japanese quail (*Coturnix japonica* v.2.0, GCA\_001577835.1, Ensembl release 103) was used for the analyses ([https://www.ensembl.org/Coturnix\\_japonica/Info/Index](https://www.ensembl.org/Coturnix_japonica/Info/Index)). It includes 8,974,565 CpG sites.

To analyse the WGBS and ONT data, nf-core/methylseq, megalodon, and an EdgeR-based R in-house pipeline were used on the Genotoul bioinformatics platform (<http://bioinfo.genotoul.fr/>).

The WGBS reads were previously cleaned by the fastp v0.21 (Chen et al., 2018) to remove artifactual polyG sequences at the end of reads. After this cleaning step, the nf-core/methylseq v1.5 tool (Ewels et al., 2020, 2022) was used to process WGBS (pair-end) reads (**Supplementary Figure S2**). Briefly, reads were trimmed with Trim Galore v0.6.4 (Krueger and Andrews, 2011) to remove adapters and low-quality read and trimmed reads were mapped to the quail reference genome by Bismark v0.22.3 (Krueger and Andrews, 2011). Bismark extracts methylation sites and generates a cytosine report for each sample (*--cytosine\_report*) considering all cytosines on both strands and reports their position, strand, trinucleotide context and methylation state. For ONT data, the methylation levels of CpG sites were extracted using the Megalodon v2.5.0 tool (Oxford Nanopore Technologies Ltd).

To detect differentially methylated cytosines between experimental groups, we developed an EdgeR-based (Robinson et al., 2010) R pipeline available at <https://forgemia.inra.fr/chloe.cerutti/bsseqmethdiffanalysis/-/tree/main>.

This pipeline analyses the outputs of nf-core/methylseq, and consists of two subparts: a pre-processing step followed by the differential analysis.

The pre-processing consisted firstly to merge methylation and coverage information from both strands. A filter was then applied to remove SNPs that could be misidentified as DMCs from the WGBS data. Longshot v0.4.1 (Edge et al., 2019) was used to detect native SNPs present directly in ONT samples. The retrieved SNP list was merged between the 23 samples with bcftools v1.14 (Danecek et al., 2021). The combined VCF file was used to filter putative SNP position (n=10,397,021) from the Bismark and Megalodon outputs. Only CpG sites with a minimum of 5 and maximum of 1000 read coverage in at least 80% of samples were used for the analysis.

To perform the differentially methylated cytosines (DMC) detection, the edgeR package (v3.30.3) (Robinson et al., 2010) was used. The differentially methylated regions (DMR) calling was performed with the *callDMR* function from the *DSS* package (v2.38.0) from the edgeR outputs. A minimum number of 3 CpGs is required to constitute a DMR and the percentage of CpG sites with significant p-values (less than 0.05) must be greater than 50%.

The linear models included the sex and treatment effects for both technologies.

The non-injected control (C) was used as the reference group in the analysis. The p-values were corrected with the Benjamini-Hochberg (BH) method (Benjamini and Hochberg, 1995), with a threshold of 5% defining significance after multiple testing correction.

For the principal component analysis (PCA), missing values were imputed with the mean methylation rate of all other samples for this position.

For the visual representations, we used *ggplot2*, *upSetR*, *pheatmap*, *qqman* and *dmrseq* R packages (**Supplementary Table S3**).

### **DMC annotation and GO over-representation analysis**

The next analysis steps have been previously described in David et al., 2019. The annotation and distribution analysis of DMCs were performed by using the R

package `GenomeFeatures` v1.3 (<https://forgemia.inra.fr/aurelien.brionne/GenomeFeatures>) using the annotation file (v103, `Coturnix_japonica.Coturnix_japonica_2.0.103.gtf`) from Ensembl. After annotation, each DMC is associated to genomic features: promoter (3 Kbp upstream and 500 bp downstream of Transcriptional Start Site - TSS), 5'UTR, 3'UTR, exons, introns, downstream (1 Kbp from Transcriptional End Site - TES), coding sequence (CDS) and intergenic region.

A Gene Ontology (GO) term over-representation analysis was performed using the R package `ViSEAGO` v1.4.0 (Brionne et al., 2019) with quail genome genes associated with CpG as background and a minimum number of genes for each GO term of 5 (`create_topGOdata` function). For each list of genes associated with DMCs identified by tissue and technology (four lists of genes), an enrichment test was performed using the exact Fisher's test (`runTest` function from the package `topGO` v2.42.0). Significant enriched GO terms ( $p\text{-value} \leq 0.01$ ) were grouped into functional clusters using hierarchical clustering based on Wang distance between GO terms and Ward's criterion (`compute_SS_distances`, `GOterms_heatmap` functions). These functional clusters were clustered using hierarchical clustering based on BMA distance between sets of GO terms and Ward's criterion (`compute_SS_distances`, `GOclusters_heatmap` functions). The  $-\log_{10}(\text{adjusted } p\text{-values})$  were used.

### **Genetic analyses**

SNP detected from ONT data were used to perform a genetic analysis. The allele frequency comparison between epilines was performed with `plink` v1.9 (Purcell et al., 2007) and the `association.test` function of the `gaston` R package (Perdry et al., 2022). The putative SNPs from the 28 autosomes ( $n=9,142,915$ ) detected from ONT samples were pruned with the `plink` v1.9. SNPs with missing call frequencies greater than 0.05 (`--geno`) were not included. SNPs with minor allele frequency lower than 0.05 (`--maf`) and with a squared correlation greater than the pairwise  $r^2$  threshold (0.1) were pruned from the window (50kb with step size of 10 SNPs, linkage disequilibrium = 0.1) (`--indep-pairwise 50 10 0.1`). A total of 182,366 SNPs were retained for further analysis and a genotype matrix was generated. The genetic relationship matrix was generated with the R package `gaston` (`GRM` function). An association test between epilines groups and SNPs was then performed with the R package `gaston` using a generalised linear mixed model (`association.test` function, `method = "lmm"`). To study the relationship between individuals, a neighbor-joining tree estimation (Saitou and Nei, 1987) was performed with the R package `ape` (`nj` function) using the distance matrix generated by `plink`. The R package `adegenet` was used to cluster the samples in two groups based on their genetic information (`find.clusters` function).

### **Data and script availability**

Sequencing data will soon be available at <https://www.ebi.ac.uk/ena/>. The differential methylation pipeline can be found here: <https://forgemia.inra.fr/chloe.cerutti/bsseqmethdiffanalysis/-/tree/main>. Scripts used

for this analysis can be found in the following git repository:  
<https://gitlab.com/Okami8/caramel>

## Results

### ***DNA methylation differences between offspring of quails injected in-ovo with genistein of not injected***

#### ***DNA methylation profiling***

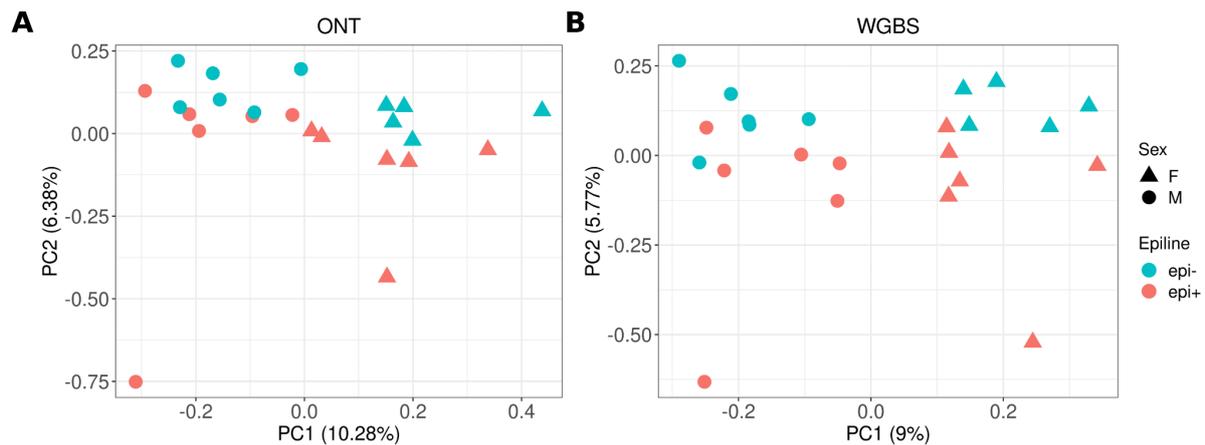
To study the DNA methylation profile, WGBS and ONT data from blood samples of different epilines (epi- and epi+ referring to non-injected control and genistein) were analysed.

After nf-core/methylseq processing of the WGBS data, the sequence data showed an average of  $284.8 \pm 47.6$  M reads per sample (**Supplementary Table S2.A**). The initial average number of CpGs was  $8,315,634 \pm 72,776$  (92.7% of the 8,974,565 CpG sites in the whole genome). A high number of putative SNPs from the VCF file were removed, resulting in an average of  $7,690,302 \pm 70,084$  CpGs (85.7%). After coverage filtering, an average of  $6,142,783 \pm 9,564$  CpG sites were selected for further analysis, corresponding to 68.4% of quail genome CpGs. Each CpG site has an average coverage of  $20.5 \pm 20.8$ , a methylation level about  $55.6 \pm 30\%$  and an average of  $34 \pm 5.1\%$  of the reads were aligned.

After megalodon processing of the ONT data, the sequence data showed an average of  $5.5 \pm 1.4$  M reads per sample (**Supplementary Table S2.B**). The initial average number of CpGs was  $8,788,404 \pm 10,082$  (98% of all CpG sites in the whole genome). A high number of putative SNPs from the VCF file were removed with an average of  $8,127,687 \pm 9,704$  CpGs (90.6%). At the end of the data preprocessing, an average of  $7,955,339 \pm 547$  CpG sites were selected for further analysis, corresponding to 88.6% of quail genome CpGs. Each CpG site has an average coverage of  $39 \pm 17.5$ , a methylation level about  $45.6 \pm 30.2\%$  and an average of  $93.2 \pm 1.1\%$  of the reads were aligned.

The mean methylation rate is higher for WGBS data, while the mean coverage is higher for ONT data. The differences observed between these two technologies are expected, as ONT sequencing utilises real-time reading technology able to generate long DNA fragments, compared to WGBS, which typically sequences shorter DNA regions.

We performed a PCA on the methylation levels of retained CpG sites from the 23 WGBS (**Figure 1.A**) and ONT (**Figure 1.B**) samples. The methylation profiles of CpG sites shows that PC1, explaining 10.28% and 9% of the variance for WGBS and ONT, respectively, separates the female and male samples. PC2, explaining 6.38% and 5.77% of the variance in WGBS and ONT data, respectively, appears to differentiate between the two epilines (epi- and epi+) for both WGBS and ONT datasets.



**Figure 1. Principal Component Analysis (PCA) of the methylation profiles of CpG sites.** CpG sites detected from (A) ONT and (B) WGBS data. The shapes differentiate the sexes: circles and triangles represent males and females, respectively. The colours differentiate the quail epilines: blue and red for the epi- (non-injected control) and epi+ (genistein) groups, respectively. Percentages indicate the proportion of variance explained by each PCA axis. Total number of CpGs sites:  $n = 6,148,705$  (A),  $n = 7,957,328$  (B).

### ***In-ovo treatments affects the offspring DNA methylome***

A total of 6,148,705 and 7,957,328 CpG sites were tested from the WGBS and ONT data, respectively. A linear model including sexe and treatment effects was applied for each CpG site. DMCs associated with the treatment and sex effects were identified (**Table 1**). This analysis identified 14,019 and 23,967 treatment-associated DMCs in WGBS and ONT data, respectively. More than one hundred thousand sex-associated DMCs were detected (**Table 1**).

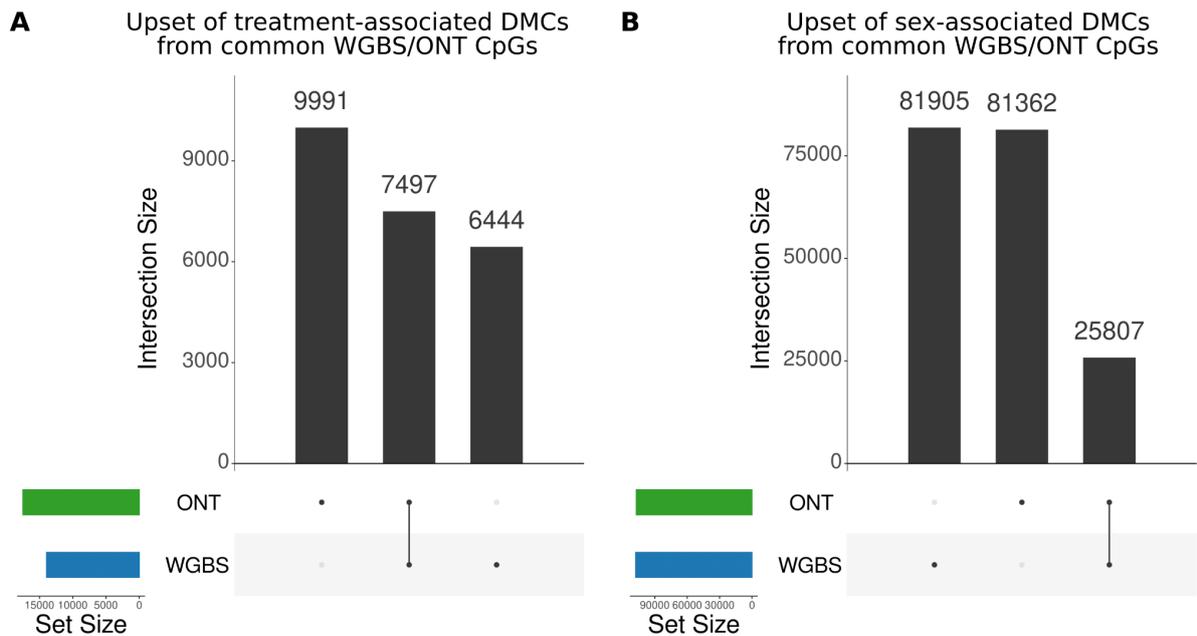
Linear model	Data type	Number of treatment DMC	Number of sex DMC	Number of treatment DMR	Number of sex DMR	Number of CpG tested
~ treatment + sex	WGBS	14,019	109,216	60	1,779	6,148,705
	ONT	23,967	120,466	201	2,223	7,957,328

**Table 1.** DMC number identified with multifactor analysis on WGBS and ONT data. A likelihood ratio test is performed for the treatment and sex effects.

A total of 6,107,897 CpG sites are common between WGBS and ONT data, (68% of all the CpG sites in the whole genome), and around 99% of WGBS CpGs are found in ONT CpGs (**Supplementary Figure S4.A**).

We highlighted an average mean correlation of  $0.89 \pm 0.02$  between WGBS and ONT data (**Supplementary Figure S4.B**). The average correlation is not significantly different between the two epilines (**Supplementary Figure S4.C**).

Among the treatment-associated DMCs, 7,497 were common between WGBS and ONT data (**Figure 2.A**). A total of 25,807 common sex-associated DMCs were found between the two technologies (**Figure 2.B**).

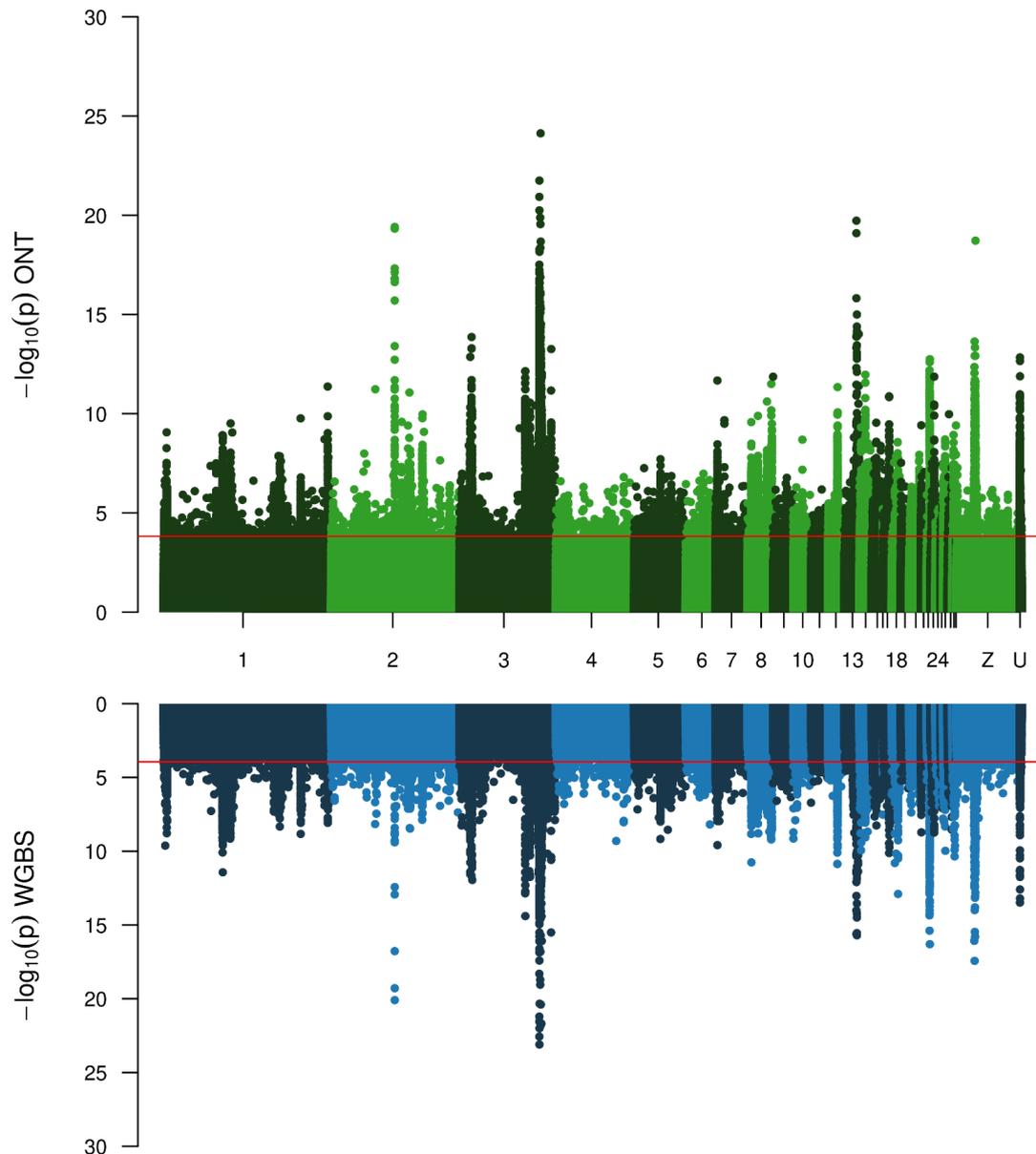


**Figure 2. Upset of DMCs detected from common WGBS and ONT CpGs sites.** Upset of (A) treatment-associated DMC and (B) sex-associated DMC detected from common CpG sites between WGBS (bleu) and ONT (green) data. For a fair comparison, only the CpG sites measured by both technologies are included.

The CpG distribution along the quail genome are similar between the two technologies (**Figure 3-4**). Treatment-associated DMCs are grouped in the same regions (**Figure 3**). The most significant sex-associated DMCs are located on the sexual chromosome (Z). Some sex-associated DMCs in the chr1 are more significant for ONT data compared to WGBS data (**Supplementary Figure S5**).

The methylation rate profile is very similar for common treatment-associated DMCs between WGBS and ONT data (**Supplementary Figure S6.A**). Despite the fact that CpGs identified as DMCs in WGBS may not overlap with DMCs in ONT (**Supplementary Figure S6.B**), and vice versa (**Supplementary Figure S6.C**), there seems to be a noticeable trend in the methylation profiles across both technologies. The same tendency was observed for sex-associated DMCs with similar methylation profiles between the two technologies (**Supplementary Figure S7.A-C**).

These analyses revealed similar methylation profiles between the two technologies. A higher CpG number was identified from ONT data than WGBS data allowing the detection of a higher DMC number.

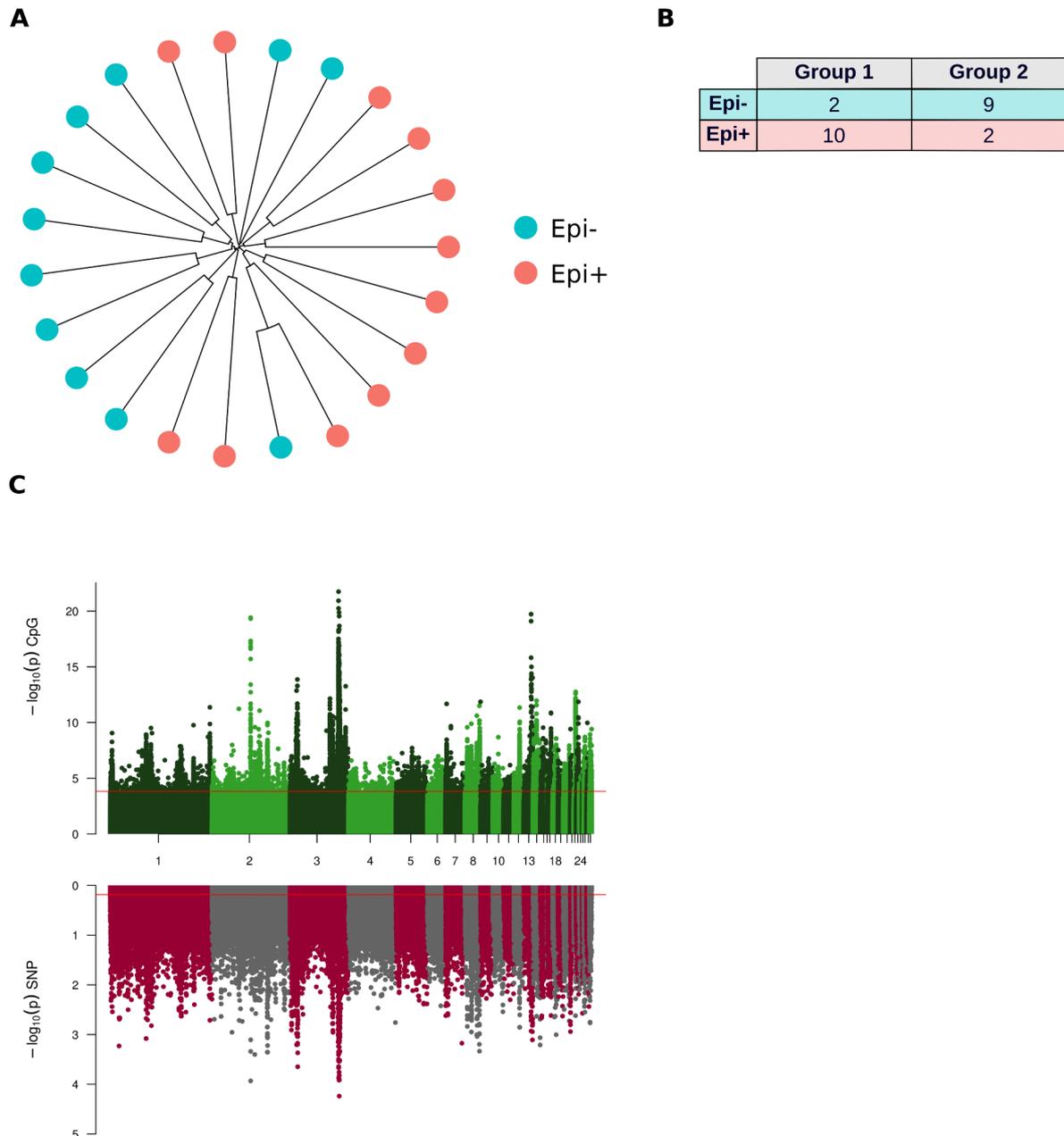


**Figure 3. Manhattan plots of treatment-associated CpGs identified from ONT and WGBS data.** Manhattan plots showing the distribution along the genome of CpGs identified in blood samples from ONT (top, in green) and WGBS (bottom, in blue) data. Each point refers to a CpG site ( $n_{\text{CpG\_ONT}}=7,957,328$ ,  $n_{\text{CpG\_WGBS}}=6,148,705$ ). The horizontal red bar represents the significance threshold for a False Discovery Rate (FDR) of 5%. X-axis: chromosome names (1-28, W, Z, U) with “W” and “Z” for sex chromosomes and “U” for unassigned contigs. Y-axis: the p-values for differential methylation levels by treatments, without multiple testing correction.

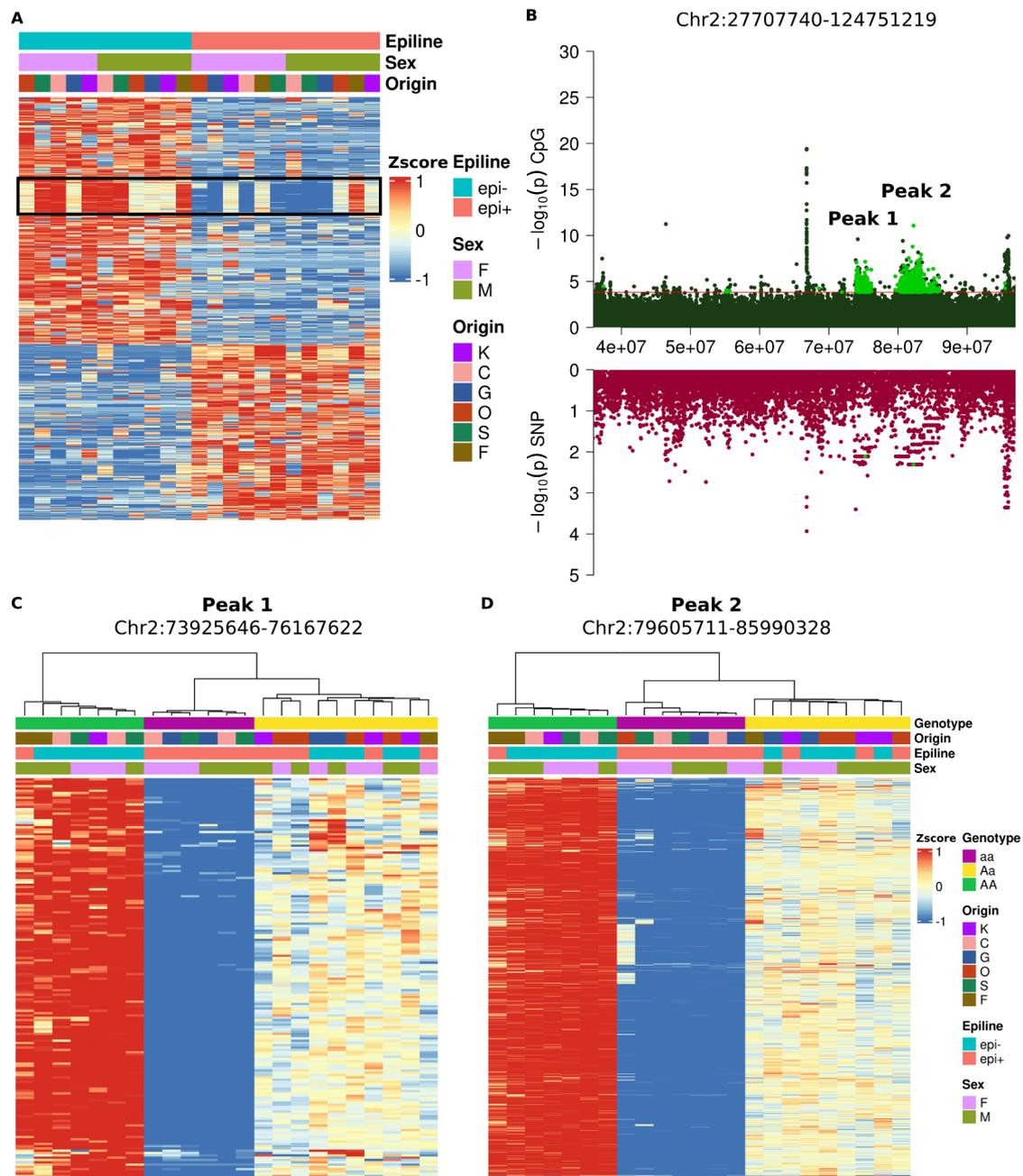
### **Correlation between genetic variations and treatment-associated DMCs**

To minimise the genetic variability between epilines, a “mirrored” crossing was performed in order to maintain similar ancestor contributions across generations in each epiline. The analysis of allelic frequencies between the two epilines (epi- and epi+) reveals that quails from the same epiline were broadly related (**Figure 4.A**). Moreover, the clustering of samples, based on their genetic information, highlights a genetic drift between epilines but not all individuals from the same epiline are consistently grouped together (**Figure 4.B**). Differential allelic frequencies between the two lines tend to be localised at genomic regions with differentially methylated cytosines (**Figure 4.C**).

Interestingly, among treatment-associated DMCs identified, a methylation pattern seems to be highlighted through the methylation profiles, with a grouping of methylation states by individuals (**Supplementary Figure S6.A-C, Figure 5.A**). The distribution of these DMCs from ONT data along the genome highlights two regions (peaks) displaying a high density of these DMCs, which appear to be co-localised with SNPs of distinct allele frequencies between the epilines (**Figure 5.B**). The clustering of DMCs from each peak by overlapping haplotypes shows an allele-specific methylation, resulting in a hypermethylation for one of the haplotype, a hypomethylation for the other haplotype, and a hemimethylation for the heterozygous animals (**Figure 5.C-D**).



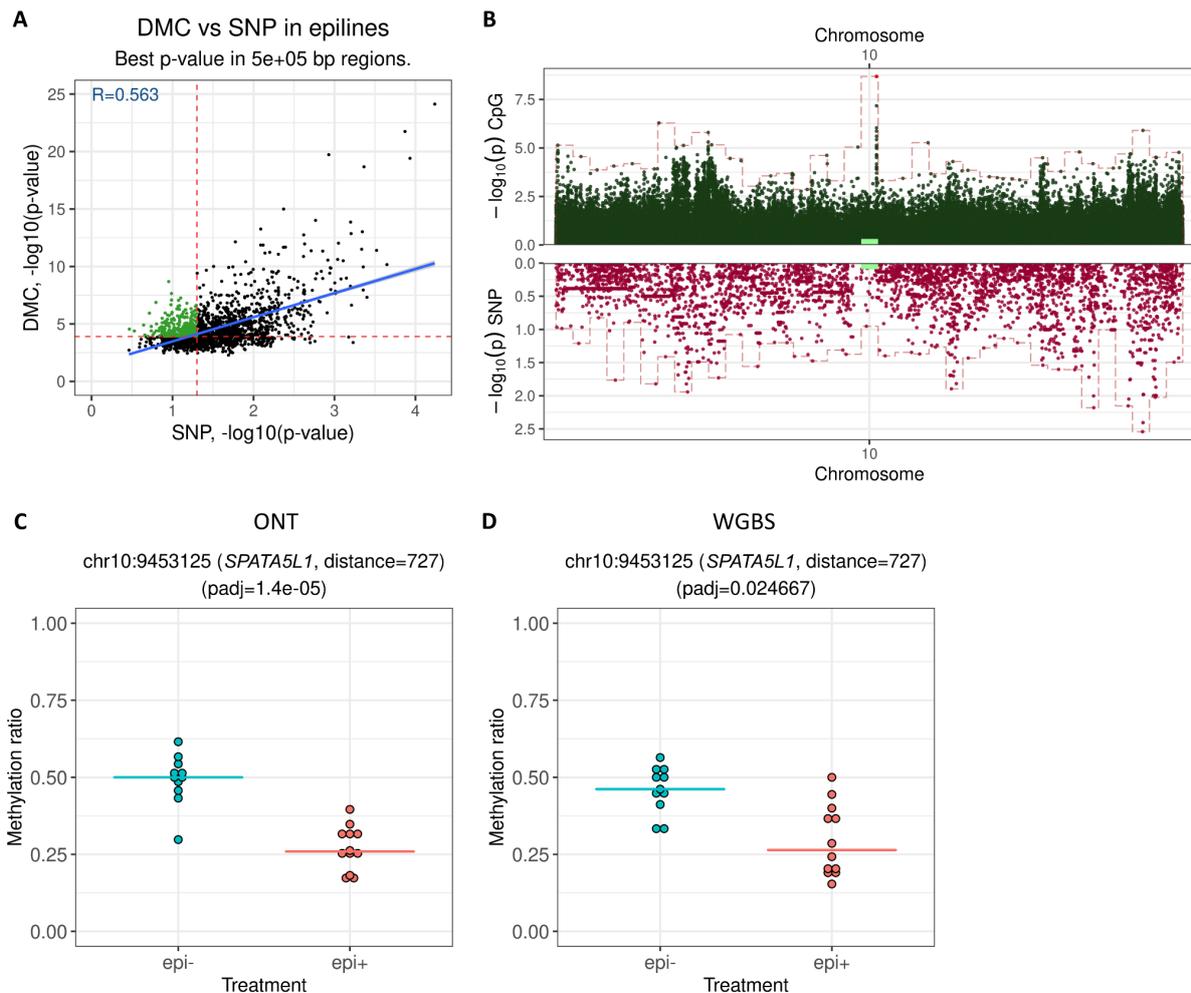
**Figure 4. Manhattan plot of treat-associated CpGs and SNPs.** (A) Estimated neighbour-joining tree between epilines based on the genotype of individuals. (B) Clustering of individuals based on their genotype separated into two groups. (C) Manhattan plots showing the distribution along the genome of CpGs (top, in green) and SNP (bottom, in red and grey) identified in blood samples from ONT data. Each point refers to a CpG site ( $n=7,957,328$ ) or a SNP ( $n=182,366$ ). The horizontal red bar represents the significance threshold for a False Discovery Rate (FDR) of 5%. X-axis: autosome names (1-28). Y-axis: the p-values for differential methylation levels by treatments (top) and p-values from the association test between epilines (bottom), without multiple testing correction.



**Figure 5. Allele-specific methylation.** (A) Heatmap of treatment-associated DMCs from ONT data (n=23,967). The black box highlights the DMCs found within the pattern (n=2,433). (B) Manhattan plots showing the distribution along the genome of ONT CpGs located in the chromosome (chr) 2 (top, in green) and SNPs (bottom, in red). Each point refers to a CpG site (top) or SNP (bottom), and light green points refer to ONT DMCs from chr 2 in the black box of the heatmap (n=1,721). The horizontal red bar represents the significance threshold for a False Discovery Rate (FDR) of 5%. X-axis: genomic coordinates along chr2. Y-axis: the p-values for differential methylation levels by treatments (top) and p-values from the association test between epilines (bottom), without multiple testing correction. (C-D) Heatmap of the methylation rate of treatment-associated DMCs from (C) peak 1 (n=175) and (D) peak 2 (n=1,514). The colour gradient (blue to red) is proportional to the z-score methylation rate: blue and red for respectively hypo-methylated and hyper-methylated cytosines.

### **Detection of treatment-associated DMCs in genomic regions with similar allelic frequencies between the epilines**

While most DMCs appeared to be near SNP with distinct allelic frequencies between the epilines, we looked for DMCs with no allelic imbalance at proximity. To do so we binned the genome in 500,000 bp regions, and computed the best p-values for DMCs and SNPs belonging to these windows between epi- and epi+ (**Figure 6.A**). A window on chromosome 10 (chr10:9453125, in red) does not appear to be strongly associated with SNP peak (**Figure 6.B**). The methylation rate of this DMC is higher for the control group (epi-) than for the injected group (epi+) for ONT and WGBS data (**Figure 6.C-D**).



**Figure 6. High confidence treatment-associated DMCs from ONT data.** (A) Scatter plot of best treatment ONT DMCs per window of 500,000pb. DMCs far from SNP in allelic imbalance between the lines ( $p\text{-value}_{\text{CpG}} > -\log_{10}(0.0001245)$  ( $\text{FDR} < 0.05$ ) and  $p\text{-value}_{\text{SNP}} < -\log_{10}(0.05)$ ) are shown in green. The blue line refers to the regression line between DMCs and SNPs ( $R = 0.563$ ). (B) Manhattan plots showing the distribution along the genome of ONT DMCs (top, in green) and SNPs (bottom, in red and red). Each point refers to a CpG site (top) or a SNP (bottom). X-axis: genomic coordinates along chr10. Y-axis: the p-values for differential methylation levels by treatments (top) and p-values from the association test between epilines (bottom), without multiple testing correction. (C-D) Dotplot of the same treatment DMCs (chr10:9453125) detected from WGBS (C) and ONT (D) data. This DMC is represented in red on the scatter plot (A) and the Manhattan plot (B, top).

### **Genes and biological functions associated with adult DMC after *in-ovo* treatments**

After the annotation of epiline-associated DMCs and CpGs from ONT data, their repartition by genomic feature was represented (**Supplementary Figure S8.A**). The DMC repartition is broadly similar to the CpG repartition with a higher count in introns and intergenic regions. The DMCs were found in 3,142 different genes.

These genes were enriched for BP GO terms grouped into 47 clusters (**Supplementary Figure S8.B**), regrouping genes involved in developmental processes (cl5,8,9-15,28,29,33), reproduction (cl6,7) and brain development (cl2,3,16,17). The annotation and GO analysis of treatment DMCs far from imbalanced SNPs highlighted genes (BP GO terms in 46 clusters) involved in developmental processes (cl3,4,6-10,12,14-17,27), reproduction (cl13,14,26,33) and brain development (cl3,14,29) (**Supplementary Figure S9**). The DMC in chr10:9453125 is associated to the *SPATA5L1* (Spermatogenesis-Associated Protein 5-Like Protein 1 = AFG2B - AFG2 AAA ATPase Homolog B), involved in the ribosome biogenesis.

### **Discussion**

In this study, we analysed the effects on the DNA methylation of blood three generations after the injection of genistein into the eggs. In contrast to mammalian blood, the avian blood contains nucleated red blood cells (Scanes, 2015), making this tissue relatively homogeneous in birds (more than 90% of the DNA isolated originates from erythrocytes, Verhulst et al., 2016). In this study, we had the opportunity to compare DNA methylation data from two different whole genome sequencing techniques (WGBS and ONT). The low methylation level observed from ONT data compared with WGBS can be partly explained by a poor read alignment in GC-rich regions (the majority being hypomethylated) by WGBS (**Supplementary Figure S10**). The analysis of the data showed comparable results but a higher number of CpG sites is detected from ONT data. The nanopore technology allows us to identify a greater number of epiline-associated DMCs than bisulfite sequencing. Moreover, ONT sequencing does not require treatment of DNA with sodium bisulfite and allows a more reliable SNP detection than WGBS.

The low methylation level observed from ONT data compared with WGBS can be partly explained by a poor read alignment in GC-rich regions (the majority being hypomethylated) by WGBS.

Genetic analyses were performed and showed that the most of the methylation level variability between the epilines seems to be co-localized with genetic variability. However, several candidate regions were showing differential methylation with no differences in allelic frequency at vicinity. While these DMCs can be the results of trans-acting methylation QTL, they could also be traces of transgenerational inheritance.

## Acknowledgments

The authors want to acknowledge C. Leterrier for providing the S+ experimental line and helpful discussions. We are grateful to the genotoul bioinformatics platform Toulouse Occitanie (GenoToul Bioinfo, 2018) for providing help, computing, and storage resources.

## Declaration of interest statement

This work was supported by the CARAMEL grant. C. Cerutti is supported by a PhD scholarship from R gion Occitanie and INRAE Animal Genetics Division.

## References

- Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Brionne, A., Juanchich, A., Hennequet-Antier, C., 2019. ViSEAGO: a Bioconductor package for clustering biological functions using Gene Ontology and semantic similarity. *BioData Mining* 12, 16. <https://doi.org/10.1186/s13040-019-0204-1>
- Cerutti, C., Leroux, S., Gourichon, D., Labrune, Y., David, I., Zerjal, T., Coustham, V., Devailly, G., Pitel, F., 2022. Short communication: Effects of *in-ovo* injection of endocrine disruptors and methyltransferase inhibitor on quail growth and egg-laying performances. *animal* 16, 100464. <https://doi.org/10.1016/j.animal.2022.100464>
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp : an ultra-fast all-in-one fastq preprocessor. *Bioinformatics*, 34(17), i884–i890.
- Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008. <https://doi.org/10.1093/gigascience/giab008>
- David I, Ricard A. A Unified Model for Inclusive Inheritance in Livestock Species. *Genetics*. 2019 Aug;212(4):1075-1099. doi: 10.1534/genetics.119.302375. Epub 2019 Jun 17.
- Edge, P., & Bansal, V. (2019). Longshot enables accurate variant calling in diploid genomes from single-molecule long read sequencing. *Nature communications*, 10(1), 4660
- Ewels, P., H tther, P., Spix, N., Miller, E., Nf-Core Bot, H rtenhuber, M., Peltzer, A., Sven, F., Alneberg, J., Di Tommaso, P., Garcia, M.U., Davenport, C., Vikesh Ajith, Krueger, F., Harshil Patel, Alessia, Devailly, G., Syme, R., VIJAY, Menden, K., Noirot C line, 2022. nf-core/methylseq: nf-core/methylseq version 2.3.0 [Pyrite Alligator]. <https://doi.org/10.5281/ZENODO.1343417>
- Ewels, P.A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M.U., Di Tommaso, P., Nahnsen, S., 2020. The nf-core framework for community-curated

bioinformatics pipelines. *Nat Biotechnol* 38, 276–278. <https://doi.org/10.1038/s41587-020-0439-x>

Feil, R., Fraga, M.F., 2012. Epigenetics and the environment: emerging patterns and implications. *Nat Rev Genet* 13, 97–109. <https://doi.org/10.1038/nrg3142>

GenoToul Bioinfo, 2018. GenoToul Bioinformatics Facility. <https://doi.org/10.15454/1.5572369328961167E12>

Krueger, F., Andrews, S.R., 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27, 1571–1572. <https://doi.org/10.1093/bioinformatics/btr167>

Leroux, S., Gourichon, D., Leterrier, C., Labrune, Y., Coustham, V., Rivière, S., Zerjal, T., Coville, J.-L., Morisson, M., Minvielle, F., Pitel, F., 2017. Embryonic environment and transgenerational effects in quail. *Genet Sel Evol* 49, 14. <https://doi.org/10.1186/s12711-017-0292-7>

Mills, A.D., Faure, J.-M., 1991. Divergent selection for duration of tonic immobility and social reinstatement behavior in Japanese quail (*Coturnix coturnix japonica*) chicks. *Journal of Comparative Psychology* 105, 25–38. <https://doi.org/10.1037/0735-7036.105.1.25>

Perdry Hervé and Dandine-Roulland Claire (2022). *gaston: Genetic Data Handling (QC, GRM, LD, PCA) & Linear Mixed Models*. R package version 1.5.9. <https://CRAN.R-project.org/package=gaston>

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007 Sep;81(3):559-75. doi: 10.1086/519795. Epub 2007 Jul 25

Rasheed, S., Rehman, K., Shahid, M., Suhail, S., Akash, M.S.H., 2022. Therapeutic potentials of genistein: New insights and perspectives. *Journal of Food Biochemistry* 46. <https://doi.org/10.1111/jfbc.14228>

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. <https://doi.org/10.1093/bioinformatics/btp616>

Roussot O, Fève K, Plisson-Petit F, Pitel F, Faure JM, Beaumont C, et al. AFLP linkage map of the Japanese quail *Coturnix japonica*. *Genet Sel Evol*. 2003;35:559–72.

Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987 Jul;4(4):406-25. doi: 10.1093/oxfordjournals.molbev.a040454. PMID: 3447015.

Scanes, C.G., 2015. Protein Metabolism, in: *Sturkie's Avian Physiology*. Elsevier, pp. 455–467. <https://doi.org/10.1016/B978-0-12-407160-5.00020-8>

Silva, L.B.A.R., Pinheiro-Castro, N., Novaes, G.M., Pascoal, G. de F.L., Ong, T.P., 2019. Bioactive food compounds, epigenetics and chronic disease prevention: Focus

on early-life interventions with polyphenols. *Food Research International* 125, 108646. <https://doi.org/10.1016/j.foodres.2019.108646>

Sinclair, K.D., Rutherford, K.M.D., Wallace, J.M., Brameld, J.M., Stöger, R., Alberio, R., Sweetman, D., Gardner, D.S., Perry, V.E.A., Adam, C.L., Ashworth, C.J., Robinson, J.E., Dwyer, C.M., 2016. Epigenetics and developmental programming of welfare and production traits in farm animals. *Reprod. Fertil. Dev.* 28, 1443. <https://doi.org/10.1071/RD16102>

Sun, R., Zhu, P., 2022. Advances in measuring DNA methylation. *Blood Science* 4, 8–15. <https://doi.org/10.1097/BS9.0000000000000098>

Wang, N., Sun, J., Pang, T., Zheng, H., Liang, F., He, X., Tang, D., Yu, T., Xiong, J., Chang, S., 2022. DNA Methylation Markers and Prediction Model for Depression and Their Contribution for Breast Cancer Risk. *Frontiers in Molecular Neuroscience* 15.

Yilmaz, B., Terekeci, H., Sandal, S., Kelestimur, F., 2020. Endocrine disrupting chemicals: exposure, effects on human health, mechanism of action, models for testing and strategies for prevention. *Rev Endocr Metab Disord* 21, 127–147. <https://doi.org/10.1007/s11154-019-09521-z>

## Transgenerational effects of *in-ovo* injection of genistein on the DNA methylation of offspring quails.

**Supplementary Table S1 - Metadata table for WGBS and ONT data.** A total of 23 blood samples were analyzed. Epi- and epi+ refers to the non-injected control group and the injected group with genistein, respectively. M and F refers to males and females, respectively.

ID ONT	ID WGBS	EpiLine	Sex
AB75069240	PRC2013090046	epi+	M
AB75005544	PRC2013090048	epi+	F
AB75230653	PRC2013090062	epi+	F
AB75007842	PRC2013090064	epi+	M
AB75070045	PRC2013090120	epi+	M
AB75069424	PRC2013090122	epi+	F
AB75007905	PRC2013090157	epi+	M
AB75005752	PRC2013090162	epi+	F
AB75070052	PRC2013090192	epi+	M
AB75007856	PRC2013090196	epi+	F
AB75007836	PRC2013090278	epi+	M
AB75069306	PRC2013090285	epi+	F
AB75069430	PRC2013090341	epi-	M
AB75005572	PRC2013090344	epi-	F
AB75005603	PRC2013090373	epi-	F
AB75007901	PRC2013090374	epi-	M
AB75230655	PRC2013090440	epi-	M
AB75005816	PRC2013090469	epi-	F
AB75070049	PRC2013090474	epi-	M
AB75070062	PRC2013090490	epi-	M
AB75070036	PRC2013090496	epi-	F
AB75005611	PRC2013090539	epi-	F
AB75007899	PRC2013090542	epi-	M

**Supplementary Table S2. Table from the MultiQC report generated by the nf-core/methylseq pipeline on WGBS data (A) and ONT data (B).** ID: pool identification, Reads (M): total number of sequenced reads (in millions), Aligned (%): percentage of aligned reads, CpG (K): total number after preprocessing of cytosines analysed (in thousands), mCpG (%): percentage of methylated cytosines in CpG context after preprocessing.

**A**

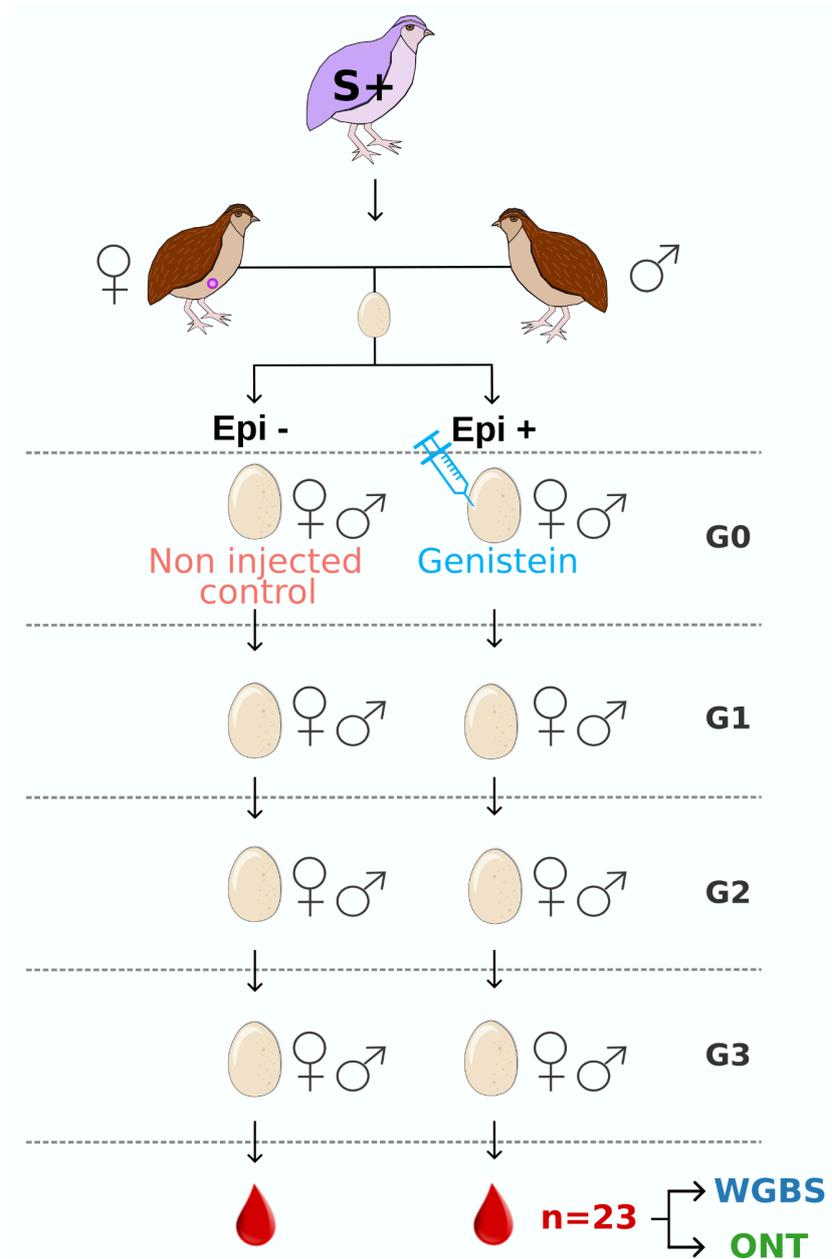
ID	Read (M)	Aligned (%)	CpG (K)	mCpG (%)
PRC2013090046	155.6	32	614.2	53.7
PRC2013090048	257.4	20.8	613.2	57.4
PRC2013090062	304.8	36.9	614.6	56.7
PRC2013090064	266.4	36.3	614.6	55.5
PRC2013090120	277.5	41.6	614.6	54.8
PRC2013090122	293.9	43.2	614.5	56.8
PRC2013090157	347.1	37.6	614.7	52.8
PRC2013090162	304.7	35.7	614.6	56.8
PRC2013090192	255.9	35.5	614.4	53.4
PRC2013090196	284.4	33.9	614.1	56.5
PRC2013090278	328	30.8	614.2	55.1
PRC2013090285	290.2	35.5	614.5	58.6
PRC2013090341	363.1	36.2	614.7	53.7
PRC2013090344	301.8	34.1	614.6	56.9
PRC2013090373	371.1	29	614.7	58.6
PRC2013090374	229.7	33.6	614.6	53.8
PRC2013090440	275.4	38.5	614.4	53.6
PRC2013090469	245	27.8	614.6	56.8
PRC2013090474	314.3	30.1	614.7	54.2
PRC2013090490	245.7	40.8	610.1	53.40
PRC2013090496	264	33.7	614.5	57.5
PRC2013090539	326.5	26.7	614.5	58.4
PRC2013090542	248	32.3	614.6	54.9

**B**

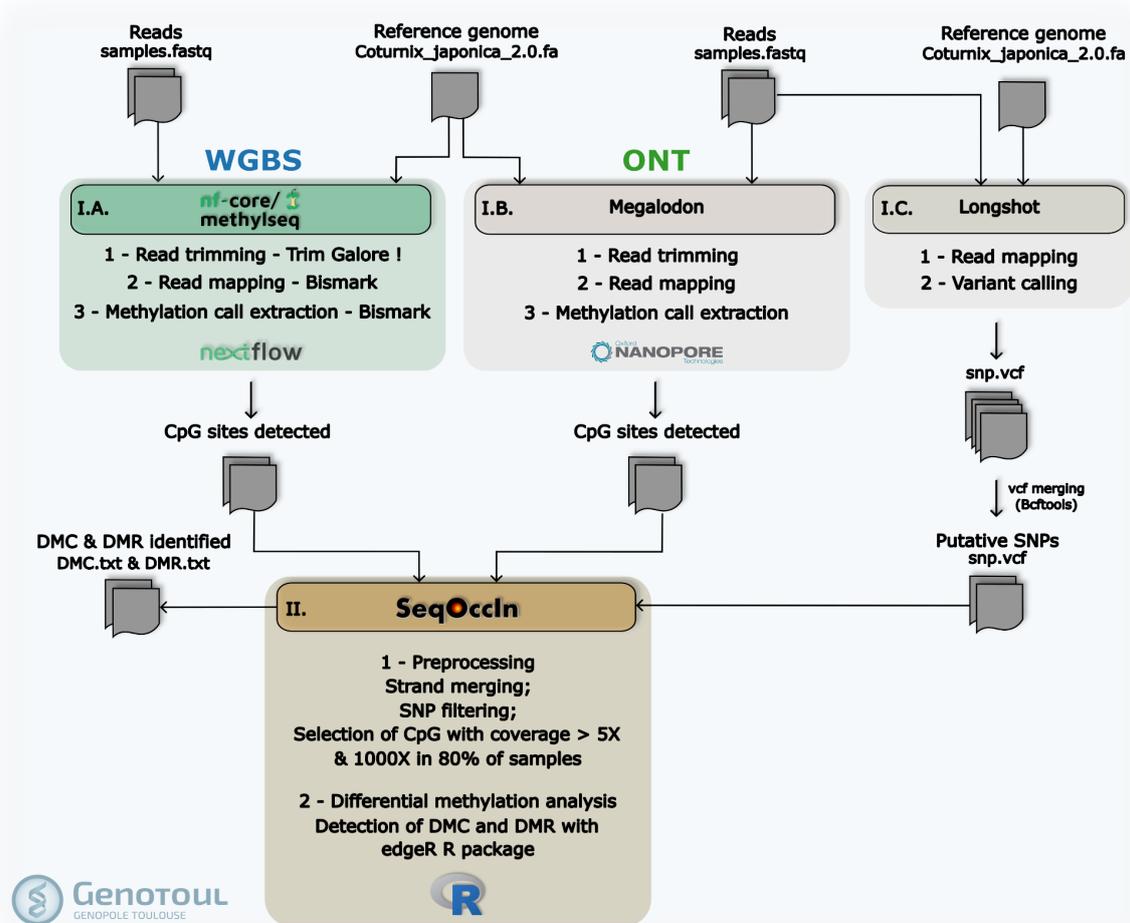
ID	Read (M)	Aligned (%)	CpG (K)	mCpG (%)
AB75069240	6.6	94.1	795.6	43.2
AB75005544	4.5	90.6	795.5	45.0
AB75230653	4.1	93.7	795.5	46.6
AB75007842	5.9	93.5	795.5	45
AB75070045	7.6	93.8	795.4	44
AB75069424	4	90.7	795.4	46.7
AB75007905	4.4	93.5	795.5	44.1
AB75005752	7.3	93.2	795.6	45.4
AB75070052	3.3	91.6	795.5	44.8
AB75007856	4.7	93.9	795.5	47
AB75007836	6.7	94.8	795.5	45.3
AB75069306	8.1	93.5	795.6	48
AB75069430	7.1	91.9	795.6	44.7
AB75005572	6.7	93.4	795.6	46.6
AB75005603	6.5	92.3	795.5	46.4
AB75007901	6.4	93.9	795.5	45.1
AB75230655	4.6	93.9	795.6	44.1
AB75005816	5.9	92.4	795.6	47
AB75070049	4.9	94	795.6	45
AB75070062	4.4	93.7	795.5	44.2
AB75070036	4.2	93	795.5	46.7
AB75005611	4	93.8	795.5	48.5
AB75007899	4.1	93.7	795.6	45.6

**Supplementary Table S3. R packages used for the visual representations.**

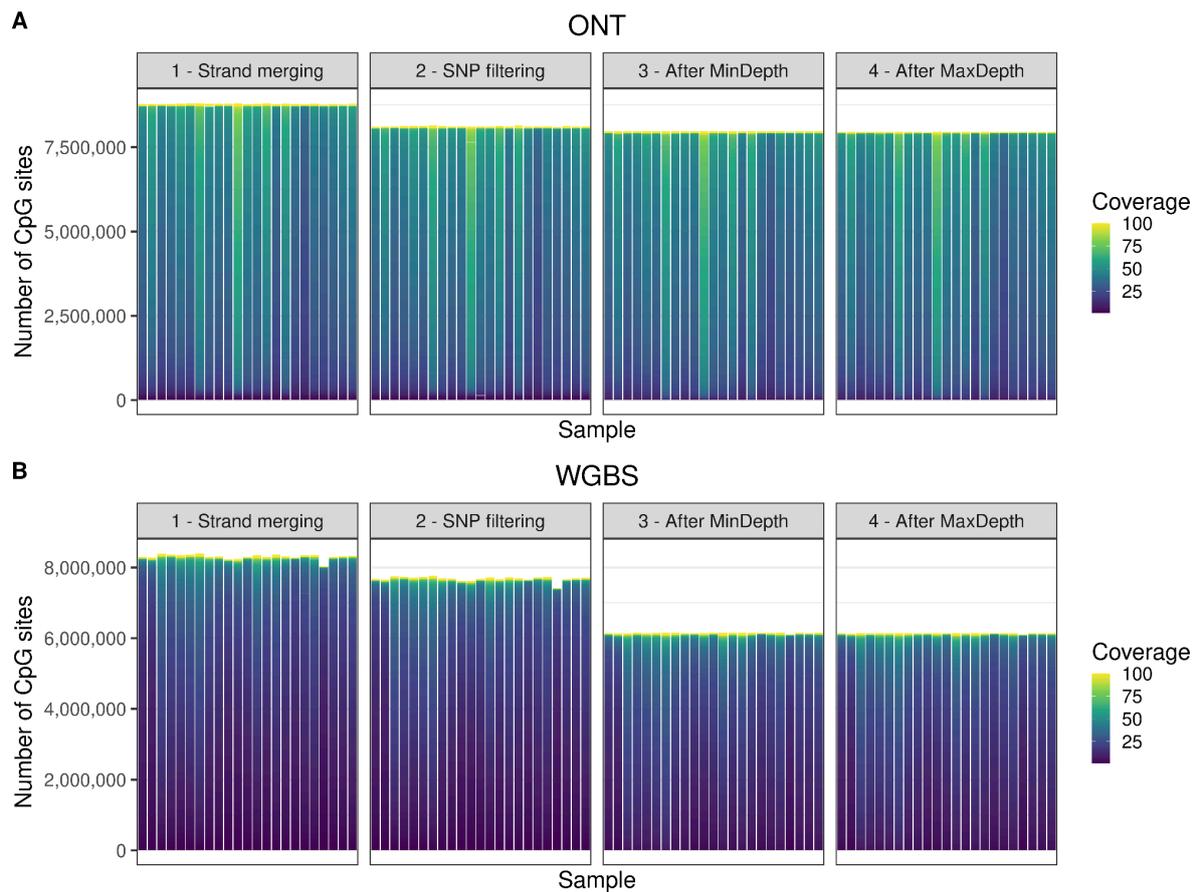
Plot types	R packages	Package version	References
Histogram / Boxplot / Scatterplot / Dotplot / Barplot / PCA	<i>ggplot2</i>	3.4.2	H. Wickham. <i>ggplot2: Elegant Graphics for Data Analysis</i> . Springer-Verlag New York, 2016
Upset	<i>UpSetR</i>	1.4.0	Nils Gehlenborg (2019). <i>UpSetR: A More Scalable Alternative to Venn and Euler Diagrams for Visualizing Intersecting Sets</i> . R package version 1.4.0. <a href="https://CRAN.R-project.org/package=UpSetR">https://CRAN.R-project.org/package=UpSetR</a>
Heatmap	<i>pheatmap</i>	1.0.12	Raivo Kolde (2019). <i>pheatmap: Pretty Heatmaps</i> . R package version 1.0.12. <a href="https://CRAN.R-project.org/package=pheatmap">https://CRAN.R-project.org/package=pheatmap</a>
Manhattan	<i>qqman</i>	0.1.8	Stephen Turner (2017). <i>qqman: Q-Q and Manhattan Plots for GWAS Data</i> . R package version 0.1.4. <a href="https://CRAN.R-project.org/package=qqman">https://CRAN.R-project.org/package=qqman</a>
DMR line plot	<i>dmrseq</i>	1.10.1	Korthauer, K., Chakraborty, S., Benjamini, Y., and Irizarry, R. A. Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing. <i>Biostatistics</i> kxy007, <a href="https://doi.org/10.1093/biostatistics/kxy007">https://doi.org/10.1093/biostatistics/kxy007</a>



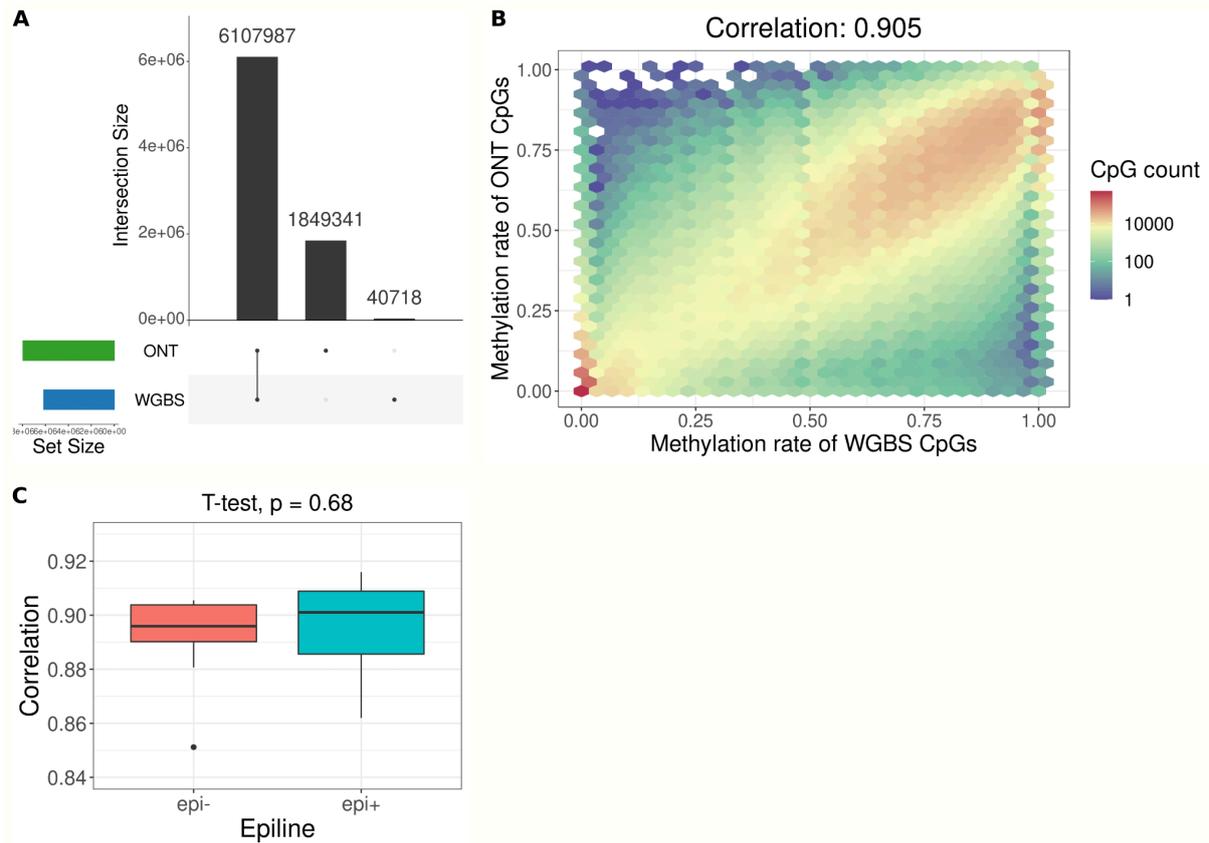
**Supplementary Figure S1. Experimental design from WGBS and ONT.** For WGBS and ONT, blood samples were collected from quails (males and females) from two groups: a control group (epi-) and an injected with genistein (epi+). The quails studied were from the experimental line S+. A total of 23 quails separated into 2 batches were recovered. Blood samples were taken from each quail.



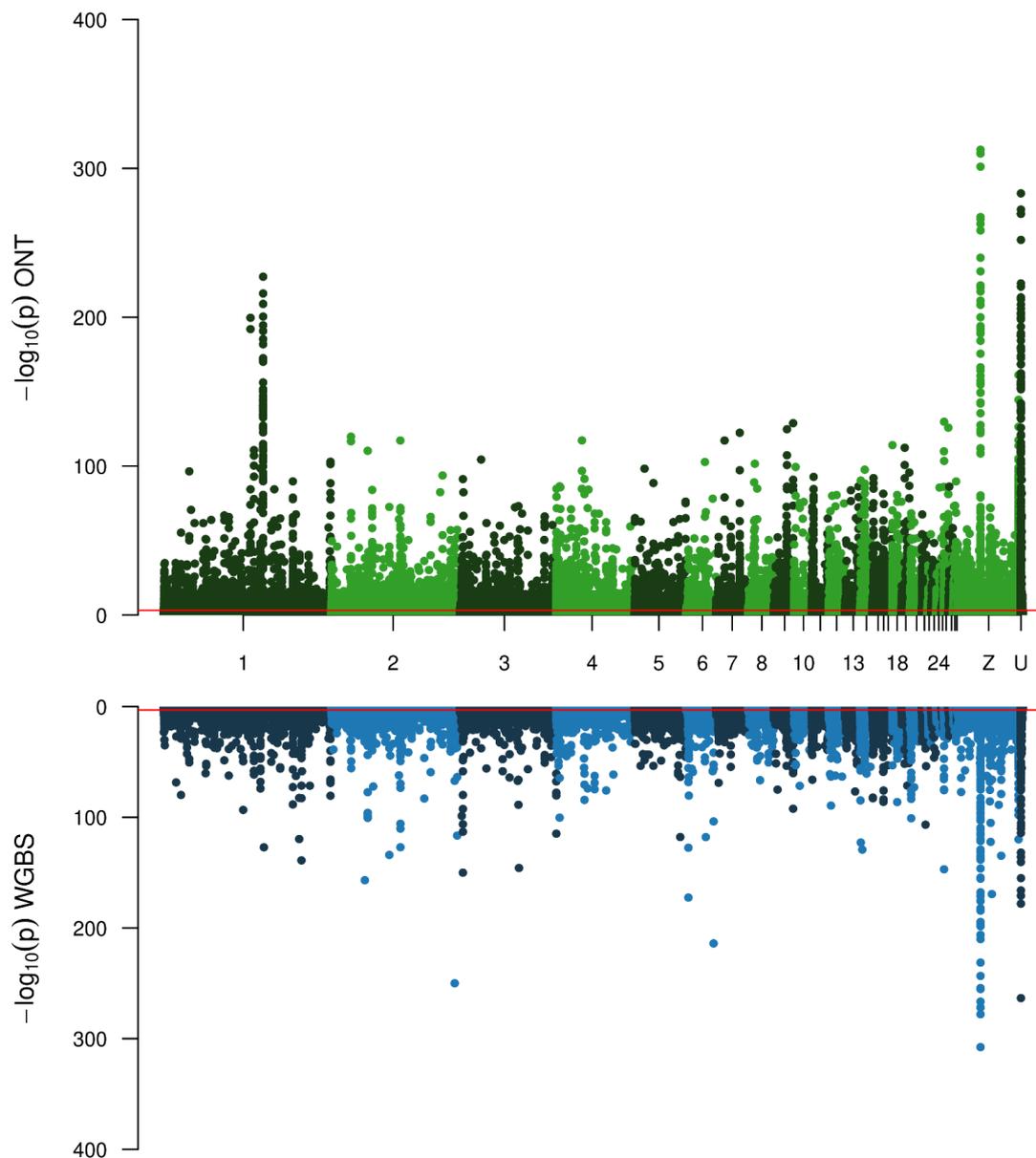
**Supplementary Figure S2. Methodology for bioinformatics analyses.** WGBS reads were processed with the (I.A) nf-core/methylseq tool. (I.A.1) Reads were trimmed with Trim Galore and trimmed reads were (I.A.2) mapped to the quail reference genome by Bismark. (I.A.3) Bismark extracts methylation sites and generates a cytosine report for each sample. ONT reads were processed with the (I.B) megalodon tool. The outputs of nf-core/methylseq and megalodon are analyzed by an (II) in-house EdgeR-based R pipeline (SeqOccin) to detect differentially methylated cytosines (DMC) between experimental groups : (II.1) a pre-processing step and (II.2) a differential methylation analysis are performed. The pre-processing step consists of merging methylation and coverage information from both strands. A filter was then applied to remove SNPs and three additional filtering were applied : CpG sites needed to have a minimum of 5 and maximum of 1000 read coverage in 80% of samples. (I.C) The longshot tool was used to detect putative SNPs present directly in ONT samples and performs the variant calling extracting SNPs. The retrieved SNP list was merged with the reference quail VCF file by bcftools. (II.1) The combined VCF file was used to filter putative SNP position from the Bismark outputs. (II.2) DMC were called by the edgeR package and differentially methylated regions (DMR) were called by *callDMR* (*DSS* package) from the edgeR outputs. The p-values were corrected with the Benjamini-Hochberg (BH) method, with a threshold of 5% defining significance after multiple testing corrections.



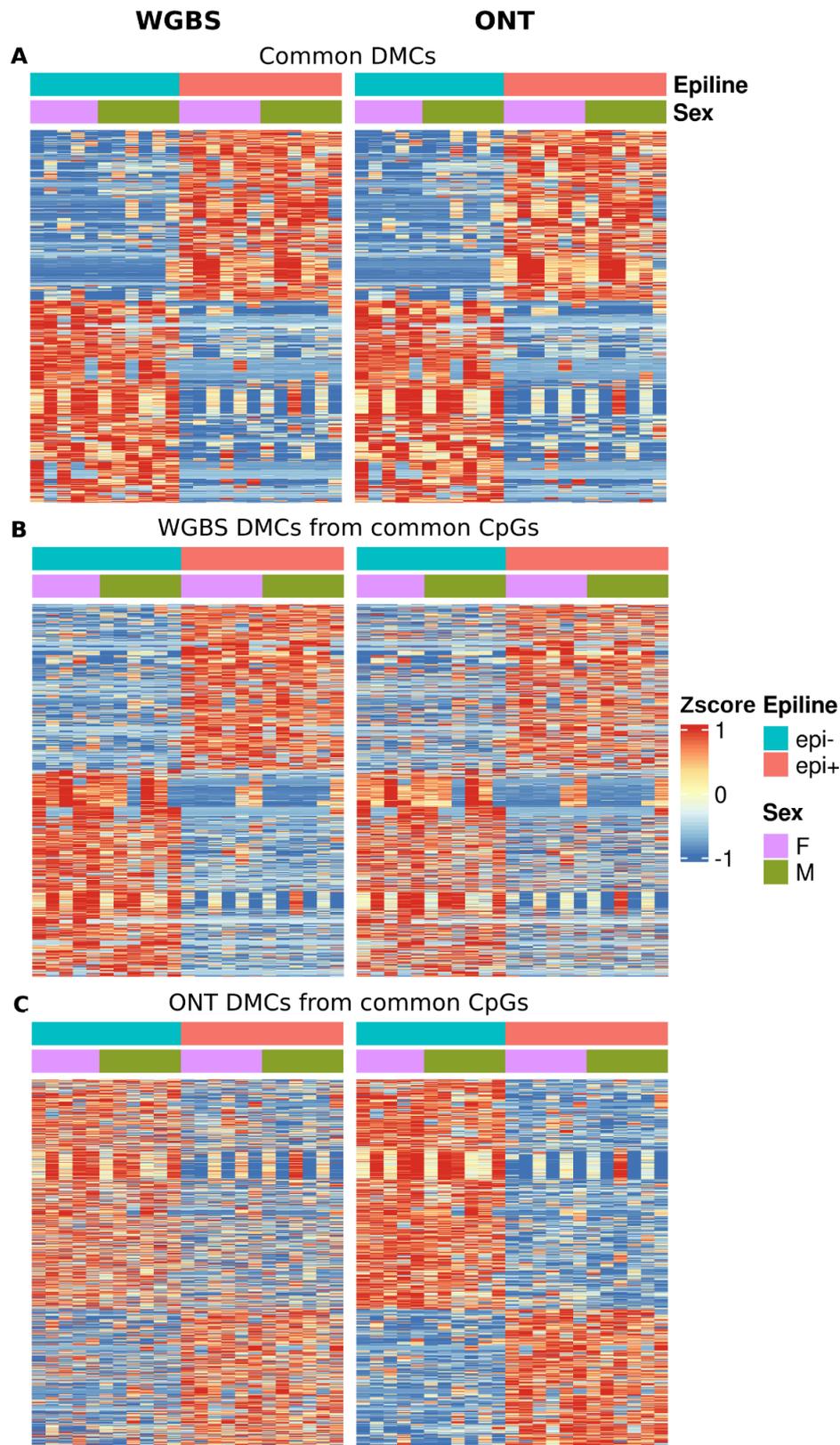
**Supplementary Figure S3. Evolution of the number of CpG sites during filtering steps for WGBS data (A) and ONT data (B).** Samples are arranged along the x-axis according to their mean CpG coverage. Y-axis : Number of CpG sites (1) at the output of the nf-core/methylseq pipeline (after deduplication of CpG sites on the minus strand) for WGBS data and after megalodon for ONT data, (2) after SNP filtering, (3) after filtering out CpG sites with a coverage of less than 5 in at least 20% of the samples, (4) after filtering out CpG sites with a coverage of more than 1,000. (A) WGBS: initial number: 77,608,925 - 183,777,730 CpGs; after (1) 8,037,454 - 8,397,640 CpGs; after (2) : 7,420,571 - 7,767,732 CpGs; after (3): 6,113,069 - 6,158,926 CpGs. At the end of preprocessing: an average of 6,142,783 CpGs were selected (68.4% of the quail genome CpGs). (B) ONT: initial number: 205,667,781 - 490,340,313 CpGs; after (1): 8,767,445 - 8,806,834 CpGs; (2): 8,107,541 - 8,145,778 CpGs; (3): 7,978,519 - 7,978,519 CpGs. At the end of preprocessing: an average of 7,955,339 CpGs were selected (88.6% of quail genome CpGs).



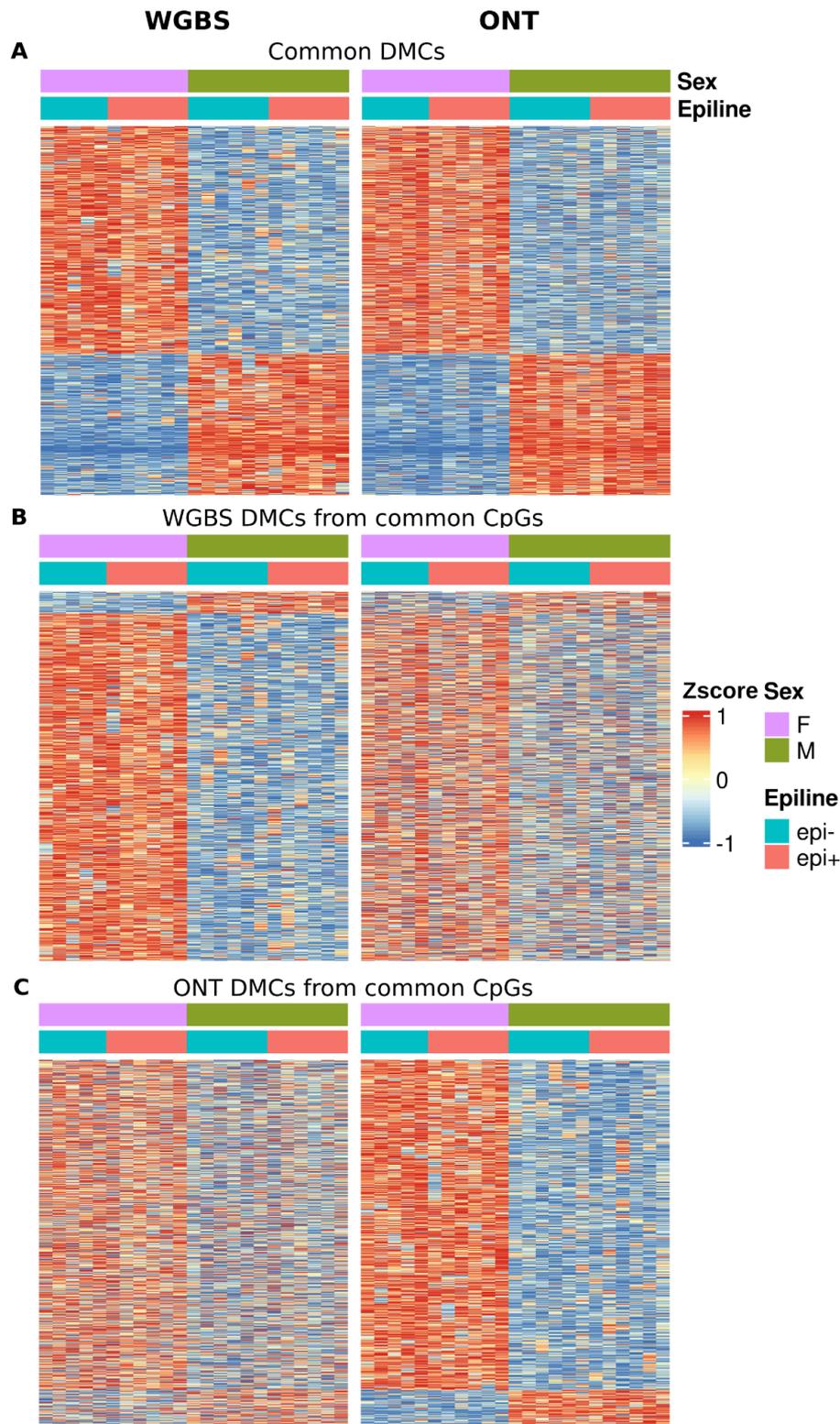
**Supplementary Figure S4. Correlation of the methylation rate between WGBS and ONT.** (A) Upset of total CpG tested from WGBS and ONT data. (B) Correlation of the methylation rate of common CpG sites detected in one sample from WGBS and ONT data ( $n=6,105,624$ ,  $\text{mean}_{\text{corr}} = 0.89$ ,  $\text{sd}_{\text{corr}} = 0.02$ ). (C) Boxplot of correlation values between WGBS and ONT data by epilene (epi-:  $\text{min}=0.85$ ,  $\text{max}=0.91$ ,  $\text{mean}=0.89$ ,  $\text{sd}=0.02$ ; epi+:  $\text{min}=0.86$ ,  $\text{max}=0.92$ ,  $\text{mean}=0.9$ ,  $\text{sd}=0.02$ ). A student test (*T-test*) was performed to compare the correlation mean between epi- and epi+ groups ( $p\text{-value} = 0.68$ ).



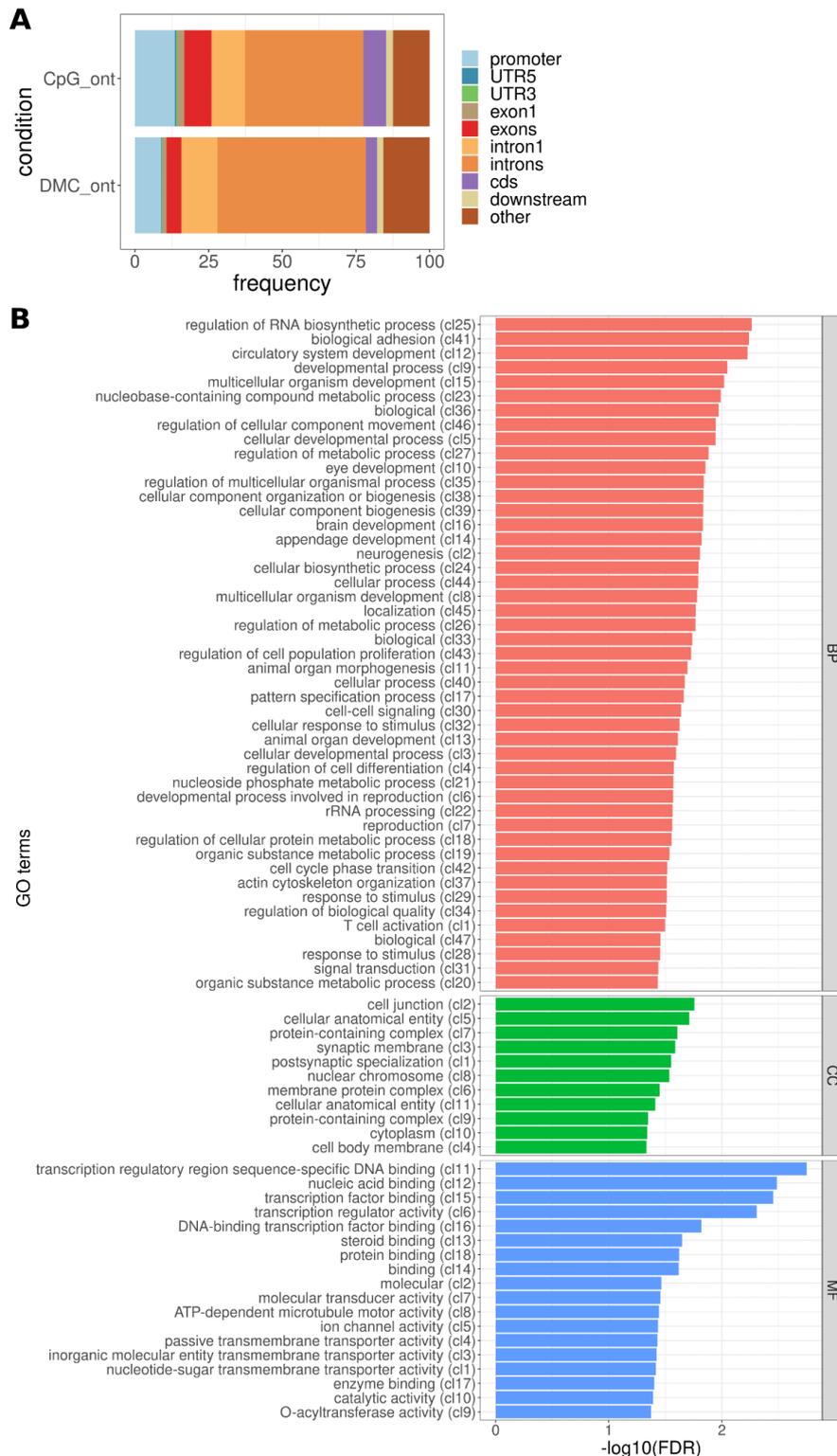
**Supplementary Figure S5. Manhattan plots of sex-associated CpGs identified from ONT and WGBS data.** Manhattan plots showing the distribution along the genome of CpGs identified in blood samples from ONT (top, in green) and WGBS (bottom, in blue) data. Each point refers to a CpG site ( $n_{\text{CpG\_ONT}}=7,957,328$   $n_{\text{CpG\_WGBS}}=6,148,705$ ). The horizontal red bar represents the significance threshold for a False Discovery Rate (FDR) of 5%. X-axis: chromosome names (1-28, W, Z, U) with "W" and "Z" for sex chromosomes and "U" for unassigned contigs. Y-axis: the p-values for differential methylation levels by treatments, without multiple testing correction.



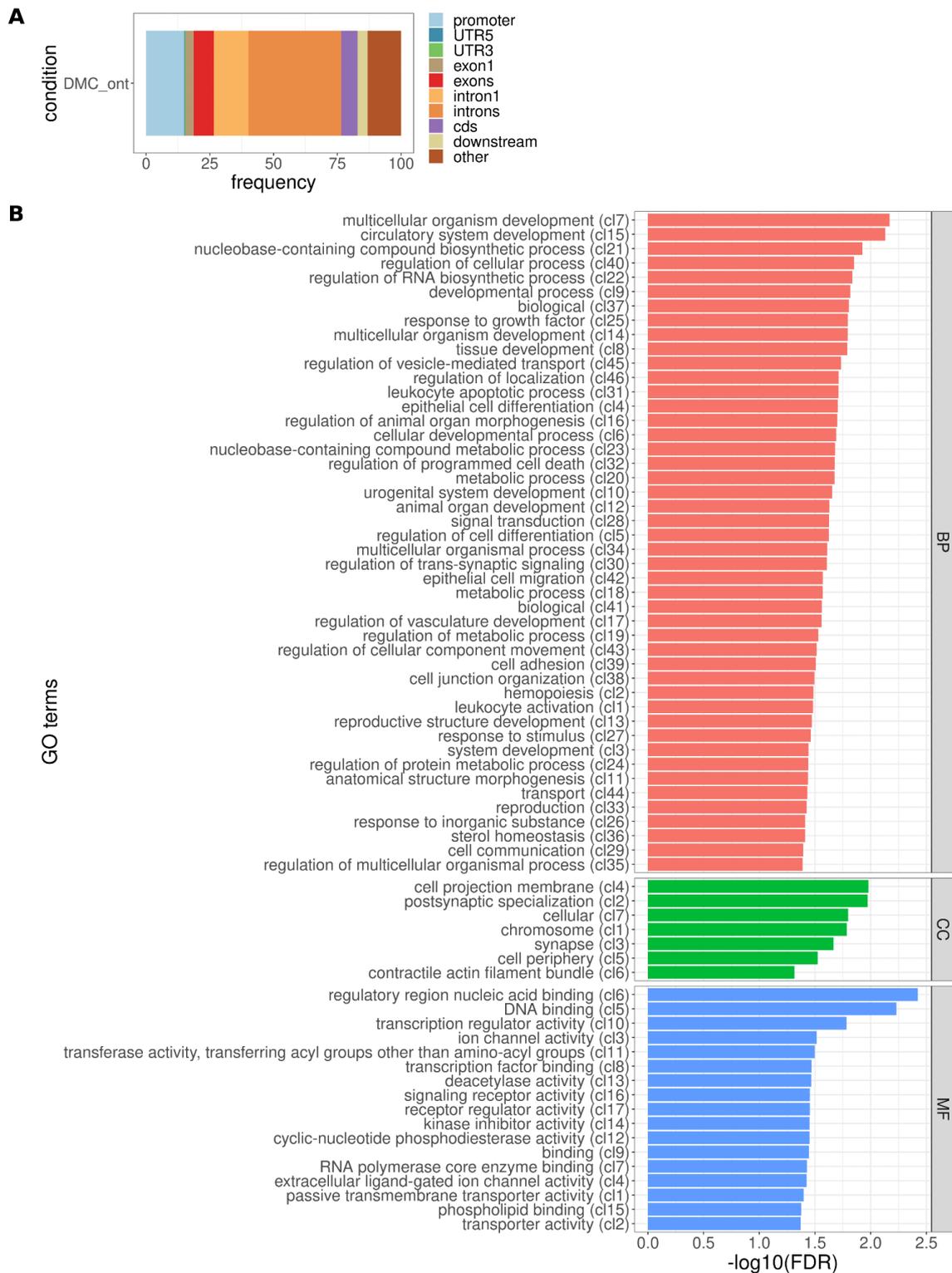
**Supplementary Figure S6. Heatmap of treatment-associated DMCs from WGBS and ONT data.** (A) Common treatment-associated DMCs ( $n = 7,497$  DMCs). (B) WGBS treatment-associated DMCs from common CpGs between WGBS and ONT data ( $n = 9,991$  DMCs). (C) ONT treatment-associated DMCs from common CpGs between WGBS and ONT data ( $n = 6,444$  DMCs). The color gradient (blue to red) is proportional to the z-score methylation rate: blue and red for respectively hypo-methylated and hyper-methylated cytosines.



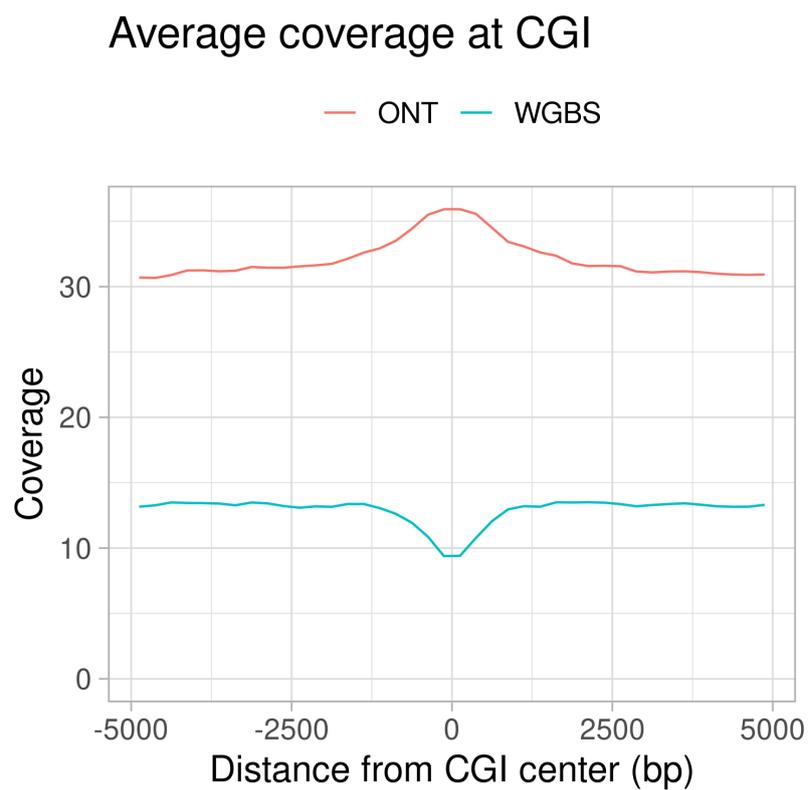
**Supplementary Figure S7. Heatmap of sex-associated DMCs from WGBS and ONT data.** (A) Common sex-associated DMCs ( $n = 10,000$  random DMCs from 25,807). (B) WGBS sex-associated DMCs from common CpGs between WGBS and ONT data ( $n = 10,000$  random DMCs from 81,905 DMCs). (C) ONT sex-associated DMCs from common CpGs between WGBS and ONT data ( $n = 10,000$  DMCs from 81,362 DMCs). The color gradient (blue to red) is proportional to the z-score methylation rate: blue and red for respectively hypo-methylated and hyper-methylated cytosines.



**Supplementary Figure S8. Annotation and Gene Ontology (GO) terms associated with the DMC treatment from ONT data.** (A) Percentage of CpGs and identified DMCs across genomic features in blood using GenomeFeatures. Promoter and downstream regions were defined upstream of Transcription Start Site (TSS) and Transcription End Site (TES), respectively. (B) Histograms showing the  $-\log_{10}(\text{FDR})$  according to the GO term for DMCs treatment identified from ONT data. Red, green and blue colors refer to the Biological Process (BP), Molecular Functions (MF) and Cellular Components (CC) GO terms, respectively.



**Supplementary Figure S9. Annotation and Gene Ontology (GO) terms associated with the best DMC treatment from ONT data.** (A) Percentage of CpGs and identified DMCs across genomic features in blood using GenomeFeatures. Promoter and downstream regions were defined upstream of Transcription Start Site (TSS) and Transcription End Site (TES), respectively. (B) Histograms showing the  $-\log_{10}(\text{FDR})$  according to the GO term for DMCs treatment identified from ONT data. Red, green and blue colors refer to the Biological Process (BP), Molecular Functions (MF) and Cellular Components (CC) GO terms, respectively.



**Supplementary Figure S10. Mean read coverage around CpG islands. WGBS vs ONT.**

## 12. Programmation thermique transgénérationnelle

### 12.1 Présentation du projet QuailHeatE

Au cours de ma thèse, j'ai également participé à l'analyse de données de méthylation de l'ADN issues de cailles japonaises ayant reçu une manipulation thermique (TM) durant le stade embryonnaire (projet QuailHeatE, <https://anr.fr/Projet-ANR-15-CE02-0009>, Vincent Coustham - INRAE, Nouzilly). Le projet QuailHeatE vise à comprendre comment l'environnement embryonnaire affecte l'épigénome des oiseaux tout au long de leur développement. Pour cela, une étude approfondie sur les effets moléculaires du traitement embryonnaire à la chaleur chez les cailles a été réalisée, en partie par l'analyse de données de méthylation de l'ADN dans le but de détecter les régions impactées par la TM. Ce projet vise également à étudier la potentielle transmission transgénérationnelle et multi-générationnelle des effets de la TM.

Une étude précédente a montré que la TM durant le développement embryonnaire améliore la tolérance à la chaleur chez les poulets de chair mâles (Piestun et al., 2008). Les premiers résultats du projet QuailHeatE démontrent que la TM pendant l'embryogenèse a un impact durable sur le développement et la physiologie de la caille japonaise, y compris un effet sur le poids et la mortalité (Vitorino Carvalho et al., 2020). Les résultats phénotypiques de l'analyse transgénérationnelle, non encore publiés, montrent que la manipulation thermique répétée pendant 4 générations a un effet sur l'éclosion, le poids corporel et le poids des œufs pondus, certains effets étant plus intenses au fil des générations. Les effets sur le poids corporel et le poids des œufs peuvent être transmis de manière transgénérationnelle, ce qui suggère des mécanismes d'hérédité non génétique. Cette hypothèse est renforcée par la réversion observée de l'effet sur la croissance après cinq générations non exposées.

L'objectif de la suite de ce projet, sur le plan moléculaire, est d'étudier les effets de la TM sur la méthylation de l'ADN chez la caille japonaise. Pour cela, des données de méthylation de l'ADN issues de séquençage WGBS et provenant de cailles de la 1ère et 4ème génération ont été analysées. Pour la première génération, les données proviennent d'échantillons de l'hypothalamus séparés en 2 groupes : un groupe contrôle n'ayant subi aucun choc thermique (C) et un groupe ayant reçu une TM (TM). Les données de la 3ème génération sont séparées en 3 groupes : un groupe contrôle représentant la descendance des cailles non traitées pendant l'incubation (C4), un groupe dont les ascendants des générations G1, G2 et G3 ont reçu une TM (TM4) et un dernier groupe dont seuls les ascendants G1 et G2 ont reçu une TM (TM2C2). Pour chacun de ces 3 groupes, des données issues de l'hypothalamus et du foie ont été extraites (**Figure 12.1**).

Ma contribution dans ce projet porte sur l'analyse de méthylation différentielle entre les différents groupes étudiés à l'aide du pipeline Seq0ccin. Pour analyser ces données, le package DSS de R

a été utilisé par comparaison des conditions deux à deux. Ces analyses ont permis de mettre en évidence des régions différenciellement méthylées entre certaines des conditions étudiées. Les résultats bruts ont été transmis à Vincent Coustham et sont en cours d'analyse approfondie, ils ne sont pas détaillés dans ce manuscrit.

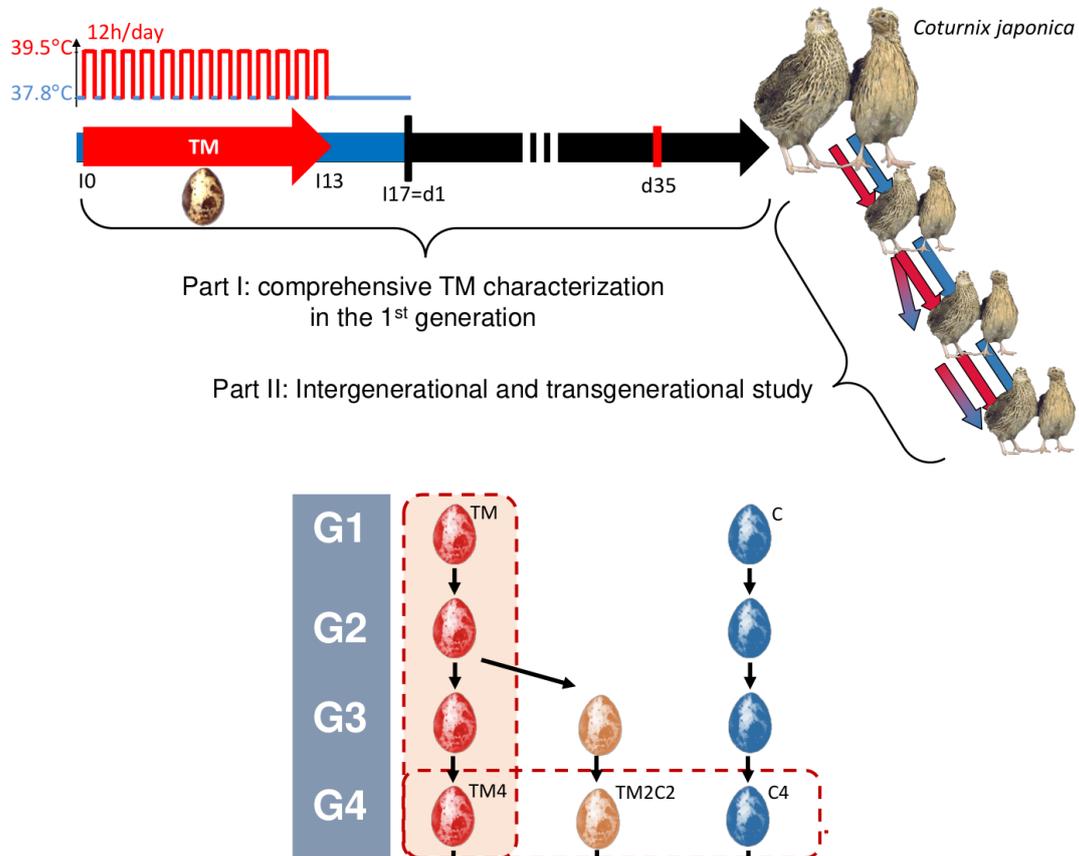


FIGURE 12.1 – **Dispositif expérimental du projet QuailHeatE.** Des données WGBS ont été extraites d'échantillons de foie et de l'hypothalamus de la première génération (G1) et de la quatrième génération (G4). Les cailles sont séparées en trois groupes : un groupe contrôle (C) n'ayant pas reçu de manipulation thermique, un groupe ayant reçu une TM à la G1 jusqu'à la G2 mais plus à partir de la G3 (TMCM), et un dernier groupe ayant reçu une TM de la G1 à la G4 (TM). Les échantillons de foie n'ont été prélevés que sur la G4. *Figure de V. Coustham.*

# VI

## Discussion générale et perspectives

<b>13</b>	<b>Rappels des principaux résultats</b> .....	<b>225</b>
<b>14</b>	<b>Dispositif expérimental</b> .....	<b>227</b>
14.1	Choix et intérêt du modèle de la Caille	
14.2	Forces et limitations des dispositifs expérimentaux étudiés	
14.3	Comparaison avec le dispositif de manipulation thermique	
14.4	Recommandations pour la création d'un dispositif expérimental pour étudier les transmissions épigénétiques inter et transgénérationnelles	
<b>15</b>	<b>Avantages et inconvénients des différentes technologies de séquençage haut-débit pour l'étude de la méthylation de l'ADN chez la caille</b> .....	<b>231</b>
15.1	Préparation des échantillons	
15.2	Nombre de CpG/DMC détectés à partir des différentes technologies	
15.3	La technologie ONT est-elle la meilleure méthode ?	
15.4	Puce de méthylation	
<b>16</b>	<b>Développement du pipeline <i>SeqOccin</i> et pistes d'améliorations</b> <b>233</b>	
16.1	Rappel de nouveaux éléments implémentés par rapport à l'ancienne version	
16.2	Normalisation des données	
16.3	Filtrage des SNP	
16.4	Nouvelles améliorations possibles du pipeline <i>SeqOccin</i>	
<b>17</b>	<b>Validation des DMC et impacts fonctionnels</b> .....	<b>237</b>
17.1	Pyroséquençage	
17.2	Edition de l'épigénome par transfection	
17.3	Lien entre l'épigénétique et les ARN non codants : rôle dans la régulation de l'expression des gènes	
17.4	Projets internationaux contribuant à l'acquisition de données et à l'amélioration des connaissances de l'annotation des génomes	
<b>18</b>	<b>Prise en compte des variations épigénétiques dans les schémas de sélection</b> .....	<b>241</b>



L'étude de phénomènes épigénétiques intergénérationnels et transgénérationnels a souvent été effectuée chez l'Homme et les animaux de laboratoire (souris, nématode), mais a été très peu réalisée chez les animaux d'élevage, notamment chez les oiseaux. La question majeure que je me suis posée dans ce projet de thèse est de savoir s'il existe des transmissions des effets de l'environnement à travers les générations chez la caille. Nous avons pu mettre en évidence une potentielle transmission des effets de l'environnement sur le développement embryonnaire chez la caille ce qui ouvre la voie pour l'étude de ce type de phénomènes chez d'autres espèces de rente.



## 13. Rappels des principaux résultats

La première partie de cette thèse consiste à étudier les effets de l'exposition précoce *in-ovo* à des perturbateurs endocriniens sur les phénotypes et la méthylation de l'ADN de cailles. L'analyse de données phénotypiques issues de G0 a pu mettre en évidence un effet significatif des différentes molécules injectées sur la croissance des cailles, les individus injectés ayant un poids plus faible que les individus non injectés. Ces molécules n'affectent pas les caractéristiques de la ponte et les effets détectés ne semblent pas dépendre du fond génétique testé. Un effet significatif de l'injection seule, pouvant être attribué à l'huile de sésame et/ou à l'éthanol (Legault et al., 2018; Wallén et al., 2021), a également été observé sur la croissance des cailles. J'ai ensuite réalisé une analyse de méthylation différentielle à partir de données issues de séquençage RRBS et WGBS, basée sur le traitement de l'ADN au bisulfite de sodium. Le séquençage RRBS ne permet d'étudier qu'une sous-représentation du génome de la caille (environ 2%), principalement au niveau des régions promotrices des gènes par l'utilisation d'une enzyme de restriction *MspI*. Ainsi, cette technique ne nous permet pas d'obtenir les informations de méthylation au niveau de la majorité du génome. Afin de pouvoir étudier l'intégralité du génome, des données de séquençage tout génome étaient également à ma disposition. Parmi ces dernières, des données sont issues du séquençage WGBS, une technique également basée sur le traitement au bisulfite de l'ADN. L'analyse de méthylation différentielle que j'ai effectuée sur ces données a également permis d'identifier des régions différentiellement méthylées entre les conditions expérimentales, mettant en évidence un effet des molécules sur des gènes impliqués dans la croissance, le comportement et la reproduction.

La deuxième partie de cette thèse a pour objectif d'étudier les effets transgénérationnels de l'exposition *in-ovo* à des perturbateurs endocriniens au niveau de la méthylation de l'ADN de la descendance. Ainsi, trois générations après l'injection de génistéine dans l'œuf, l'analyse de méthylation différentielle a été réalisée à partir de données issues de séquençage WGBS et une autre technique de séquençage ne nécessitant pas de traitement de l'ADN au bisulfite de sodium a été utilisée : le séquençage nanopore (ONT). L'analyse de ces données m'a permis d'identifier des régions différentiellement méthylées entre les deux épilignées. Même si les effets génétiques ont été minimisés, une cause génétique à certains différentiels de méthylation observés n'a pas pu être écartée. La distribution de ces DMCs à partir des données ONT le long du génome a pu mettre en évidence des régions présentant une forte densité de DMC, qui semblent être co-localisées avec des SNP de fréquences alléliques distinctes entre les épilignées. Des phénomènes de méthylation allèle-spécifiques (ASM) ont été identifiés, le niveau de méthylation de ces DMC étant spécifique à chaque allèle d'un locus. Cependant, même si la plupart des différentiels de

méthylation de l'ADN entre les épilignées semblent être liées à des différences génétiques, certaines des régions identifiées semblent bien être liées à une transmission épigénétique transgénérationnelle des effets de l'environnement. L'injection de génistéine semble impacter des gènes liés aux voies métaboliques, du développement du cerveau et de la reproduction. A partir de données issues de cailles ayant reçu une manipulation thermique au stade embryonnaire (projet QuailHeatE), j'ai pu mettre en évidence un effet transgénérationnel de la TM. Ces analyses ont donc permis de caractériser la réponse épigénétique aux variations de l'environnement embryonnaire et de mettre en évidence des régions du génome affectées par la transmission de ces effets chez la caille.

## 14. Dispositif expérimental

### 14.1 Choix et intérêt du modèle de la Caille

La caille est une espèce très intéressante en partie pour son intervalle de génération qui est très court par rapport à d'autres espèces ainsi que pour son génome qui est assez proche de celui de la poule, ce dernier étant très bien annoté (Morris et al., 2020).

### 14.2 Forces et limitations des dispositifs expérimentaux étudiés

Pour caractériser la réponse de l'organisme aux effets de l'environnement embryonnaire, divers dispositifs expérimentaux ont été utilisés entre les différentes générations. Les données phénotypiques de la G0 ainsi que les données RRBS sont issues du même dispositif expérimental permettant ainsi de faire directement le lien entre les phénotypes impactés par les molécules et les régions présentant des différentiels de méthylation. Dans ce dispositif, deux groupes contrôles, un non injecté et un injecté avec de l'huile de sésame (étant le dilueur des différentes molécules injectées), ont été utilisés, ce qui nous a permis de distinguer les effets spécifiques de chaque molécule de celle de l'injection seule sur les phénotypes étudiés. Pour l'analyse de méthylation différentielle, le groupe contrôle injecté a également été utilisé comme groupe de référence. Ainsi, les différentiels de méthylation observés seraient bien induits par la molécule injectée et non pas par l'acte de l'injection seule. Concernant les données WGBS, ces dernières ont été issues d'un sous-ensemble d'individus du dispositif du RRBS. Seulement une lignée (S+), un tissu (sang) et quatre groupes de traitement (contrôle non injecté, génistéine, bisphénol A et 5Aza) ont été analysés, sans groupe contrôle injecté. Ainsi, le groupe non injecté étant utilisé comme groupe de référence, il est probable que les différences observées ne soient pas nécessairement induites par la molécule mais par l'injection seule. L'utilisation d'un contrôle non injecté est justifiée par l'objectif initial de mieux comprendre la transmission transgénérationnelle des modifications de la méthylation de l'ADN, quelle que soit la cause des différences du profil de méthylation observées entre les différentes conditions. De plus, contrairement aux données phénotypiques qui ont été récupérées chez chaque individu, les données de méthylation ont été groupées par condition (même tissu, lignée, traitement) pour des raisons budgétaires, ne permettant pas de comparer les niveaux de méthylation à l'échelle de l'individu. Concernant les données de méthylation de la G3, seulement deux groupes ont été étudiés, un groupe injecté avec de la génistéine et un groupe contrôle non injecté, issus d'une seule lignée et d'un tissu. Comme pour la G0, des différentiels de méthylation de l'ADN observés entre les deux épilignées pourraient être attribués à l'effet de la génistéine ou bien de l'injection seule.

L'information du sexe pour les données WGBS de la G0 et de la G3 était disponible contraire-

ment au RRBS. Chez les mammifères mais aussi chez les oiseaux, certains des effets environnementaux précoces sur le phénotype de la descendance ainsi que le profil de méthylation de certaines régions du génome peuvent être spécifiques au sexe des individus (A. Gabory et al., 2011b; Brun et al., 2015; Leroux et al., 2017). L'information du sexe a été prise en compte dans les analyses de méthylation différentielle pour les données WGBS (G0 et G3) et ONT (G3). Concernant les données RRBS, les données issues des mâles et femelles ayant été regroupées en pool, l'information du sexe n'a pas pu être prise en compte dans les modèles d'analyse différentielle. Pour chaque dispositif expérimental, un nombre équilibré de mâles et de femelles a été utilisé.

Concernant l'étude d'une transmission transgénérationnelle des effets de l'environnement maternel, des données de méthylation de l'ADN issues de la G2 auraient pu être étudiées, la transmission étant qualifiée de transgénérationnelle à partir de la G2. En effet, l'exposition d'une caille (G0) porteuse de cellules germinales (G1) à un stress environnemental peut impacter ses marques épigénétiques, dont la méthylation de l'ADN, ainsi que celles de la descendance G1 : les G0 et G1 sont considérées comme étant directement impactées. La G2 n'est en revanche pas directement impactée.

Pour la G3, les données issues de plusieurs technologies de séquençage ont pu être comparées (ONT et WGBS). Ces données étant issues des mêmes échantillons de sang, une comparaison des deux technologies a pu être effectuée dans les meilleures conditions.

En complément des données de méthylation, il aurait été intéressant d'effectuer une analyse d'expression différentielle à partir de données issues de séquençage ARN issues des mêmes échantillons et pour l'ensemble des dispositifs étudiés. Cela nous aurait permis d'étudier le niveau d'expression de gènes sur lesquels des différentiels de méthylation ont été identifiés et potentiellement confirmer l'impact d'une partie des DMC identifiées. Nos résultats ne permettent donc pas de différencier des marques épigénétiques associées à un différentiel d'expression simultané de marques potentiellement impliquées dans la "programmation" d'une réponse plus tardive.

### 14.3 Comparaison avec le dispositif de manipulation thermique

Une étude portant sur l'effet de la manipulation thermique (TM) chez la caille japonaise a été réalisée (projet QuaiHeatE). Après deux générations de cailles ayant reçu une MT, le poids des cailles à 4 semaines a été significativement réduit par rapport au groupe contrôle. Ce phénomène est également observé sur des cailles de la 3ème génération âgées de 4 et 5 semaines (Carvalho et al., 2023). Cela suggère l'existence d'un mécanisme de compensation de croissance pour les deux premières générations ayant reçu une MT (Vitorino Carvalho et al., 2020) qui ne serait pas suffisante après plusieurs répétitions de la MT au fil des générations, mettant ainsi en évidence un effet cumulatif de la TM. Par rapport au dispositif du projet Caramel, celui du projet QuailHeatE permet ainsi d'étudier l'effet cumulatif de la TM sur plusieurs générations. Il serait intéressant de la prendre en compte dans la suite de nos analyses. Il est intéressant de se demander si une exposition répétée *in-ovo* pendant plusieurs générations peut augmenter les effets observés et si les effets directs, intergénérationnels et transgénérationnels se cumulent au fil des générations.

## **14.4 Recommandations pour la création d'un dispositif expérimental pour étudier les transmissions épigénétiques inter et transgénérationnelles**

Afin d'étudier les transmissions intergénérationnelles et transgénérationnelles des effets de l'environnement, des données issues des mêmes groupes de traitements, tissus, lignées et sexes pourraient être analysées et cela pour chaque génération (G0 à G3 et même au-delà). En fonction des objectifs scientifiques, l'utilisation d'un contrôle injecté permettrait d'identifier des différentiels liés aux molécules injectées. Un effectif suffisant à chaque génération pour réaliser des études génétiques (plusieurs centaines d'animaux) serait très informatif sur les transmissions inter et transgénérationnelles.



# 15. Avantages et inconvénients des différentes technologies de séquençage haut-débit pour l'étude de la méthylation de l'ADN chez la caille

## 15.1 Préparation des échantillons

Les technologies de séquençage utilisées nécessitent des échantillons d'ADN de très bonne qualité afin de permettre une détection optimale des différentiels de méthylation. C'est notamment le cas de la technologie Nanopore dont l'utilisation est optimale avec de l'ADN génomique de haut poids moléculaire. Le traitement de l'ADN au bisulfite de sodium peut fragiliser l'ADN en introduisant notamment des cassures (Grunau et al., 2001). La technologie de conversion enzymatique EM-seq, quant à elle, fragilise moins l'ADN que les technologies basées sur l'utilisation du bisulfite (Han et al., 2022).

## 15.2 Nombre de CpG/DMC détectés à partir des différentes technologies

Très peu de DMC liées à l'effet des traitements ont été identifiées à partir des données RRBS de la G0 par rapport au WGBS G0. Les régions identifiées par le RRBS seraient plutôt impactées par les molécules injectées, alors que parmi celles détectées en WGBS, il est probable d'observer des effets de l'injection seule.

Le WGBS semble donc être un choix plus efficace par rapport au RRBS, bien que plus onéreux, en permettant de détecter plus de régions impactées, en particulier par l'injection seule. Concernant la G3, des milliers de DMC ont été détectés à partir des données WGBS et ONT. Un nombre plus important de sites CpG et de DMC ont été identifiés à partir des données ONT.

La méthode de détection des régions différentiellement méthylées entre les conditions étudiées ne permet pas de récupérer les effets spécifiques des différentes molécules testées. Nous savons que les régions identifiées sont impactées par l'injection des molécules sans savoir de quelle molécule exactement dépend cet effet. Une analyse post-hoc pourrait être réalisée si on s'intéressait à une molécule particulière.

## 15.3 La technologie ONT est-elle la meilleure méthode ?

L'analyse de données issues de séquençage ONT permet d'étudier l'intégralité du génome sans traitement de l'ADN au bisulfite de sodium et également de détecter directement les SNP, ces derniers étant plus fiables que ceux identifiés à partir de données WGBS. Dans le cadre de cette thèse et au vu des résultats obtenus, l'ONT semble être une bonne alternative mais reste une technologie assez onéreuse. Une autre méthode de séquençage basée sur l'ONT et moins onéreuse existe : le séquençage EpiGLow (*Epi-Genotype Low pass sequencing*). Cette dernière est basée sur une profondeur plus faible (Lamb et al., 2021 ; González-Recio et al., 2023) et pourrait être utilisée dans le cadre d'études ultérieures.

## 15.4 Puce de méthylation

La réalisation d'un séquençage du génome complet sur un nombre important d'individus reste limitée par des coûts élevés. Le développement et l'utilisation de puce de méthylation est une alternative moins onéreuse qui permet de quantifier les variations pour un sous-ensemble de marques épigénétiques (27 000 à 1 000 000 de CpGs) à partir d'une sélection de marqueurs épigénétiques représentatifs et informatifs issus d'analyses de génomes entiers.

L'épigénotypage par puce peut se réaliser en routine et être déployé à grande échelle permettant ainsi d'épigénotyper plusieurs milliers d'individus. La première puce qui a vu le jour a été développée chez l'homme (HumanMethylation27 BeadChip - (Thirlwell et al., 2010)) en 2008 couvrant 27 578 loci associés aux promoteurs de plus de 14 000 gènes. L'analyse des marques épigénétiques est très utilisée en cancérologie. A ce jour, ces puces ont été développées chez l'homme (Infinium MethylationEPIC v2.0 935 000 CpG, Infinium-methylation) et la souris (Infinium Mouse Methylation BeadChip avec 285 000 markers CpG, Infinium-mouse-methylation) (Arneson et al., 2022 ; Pidsley et al., 2013, 2016). Une puce multi-espèce pan-mammifère (Infinium-methylation) a également été développée (Arneson et al., 2022) comprenant 36 000 CpG bien conservés chez plus de 200 espèces de mammifères.

A ce jour, des puces d'épigénotypage (EDMA - *EmbryoGENE DNA methylation Array*) sont disponibles pour les espèces de rente telles que le porc (Shojaei Saadi et al., 2014) et la construction d'une puce est en cours au sein d'un projet européen en bovins (RUMIGEN). L'investissement initial dans ce type d'outils est très important, et on ne peut imaginer son utilisation que sur des espèces où l'amortissement est certain grâce à un nombre très importants d'épigénotypages, qui permet un coût raisonnable par individu. C'est le cas pour les humains et les bovins. Cependant, il est extrêmement peu probable que cela soit un jour développé chez la caille.

# 16. Développement du pipeline SeqOccin et pistes d'améliorations

## 16.1 Rappel de nouveaux éléments implémentés par rapport à l'ancienne version

La méthode de comparaison entre les groupes deux à deux (*DSS singlefactor*) a été initialement implémentée par Marjorie Mersch dans le cadre de sa thèse (2018). Au cours de ma thèse, j'ai implémenté l'analyse multifactorielle de *DSS* (*DSS multifactor*) basée sur l'utilisation d'un modèle linéaire permettant de tester chaque paramètre du modèle en prenant en compte d'autres paramètres tels que par exemple la lignée ou bien le sexe des individus. Les données issues du projet QuailHeatE ont été analysées avec l'outil *DSS* en *singlefactor*.

Les analyses effectuées sur les données ONT avec l'analyse multifactorielle de *DSS* ne donnant pas de résultats comparables à celles effectuées sur les données WGBS, certaines fonctionnalités de l'outil *edgeR* permettant également d'effectuer une analyse multifactorielle ont été implémentées dans le pipeline SeqOccin. Les outils *DSS singlefactor* et *edgeR* sur l'ensemble des données permettent une bonne détection des différentiels de méthylation de l'ADN entre les conditions expérimentales étudiées. *EdgeR* a été initialement développé pour l'analyse d'expression différentielle à partir de données RNAseq. Cet outil permet également d'identifier des régions différentiellement méthylées à partir de données issues de séquençage haut débit. L'analyse multifactorielle de *DSS* à partir des données ONT n'a pas permis d'identifier des différentiels de méthylation entre les deux groupes alors que de nombreuses DMC ont été identifiées à partir des données WGBS. Les profils de méthylation entre les deux technologies étant assez similaires, la cause de la détection quasi nulle de DMC à partir de l'ONT reste à élucider.

## 16.2 Normalisation des données

Pour normaliser les données, les auteurs de l'outil *DSS* ont fait le choix d'utiliser la fonction de lien arcsine (G. Yu, 2009) au lieu d'une régression logistique, couramment utilisée, car elle traiterai mieux les données réparties au niveau des limites (taux de méthylation proches de 0 ou 1) (Park & Wu, 2016). Ces taux de méthylation transformés ont ensuite été ajustés par la méthode des moindres carrés. Sur des données simulées, cette normalisation semble bien fonctionner. Cependant, en regardant l'effet de cette normalisation sur les données réelles, nous constatons qu'elle n'a pas l'effet escompté, les données ne semblant pas suivre une distribution normale. Cela dit, le test étant effectué par site CpG, la normalité des données est importante pour chaque site CpG indépendamment, et non en mélangeant tous les sites CpG du génome. Même si l'utilisation de cette normalisation est communément admise pour ce type d'analyse (Binder & Michels, 2013), il serait intéressant d'avoir accès à une validation plus approfondie de cette méthode de normalisation sur des données de méthylation dans la littérature.

Contrairement à DSS, edgeR permet de traiter séparément les lectures méthylées (Me, contenant un C) et non méthylées (Un, contenant un T) pour chaque échantillon. Pour que les lectures Me et Un d'un même échantillon soient traitées à la même échelle, les données sont normalisées par la taille totale de la librairie, les paires de librairies Me et Un ont donc une taille égale ( $C + T$ ) (Y. Chen et al., 2017).

D'autres outils tels que limma (Ritchie et al., 2015) ou DMRcate (Peters et al., 2021) peuvent aussi être utilisés pour effectuer une analyse de méthylation différentielle à partir de données de méthylation.

### 16.3 Filtrage des SNP

Comme expliqué dans la [Chapitre 5.1.4](#), certains SNP peuvent être identifiés à tort comme DMC. Afin de limiter ce biais, nous avons filtré de nos analyses les sites CpG qui coïncidaient avec des SNP fréquents chez la caille combinés aux SNP détectés à l'aide de l'outil gemBS (Merkel, Fernández-Callejo, et al., 2019) (G0), ou bien des SNP détectés directement à partir des données issues du séquençage ONT (G3). Malgré ce filtre, parmi les DMC identifiées en fin d'analyse différentielle, certaines avaient un taux de méthylation moyen à 0 pour une lignée donnée et/ou un tissu donné. Un taux de méthylation systématiquement nul au niveau d'un site CpG est peu probable et correspond souvent à un SNP, nous l'avons d'ailleurs observé en visualisant les lectures sous IGV à quelques positions de ce type. Ces DMC ont donc été filtrées a posteriori de l'analyse différentielle. De plus, même si l'outil gemBS a été développé pour rechercher des variants à partir de séquences génomiques traitées au bisulfite de sodium, les séquences génomiques étant modifiées par le traitement, cela rend la détection des SNP moins fiable que si cette détection avait été faite après séquençage classique d'ADN non traité. D'autres méthodes de séquençage existent pour analyser la méthylation de l'ADN ne nécessitant pas de traitement préalable de l'ADN au bisulfite, notamment le séquençage ONT (Lu et al., 2016). La détection de variants à partir de ce type de données est potentiellement plus fiable mais cette technique de séquençage est bien plus onéreuse que le RRBS ou le WGBS, et son utilisation dans le domaine agronomique en est encore à ses débuts. Le logiciel gemBS a été choisi pour filtrer les SNP après une analyse exploratoire pour identifier le meilleur outil en comparant gemBS, biscuit et BisSNP pour la G0. Cette liste a été enrichie par les SNP potentiels fréquemment rencontrés dans le génome de la caille grâce à un fichier VCF élaboré à partir de données publiées (Bardou, 2022). Concernant la G3, des données WGBS et ONT issues des mêmes individus étant à notre disposition, les SNP ont pu être directement détectés à partir des données ONT à l'aide de l'outil longshot. Ces derniers ont été exfiltrés de la liste des CpG identifiés à partir des données WGBS et ONT.

### 16.4 Nouvelles améliorations possibles du pipeline SeqOccin

Pour faciliter l'utilisation du pipeline SeqOccin, il serait pertinent d'utiliser un gestionnaire de workflow intégrant les étapes de preprocessing et d'analyse différentielle. Même si l'étape de preprocessing est assez rapide, le temps computationnel peut être également amélioré en particulier pour la réalisation des graphiques dans la partie d'analyse différentielle.

Malgré une bonne corrélation entre les données WGBS et ONT de la G3, l'outil DSS *multifactor* a

déte t  un nombre de diff rentiels de m thylation extr mement faible   partir des donn es ONT par rapport au WGBS. Par cons quent, il serait int ressant d'approfondir par la suite l'origine de cette diff rence de d tection.



## 17. Validation des DMC et impacts fonctionnels

Les régions d'intérêt mises en évidence pourront être par la suite validées par pyroséquençage et leur éventuel impact fonctionnel pourrait être étudié par édition de l'épigénome. L'information de ces marques épigénétiques pourrait être prise en compte dans les schémas de sélection par l'utilisation d'un "modèle de transmissibilité" (David & Ricard, 2019) incluant les informations génétiques et épigénétiques de chaque individu.

### 17.1 Pyroséquençage

Le pyroséquençage est une technique développée en 1985 permettant de séquencer de courts fragments d'ADN et d'analyser notamment les polymorphismes de type SNP. Contrairement au séquençage Sanger, le pyroséquençage permet de quantifier les pics lumineux. Son emploi a été détourné pour son utilisation en épigénétique, en analysant des fragments d'ADN préalablement traités au bisulfite de sodium. Dans le cas des CpG, le traitement au bisulfite crée artificiellement un "polymorphisme" C>T. La technique repose sur le "séquençage par synthèse", c'est-à-dire que les nucléotides sont ajoutés un à un et/ou dégradés avant l'ajout du nucléotide suivant. L'ajout d'un nucléotide complémentaire au brin d'ADN par l'ADN polymérase conduit à produire un pyrophosphate (PPi). Ensuite, l'ADN sulfurylase va transformer le PPi en ATP qui va être utilisé par la luciférase pour oxyder la luciférine, réaction produisant de la lumière. Ainsi, au niveau d'un site CpG, la comparaison du pic d'émission lumineuse de l'incorporation de C ou T dans l'amplicon donne une mesure précise de la quantité de méthylation à cette position dans l'échantillon (Delaney et al., 2015).

Au cours de la thèse, le délai pour faire les analyses et validations étant assez restreint, ces dernières n'ont pas pu être réalisées. Lorsque la liste des régions d'intérêt sera affinée, une étape de validation pourra être effectuée par pyroséquençage. Toutefois, l'analyse par pyroséquençage ne peut se faire que sur un nombre restreint d'individus et ne permet de cibler que quelques CpG. D'autres techniques comme le séquençage d'amplicons sur MiSeq (Illumina) après traitement au bisulfite pourraient être envisagées pour valider un grand nombre de positions.

### 17.2 Edition de l'épigénome par transfection

En laboratoire, il est possible d'étudier l'impact fonctionnel d'un site CpG par édition de génome. Le système CRISPR (*Clustered Regulatory Interspaced Short Palindromic Repeat*)-Cas est un système naturel, initialement découvert chez les bactéries pour se protéger des infections virales, ayant mené au prix Nobel de Charpentier et Doudna, et trouve aujourd'hui son application en biologie moléculaire pour activer/inhiber la transcription d'un gène *in vitro*. Ce système agit

comme un "ciseau moléculaire" en coupant de manière précise la séquence cible. Il est composé d'un (i) ARN guide (ARNg) qui est un fragment d'ADN qui reconnaît la séquence spécifique sur l'ADN et (ii) d'une nucléase (la Cas9) qui coupe les deux brins d'ADN à l'endroit précis. En épigénétique, on utilise une Cas9 catalytiquement inactive (dCas9) qui va se lier à l'ADN sans cliver le brin d'ADN. En couplant ce système avec une protéine méthyltransférase *writer* de type DNMT3A, le complexe dCas9-DNMT3A permet de réaliser une méthylation ciblée à l'endroit de fixation de l'ARNg. Au contraire, en couplant le système avec une protéine déméthylase *eraser* de type TET1, le complexe dCas9-TET1 permet une déméthylation. La transfection de ces complexes dans ces cellules eucaryotes et la réalisation de mesure au pyroséquenceur permettent de caractériser l'état méthylé/non-méthylé des constructions. De plus, la réalisation de transcription *in vitro* suivie de RT-qPCR permet de quantifier les transcrits obtenus dans chaque type de construction (projets RUMIGEN et GeneSwitch) (Tadić et al., 2019; X. S. Liu et al., 2016; Stepper et al., 2017).

Pour réaliser ses épi-éditions, il est souhaitable de disposer d'excellentes efficacités de transfection afin qu'une majorité des cellules soit épi-éditée. Cela nous contraint dans le choix de modèles d'études, se résumant souvent à l'utilisation de lignées cellulaires immortalisées se transfectant bien. L'électroporation ou la transfection d'ARNm à la place de plasmide permettent d'obtenir de meilleures efficacités de transfection que la lipofection de plasmides, et permet donc potentiellement d'induire des épi-mutations dans une plus grande gamme de modèle cellulaires : lignées primaires, organoïde (Fève et al., 2022). Une autre possibilité est de trier (FACS) ou sélectionner (traitement antibiotique et transfection d'un gène de résistance) les cellules transfectées.

### 17.3 Lien entre l'épigénétique et les ARN non codants : rôle dans la régulation de l'expression des gènes

La méthylation de l'ADN peut jouer un rôle dans la régulation de l'expression d'ARN non codants, tels que les micro-ARN, notamment dans le cancer chez l'Homme (Beaujean et al., 2020; Choudhry & Catto, 2011). Les ARN non codants comprennent également les ARN circulaires, ces derniers étant des sous-produits de l'épissage de gènes. Au cours de mon stage de M2 de bioinformatique au sein de la même équipe de recherche (GenEpi), sous la direction de Annie Robic et Thomas Faraut, j'ai développé un pipeline pour détecter et annoter des ARN circulaires à partir de données Total RNAseq (**Annexe 20**). Certains de ces ARN circulaires sont également impliqués dans la régulation de l'expression de gènes. Une étude a mis en évidence qu'une seule injection de dexaméthasone peut avoir des conséquences sur le phénotype des descendants et induire des modifications épigénétiques ainsi que des changements dans les circRNAs des spermatozoïdes matures (Gapp et al., 2021).

### 17.4 Projets internationaux contribuant à l'acquisition de données et à l'amélioration des connaissances de l'annotation des génomes

Le consortium FAANG (*Functional Annotation of ANimal Genomes*) est une collaboration internationale entre des communautés de recherche qui ont pour objectif d'accélérer la compréhension de la relation "*genotype-to-phenotype*" (G2P) chez les animaux d'élevage (Andersson et

al., 2015; Clark et al., 2020). Il repose sur une politique de partage de données ouvertes. Une composante européenne s'est organisée en 2020, EuroFAANG, par l'association de trois projets financés par l'UE : AQUA-FAANG pour les poissons, BovReg pour les bovins et GENE-SWitCH pour le porc et la poule afin d'harmoniser la recherche FAANG en Europe et d'en maximiser l'impact. Cette communauté regroupe diverses expertises en biologie animale, génomique, bioinformatique, phénotypage, et modélisation *in silico*. Le projet BovReg s'intéresse notamment à la transmission transgénérationnelle des effets du stress thermique chez les bovins. Deux nouveaux projets européens, GEroNIMO et RUMIGEN, travaillant respectivement sur les monogastriques et les bovins, ont été associés à EuroFAANG en 2021, rejoints récemment par HoloRuminant, plus particulièrement orienté vers les microbiomes des ruminants. Ces projets visent à décrire les éléments codants et régulateurs du génome pour comprendre les mécanismes sous-jacents gouvernant la variabilité des caractères importants pour une production animale durable. Ils se concentrent sur différentes espèces animales (porcs, poules, poissons, ruminants) et contribuent à identifier les régions fonctionnelles du génome, en particulier en utilisant les nouvelles technologies de séquençage. De plus, les projets ont pour but d'explorer la manière dont le génome répond à des signaux environnementaux (qualité de l'aliment, alimentation de la mère, saison, température, stress de transport en début de vie...) par le biais de mécanismes épigénétiques. La question de la transmissibilité transgénérationnelle est également étudiée dans certains de ces projets. A titre d'exemple, des effets transgénérationnels éventuels d'un stress de chaleur sur la vache en gestation seront recherchés sur la carrière de son veau (projet RUMIGEN).

En lien avec ces projets, VizFaDa permet une visualisation interactive des données de FAANG (RNA-Seq, Chip-seq et de méthylation) (Morel et al., 2022). EuroFAANG vise également à standardiser les protocoles d'annotation fonctionnelle et à encourager la recherche sur l'annotation fonctionnelle dans d'autres espèces animales en Europe.

En combinant les contributions des différents projets internationaux, EuroFAANG facilite l'annotation fonctionnelle de haute qualité des génomes animaux, soutenant ainsi une production animale durable et rentable en Europe grâce à l'exploitation future dans les programmes d'élevage et les outils de gestion des systèmes d'élevage.

Les données que nous mettons à disposition de la communauté internationale pourront être utiles dans ce cadre, notamment car les impacts fonctionnels trouvés chez la caille peuvent être retrouvés chez la poule, les espèces étant proches (F. Minvielle, 2004).



## 18. Prise en compte des variations épigénétiques dans les schémas de sélection

La prise en compte des variations épigénétiques dans les schémas de sélection est possible en unités expérimentales (UE) mais reste coûteuse en élevage. En routine, les animaux des schémas de sélection sont déjà génotypés sur puce SNP (technologie peu coûteuse et réalisées en routine dans les schémas de sélection pratiquant la sélection génomique). Les épi-génotypages à haut-débit ne sont en revanche pas encore réalisés chez les animaux d'élevage, alors que ce type d'analyses est bien connu chez l'homme par exemple, avec des études EWAS (*epigenome-wide association studies* (Rakyan et al., 2011)) mettant en avant la part de variance de caractères associée à une marque épigénétique, en l'occurrence la méthylation de l'ADN.

Bien que l'ampleur du rôle de l'épigénétique dans la variabilité des caractères puisse être limitée (Goddard & Whitelaw, 2014), son importance devrait être étudiée de manière plus approfondie, et à grande échelle, afin de décider dans quelle mesure les informations épigénétiques devraient ou non être incluses dans les modèles génétiques. En effet, les performances d'un individu pour un caractère donné se décomposent en une composante génétique et une composante environnementale. Cette composante génétique peut se décomposer en effet additif des gènes sur le caractère cumulé à des effets de dominance et d'épistasie. En sélection génétique, on cherche, pour le moment, à sélectionner les animaux avec les meilleures valeurs génétiques additives. Pour estimer ces valeurs génétiques additives, un modèle de prédiction appelé BLUP (*Best Linear Unbiased Predictor* - modèle animal) est utilisé. Par cette méthode, la part de la variance génétique qui explique le phénotype ne repose que sur la transmission génétique. Or, nous avons vu au cours de cette thèse, que l'épigénétique contribue également à moduler les phénotypes. Des modèles incluant la transmission épigénétique et les interactions environnementales ont déjà été proposés (Tal et al., 2010; Johannes & Colomé-Tatché, 2011; Furrow et al., 2011). Plus récemment, David et Ricard (2019) proposent un nouveau modèle appelé "modèle de transmissibilité" qui permet de prendre en compte à la fois les effets génétiques directs liés aux gènes et les transmissions d'informations non-génétiques liées à l'épigénétique, au microbiote, au comportement, ou encore au système d'élevage. Plus précisément, une matrice de relation épigénétique pourrait être construite à partir des données issues des analyses haut-débit (méthylation de l'ADN principalement), permettant d'identifier les similitudes entre les échantillons et pourrait aider à prédire la valeur épigénétique transmise des individus. L'étude par simulation de ce nouveau modèle de transmissibilité a montré des résultats similaires au modèle animal lorsque la transmission était d'origine génétique uniquement, mais a surpassé le modèle animal pour estimer les covariances entre parents lorsque la transmission n'était pas uniquement d'origine génétique. L'intégration d'informations épigénétiques dans les modèles génétiques devrait donc améliorer à la fois la précision des modèles de prédiction et leur

efficacité dans la sélection des animaux (González-Recio et al., 2015 ; Varona et al., 2015).

Cependant, l'utilisation de ce type de modèles nécessite que l'information épigénétique pour un grand nombre de sites CpG soit disponible. Même si les puces de méthylation humaines, largement utilisées pour l'EWAS dans cette espèce, peuvent être efficaces chez les mammifères tels que les souris (Wong et al., 2013) ou les bovins (Takeda et al., 2017), leur efficacité n'est pas à la hauteur des puces spécifiques. Chez les bovins, la démarche de création d'une telle puce est en cours (RUMIGEN). Dans les espèces avicoles, s'il est possible d'imaginer le développement d'une puce sur l'espèce poule, des solutions innovantes, basées sur du séquençage, constituent probablement l'avenir de l'épigénotypage à haut-débit.





# VIII

## Annexes

<b>19</b>	<b>Epigénétique .....</b>	<b>247</b>
<b>20</b>	<b>Continuité du stage de M2 : analyse des circRNA .....</b>	<b>283</b>



## 19. Epigénétique

**Article Production Animale** (soumis) - Coustham et al., 2023 - *Animal* - **Epigénétique, gènes et environnement : quelle importance pour les pratiques d'élevage et les méthodes de sélection des volailles ?**

**Communication orale** - Cerutti et al., 2021 - *EpiPHASE* - **Effet de l'injection *in-ovo* de perturbateurs endocriniens ou d'un inhibiteur de méthyltransférases sur la croissance et la reproduction des cailles.** hal-03617387.

**Poster** - Cerutti et al., 2022 - *ECCB2022 (Sitges, Espagne)* - **Transgenerational epigenetics in quail : whole genome DNA methylation analysis.** hal-03935584.

**Poster** - Vitorino Carvalho et al., 2022 - *EpiPHASE* - **Effets transgénérationnels d'une exposition embryonnaire à la chaleur sur le transcriptome et l'épigénome de la caille japonaise.** hal-03726346.

**Communication orale** - Cerutti et al., 2022 - *Journée Bioinfo/Biostat GenoToul (Toulouse)* - **Transgenerational epigenetics in quail : whole genome DNA methylation analysis.** hal-03957296.

**Poster** - Cerutti et al., 2023 - *Epigenetic Inheritance (Zurich)* - **Transgenerational epigenetics in quail : whole genome DNA methylation analysis.**

# Epigénétique, gènes et environnement : quelle importance pour les pratiques d'élevage et les méthodes de sélection des volailles ?

---

*Vincent COUSTHAM<sup>1</sup>, Charlotte ANDRIEUX<sup>1</sup>, Chloé CERUTTI<sup>2</sup>, Anne COLLIN<sup>3</sup>, Ingrid DAVID<sup>2</sup>, Julie DEMARS<sup>2</sup>, Guillaume DEVAILLY<sup>2</sup>, Mireille MORISSON<sup>2</sup>, Marianne HOUSSEIER<sup>1</sup>, Sandrine LAGARRIGUE<sup>4</sup>, Sonia METAYER-COUSTARD<sup>3</sup>, Sandrine MIGNON-GRASTEAU<sup>3</sup>, Stéphane PANSEERAT<sup>1</sup>, Angélique PETIT<sup>3</sup>, Anaïs VITORINO CARVALHO<sup>5</sup>, Tatiana ZERJAL<sup>6</sup>, Frédérique PITEL<sup>2</sup>*

<sup>1</sup> INRAE, Université de Pau et des Pays de l'Adour, E2S UPPA, NUMEA, 64310 Saint-Pée-sur-Nivelle, France

<sup>2</sup> GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet-Tolosan, France

<sup>3</sup> INRAE, Université de Tours, BOA, 37380, Nouzilly, France

<sup>4</sup> INRAE, Institut Agro, PEGASE, 35590, Saint-Gilles, France

<sup>5</sup> CNRS, IFCE, INRAE, Université de Tours, PRC, 37380, Nouzilly, France

<sup>6</sup> INRAE, AgroParisTech, Université Paris-Saclay, GABI, Jouy-en-Josas, France

*Courriel : [vincent.cousthamaal@inrae.fr](mailto:vincent.cousthamaal@inrae.fr), [frederique.pitel@inrae.fr](mailto:frederique.pitel@inrae.fr)*

## Chapeau

L'épigénétique, la couche d'information qui se superpose à la génétique pour réguler l'expression du génome, offre des perspectives nouvelles d'amélioration des phénotypes dans le domaine de l'élevage. Dans cette synthèse, nous verrons comment l'épigénétique pourrait contribuer à la sélection de volailles plus performantes et en meilleure santé ainsi qu'au développement de stratégies innovantes de programmation des phénotypes<sup>1</sup>.

---

<sup>1</sup> Cette revue est adaptée de la synthèse « L'épigénétique à l'interface entre l'environnement et l'ADN : quelle importance pour les pratiques d'élevage et les méthodes de sélection ? » présentée aux Journées de la Recherche Avicole (Cousthamaal et al, 2022)

## Résumé

L'épigénétique est communément définie comme l'étude de l'ensemble des mécanismes moléculaires impliqués dans la régulation de l'expression des gènes qui sont réversibles et transmissibles au cours du développement et parfois entre générations, sans altérer la séquence de l'ADN. Plusieurs mécanismes épigénétiques sont maintenant bien connus, comme la méthylation de l'ADN, les variants et modifications post-traductionnelles des histones, ainsi que certains ARN non codants. Grâce au développement technologique du séquençage tout-génome, les modifications épigénétiques peuvent être étudiées à l'échelle du génome entier. Il est aujourd'hui clairement établi que l'épigénome est sensible aux fluctuations de l'environnement, notamment la température ou la nutrition. Des stratégies de programmation précoce des phénotypes reposant sur ces mécanismes épigénétiques sont ainsi envisagées comme levier pour coadapter le phénotype ultérieur des individus à leurs conditions de vie. Par ailleurs, au cours des dernières décennies, la sélection génétique a contribué à l'amélioration considérable des performances des animaux. Bien que la composante génétique puisse être estimée avec précision, une grande partie de la variabilité phénotypique n'est pas directement accessible par les approches actuelles. Dans un contexte de diversification des environnements de production (changement climatique, modes de production plus respectueux du bien-être et de l'environnement), il est nécessaire de comprendre l'impact de l'environnement sur la variabilité phénotypique via les marques épigénétiques, pour optimiser les systèmes d'élevage et mieux prédire le phénotype d'un animal. Comme la sélection génomique il y a quelques années, l'apport de la recherche épigénétique pourrait contribuer à rendre les systèmes de production avicole plus efficaces et plus durables.

## Abstract

### ***Epigenetics, genes and environment: what importance for breeding practices and selection methods in poultry?***

Epigenetics is commonly defined as the set of molecular mechanisms involved in gene expression regulation that are reversible and transmissible during development, and sometimes between generations, without altering the DNA sequence. Several epigenetic marks involved in chromatin regulation are now well known, such as the methylation of DNA cytosines, histone post-translational modifications, and non-coding RNAs. Thanks to the technological development of high-throughput DNA sequencing, epigenetic marks can now be investigated at the whole-genome scale. It is also well known that the epigenome is sensitive to environmental fluctuations, including temperature or nutrition, and can be considered as a mediator between the genes and the environment. Phenotypic programming strategies based on epigenetic mechanisms could thus be used as a lever to adjust the phenotype of individuals with their living conditions. Moreover, over the last decades, genetic selection has contributed to a considerable improvement in animal performances. Although the genetic component can be estimated with some accuracy, a large part of the phenotypic variability, considered to be of environmental origin, is not directly accessible by genetic approaches. In a context of diversification of production environments (climate change, production methods that are more respectful of welfare and the environment...), it is necessary to understand the impact of the environment on phenotypic variation via epigenetic marks, to optimize breeding systems and to predict the phenotype of an animal. Like genomic selection

a few years ago, the contribution of epigenetic research could contribute to more efficient and sustainable production systems in poultry farming.

## Introduction

### 1. L'épigénétique, élément clé de l'expression du génome

#### 1.1. Définition et concepts actuels

Dans les années 1950, l'embryologiste C. H. Waddington a proposé qu'une cellule en développement pouvait avoir différentes fonctions finales (Waddington 1957). Ainsi, alors que l'ADN, support de l'information génétique, est unique et identique dans chacune des cellules (hors événements de recombinaison et de mutations). Les transcrits de gènes, pouvant pour certains être traduits en protéines, vont permettre une grande diversité de types cellulaires en fonction de la combinatoire de gènes qui s'expriment. De ce constat a découlé la définition moderne de l'épigénétique, "l'étude des changements dans l'expression des gènes, hérités par mitose et éventuellement par méiose, qui ne peuvent être expliqués par des changements dans la séquence d'ADN" (Riggs *et al.*, 1996). Ainsi, l'épigénétique est communément définie comme l'ensemble des mécanismes moléculaires impliqués dans la régulation de l'expression des gènes, mécanismes qui sont réversibles et transmissibles au cours du développement, et parfois entre générations, sans altération de la séquence d'ADN.

Le cœur des mécanismes épigénétiques repose sur la compaction de l'ADN au sein de la cellule eucaryote, et en conséquence son accessibilité à la machinerie transcriptionnelle. Plusieurs niveaux de compaction existent, le premier correspondant à l'enroulement de deux tours d'ADN (146 paires de bases) autour d'un octamère de protéines histones. Cette association de l'ADN avec les histones se nomme le nucléosome. Les nucléosomes sont eux-mêmes enroulés en fibres de chromatine qui peuvent elles-mêmes s'enrouler sur plusieurs niveaux jusqu'à aboutir à l'état final de compaction, le chromosome. En conséquence, pour qu'un gène puisse être transcrit, il y a nécessité de mécanismes rendant les gènes accessibles à la machinerie transcriptionnelle. Des marques épigénétiques vont favoriser ou non l'ouverture de la chromatine, mettre en relation des domaines chromatiniens distants, ou encore favoriser le recrutement de facteurs nucléaires environnants.

L'expression d'un gène est la résultante de mécanismes complexes qui repose d'une part sur la machinerie transcriptionnelle et d'autre part sur une combinatoire de marques épigénétiques favorisant ou non l'accessibilité de ce gène à la machinerie transcriptionnelle. Ainsi, la présence d'une marque épigénétique au niveau d'un gène ne conditionne pas forcément son niveau d'expression : des gènes porteurs de marques généralement associées avec l'expression génique peuvent ne pas être exprimés, et vice-versa. Ainsi, la mise en évidence d'une relation de causalité entre la présence d'une marque épigénétique et l'expression du gène associé doit être démontrée par des approches moléculaires directes impliquant par exemple des stratégies d'édition du génomes ou de l'épigénome (type CRISPR/Cas9) permettant d'évincer spécifiquement une marque et d'en démontrer l'effet (Nakamura *et al.*, 2021).

#### 1.2. Principales marques épigénétiques

Plusieurs mécanismes épigénétiques impliqués dans la régulation de la chromatine sont aujourd'hui connus. Ceux-ci sont extrêmement conservés chez les eucaryotes. L'une des modifications chimiques les plus étudiées est la liaison covalente d'un groupement méthyle

sur les cytosines de l'ADN, appelée méthylation des cytosines de l'ADN (5mC). La 5mC est mise en place par des enzymes spécifiques nommées méthyltransférases ADN. Elle peut être retirée de manière passive (par dilution au cours de la réplication de l'ADN) ou par des mécanismes actifs. Lorsqu'elle est enrichie dans les îlots CpG de la région promotrice des gènes, la 5mC est en général liée à la répression de leur expression (Jones 2012). Cependant, de nombreuses autres fonctions ont été identifiées, comme son rôle dans l'inactivation du X (extinction des gènes de l'un des chromosomes X de la paire présente chez les femelles), l'empreinte parentale (faculté qu'ont certains gènes de n'être exprimés qu'à partir d'un seul allèle en fonction de son origine parentale), l'épissage alternatif des transcrits ou dans la répression des événements de transcription indésirables initiés à partir de rétrotransposons ou de séquences dans les corps de gènes (Tirado-Magallanes *et al.*, 2017).

Les modifications post-traductionnelles des histones (MPTH) sont également très étudiées. En effet, les extrémités des histones sont accessibles en dehors de la structure nucléosomique et peuvent être la cible privilégiée de nombreuses modifications telles que la méthylation, la phosphorylation ou encore la sumoylation. Ainsi, certaines MPTH, comme l'acétylation des lysines, peuvent agir sur la charge des nucléosomes (positive, alors que l'ADN est chargé négativement) et les faire coulisser, exposant des séquences d'ADN à la machinerie transcriptionnelle. D'autre part, les MPTH exposées peuvent aussi servir de point d'ancrage pour des facteurs qui régulent directement ou indirectement la structure chromatinienne et l'expression des gènes. Par exemple, la triméthylation de la lysine en position 4 sur l'histone H3 (H3K4me3) peut favoriser l'ouverture de la chromatine et permettre le recrutement de facteurs de transcription au niveau des séquences régulatrices des gènes (Zhang *et al.*, 2015). A l'opposé, d'autres MPTH, comme la triméthylation de la lysine en position 27 de l'histone H3 (H3K27me3), peuvent conduire à un état compacté de la chromatine défavorable à l'expression des gènes.

Un autre type de mécanisme épigénétique communément étudié est la famille des acides ribonucléiques (ARN) non codant, et notamment les petits ARN. Ces ARN de diverses natures sont des transcrits de l'ADN qui ne sont pas traduits en protéines. Ils sont de tailles variées, allant de quelques dizaines de nucléotides (appelés alors "petits ARN") jusqu'à plusieurs centaines à milliers de bases (appelés alors "ARN longs non codants"), ces derniers étant codés par des gènes aussi nombreux que les gènes codant des protéines (Lagarigue *et al.*, 2021). Leurs fonctions sont diverses au sein du noyau, telles que la régulation de la conformation de l'ADN, la dégradation de transcrits ou l'interférence avec la machinerie permettant l'expression des gènes. Ces ARN peuvent agir alors localement sur l'expression des gènes voisins du gène non codant mais également - de par leur nature - à distance (Li *et al.*, 2021). Certains peuvent même circuler dans les tissus pour effectuer des actions à distance voire transmettre de manière intergénérationnelle des informations épigénétiques d'un individu à l'autre, par exemple via le sperme (Grandjean *et al.*, 2015, Chen *et al.*, 2016) ou le lait maternel (Melnik *et al.*, 2021).

### 1.3. Stratégies d'étude des marques épigénétiques : des approches ciblées aux approches tout-génome

Diverses méthodes permettent d'étudier la présence de marques épigénétiques. Par exemple, la conversion chimique par déamination des cytosines non méthylées en présence de bisulfite de sodium, qui n'affecte pas les cytosines méthylées, permet de connaître par déduction l'état de méthylation de l'ADN par séquençage. D'autres méthodes existent, notamment par immuno-précipitation ou digestion par des enzymes sensibles ou non à la méthylation de l'ADN, chacune ayant ses caractéristiques propres en termes de précision du signal, couverture des régions génomiques analysées, coût et durée d'analyse (David *et al.*, 2017,

Mersch *et al.*, 2018). Il est plus complexe d'étudier la localisation d'une MPTH car la modification est portée sur l'histone et non sur l'ADN, et nécessite l'utilisation d'anticorps spécifiquement développé contre la MPTH étudiée. Brièvement, le protocole d'immunoprécipitation de la chromatine ("ChIP") le plus courant consiste à figer les interactions ADN-protéines (dont les histones, porteuses des MPTH) avec un agent chimique (souvent le formaldéhyde) puis de fragmenter la chromatine et de l'incuber avec des anticorps dirigés spécifiquement contre la MPTH d'intérêt. L'ADN associé au complexe ADN-protéine-anticorps est par la suite isolé et amplifié par PCR quantitative ou séquencé pour déterminer sa distribution génomique et son abondance. Plus récemment, une méthode nommée "CUT&RUN" a été développée comme alternative à l'immunoprécipitation directement dans le noyau, permettant de réduire le matériel de départ requis et le bruit de fond expérimental (Skene et Henikoff 2017). Nous n'aborderons pas dans cette revue les approches pour étudier des ARN non codants, mais celles-ci sont également en constante évolution (voir par exemple (Sun et Chen 2020)).

Depuis une dizaine d'années, grâce au développement technologique des séquenceurs d'ADN dits "à haut débit", les marques épigénétiques peuvent être étudiées sur l'ensemble du génome. Il est notamment possible d'étudier l'expression de tous les gènes (transcriptome) et l'état de méthylation de tous les dinucléotides CpG (méthylome) dans une seule et même cellule d'un organisme (Lee *et al.*, 2020). Un séquençage de génome aviaire de bonne qualité coûte aujourd'hui entre 500 et 1000 €, et un transcriptome environ 200 € par échantillon, ce qui rend accessible ce type d'approche tout-génome. Cependant, la préparation des bibliothèques de séquençage (ensemble des fragments d'ADN que l'on veut séquencer) et les analyses bioinformatiques sont des étapes chronophages, ces dernières requérant des compétences spécifiques en bioinformatique et en statistiques ainsi que des infrastructures adaptées pour analyser et stocker les importantes quantités de données. Par ailleurs, le succès des analyses basées sur le séquençage tout-génome dépend également de la qualité de l'assemblage du génome de référence et de son annotation, qui varie principalement selon la popularité du modèle étudié. A ce jour, le génome de référence du poulet (bGalGal1.mat.broiler.GRCg7b) est de très bonne qualité, comme le suggèrent plusieurs indicateurs tels qu'une taille de contig N50 de 19 Mb et peu de séquences manquantes (<https://www.ncbi.nlm.nih.gov/grc/chicken>). D'une manière générale, les génomes aviaires récents tels que celui de la caille japonaise (Cotja 2.0, 2016), à base de séquences longues (type PacBio) et courtes (type Illumina), ont des assemblages de qualité suffisante pour des études épigénétiques tout-génome.

## 2. Épigénétique et programmation environnementale des phénotypes

Il est aujourd'hui clairement établi que l'épigénome est sensible aux variations de l'environnement précoce, subies lors du développement embryonnaire ou dans la période périnatale, et qu'il peut être impacté par différents facteurs tels que la température, la nutrition, les molécules chimiques, ou encore les interactions sociales (Feil et Fraga 2012, Alabduljabbar *et al.*, 2021, Di Criscio *et al.*, 2023). La phase de développement précoce (embryonnaire et post-natal) est une période de forte plasticité au cours de laquelle l'environnement peut influencer de manière durable et significative la santé, le bien-être, la robustesse et les performances ultérieures de l'animal en développement (Ho *et al.*, 2011, Reed et Clark 2011). La compréhension des mécanismes épigénétiques résultants de ces effets

précoces de l'environnement pourrait donc être utilisée pour programmer<sup>2</sup> le phénotype des individus et leurs performances zootechniques.

De nombreux effets à long terme de l'environnement précoce ont en effet été démontrés chez les oiseaux (Feeney *et al.*, 2014, Loyau *et al.*, 2015). Ces effets peuvent être mis directement en évidence chez les espèces ovipares car l'environnement de l'embryon (dans l'œuf) peut être directement manipulé. Cette particularité a ouvert de nouvelles possibilités de programmation phénotypique en productions avicoles. Nous recensons dans la littérature la démonstration d'effets bénéfiques sur les animaux par différents stimuli pendant l'embryogénèse tels que la lumière verte (Rozenboim *et al.*, 2004), la température (Loyau *et al.*, 2015), la nutrition (Uni *et al.*, 2005, Cherian 2011) ou les changements de régime alimentaire maternel (Hynd *et al.*, 2016, Baeza *et al.*, 2017, Thanabalan et Kiarie 2021). Dans cette revue seront abordés plus spécifiquement les effets précoces en lien avec la nutrition et la température.

## 2.1. Programmation nutritionnelle

Deux leviers nutritionnels principaux peuvent être envisagés pour modifier la teneur en nutriments des œufs afin d'améliorer la robustesse, la croissance et la composition corporelle des poussins : la nutrition de la mère et la supplémentation directe *in ovo*.

Les liens entre la nutrition de la poule, la composition de l'œuf et le comportement ultérieur de l'animal, ses performances et sa sensibilité aux maladies sont aujourd'hui bien établis (Aigueperse *et al.*, 2013, Bergoug *et al.*, 2013). La nutrition de la poule peut en effet moduler les niveaux de nutriments essentiels dans l'œuf qui, à leur tour, peuvent avoir un impact sur la santé des poussins éclos et leurs phénotypes ultérieurs (voir (Ruhl 2007, Morisson *et al.*, 2017). Par exemple, l'administration de bêtaïne aux poules (un composant du cycle de la méthionine), également considérée comme un agent antioxydant et un donneur de méthyle efficace, entraîne des modulations épigénétiques modifiant l'expression hépatique de plusieurs gènes chez les poussins (Hou *et al.*, 2018, Hu *et al.*, 2020). Des études menées chez le canard mulard ont montré qu'un régime alimentaire maternel restreint en méthionine impacte le métabolisme énergétique (Bodin *et al.*, 2019). Il modifie l'expression hépatique de gènes du métabolisme énergétique et du métabolisme du monocarbonate ainsi que celle de quelques gènes impliqués dans des mécanismes épigénétiques tels que la méthylation de l'ADN ou la désacétylation des histones (Sécula *et al.*, 2022a, Sécula *et al.*, 2022b) chez les canetons à l'éclosion. Les effets sur le métabolisme hépatique perdurent chez les adultes et altèrent la production de foie gras (Bodin *et al.*, 2022).

L'alimentation *in ovo* est un moyen plus direct d'influencer le phénotype des poussins. Plusieurs études ont décrit l'utilisation de la supplémentation en nutriments *in ovo* pour réduire la fenêtre d'éclosion et améliorer la santé, le statut immunitaire après l'éclosion, l'éclosabilité, le poids du poussin éclos, les performances de croissance et la qualité de la viande (Uni et Ferket 2004, Wei *et al.*, 2011, Kadam *et al.*, 2013, Roto *et al.*, 2016, Gao *et al.*, 2017, Peebles 2018, Taha-Abdelaziz *et al.*, 2018, Jha *et al.*, 2019, Kalantar *et al.*, 2019, Ayansola *et al.*, 2021). Les nouvelles stratégies *in ovo* visent également à relever de nouveaux défis tels que la recherche d'alternatives à l'utilisation des antibiotiques par l'injection de probiotiques (Oladokun et Adewole 2020).

---

<sup>2</sup> Nous utilisons dans cette synthèse le terme de « programmation » aussi bien pour parler des effets nécessitant un stimulus ultérieur pour être révélés que des effets survenant dans un environnement stable. Lorsqu'un stimulus est nécessaire, certains auteurs préfèrent utiliser le terme de « conditionnement » (voir Hanson et Gluckman 2014) sans qu'un consensus n'existe dans la littérature.

Chez les oiseaux, la plupart des études ne rapportent que les résultats phénotypiques des injections *in ovo*. C'est le cas d'une étude récente portant sur l'effet de l'injection *in ovo* de perturbateurs endocriniens (génistéine et bisphénol A) et d'un inhibiteur de l'ADN méthyltransférase (5-Aza-2'-deoxycytidine) sur des caractères de production de cailles japonaises (croissance et performances de ponte) qui a pu mettre en évidence un effet significatif de ces molécules sur la croissance des cailles (Cerutti *et al.*, 2022a). Seules quelques études visent à déchiffrer les mécanismes impliqués dans ces changements phénotypiques. Des changements épigénétiques pourraient être impliqués, notamment lorsque des donneurs de groupes méthyles, comme la méthionine, sont injectés (Anderson *et al.*, 2012, Donohoe et Bultman 2012, Veron *et al.*, 2018). Ainsi, la manipulation de la teneur en acides aminés soufrés pourrait induire des changements dans la fonction cellulaire avec des implications sur le développement, la croissance à long terme et la santé de l'animal. Il a par ailleurs été montré que la supplémentation en folates améliore les performances de croissance, la fonction immunitaire et le métabolisme du folate des poulets de chair par la régulation épigénétique des gènes immunitaires en modifiant l'accessibilité de la chromatine et les marques épigénétiques telles que la méthylation des histones (Li *et al.*, 2016). Enfin, des travaux montrent que l'injection de bétaïne affecte le métabolisme hépatique du cholestérol par la régulation épigénétique des gènes chez les poussins nouvellement éclos (Hu *et al.*, 2015). Au bilan, la détermination des mécanismes par lesquels les nutriments de l'œuf régulent le métabolisme cellulaire, la signalisation, l'expression et la fonction des gènes est essentielle pour améliorer l'utilisation des nutriments, l'efficacité de la production avicole et la robustesse des animaux.

## 2.2. Programmation thermique

Les poulets de chair ont longtemps été génétiquement sélectionnés sur des caractères de performance de croissance, avec des progrès significatifs au cours des dernières décennies. L'augmentation de la masse corporelle du poulet et de la croissance du muscle n'ont pas été associées à un développement proportionnel de leur appareil cardio-respiratoire (Havenstein *et al.*, 2003). Ceci nuit à leurs capacités de thermorégulation, les rendant très sensibles à une augmentation de la température ambiante avec un impact direct sur le bien-être, les performances et la mortalité (Yahav et McMurtry 2001). Le réchauffement climatique nécessite d'ajuster nos pratiques d'élevage pour que les animaux soient plus robustes face à des vagues de chaleur de plus en plus fréquentes et intenses. Dans ce contexte, une stratégie de programmation thermique précoce a été mise au point pour améliorer la résistance à la chaleur, initialement chez le poulet de chair. Le traitement de manipulation thermique embryonnaire à la chaleur (MT) chez le poulet consiste à augmenter la température d'incubation des œufs, par exemple de + 1,7 °C, pendant 12 heures par jour, des jours 7 à 16 de l'incubation. Il a été montré qu'un tel traitement cyclique permet une diminution de 50% de la mortalité des poulets mâles soumis à un stress thermique à l'âge d'abattage (Piestun *et al.*, 2008). Bien que le poulet soit l'espèce la plus étudiée, des protocoles de MT ont été appliqués sur différentes espèces aviaires (voir Loyau *et al.*, 2015). Notamment, une étude récente chez la caille japonaise montre qu'un protocole de MT réalisé du premier jour de l'embryogenèse jusqu'au 13<sup>ème</sup> jour (+ 1,7 °C, 12h par jour) permet de réduire les mortalités post-éclosion (jours 3-6) en cas de coup de chaleur (Vitorino Carvalho *et al.*, 2023).

Une étude de l'expression des gènes du muscle pectoral, tissu responsable d'une partie significative de la production de chaleur, a été réalisée chez des poulets de chair mâles témoins et MT par des approches ciblées et d'analyse à haut-débit (transcriptomique). Ces

travaux montrent que la MT affecte l'expression de gènes dans le muscle à 35 jours d'âge (J35), pour certains impliqués dans le métabolisme énergétique et la fonction mitochondriale en relation avec la production de chaleur (Loyau *et al.*, 2014, Loyau *et al.*, 2016). Par ailleurs, l'analyse différentielle du transcriptome des animaux incubés en condition standard (C) et MT ayant subi (CC) ou non (RT) un coup de chaleur (32°C, 5 h à J34) a montré 6 fois plus de gènes différentiels en réponse à la chaleur chez les animaux MT (MT-RT vs MT-CC, caractérisés par des températures internes moins élevées que les poulets témoins) que la même comparaison chez les témoins (C-RT vs C-CC ; Loyau *et al.*, 2016). Ainsi, les effets de la MT sont révélés par une exposition à la chaleur et suggèrent que la MT pourrait permettre la mobilisation d'un plus grand nombre de gènes en cas de réponse à une exposition thermique ultérieure, ou changerait l'intensité de la réponse génique ou encore le timing de cette réponse génique. Des observations similaires ont été rapportées dans l'hypothalamus des cailles, tissu impliqué dans la régulation de la température interne (Vitorino Carvalho *et al.*, 2021). De manière intéressante, la MT a affecté principalement la réponse transcriptomique des cailles femelles, en accord avec l'observation d'un effet phénotypique prépondérant de la MT des femelles chez cette espèce (Vitorino Carvalho *et al.*, 2020). Ainsi, des différences importantes de réponse au traitement existent en fonction des espèces et du sexe de l'animal.

Une étude a exploré les mécanismes épigénétiques de la MT chez le poulet de chair. Cette étude s'est focalisée sur deux modifications post-traductionnelles des histones dans le muscle et l'hypothalamus : les marques H3K4me3 et H3K27me3 (associées respectivement avec l'activation et la répression de l'expression des gènes). Les résultats de cette étude montrent que les deux marques sont impactées par la MT dans l'hypothalamus, avec un effet prépondérant sur la MPTH H3K4me3, pour laquelle près de 800 régions différentiellement enrichies ont pu être identifiées à 35 jours d'âge entre poulets MT et contrôles (David *et al.*, 2019b). La fonction des gènes à proximité de ces régions a été analysée et suggère que ces marques pourraient contribuer à programmer la réponse d'un certain nombre de gènes en lien avec le métabolisme des poulets. Ainsi, la MT pourrait induire une reprogrammation épigénétique de l'expression de certains gènes contrôlant des processus métaboliques liés à la thermorégulation. Ces changements épigénétiques seraient maintenus au cours de la croissance du poulet permettant une réponse métabolique adaptée en cas d'exposition ultérieure à la chaleur (Figure 1). Cette hypothèse est corroborée par des études montrant que plusieurs marques épigénétiques sont impliquées dans la réponse à un traitement thermique post-natal (3 jours après éclosion) lors de l'établissement du contrôle thermique dans le cerveau (Yossifoff *et al.*, 2008, Kisliouk *et al.*, 2010).

Chez les autres espèces aviaires pour lesquelles des protocoles de MT ont été étudiés, les mécanismes sont pour l'heure encore inexplorés. Cependant, plusieurs indices suggèrent que les mécanismes épigénétiques pourraient être également impliqués. Chez le canard, il a été montré que la MT affecte le niveau d'expression génique des enzymes de méthylation, ce qui suggère que la température d'incubation peut influencer la méthylation de l'ADN chez cette espèce au cours du développement précoce (Yan *et al.*, 2015).

Il est à noter que l'exposition des œufs à de basses températures d'incubation affecte également la physiologie, le comportement, le métabolisme, la santé et le bien-être des poussins et des poulets à plus long terme (Shinder *et al.*, 2011, Nyuiadzi *et al.*, 2020). Cependant, les mécanismes épigénétiques associés aux expositions à plus basse température n'ont pas encore été explorés.

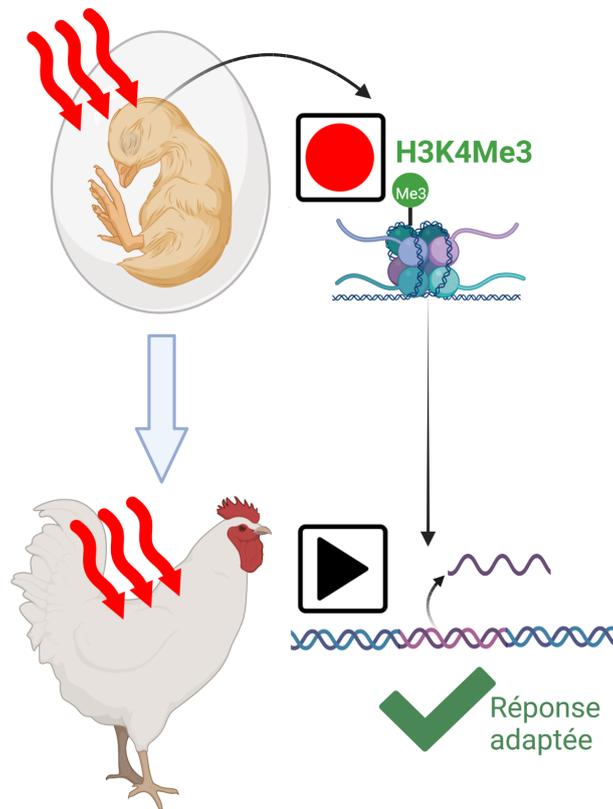


Figure 1. Mécanismes épigénétiques de la manipulation thermique à la chaleur chez l'oiseau.

*Le traitement de manipulation thermique (MT) du poulet de chair consiste à augmenter la température d'incubation des œufs de 1,7°C de manière cyclique (12 h / j) des jours 7 à 16 d'incubation. L'hypothèse épigénétique de la MT suggère que cette exposition précoce pourrait être « enregistrée » au niveau de la lysine 4 triméthylée de l'histone H3 (H3K4me3) sur plusieurs régions du génome dont certaines à proximité de gènes d'intérêt pour la thermorégulation. Cette reprogrammation épigénétique localisée persisterait au cours de la vie et permettrait la « lecture » d'une réponse génique adaptée des poulets en cas d'exposition ultérieure à un coup de chaleur. Illustration créée avec BioRender.com*

### 2.3. Phénomènes transgénérationnels

L'hypothèse d'un mécanisme de transmission non-génétique d'un effet de l'environnement fait l'objet d'une grande attention grâce à l'avènement de l'épigénétique, tout en restant controversée car il n'existe actuellement que peu d'études qui fournissent des indications claires sur les mécanismes en jeu chez les vertébrés. C'est une question sociétale importante, car il semble que les animaux ne transmettent pas que leur capital génétique à leur descendance (Miska et Ferguson-Smith 2016, David *et al.*, 2019a). En effet, ils pourraient aussi transmettre des facteurs non-génétiques, caractéristiques acquises par ce qu'ils ingèrent (alimentation, médicaments, etc.), ce qu'ils respirent, ce qu'ils font (activité physique, etc.) ou encore ce qu'ils vivent (stresseurs, rayonnements, etc.). On utilise en général le terme "intergénérationnel" pour les transmissions entre parents et descendants, et

"transgénérationnel" pour les phénomènes observés uniquement à partir de la première génération non directement exposée au stimulus, même à l'état de gamète, soit la troisième génération pour une transmission par voie maternelle. Les mécanismes épigénétiques pourraient constituer l'un des vecteurs de cette transmission.

Chez l'animal de rente, la compréhension des mécanismes par lesquels les modifications épigénétiques transmettent cette information est essentielle pour anticiper l'impact de nos pratiques d'élevage sur les descendants, et aiderait au développement de stratégies innovantes permettant d'améliorer la santé, le bien-être et les performances sur le long terme. En outre, identifier les marques épigénétiques transmissibles d'origine environnementale permettrait d'ouvrir de nouvelles perspectives en sélection : tenir compte des facteurs génétiques et non génétiques peut améliorer l'estimation du potentiel transmissible d'un futur reproducteur.

Cependant, de par la diversité des modes d'élevage, des conditions environnementales et, surtout, à cause de la difficulté à discerner les parts génétique et non génétique dans la construction des phénotypes des animaux de rente, cette question est très difficile à appréhender. Il existe donc peu d'études réellement transgénérationnelles chez les oiseaux, la première à avoir démontré un effet transgénérationnel concerne la caille japonaise (Leroux *et al.*, 2017). Deux "épilignées" de cailles ont été produites (populations issues des mêmes fondateurs, de même fond génétique, mais qui diffèrent sur le plan épigénétique) : avant l'incubation, de la génistéine a été injectée (épilignée traitée) ou non (épilignée contrôle) dans les œufs. La génistéine est un phyto-œstrogène naturel présent dans le soja, notamment modificateur de la méthylation de l'ADN (Silva *et al.*, 2019). Après 3 générations parallèles de reproduction, sans autre injection, des mesures phénotypiques ont révélé plusieurs résultats intéressants. En effet, la comparaison des performances entre les descendants issus des épilignées traitée et non traitée montre, entre autres, une diminution du poids corporel, une augmentation du poids de gras abdominal et un âge au premier œuf retardé chez les descendants de la lignée traitée. Des caractères de comportement ont également été impactés. Les différences génétiques entre épilignées ayant été minimisées grâce à un dispositif de croisements permettant d'équilibrer les origines génétiques des fondateurs, les différences observées sont probablement dues, au moins en partie, à une transmission épigénétique transgénérationnelle. Pour mieux caractériser la transmission transgénérationnelle des effets de l'environnement, des données de méthylation de l'ADN issues d'échantillons de sang de cailles japonaises de la 3<sup>ème</sup> génération sont en cours d'analyse. Ces données ont été générées par deux méthodes de séquençage tout génome sur les mêmes individus : WGBS (Whole Genome Bisulfite Sequencing, lectures courtes) et ONT (Oxford Nanopore Technologies, lectures longues). Grâce à l'ensemble de ces données, une analyse préliminaire de méthylation différentielle réalisée par comparaison des deux épilignées (epi-/epi+) semble confirmer l'existence d'une transmission transgénérationnelle des effets de la génistéine sur le méthylome des cailles (Cerutti *et al.*, 2022b).

Plus récemment, une étude a été menée pour étudier les effets de la MT sur 7 générations de cailles japonaises (Vitorino Carvalho *et al.*, 2023). Dans cette étude, les effets multigénérationnels (répétition de la MT) et transgénérationnels (persistance des effets chez les grand-descendants suite à l'arrêt du traitement) ont été étudiés. Ces travaux montrent que la MT a un effet négatif sur le poids des cailles et de leurs œufs pondus qui semble apparaître ou croître en fonction de l'âge (4 ou 5 semaines post-éclosion) au fur et à mesure que le traitement est répété au fil des générations. En outre, l'étude a montré que ces effets sont transmis de manière transgénérationnelle, et que les effets sont partiellement réversibles en fonction de l'âge et du sexe observé, à partir de 3 générations consécutives incubées en condition standard après la première observation de l'effet transgénérationnel (Vitorino Carvalho *et al.*, 2023). Ces observations renforcent l'hypothèse de mécanismes épigénétiques de la MT chez la caille japonaise (les effets épigénétiques étant par nature potentiellement

transmissibles et réversibles), bien que les mécanismes précis soient pour l'heure encore inconnus. Des effets positifs de la MT ont par ailleurs pu être observés sur la survie des cailleaux soumis à un coup de chaleur quelques jours après l'éclosion, mais ceux-ci ne semblent pas être transgénérationnels (Vitorino Carvalho *et al.*, 2023). Cette observation met en évidence les limites de l'utilisation systématique de protocoles de programmation des phénotypes, dont les effets bénéfiques au sein d'une génération pourraient être contrebalancés par des effets délétères à plus long terme si les protocoles sont appliqués à chaque génération.

### 3. Épigenétique, génétique et interactions génotype x environnement

Les animaux d'élevage sont confrontés à de multiples contraintes au cours de leur vie et doivent s'adapter à différentes perturbations telles que des changements dans l'environnement d'élevage, des transitions alimentaires, des variations climatiques ou des maladies. Ces événements potentiellement stressants peuvent avoir un effet important sur la production, la santé et le bien-être, dont l'ampleur varie selon les individus, notamment en fonction de leur génotype (Nauta *et al.*, 2006, Cardoso et Tempelman 2012). Au cours des dernières décennies, la sélection génétique a contribué à améliorer considérablement les performances des animaux : en volaille, l'augmentation des performances a été remarquable tant chez les poulets de chair que chez les poules pondeuses (voir (Aggrey *et al.*, 2020)). Le modèle d'évaluation génétique décompose le phénotype en effets génétiques (G) et environnementaux (E). Du fait d'une certaine standardisation des conditions environnementales intra-race, le modèle suppose le plus souvent l'absence d'interaction génotype-environnement. Cependant, du fait du changement climatique, de la transition agroécologique et des nouvelles demandes sociétales, les animaux vont être élevés dans des environnements plus fluctuants (par exemple l'élevage de plein air) et faire face à une plus grande variabilité de leur alimentation, fluctuations auxquelles ils devront s'adapter. Il est donc souhaitable de mieux comprendre la composante GxE pour être en mesure de sélectionner des génotypes mieux adaptés (Figure 2).

#### 3.1. Épigenétique et génétique

##### a. La génétique influence l'épigenétique

Il est désormais bien établi que les polymorphismes de l'ADN peuvent affecter les marques épigénétiques (Bell *et al.*, 2011, Do *et al.*, 2017). Par exemple, chez la poule, il a été montré récemment dans un tissu, le foie, qu'environ 30% des gènes longs non-codants (lncRNA) annotés chez la poule avaient leur expression significativement associée à au moins un variant génétique (Jehl *et al.*, 2021). Sur une étude portant sur plus de 50 tissus humains et plus d'individus, cette proportion de gènes lncRNA régulés génétiquement dans au moins un tissu dépassait les 65% (de Goede *et al.*, 2021). De même, jusqu'à 80 % de la variation de la méthylation de l'ADN peut être expliquée par le génotype (Gertz *et al.*, 2011). Une étude chez l'homme a démontré que la méthylation CpG était parfois significativement associée avec des polymorphismes génétiques distants des sites régulés (Lemire *et al.*, 2015). Do *et al.* (Do *et al.*, 2017) ont passé en revue de nombreuses études révélant l'existence d'une grande variabilité épigénétique d'origine génétique. Un mécanisme expliquant l'influence des génotypes sur l'état épigénétique est le fait que la séquence d'ADN à des sites de liaison spécifiques peut affecter la liaison des facteurs de transcription, certains de ces facteurs de transcription étant capables de modifier le niveau de méthylation de l'ADN à proximité (Feldmann *et al.*, 2013). Enfin, la méthylation de l'ADN est absente dès qu'un SNP détruit un site CpG, ce qui peut avoir un effet similaire à une déméthylation de ce site.

### b. L'épigénétique influence la génétique

D'un autre côté, les taux de mutation peuvent également être affectés par les états épigénétiques : les dinucléotides CpG ont un taux de mutation environ 12 fois plus élevé que les autres types de transition, car les méthylcytosines sont plus susceptibles de muter que les bases non méthylées (Sved et Bird 1990). Le contenu en CpG peut même affecter le taux de mutation des non-CpG (Walser et Furano 2010). Plus généralement, des associations entre la structure de la chromatine et les taux de mutation ont été rapportées (Makova et Hardison 2015). Ainsi, les expositions environnementales pourraient non seulement exercer une pression de sélection sur des allèles préexistants, mais aussi jouer un rôle "direct" sur les polymorphismes de l'ADN, même si les études actuellement publiées reposent sur des corrélations, non démonstratives d'une action causale. Ces phénomènes, lorsqu'ils sont déclenchés par l'environnement depuis des générations via des marques épigénétiques conduisant à des modifications génétiques, ont été appelés "assimilation mutationnelle facilitée par l'épigénétique" (Danchin *et al.*, 2019).

Les mécanismes épigénétiques sont donc directement liés à la fois à l'environnement et à la génétique d'un organisme, acteurs de l'interaction GxE.

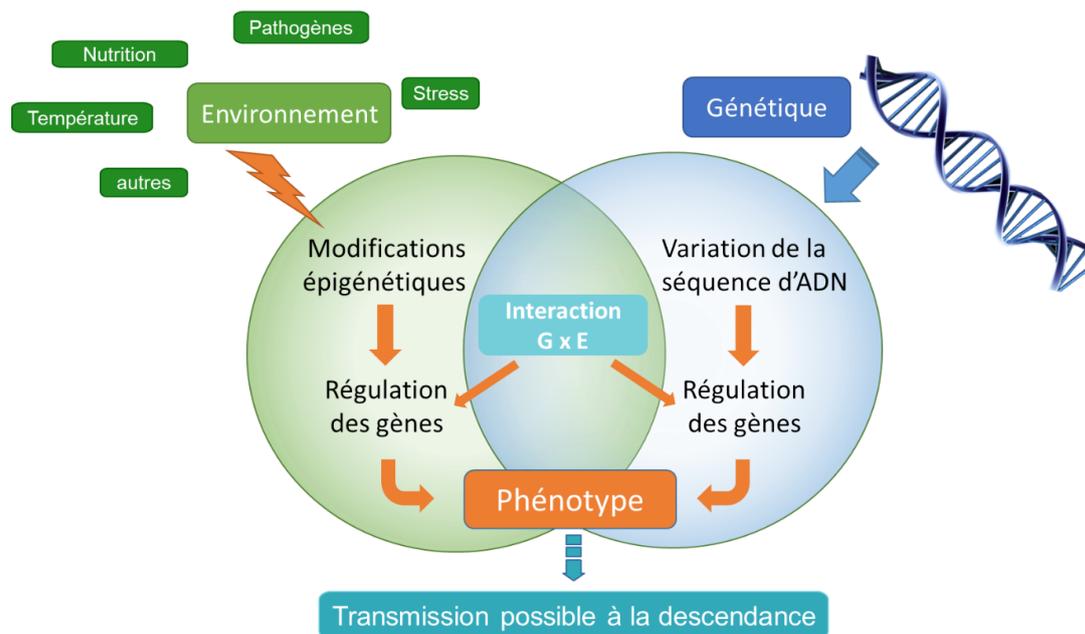


Figure 2. Interaction entre génétique et environnement dans la construction des phénotypes.

*G* : génétique ; *E* : environnement.

### 3.2. Interactions GxE

Lorsque l'effet d'une variation d'environnement diffère en fonction du génotype, il existe une interaction génotype-environnement. Les environnements peuvent être différents en termes de

facteurs biotiques (concentration d'hormones, stades de développement ou physiologiques, maladies, etc.) ou abiotiques (température, exposition aux contaminants chimiques, etc.).

Ces interactions peuvent modifier le classement, en fonction des performances, de différents individus lorsqu'ils sont exposés à différents environnements (Haldane 1946).

Les interactions GxE ont été étudiées de manière approfondie chez les bovins (par exemple, (Hayes *et al.*, 2009, Bernabucci *et al.*, 2014). Dans les programmes de sélection tels que ceux rencontrés chez le porc et la volaille, une perte de gain génétique due à l'interaction GxE peut également être observée (Mulder et Bijma 2005). L'interaction GxE a été observée pour plusieurs caractères tant dans les productions de poulets de chair (croissance lente, (N'Dri A *et al.*, 2007) que de poules pondeuses (Mathur et Horst 1994, Romé *et al.*, 2015). Par exemple, il a été démontré des interactions génétique x régime et génétique x âge pour différents caractères de production et de qualité des œufs (Romé *et al.*, 2015).

Comme décrit par Lillehammer *et al.* (Lillehammer *et al.*, 2009) chez les bovins laitiers, les polymorphismes génétiques peuvent provoquer ces interactions GxE. Certains allèles peuvent affecter à la fois la production et la sensibilité à l'environnement. Les allèles qui maintiennent ou améliorent la production tout en réduisant la sensibilité à l'environnement pourraient être de bons candidats pour la sélection assistée par marqueurs pour la robustesse (capacité d'un animal à bien s'adapter à différents environnements). D'autres gènes peuvent contribuer à la modification du classement des animaux dans divers environnements. Dans ce cas, la sélection de l'allèle améliorateur dans un environnement peut entraîner une perte de performance dans un autre environnement. Un exemple de ce mécanisme impliquant différents gènes dans plusieurs environnements a été démontré pour l'efficacité alimentaire chez la volaille (Mignon-Grasteau *et al.*, 2010) : avec un régime maïs-soja facilement digéré, la composante génétique de l'efficacité digestive sera peu impliquée dans la variabilité de l'efficacité alimentaire, alors qu'avec un régime difficile à digérer, cette composante digestive sera plus importante, et impliquera des gènes différents, régulant plusieurs voies métaboliques.

On trouve fréquemment une corrélation génétique positive entre la production générale et la sensibilité à l'environnement ; en conséquence, la sensibilité à l'environnement augmente avec la sélection pour une haute performance (Rauw et Gomez-Raya 2015). Ce lien néfaste peut être dû aux compromis entre les fonctions, ou "trade-offs" (Friggens *et al.*, 2017).

### 3.3. Importance en sélection

Les sélectionneurs sont donc confrontés à la question de savoir comment gérer ces possibles changements de classement de leurs lignées en fonction de l'environnement : est-il préférable de sélectionner les meilleurs animaux dans un environnement donné en prenant le risque que leurs performances soient fortement modifiées si les conditions environnementales changent, ou est-il préférable de sélectionner des animaux aux performances plus modestes mais plus plastiques et plus robustes, c'est-à-dire plus stables sur une grande variété d'environnements ?

Dans le cas des volailles dont les élevages de sélection et de production sont structurés en schéma pyramidal, si l'importance de l'interaction GxE est assez limitée pour les lignées en sélection qui sont élevées dans des conditions standardisées, elle devient pertinente pour les animaux en production, qui sont issus des croisements des lignées en sélection et qui sont élevés dans différentes parties du monde aux climats et systèmes d'élevage variés. C'est également le cas pour les productions alternatives (bio, label), qui sont élevées avec un accès extérieur, alors que les reproducteurs sont souvent élevés dans des bâtiments fermés. La quantification de l'interaction entre génétique et environnement devient donc nécessaire, notamment pour fournir des reproducteurs pouvant produire des individus capables

d'exprimer des performances optimales dans une large gamme d'environnements de production. Pour y parvenir, une stratégie consisterait à prendre en compte les performances sur le terrain des individus croisés dans l'amélioration génétique des lignées pures (Esfandyari *et al.*, 2016, Duenk *et al.*, 2019). Par ailleurs, l'adaptation des animaux pendant une génération devrait conduire à une meilleure performance de la descendance finale grâce à la transmissibilité des effets de l'environnement par des voies épigénétiques. L'étude de la variabilité épigénétique à grande échelle - l'un des objectifs du projet européen GeroNIMO (<https://www.geronimo-h2020.eu/>) -, notamment par l'analyse d'un même fond génétique dans différents environnements, devrait nous amener à mieux prédire les effets de l'environnement. Ces avancées devraient donc, à terme, permettre d'optimiser les conditions de production, voire d'améliorer la sélection pour produire des animaux commerciaux mieux adaptés à différents environnements de production.

Il est donc nécessaire de démêler l'impact de l'environnement sur la variation phénotypique, largement médiée par les interactions entre les composants génétiques et non génétiques, pour atteindre l'objectif ultime d'une meilleure prédiction du phénotype d'un animal. Estimer et incorporer les effets épigénétiques dans les modèles de prédiction phénotypique en améliorera la précision. Utiliser de nouvelles connaissances sur la transmission des marques épigénétiques apposées en réponse à l'environnement permettra d'améliorer l'estimation de la valeur transmissible d'un reproducteur. Le progrès "génétique" pourrait alors être accéléré, en particulier pour les caractères peu héréditaires pour lesquels les modèles actuels ne considèrent qu'une partie de la variabilité phénotypique.

## Conclusion

L'épigénétique est un domaine de recherche émergent dans les productions animales. Il est désormais reconnu que les changements environnementaux auxquels l'animal est confronté induisent des modifications épigénétiques qui ont un impact sur les phénotypes à long terme. Dans un contexte de diversification des environnements de production (changement climatique, méthodes de production plus respectueuses du bien-être et de l'environnement...), il est nécessaire de comprendre l'impact de l'environnement sur la variabilité phénotypique, pour optimiser les systèmes d'élevage et mieux prédire le phénotype d'un animal, voire de sa descendance, à partir des informations génétiques et environnementales. Une meilleure prédiction des phénotypes permettrait également de mieux anticiper l'adaptation d'un individu à un environnement particulier, notamment en développant des stratégies innovantes de programmation des phénotypes reposant sur ces empreintes épigénétiques précoces.

Le transfert de ces connaissances vers les industries de sélection permettrait une évaluation précise des effets épigénétiques grâce à des analyses sur de grandes populations. Ceci nécessite le développement de nouvelles méthodes de caractérisation de l'épigénome, compatibles avec une utilisation à grande échelle. Comme la sélection génomique il y a quelques années, l'apport de la recherche épigénétique pourrait contribuer à des systèmes de production plus efficaces et plus durables en aviculture.

## Remerciements

La thèse de Chloé Cerutti est financée par la Région Occitanie et le Département de Génétique Animale (INRAE). La thèse de Charlotte Andrieux a été financée par le Comité

Département des Landes (CD40). La thèse d'Angélique Petit a été financée par l'Université de Tours.

## Références

- Aggrey S.E., Zhou H., Tixier-Boichard M., Weigend S., Rhoads D.D., Andersson L., Bed'hom B., Chuong C.M., Inaba M., Okimoto R., 2020. In: *Advances in Poultry Genetics and Genomics*. Burleigh Dodds Science Publishing Limited.
- Aigueperse N., Calandreau L., Bertin A., 2013. Maternal diet influences offspring feeding behavior and fearfulness in the precocial chicken. *PLoS One*, 8, e77583. [10.1371/journal.pone.0077583](https://doi.org/10.1371/journal.pone.0077583)
- Alabduljabbar S., Zaidan S.A., Lakshmanan A.P., Terranegra A., 2021. Personalized Nutrition Approach in Pregnancy and Early Life to Tackle Childhood and Adult Non-Communicable Diseases. *Life (Basel)*, 11. [10.3390/life11060467](https://doi.org/10.3390/life11060467)
- Anderson O.S., Sant K.E., Dolinoy D.C., 2012. Nutrition and epigenetics: an interplay of dietary methyl donors, one-carbon metabolism and DNA methylation. *J Nutr Biochem*, 23, 853-859. [10.1016/j.jnutbio.2012.03.003](https://doi.org/10.1016/j.jnutbio.2012.03.003)
- Ayansola H., Liao C., Dong Y., Yu X., Zhang B., Wang B., 2021. Prospect of early vascular tone and satellite cell modulations on white striping muscle myopathy. *Poult Sci*, 100, 100945. [10.1016/j.psj.2020.12.042](https://doi.org/10.1016/j.psj.2020.12.042)
- Baeza E., Chartrin P., Bordeau T., Lessire M., Thoby J.M., Gigaud V., Blanchet M., Alinier A., Leterrier C., 2017. Omega-3 polyunsaturated fatty acids provided during embryonic development improve the growth performance and welfare of Muscovy ducks (*Cairina moschata*). *Poult Sci*, 96, 3176-3187. [10.3382/ps/pex147](https://doi.org/10.3382/ps/pex147)
- Bell J.T., Pai A.A., Pickrell J.K., Gaffney D.J., Pique-Regi R., Degner J.F., Gilad Y., Pritchard J.K., 2011. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol*, 12, R10. [10.1186/gb-2011-12-1-r10](https://doi.org/10.1186/gb-2011-12-1-r10)
- Bergoug H., Guinebretiere M., Tong Q., Roulston N., Romanini C.E., Exadaktylos V., Berckmans D., Garain P., Demmers T.G., McGonnell I.M., Bahr C., Burel C., Etteradossi N., Michel V., 2013. Effect of transportation duration of 1-day-old chicks on postplacement production performances and pododermatitis of broilers up to slaughter age. *Poult Sci*, 92, 3300-3309. [10.3382/ps.2013-03118](https://doi.org/10.3382/ps.2013-03118)
- Bernabucci U., Biffani S., Buggiotti L., Vitali A., Lacetera N., Nardone A., 2014. The effects of heat stress in Italian Holstein dairy cattle. *J. Dairy Sci.*, 97, 471-486. <https://doi.org/10.3168/jds.2013-6611>
- Bodin L., Secula A., Chapuis H., Cornuez A., Bernadet M.-D., Cobo E., Barrieu J., Manse H., Lessire M., Mercierand F., Le Bourhis M.-C., Brun J.-M., Martin X., Bonnefont C., M. D., Morisson M., 2022. Programmation nutritionnelle du métabolisme hépatique chez le canard mulard : Données zootechniques. Quatorzièmes Journées de la Recherche Avicole et Palmipèdes à Foie Gras, Tours, France
- Bodin L., Sécula A., Chapuis H., Cornuez A., Lessire M., Cobo E., Marie-Louise S., Bonnefont C.M.D., Barrieu J., Mercierand F., Bravo C., Manse H., Le Bourhis M.C., Martin X., Pitel F., Brun J.M., Morisson M., 2019. Dietary methionine deficiency reduces laying performances of female common ducks and impacts traits of interest of their mule ducklings. *Poult Sci*, 98, 5590-5600. [10.3382/ps/pez315](https://doi.org/10.3382/ps/pez315)
- Cardoso F.F., Tempelman R.J., 2012. Linear reaction norm models for genetic merit prediction of Angus cattle under genotype by environment interaction1. *J Anim Sci*, 90, 2130-2141. [10.2527/jas.2011-4333](https://doi.org/10.2527/jas.2011-4333)

- Cerutti C., Leroux S., Gourichon D., Labrune Y., David I., Zerjal T., Cousthama V., Devailly G., Pitel F., 2022. Short communication: Effects of in-ovo injection of endocrine disruptors and methyltransferase inhibitor on quail growth and egg-laying performances. *Animal*, 16, 100464. [10.1016/j.animal.2022.100464](https://doi.org/10.1016/j.animal.2022.100464)
- Cerutti C., Leroux S., Terzian P., Lledo J., Gourichon D., Serre R.-F., Vandecasteele C., Klopp C., C., Iampietro C., Gaspin C., Milan D., Donnadiou C., Pitel F., Devailly G., 2022. Transgenerational epigenetics in quail whole genome DNA methylation analysis. ECCB2022 - 21st European Conference on Computational Biology, Sitges, Spain
- Chen Q., Yan W., Duan E., 2016. Epigenetic inheritance of acquired traits through sperm RNAs and sperm RNA modifications. *Nature reviews. Genetics*, 17, 733-743. [10.1038/nrg.2016.106](https://doi.org/10.1038/nrg.2016.106)
- Cerian G., 2011. Essential fatty acids and early life programming in meat-type birds. *World's Poultry Science Journal*, 67, 599-614. [10.1017/S0043933911000705](https://doi.org/10.1017/S0043933911000705)
- Cousthama V., Andrieux C., Collin A., David I., Demars J., Houssier M., Lagarrigue S., Metayer-Coustard S., Mignon-Grasteau S., Petit A., Vitorino Carvalho A., Zerjal T., Pitel F., 2022. Epigenetics at the interface between environment and DNA: what importance for breeding practices and selection methods? *L'épigénétique à l'interface entre l'environnement et l'ADN : quelle importance pour les pratiques d'élevage et les méthodes de sélection ?* In: 14. Journées de la Recherche Avicole et Palmipèdes à Foie Gras. Elsevier, Tours, France, 628-629. [10.1016/j.anscip.2022.05.028](https://doi.org/10.1016/j.anscip.2022.05.028)
- Danchin E., Pocheville A., Rey O., Pujol B., Blanchet S., 2019. Epigenetically facilitated mutational assimilation: epigenetics as a hub within the inclusive evolutionary synthesis. *Biol Rev*, 94, 259-282. <https://doi.org/10.1111/brv.12453>
- David I., Canario L., Combes S., Demars J., 2019. Intergenerational Transmission of Characters Through Genetics, Epigenetics, Microbiota, and Learning in Livestock. *Front Genet*, 10, 1058. [10.3389/fgene.2019.01058](https://doi.org/10.3389/fgene.2019.01058)
- David S.A., Vitorino Carvalho A., Gimonnet C., Brionne A., Hennequet-Antier C., Piegu B., Crochet S., Courousse N., Bordeau T., Bigot Y., Collin A., Cousthama V., 2019. Thermal Manipulation During Embryogenesis Impacts H3K4me3 and H3K27me3 Histone Marks in Chicken Hypothalamus. *Front Genet*, 10, 1207. [10.3389/fgene.2019.01207](https://doi.org/10.3389/fgene.2019.01207)
- David S.-A., Mersch M., Foissac S., Collin A., Pitel F., Cousthama V., 2017. Genome-Wide Epigenetic Studies in Chicken: A Review. *Epigenomes*, 1, 20.
- de Goede O.M., Nachun D.C., Ferraro N.M., Gludemans M.J., Rao A.S., Smail C., Eulalio T.Y., Aguet F., Ng B., Xu J., Barbeira A.N., Castel S.E., Kim-Hellmuth S., Park Y., Scott A.J., Strober B.J., Anand S., Gabriel S., Getz G.A., Graubert A., Hadley K., Handsaker R.E., Huang K.H., Li X., MacArthur D.G., Meier S.R., Nedzel J.L., Nguyen D.T., Segrè A.V., Todres E., Balliu B., Bonazzola R., Brown A., Conrad D.F., Cotter D.J., Cox N., Das S., Dermitzakis E.T., Einson J., Engelhardt B.E., Eskin E., Flynn E.D., Fresard L., Gamazon E.R., Garrido-Martín D., Gay N.R., Guigó R., Hamel A.R., He Y., Hoffman P.J., Hormozdiari F., Hou L., Jo B., Kasela S., Kashin S., Kellis M., Kwong A., Li X., Liang Y., Mangul S., Mohammadi P., Muñoz-Aguirre M., Nobel A.B., Oliva M., Park Y., Parsana P., Reverter F., Rouhana J.M., Sabatti C., Saha A., Stephens M., Stranger B.E., Teran N.A., Viñuela A., Wang G., Wright F., Wucher V., Zou Y., Ferreira P.G., Li G., Melé M., Yeger-Lotem E., Bradbury D., Krubit T., McLean J.A., Qi L., Robinson K., Roche N.V., Smith A.M., Tabor D.E., Undale A., Bridge J., Brigham L.E., Foster B.A., Gillard B.M., Hasz R., Hunter M., Johns C., Johnson M., Karasik E., Kopen G., Leinweber W.F., McDonald A., Moser M.T., Myer K., Ramsey K.D., Roe B., Shad S., Thomas J.A., Walters G., Washington M., Wheeler J., Jewell S.D., Rohrer D.C., Valley D.R., Davis D.A., Mash D.C., Barcus M.E., Branton P.A., Sobin L., Barker L.K., Gardiner H.M., Mosavel M., Siminoff L.A., Flicek P., Haeussler M., Juettemann

- T., Kent W.J., Lee C.M., Powell C.C., Rosenbloom K.R., Ruffier M., Sheppard D., Taylor K., Trevanion S.J., Zerbino D.R., Abell N.S., Akey J., Chen L., Demanelis K., Doherty J.A., Feinberg A.P., Hansen K.D., Hickey P.F., Jasmine F., Jiang L., Kaul R., Kibriya M.G., Li J.B., Li Q., Lin S., Linder S.E., Pierce B.L., Rizzardi L.F., Skol A.D., Smith K.S., Snyder M., Stamatoyannopoulos J., Tang H., Wang M., Carithers L.J., Guan P., Koester S.E., Little A.R., Moore H.M., Nierras C.R., Rao A.K., Vaught J.B., Volpi S., Brown C.D., Wen X., Hall I.M., Battle A., Lappalainen T., Im H.K., Ardlie K.G., Mostafavi S., Quertermous T., Kirkegaard K., Montgomery S.B., 2021. Population-scale tissue transcriptomics maps long non-coding RNAs to complex disease. *Cell*, 184, 2633-2648. <https://doi.org/10.1016/j.cell.2021.03.050>
- de Lamarck J.B.P.A.d.M., 1809. Philosophie zoologique, ou Exposition des considérations relative à l'histoire naturelle des animaux. Dentu, et l'auteur.
- Di Criscio M., Lodahl J.E., Stamatakis A., Kitraki E., Bakoyiannis I., Repouskou A., Bornehag C.G., Gennings C., Lupu D., Rüegg J., 2023. A human-relevant mixture of endocrine disrupting chemicals induces changes in hippocampal DNA methylation correlating with hyperactive behavior in male mice. *Chemosphere*, 313, 137633. [10.1016/j.chemosphere.2022.137633](https://doi.org/10.1016/j.chemosphere.2022.137633)
- Do C., Shearer A., Suzuki M., Terry M.B., Gelernter J., Greally J.M., Tycko B., 2017. Genetic-epigenetic interactions in cis: a major focus in the post-GWAS era. *Genome Biol*, 18, 120. [10.1186/s13059-017-1250-y](https://doi.org/10.1186/s13059-017-1250-y)
- Donohoe D.R., Bultman S.J., 2012. Metaboloepigenetics: interrelationships between energy metabolism and epigenetic control of gene expression. *J Cell Physiol*, 227, 3169-3177. [10.1002/jcp.24054](https://doi.org/10.1002/jcp.24054)
- Duenk P., Calus M.P.L., Wientjes Y.C.J., Breen V.P., Henshall J.M., Hawken R., Bijma P., 2019. Estimating the purebred-crossbred genetic correlation of body weight in broiler chickens with pedigree or genomic relationships. *Genet Sel Evol*, 51, 6. [10.1186/s12711-019-0447-9](https://doi.org/10.1186/s12711-019-0447-9)
- Esfandyari H., Bijma P., Henryon M., Christensen O.F., Sørensen A.C., 2016. Genomic prediction of crossbred performance based on purebred Landrace and Yorkshire data using a dominance model. *Genet Sel Evol*, 48, 40. [10.1186/s12711-016-0220-2](https://doi.org/10.1186/s12711-016-0220-2)
- Feeney A., Nilsson E., Skinner M.K., 2014. Epigenetics and transgenerational inheritance in domesticated farm animals. *J Anim Sci Biotechnol*, 5, 48. [10.1186/2049-1891-5-48](https://doi.org/10.1186/2049-1891-5-48)
- Feil R., Fraga M.F., 2012. Epigenetics and the environment: emerging patterns and implications. *Nat Rev Genet*, 13, 97-109. [10.1038/nrg3142](https://doi.org/10.1038/nrg3142)
- Feldmann A., Ivanek R., Murr R., Gaidatzis D., Burger L., Schübeler D., 2013. Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS genetics*, 9, e1003994-e1003994. [10.1371/journal.pgen.1003994](https://doi.org/10.1371/journal.pgen.1003994)
- Friggens N.C., Blanc F., Berry D.P., Puillet L., 2017. Review: Deciphering animal robustness. A synthesis to facilitate its use in livestock breeding and management. *Animal*, 11, 2237-2251. <https://doi.org/10.1017/S175173111700088X>
- Gao T., Zhao M., Zhang L., Li J., Yu L., Lv P., Gao F., Zhou G., 2017. Effect of in ovo feeding of l-arginine on the hatchability, growth performance, gastrointestinal hormones, and jejunal digestive and absorptive capacity of posthatch broilers. *J Anim Sci*, 95, 3079-3092. [10.2527/jas.2016.0465](https://doi.org/10.2527/jas.2016.0465)
- Gertz J., Varley K.E., Reddy T.E., Bowling K.M., Pauli F., Parker S.L., Kucera K.S., Willard H.F., Myers R.M., 2011. Analysis of DNA Methylation in a Three-Generation Family Reveals Widespread Genetic Influence on Epigenetic Regulation. *PLOS Genetics*, 7, e1002228. [10.1371/journal.pgen.1002228](https://doi.org/10.1371/journal.pgen.1002228)

- Grandjean V., Fourré S., De Abreu D.A., Derieppe M.A., Remy J.J., Rassoulzadegan M., 2015. RNA-mediated paternal heredity of diet-induced obesity and metabolic disorders. *Sci Rep*, 5, 18193. 10.1038/srep18193
- Haldane J.B.S., 1946. The interaction of Nature and Nurture. *Annals of Eugenics*, 13, 197-205. <https://doi.org/10.1111/j.1469-1809.1946.tb02358.x>
- Hanson M.A., Gluckman P.D., 2014. Early developmental conditioning of later health and disease: physiology or pathophysiology? *Physiol Rev*, 94, 1027-1076. 10.1152/physrev.00029.2013
- Havenstein G.B., Ferket P.R., Qureshi M.A., 2003. Growth, livability, and feed conversion of 1957 versus 2001 broilers when fed representative 1957 and 2001 broiler diets. *Poult Sci*, 82, 1500-1508. 10.1093/ps/82.10.1500
- Hayes B.J., Bowman P.J., Chamberlain A.J., Savin K., van Tassell C.P., Sonstegard T.S., Goddard M.E., 2009. A validated genome wide association study to breed cattle adapted to an environment altered by climate change. *PLoS One*, 4, e6676. 10.1371/journal.pone.0006676
- Ho D.H., Reed W.L., Burggren W.W., 2011. Egg yolk environment differentially influences physiological and morphological development of broiler and layer chicken embryos. *J Exp Biol*, 214, 619-628. 10.1242/jeb.046714
- Hou Z., Sun Q., Hu Y., Yang S., Zong Y., Zhao R., 2018. Maternal betaine administration modulates hepatic type 1 iodothyronine deiodinase (Dio1) expression in chicken offspring through epigenetic modifications. *Comp Biochem Physiol B Biochem Mol Biol*, 218, 30-36. 10.1016/j.cbpb.2018.01.008
- Hu Y., Feng Y., Ding Z., Lv L., Sui Y., Sun Q., Abobaker H., Cai D., Zhao R., 2020. Maternal betaine supplementation decreases hepatic cholesterol deposition in chicken offspring with epigenetic modulation of SREBP2 and CYP7A1 genes. *Poult Sci*, 99, 3770. 10.1016/j.psj.2020.06.002
- Hu Y., Sun Q., Li X., Wang M., Cai D., Li X., Zhao R., 2015. In Ovo injection of betaine affects hepatic cholesterol metabolism through epigenetic gene regulation in newly hatched chicks. *PLoS One*, 10, e0122643. 10.1371/journal.pone.0122643
- Hynd P.I., Weaver S., Edwards N.M., Heberle N.D., Bowling M., 2016. Developmental programming: a new frontier for the poultry industry? *Animal Production Science*, 56, 1233-1238. <https://doi.org/10.1071/AN15373>
- Jehl F., Degalez F., Bernard M., Lecerf F., Lagoutte L., Désert C., Coulée M., Bouchez O., Leroux S., Abasht B., Tixier-Boichard M., Bed'hom B., Burlot T., Gourichon D., Bardou P., Acloque H., Foissac S., Djebali S., Giuffra E., Zerjal T., Pitel F., Klopp C., Lagarrigue S., 2021. RNA-Seq Data for Reliable SNP Detection and Genotype Calling: Interest for Coding Variant Characterization and Cis-Regulation Analysis by Allele-Specific Expression in Livestock Species. *Front Genet*, 12, 655707. 10.3389/fgene.2021.655707
- Jha R., Singh A.K., Yadav S., Berrocoso J.F.D., Mishra B., 2019. Early Nutrition Programming (in ovo and Post-hatch Feeding) as a Strategy to Modulate Gut Health of Poultry. *Front Vet Sci*, 6, 82. 10.3389/fvets.2019.00082
- Jones P.A., 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*, 13, 484-492. 10.1038/nrg3230
- Kadam M.M., Barekatin M.R., Bhanja S.K., Iji P.A., 2013. Prospects of in ovo feeding and nutrient supplementation for poultry: the science and commercial applications--a review. *J Sci Food Agric*, 93, 3654-3661. 10.1002/jsfa.6301
- Kalantar M., Hosseini S.M., Hosseini M.R., Kalantar M.H., Farmanullah, Yang L.G., 2019. Effects of In Ovo Injection of Coenzyme Q10 on Hatchability, Subsequent Performance, and Immunity of Broiler Chickens. *Biomed Res Int*, 2019, 7167525. 10.1155/2019/7167525

- Kisliouk T., Ziv M., Meiri N., 2010. Epigenetic control of translation regulation: alterations in histone H3 lysine 9 post-translation modifications are correlated with the expression of the translation initiation factor 2B (Eif2b5) during thermal control establishment. *Dev Neurobiol*, 70, 100-113. 10.1002/dneu.20763
- Lagarigue S., Lorthiois M., Degalez F., Gilot D., Derrien T., 2021. LncRNAs in domesticated animals: from dog to livestock species. *Mamm Genome*. 10.1007/s00335-021-09928-7
- Lee J., Hyeon D.Y., Hwang D., 2020. Single-cell multiomics: technologies and data analysis methods. *Exp Mol Med*, 52, 1428-1442. 10.1038/s12276-020-0420-2
- Lemire M., Zaidi S.H.E., Ban M., Ge B., Aissi D., Germain M., Kassam I., Wang M., Zanke B.W., Gagnon F., Morange P.-E., Trégouët D.-A., Wells P.S., Sawcer S., Gallinger S., Pastinen T., Hudson T.J., 2015. Long-range epigenetic regulation is conferred by genetic variation located at thousands of independent loci. *Nat Commun*, 6, 6326. 10.1038/ncomms7326
- Leroux S., Gourichon D., Leterrier C., Labrune Y., Coustham V., Riviere S., Zerjal T., Coville J.L., Morisson M., Minvielle F., Pitel F., 2017. Embryonic environment and transgenerational effects in quail. *Genet Sel Evol*, 49, 14. 10.1186/s12711-017-0292-7
- Li C., Ni Y.Q., Xu H., Xiang Q.Y., Zhao Y., Zhan J.K., He J.Y., Li S., Liu Y.S., 2021. Roles and mechanisms of exosomal non-coding RNAs in human health and diseases. *Signal Transduct Target Ther*, 6, 383. 10.1038/s41392-021-00779-x
- Li S., Zhi L., Liu Y., Shen J., Liu L., Yao J., Yang X., 2016. Effect of in ovo feeding of folic acid on the folate metabolism, immune function and epigenetic modification of immune effector molecules of broiler. *Br J Nutr*, 115, 411-421. 10.1017/S0007114515004511
- Lillehammer M., Hayes B.J., Meuwissen T.H.E., Goddard M.E., 2009. Gene by environment interactions for production traits in Australian dairy cattle. *Journal of Dairy Science*, 92, 4008-4017. <https://doi.org/10.3168/jds.2008-1484>
- Loyau T., Bedrani L., Berri C., Metayer-Coustard S., Praud C., Coustham V., Mignon-Grasteau S., Duclos M.J., Tesseraud S., Rideau N., Hennequet-Antier C., Everaert N., Yahav S., Collin A., 2015. Cyclic variations in incubation conditions induce adaptive responses to later heat exposure in chickens: a review. *Animal*, 9, 76-85. 10.1017/S1751731114001931
- Loyau T., Hennequet-Antier C., Coustham V., Berri C., Leduc M., Crochet S., Sannier M., Duclos M.J., Mignon-Grasteau S., Tesseraud S., Brionne A., Metayer-Coustard S., Moroldo M., Lecardonnel J., Martin P., Lagarigue S., Yahav S., Collin A., 2016. Thermal manipulation of the chicken embryo triggers differential gene expression in response to a later heat challenge. *BMC Genomics*, 17, 329. 10.1186/s12864-016-2661-y
- Loyau T., Metayer-Coustard S., Berri C., Crochet S., Cailleau-Audouin E., Sannier M., Chartrin P., Praud C., Hennequet-Antier C., Rideau N., Courousse N., Mignon-Grasteau S., Everaert N., Duclos M.J., Yahav S., Tesseraud S., Collin A., 2014. Thermal manipulation during embryogenesis has long-term effects on muscle and liver metabolism in fast-growing chickens. *PLoS One*, 9, e105339. 10.1371/journal.pone.0105339
- Makova K.D., Hardison R.C., 2015. The effects of chromatin organization on variation in mutation rates in the genome. *Nat Rev Genet*, 16, 213-223. 10.1038/nrg3890
- Mathur P.K., Horst P., 1994. Genotype by environment interactions in laying hens based on relationship between breeding values of sires in temperate and tropical environments. *Poult Sci*, 73, 1777-1784. 10.3382/ps.0731777
- Melnik B.C., Stremmel W., Weiskirchen R., John S.M., Schmitz G., 2021. Exosome-Derived MicroRNAs of Human Milk and Their Effects on Infant Health and Development. *Biomolecules*, 11. 10.3390/biom11060851

- Mersch M., David S.-A., Vitorino Carvalho A., Foissac S., Collin A., Pitel F., Cousthama V., 2018. Apports du séquençage haut-débit sur la connaissance de l'épigénome aviaire. *INRA Productions Animales*, 31, 325-336. [10.20870/productions-animales.2018.31.4.2372](https://doi.org/10.20870/productions-animales.2018.31.4.2372)
- Mignon-Grasteau S., Juin H., Bastianelli D., Gomez J., Carré B., 2010. Genetic parameters of digestibility of wheat- or corn-based diets in chickens. In: *World Congress on Genetics Applied to Livestock Production*. Gft D.E.U. (Ed). Gesellschaft für Tierzuchtwissenschaften, Leipzig, Germany, 4 p.
- Miska E.A., Ferguson-Smith A.C., 2016. Transgenerational inheritance: Models and mechanisms of non-DNA sequence-based inheritance. *Science*, 354, 59-63. [doi:10.1126/science.aaf4945](https://doi.org/10.1126/science.aaf4945)
- Morisson M., Cousthama V., FRESARD L., Collin A., Zerjal T., Metayer-Coustard S., Bodin L., Minvielle F., Brun J.-M., Pitel F., 2017. Nutritional Programming and Effect of Ancestor Diet in Birds. In: *Handbook of Nutrition, Diet, and Epigenetics*. Springer International Publishing, np.
- Mulder H.A., Bijma P., 2005. Effects of genotype x environment interaction on genetic gain in breeding programs. *J Anim Sci*, 83, 49-61. [10.2527/2005.83149x](https://doi.org/10.2527/2005.83149x)
- Nakamura M., Gao Y., Dominguez A.A., Qi L.S., 2021. CRISPR technologies for precise epigenome editing. *Nat Cell Biol*, 23, 11-22. [10.1038/s41556-020-00620-7](https://doi.org/10.1038/s41556-020-00620-7)
- Nauta W.J., Veerkamp R.F., Brascamp E.W., Bovenhuis H., 2006. Genotype by Environment Interaction for Milk Production Traits Between Organic and Conventional Dairy Cattle Production in The Netherlands. *J Dairy Sci*, 89, 2729-2737. [https://doi.org/10.3168/jds.S0022-0302\(06\)72349-9](https://doi.org/10.3168/jds.S0022-0302(06)72349-9)
- N'Dri A L., Sellier N., Tixier-Boichard M., Beaumont C., Mignon-Grasteau S., 2007. Genotype by environment interactions in relation to growth traits in slow growing chickens. *Genet Sel Evol*, 39, 513-528. [10.1186/1297-9686-39-5-513](https://doi.org/10.1186/1297-9686-39-5-513)
- Nyuiadzi D., Berri C., Dusart L., Travel A., Meda B., Bouvarel I., Guilloteau L.A., Chartrin P., Cousthama V., Praud C., Le Bihan-Duval E., Tona J.K., Collin A., 2020. Short cold exposures during incubation and postnatal cold temperature affect performance, breast meat quality, and welfare parameters in broiler chickens. *Poult Sci*, 99, 857-868. [10.1016/j.psj.2019.10.024](https://doi.org/10.1016/j.psj.2019.10.024)
- Oladokun S., Adewole D.I., 2020. In ovo delivery of bioactive substances: an alternative to the use of antibiotic growth promoters in poultry production—a review. *Journal of Applied Poultry Research*, 29, 744-763. <https://doi.org/10.1016/j.japr.2020.06.002>
- Peebles E.D., 2018. In ovo applications in poultry: A review. *Poult Sci*, 97, 2322-2338. [10.3382/ps/pey081](https://doi.org/10.3382/ps/pey081)
- Piestun Y., Shinder D., Ruzal M., Halevy O., Brake J., Yahav S., 2008. Thermal manipulations during broiler embryogenesis: effect on the acquisition of thermotolerance. *Poult Sci*, 87, 1516-1525. [10.3382/ps.2008-00030](https://doi.org/10.3382/ps.2008-00030)
- Rauw W.M., Gomez-Raya L., 2015. Genotype by environment interaction and breeding for robustness in livestock. *Front Genet*, 6, 310. [10.3389/fgene.2015.00310](https://doi.org/10.3389/fgene.2015.00310)
- Reed W.L., Clark M.E., 2011. Beyond maternal effects in birds: responses of the embryo to the environment. *Integr Comp Biol*, 51, 73-80. [10.1093/icb/icr032](https://doi.org/10.1093/icb/icr032)
- Riggs A.D., Martienssen R.A., Russo V.E.A., 1996. Epigenetic mechanisms of gene regulation - Introduction. CSHL Press (N.Y.), Cold Spring Harbor Monograph Archive.
- Romé H., Varenne A., Hérault F., Chapuis H., Alleno C., Dehais P., Vignal A., Burlot T., Le Roy P., 2015. GWAS analyses reveal QTL in egg layers that differ in response to diet differences. *Genet Sel Evol*, 47, 83. [10.1186/s12711-015-0160-2](https://doi.org/10.1186/s12711-015-0160-2)

- Roto S.M., Kwon Y.M., Ricke S.C., 2016. Applications of In Ovo Technique for the Optimal Development of the Gastrointestinal Tract and the Potential Influence on the Establishment of Its Microbiome in Poultry. *Front Vet Sci*, 3, 63. 10.3389/fvets.2016.00063
- Rozenboim I., Piestun Y., Mobarkey N., Barak M., Hoyzman A., Halevy O., 2004. Monochromatic light stimuli during embryogenesis enhance embryo development and posthatch growth. *Poult Sci*, 83, 1413-1419. 10.1093/ps/83.8.1413
- Ruhl R., 2007. Effects of dietary retinoids and carotenoids on immune development. *Proc Nutr Soc*, 66, 458-469. 10.1017/S002966510600509X
- Sécuela A., Bluy L.E., Chapuis H., Bonnet A., Collin A., Gress L., Cornuez A., Martin X., Bodin L., Bonnefont C.M.D., Morisson M., 2022. Maternal dietary methionine restriction alters hepatic expression of one-carbon metabolism and epigenetic mechanism genes in the ducklings. *BMC Genomics*, 23, 823. 10.1186/s12864-022-09066-7
- Sécuela A., Chapuis H., Collin A., Bluy L.E., Bonnet A., Bodin L., Gress L., Cornuez A., Martin X., Bonnefont C.M.D., Morisson M., 2022. Maternal dietary methionine restriction alters the expression of energy metabolism genes in the duckling liver. *BMC Genomics*, 23, 407. 10.1186/s12864-022-08634-1
- Shinder D., Ruzal M., Giloh M., Druyan S., Piestun Y., Yahav S., 2011. Improvement of cold resistance and performance of broilers by acute cold exposure during late embryogenesis. *Poult Sci*, 90, 633-641. 10.3382/ps.2010-01089
- Silva L., Pinheiro-Castro N., Novaes G.M., Pascoal G.F.L., Ong T.P., 2019. Bioactive food compounds, epigenetics and chronic disease prevention: Focus on early-life interventions with polyphenols. *Food Res Int*, 125, 108646. 10.1016/j.foodres.2019.108646
- Skene P.J., Henikoff S., 2017. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife*, 6. 10.7554/eLife.21856
- Sun Y.M., Chen Y.Q., 2020. Principles and innovative technologies for decrypting noncoding RNAs: from discovery and functional prediction to clinical application. *J Hematol Oncol*, 13, 109. 10.1186/s13045-020-00945-8
- Sved J., Bird A., 1990. The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. *Proc Natl Acad Sci U S A*, 87, 4692-4696. 10.1073/pnas.87.12.4692
- Taha-Abdelaziz K., Hodgins D.C., Lammers A., Alkie T.N., Sharif S., 2018. Effects of early feeding and dietary interventions on development of lymphoid organs and immune competence in neonatal chickens: A review. *Vet Immunol Immunopathol*, 201, 1-11. 10.1016/j.vetimm.2018.05.001
- Thanabalan A., Kiarie E.G., 2021. Influence of Feeding Omega-3 Polyunsaturated Fatty Acids to Broiler Breeders on Indices of Immunocompetence, Gastrointestinal, and Skeletal Development in Broiler Chickens. *Front Vet Sci*, 8, 653152. 10.3389/fvets.2021.653152
- Tirado-Magallanes R., Rebbani K., Lim R., Pradhan S., Benoukraf T., 2017. Whole genome DNA methylation: beyond genes silencing. *Oncotarget*, 8, 5629-5637. 10.18632/oncotarget.13562
- Uni Z., Ferket P.R., Tako E., Kedar O., 2005. In ovo feeding improves energy status of late-term chicken embryos. *Poult Sci*, 84, 764-770. 10.1093/ps/84.5.764
- Uni Z., Ferket R.P., 2004. Methods for early nutrition and their potential. *World's Poultry Science Journal*, 60, 101-111. 10.1079/WPS20038
- Veron V., Marandel L., Liu J., Velez E.J., Lepais O., Panserat S., Skiba S., Seiliez I., 2018. DNA methylation of the promoter region of *bnip3* and *bnip3l* genes induced by metabolic programming. *BMC Genomics*, 19, 677. 10.1186/s12864-018-5048-4
- Vitorino Carvalho A., Hennequet-Antier C., Rouger R., Delaveau J., Bordeau T., Crochet S., Couroussé N., Pitel F., Collin A., Coustham V., 2023. Thermal conditioning of

- quail embryos has transgenerational and reversible long-term effects. *bioRxiv* 2023.02.02.526800; doi: <https://doi.org/10.1101/2023.02.02.526800>
- Vitorino Carvalho A., Hennequet-Antier C., Brionne A., Crochet S., Jimenez J., Courousse N., Collin A., Cousthama V., 2021. Embryonic thermal manipulation impacts the postnatal transcriptome response of heat-challenged Japanese quails. *BMC Genomics*, 22, 488. [10.1186/s12864-021-07832-7](https://doi.org/10.1186/s12864-021-07832-7)
- Vitorino Carvalho A., Hennequet-Antier C., Crochet S., Bordeau T., Courousse N., Cailleau-Audouin E., Chartrin P., Darras V.M., Zerjal T., Collin A., Cousthama V., 2020. Embryonic thermal manipulation has short and long-term effects on the development and the physiology of the Japanese quail. *PLoS One*, 15, e0227700. [10.1371/journal.pone.0227700](https://doi.org/10.1371/journal.pone.0227700)
- Waddington C.H., 1957. *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology*. In: *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology*. London: George Allen & Unwin, Ltd.
- Walser J.C., Furano A.V., 2010. The mutational spectrum of non-CpG DNA varies with CpG content. *Genome Res*, 20, 875-882. [10.1101/gr.103283.109](https://doi.org/10.1101/gr.103283.109)
- Wei X.J., Ni Y.D., Lu L.Z., Grossmann R., Zhao R.Q., 2011. The effect of equol injection in ovo on posthatch growth, meat quality and antioxidation in broilers. *Animal*, 5, 320-327. [10.1017/S1751731110001771](https://doi.org/10.1017/S1751731110001771)
- Yahav S., McMurtry J.P., 2001. Thermotolerance acquisition in broiler chickens by temperature conditioning early in life--the effect of timing and ambient temperature. *Poult Sci*, 80, 1662-1666. [10.1093/ps/80.12.1662](https://doi.org/10.1093/ps/80.12.1662)
- Yan X.P., Liu H.H., Liu J.Y., Zhang R.P., Wang G.S., Li Q.Q., Wang D.M., Li L., Wang J.W., 2015. Evidence in duck for supporting alteration of incubation temperature may have influence on methylation of genomic DNA. *Poult Sci*, 94, 2537-2545. [10.3382/ps/pev201](https://doi.org/10.3382/ps/pev201)
- Yossifoff M., Kisliouk T., Meiri N., 2008. Dynamic changes in DNA methylation during thermal control establishment affect CREB binding to the brain-derived neurotrophic factor promoter. *Eur J Neurosci*, 28, 2267-2277. [10.1111/j.1460-9568.2008.06532.x](https://doi.org/10.1111/j.1460-9568.2008.06532.x)
- Zhang T., Cooper S., Brockdorff N., 2015. The interplay of histone modifications - writers that read. *EMBO Rep*, 16, 1467-1481. [10.15252/embr.201540945](https://doi.org/10.15252/embr.201540945)

## EFFET DE L'INJECTION IN-OVO DE PERTURBATEURS ENDOCRINIENS OU D'UN INHIBITEUR DE MÉTHYLTRANSFÉRASES SUR LA CROISSANCE ET LA REPRODUCTION DES CAILLES

**Cerutti C.**<sup>1</sup>, Leroux S.<sup>1</sup>, Gourichon D.<sup>2</sup>, Labrune Y.<sup>1</sup>, Zerjal T.<sup>4</sup>, Coustham V.<sup>3,5</sup>, Devailly G.<sup>1</sup>, Pitel F.<sup>1</sup>

<sup>1</sup>GenPhySE, Université de Toulouse, INRAE, ENVT, 31320, Castanet Tolosan, France;

<sup>2</sup>INRAE, PEAT, 37380, Nouzilly, France; <sup>3</sup>INRAE, Université de Tours, BOA, 37380,

Nouzilly, France; <sup>4</sup>AgroParisTech, INRAE, Université Paris-Saclay, GABI, 78350, Jouy-en-

Josas, France; <sup>5</sup> Université de Pau et des Pays de l'Adour, E2S UPPA, INRAE, NUMEA; F-64310, Saint-Pée-sur-Nivelle, France.

Les expériences précoces, incluant l'environnement prénatal, sont connues pour influencer une grande variété de mécanismes impliqués dans la construction de l'individu [1]. Nous avons étudié l'effet de l'ajout de perturbateurs endocriniens ou d'un inhibiteur de méthyltransférases de l'ADN au cours du développement embryonnaire des cailles, sur leur croissance et leurs performances de ponte. Cinquante-quatre paires de parents par lignée ont été utilisées et les œufs fécondés de chaque paire ont été répartis aléatoirement en 5 groupes : un groupe témoin sans aucune injection, un groupe témoin traité par injection dans l'œuf d'huile de sésame (utilisé comme véhicule pour injecter les molécules), et 3 groupes traités par injection de génistéine, de bisphénol A (BPA) ou de 5-Aza-2'-déoxycytidine (5-Aza). Toutes les cailles issues de ces œufs (n=1979) ont été pesées individuellement à 8, 21, 36 et 78 jours et l'âge de la première ponte et le nombre d'œufs pondus ont été enregistrés pour les femelles. Ces analyses ont révélé un impact significatif du traitement sur la croissance. Les effets observés semblent spécifiques à chaque molécule, affectant de manière significative le poids des individus en interaction avec leur âge. Aucune interaction entre la lignée et le traitement n'a été détectée. En conclusion, l'injection de perturbateurs endocriniens ou d'inhibiteur de méthyltransférase dans l'œuf semble avoir des effets significatifs sur le développement des cailles. Les effets éventuels de ces traitements sur la méthylation de l'ADN restent à étudier afin de mettre en évidence les régions du génome affectées par les traitements et ainsi mieux comprendre les mécanismes mis en jeu.

*Financement : Ce travail a été soutenu par le programme EPIQUES (INRAE, Métaprogramme SelGen). CC est financée par la région Occitanie et le département de Génétique Animale d'INRAE.*

*Références : [1] Sinclair, K.D., Rutherford, K.M.D., Wallace, J.M., Brameld, J.M., Stöger, R., Alberio, R., Sweetman, D., Gardner, D.S., Perry, V.E.A., Adam, C.L., Ashworth, C.J., Robinson, J.E., Dwyer, C.M., 2016. Epigenetics and developmental programming of welfare and production traits in farm animals. *Reprod. Fertil. Dev.* 28, 1443.*

*<https://doi.org/10.1071/RD16102>*





## Transgenerational epigenetics in quail

Chloé Cerutti, Sophie Leroux, Paul Terzian, Joanna Lledo, David Gourichon,  
Rémy-Félix Serre, Céline Vandecasteele, Christophe C. Klopp, Carole  
Iampietro, Christine Gaspin, et al.

### ► To cite this version:

Chloé Cerutti, Sophie Leroux, Paul Terzian, Joanna Lledo, David Gourichon, et al.. Transgenerational epigenetics in quail: whole genome DNA methylation analysis. ECCB2022 - 21st European Conference on Computational Biology, Sep 2022, Sitges, Spain. hal-03935584

**HAL Id: hal-03935584**

**<https://hal.inrae.fr/hal-03935584>**

Submitted on 12 Jan 2023

**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.

# Transgenerational epigenetics in quail: whole genome DNA methylation analysis

Chloé Cerutti<sup>1</sup>, Sophie Leroux<sup>1</sup>, Paul Terzian<sup>2</sup>, Joanna Lledo<sup>3</sup>, David Gourichon<sup>4</sup>, Rémy-Félix Serre<sup>3</sup>, Céline Vandecasteele<sup>3</sup>, Christophe Klopp<sup>2</sup>, Carole Iampietro<sup>3</sup>, Christine Gaspin<sup>2</sup>, Denis Milan<sup>1</sup>, Cécile Donnadieu<sup>3</sup>, Frédérique Pitel<sup>1</sup>, Guillaume Devailly<sup>1</sup>

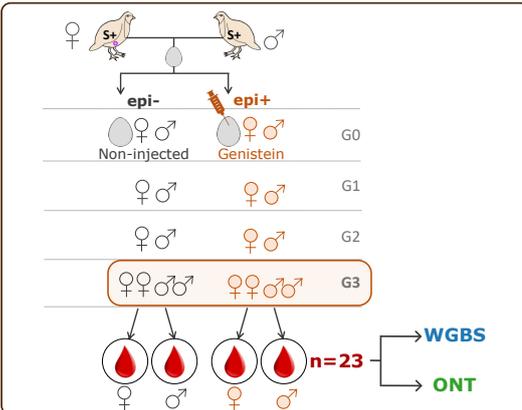
<sup>1</sup> GenPhySE, Université de Toulouse, INRAE, ENVT, 31320 Castanet-Tolosan, France  
<sup>2</sup> MIAT, PF Bioinfo GenoToul, Université de Toulouse, INRAE, Chemin de Borde Rouge, 31320 Castanet-Tolosan, France  
<sup>3</sup> GeT-PlaGe - Génome et Transcriptome - Plateforme Génomique, GET - INRAE, 31320 Castanet-Tolosan, France  
<sup>4</sup> INRAE, PEAT, 37380 Nouzilly, France



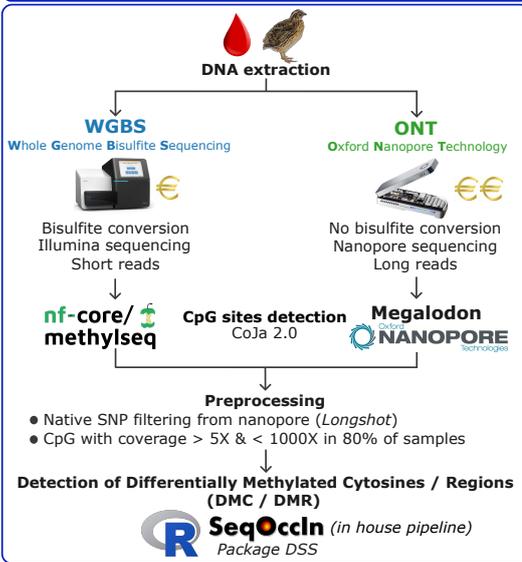
Contact : chloe.cerutti@inrae.fr

The influence of the **prenatal environment** on the adult **phenotype development** is partially mediated by **epigenetic phenomena**. Recently, an increasing number of studies highlighted the **transmission of epigenetic marks between generations** following an environmental exposure. However, there is much debate about their acquired transmission beyond the exposed individuals. Recent studies revealed that **non-genetic inheritance** was probably present in **avian species**. In one of them, fertilized eggs were injected with an endocrine disruptor (**genistein**) and after **three generations** without any other injection, several traits were impacted by the ancestor treatment such as the **reproduction** and the **behavior**. Here we analyse the **DNA methylation** between **control** and **treated lines** of the **third generation**.

## 1 - Experimental design



## 2 - Methods



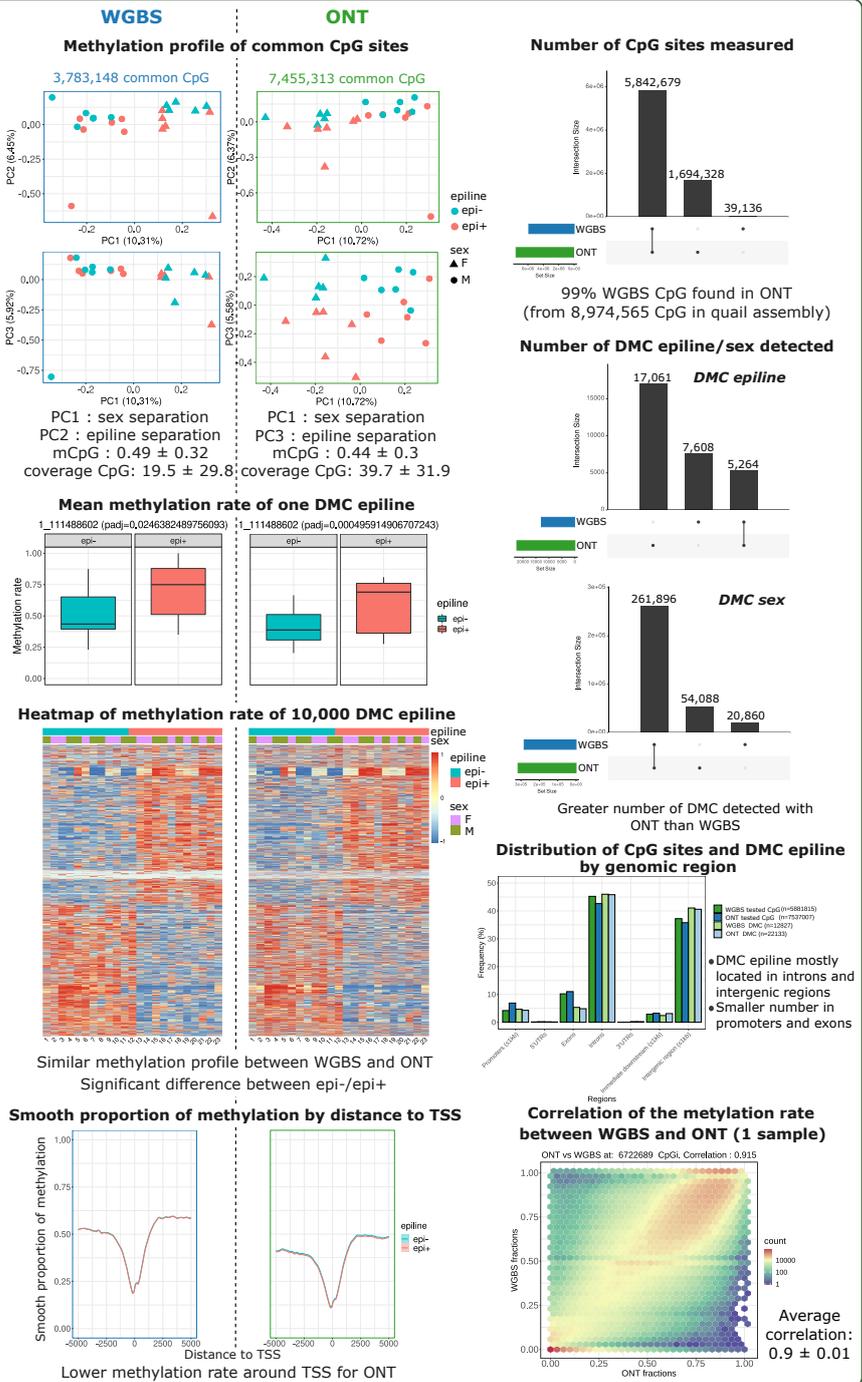
## Conclusion

- Thousands of DMC detected between epilines
- Transgenerational effects of genistein on the DNA methylation
- ONT vs WGBS: comparable results and best detection of transgenerational effects from ONT data

## Perspectives

- Optimising SNP filtering and minimising false positives:
  - SNP calling with GemBS for WGBS data
- Gene Ontology (GO) analysis to be finalised:
  - VISEAGO
- Exclude the genetic causes for the observed differences between epilines
- Identification of the biological mechanisms involved

## 3 - Results





**HAL**  
open science

## Effets transgénérationnels d'une exposition embryonnaire à la chaleur sur le transcriptome et l'épigénome de la caille japonaise

A Vitorino Carvalho, Benoît Piégu, Christelle Hennequet-Antier, Nathalie Couroussé, Sabine Crochet, Thierry Bordeau, Romuald Rouger, Aurélien Brionne, Chloé Cerutti, Guillaume Devailly, et al.

### ► To cite this version:

A Vitorino Carvalho, Benoît Piégu, Christelle Hennequet-Antier, Nathalie Couroussé, Sabine Crochet, et al.. Effets transgénérationnels d'une exposition embryonnaire à la chaleur sur le transcriptome et l'épigénome de la caille japonaise. *épiPhase - 7. Journées d'Animation Scientifique autour de l'Épigénétique*, Jun 2022, Rennes, France. hal-03726346

**HAL Id: hal-03726346**

**<https://hal.inrae.fr/hal-03726346>**

Submitted on 18 Jul 2022

**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.

## EFFETS TRANSGENERATIONNELS D'UNE EXPOSITION EMBRYONNAIRE A LA CHALEUR SUR LE TRANSCRIPTOME ET L'EPIGENOME DE LA CAILLE JAPONAISE

A. Vitorino Carvalho<sup>1,2</sup>, B. Piégu<sup>2</sup>, C. Hennequet-Antier<sup>1,3</sup>, N. Couroussé<sup>1</sup>, S. Crochet<sup>1</sup>, T. Bordeau<sup>1</sup>, R. Rouger<sup>4</sup>, A. Brionne<sup>1,5</sup>, C. Cerutti<sup>6</sup>, G. Devailly<sup>6</sup>, F. Pitel<sup>6</sup>, A. Collin<sup>1</sup>, **V. Coustham**<sup>1,7\*</sup>

<sup>1</sup>INRAE, Université de Tours, BOA, 37380, Nouzilly, France

<sup>2</sup>CNRS, IFCE, INRAE, Université de Tours, PRC, 37380, Nouzilly, France

<sup>3</sup>Université Paris-Saclay, INRAE, MIGALE, Jouy-en-Josas, 78350, France.

<sup>4</sup>SYSAAF, 37380 Nouzilly, France

<sup>5</sup>INRAE, LPGP, 35000, Rennes, France.

<sup>6</sup>GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet-Tolosan, France

<sup>7</sup>Université de Pau et des Pays de l'Adour, E2S UPPA, INRAE, NUMEA, 64310, Saint-Pée-sur-Nivelle, France

[\\*vincent.coustham@inrae.fr](mailto:vincent.coustham@inrae.fr)

Les travaux précédents de l'équipe MOQA (UMR BOA) et collaborateurs ont montré que la tolérance à la chaleur des poulets commerciaux mâles était améliorée en augmentant de façon cyclique la température d'incubation des œufs. Cette procédure, appelée manipulation thermique embryonnaire (MT), est associée à une réponse accrue du transcriptome musculaire lorsque les animaux sont soumis à un challenge thermique à l'âge d'abattage, 35 jours après l'éclosion (J35). Nous avons également montré que la MT est associée au changement de deux marques épigénétiques dans l'hypothalamus de poulets à J35 qui pourraient contribuer à la réponse génique induite par la chaleur chez les animaux MT (David, Vitorino Carvalho et al., 2019).

Pour explorer l'impact sur plusieurs générations de la MT sur le phénotype, le transcriptome et l'épigénome, nous avons utilisé une lignée consanguine de cailles japonaises (*Coturnix japonica*). Le protocole de MT a été transposé chez la caille en élevant la température d'incubation de 37,8°C à 39,5°C pendant 12 heures par jour à partir de la 12ème heure d'incubation jusqu'à I13 (I0-13 ; Vitorino Carvalho et al., 2020). Le traitement MT a été répété sur les descendants des animaux MT pendant quatre générations consécutives afin d'évaluer l'impact multigénérationnel du traitement, en parallèle de 4 générations d'animaux témoins au pedigree similaire incubés dans des conditions standard. De plus, après 2 générations MT, nous avons dérivé une lignée dont les descendants ont été incubés en condition standard lors des deux générations suivantes (2 générations incubation MT + 2 générations incubation standard) pour évaluer les effets transgénérationnels de la TM. Une caractérisation phénotypique approfondie a révélé que la MT a des effets transgénérationnels et réversibles à long terme sur le poids des animaux et des œufs pondus (Vitorino Carvalho et al., communication épiPhase 2021).

L'étude des mécanismes moléculaires a été réalisée en 1<sup>ère</sup> et en 4<sup>ème</sup> génération en utilisant des approches de séquençage tout-génome pour étudier sans *a priori* l'expression des gènes, la méthylation de l'ADN et les modifications post-traductionnelles des histones dans l'hypothalamus et le foie pour certaines conditions. En première génération, un effet sur l'expression des gènes en réponse à un coup de chaleur a été observé dans l'hypothalamus des femelles MT (Vitorino Carvalho et al., 2021), mais aucune région différentiellement méthylée dans l'hypothalamus a pu être observée quel que soit le sexe. En quatrième génération, en l'absence de coup de chaleur, aucun effet évident de la MT répétée sur 4 générations n'a pu être détecté sur les transcriptomes et les méthylomes hypothalamiques et hépatiques quel que soit le sexe malgré les différences phénotypiques observées. Cependant, un effet de la MT a pu être observé en 4<sup>ème</sup> génération sur la modification post-traductionnelle H3K4me3 (méthylation de la lysine 4 de l'histone H3) dans l'hypothalamus suggérant, comme chez le poulet, un effet épigénétique de la MT sur cette marque épigénétique. Par ailleurs, le profil de H3K4me3 des animaux transgénérationnels, intermédiaire entre celui des animaux MT de 4<sup>ème</sup> génération et témoins de 4<sup>ème</sup> génération, suggère que ceux-ci ont conservé une partie des marques acquises lors de l'exposition à la MT pendant les deux premières générations. Ces résultats seront discutés en regard des challenges techniques qui ont dû être levés pour aboutir à des conclusions fiables.

**Financement:** ANR JCJC "QuailHeatE" [ANR-15-CE02-0009-01].

**Références :**

David S-A, Vitorino Carvalho A, Gimmonet C, Brionne A, Hennequet-Antier C, Piégu B, Crochet S, Couroussé N, Bordeau T, Bigot Y, Collin A and Coustham V. *Thermal Manipulation During Embryogenesis Impacts H3K4me3 and H3K27me3 Histone Marks in Chicken Hypothalamus*. *Front Genet.*, **2019**, 10:1207.

Vitorino Carvalho A, Hennequet-Antier C, Crochet S, Bordeau T, Couroussé N, Cailleau-Audouin E, Chartrin P, Darras VM, Zerjal T, Collin A and Coustham V. *Embryonic thermal manipulation has short and long-term effects on the development and the physiology of the Japanese quail*. *PLoS One*, **2020**, 15(1):e0227700.

Vitorino Carvalho A, Hennequet-Antier C, Brionne A, Crochet S, Jimenez J, Couroussé N, Collin A and Coustham V. *Embryonic thermal manipulation impacts the postnatal transcriptome response of heat-challenged Japanese quails*. *BMC Genomics*, **2021**, 22(1):488.



**HAL**  
open science

## Transgenerational epigenetics in quail

Chloé Cerutti, Sophie Leroux, Paul Terzian, Joanna Lledo, David Gourichon, Rémy-Félix Serre, Céline Vandecasteele, Christophe Klopp, Carole Iampietro, Christine Gaspin, et al.

► **To cite this version:**

Chloé Cerutti, Sophie Leroux, Paul Terzian, Joanna Lledo, David Gourichon, et al.. Transgenerational epigenetics in quail: whole genome DNA methylation analysis. Genotoul Biostat Bioinfo Day, Dec 2022, Castanet-Tolosan, France. hal-03957296

**HAL Id: hal-03957296**

**<https://hal.inrae.fr/hal-03957296>**

Submitted on 26 Jan 2023

**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.

# Transgenerational epigenetics in quail : whole genome DNA methylation analysis

Chloé Cerutti<sup>1</sup>, Sophie Leroux<sup>1</sup>, Paul Terzian<sup>2</sup>, Joanna Lledo<sup>3</sup>, David Gourichon<sup>4</sup>, Rémy-Félix Serre<sup>3</sup>, Céline Vandecasteele<sup>3</sup>, Christophe Klopp<sup>2</sup>, Carole Iampietro<sup>3</sup>, Christine Gaspin<sup>2</sup>, Denis Milan<sup>1</sup>, Cécile Donnadiou<sup>3</sup>, Frédérique Pitel<sup>1</sup> and Guillaume Devailly<sup>1</sup>

<sup>1</sup> GenPhySE, Université de Toulouse, INRAE, ENVT, 31326, Castanet Tolosan, France  
{chloe.cerutti, sophie.leroux1, denis.milan, frederique.pitel, guillaume.devailly}@inrae.fr

<sup>2</sup> MIAT, UR875 INRAE, 31326, Castanet-Tolosan, France

{paul.terzian, christophe.klopp, christine.gaspin}@inrae.fr

<sup>3</sup> GeT-PlaGe - Génome et Transcriptome - Plateforme Génomique, US1426 GET - INRAE, 31326, Castanet-Tolosan, France

{joanna.lledo, remy-felix.serre, celine.vandecasteele, carole.iampietro, cecile.donnadiou}@inrae.fr

<sup>4</sup> PEAT, UE1295 - INRAE Centre Val de Loire, 37380, Nouzilly, France

{david.gourichon}@inrae.fr

## Abstract

Prenatal environment influence on adult phenotype is partially mediated by epigenetics phenomena. One study highlighted significant effects of the *in-ovo* injection of endocrine disruptors or DNA methyltransferase inhibitor on quail development, significantly reducing their weight [1]. Recently, an increasing number of studies highlighted the transmission of epigenetics marks between generations following an environmental exposure. However, there is much debate about their transmission beyond the exposed individuals. Recent studies revealed that non-genetics inheritance was probably present in avian species. In one of them [2], fertilized eggs were divided into two groups: one group injected with an endocrine disruptor, Genistein, and a non-injected control group. After three generations without any other injection, several traits were impacted by the ancestor treatment including reproduction and behavior. This pilot study highlighted the potential existence of transgenerational transmission of environmental effects in quails. To better understand the transgenerational transmission of these environmental effects, we generated DNA methylation data from blood samples of the third generation using WGBS (Whole Genome Bisulfite Sequencing) and ONT (Oxford Nanopore Technologies). To analyse the DNA methylation states, we developed a bioinformatics pipeline in order to detect differentially methylated cytosines (DMCs) due to transgenerational phenomena. We detected thousands of DMCs between both groups (Table 1). In addition to these analyses, these data allowed us to perform a comparative analysis between the WGBS and ONT sequencing technologies (Fig. 1).

	measured CpGs sites	epiline DMCs detected
WGBS	5 881 815	12 872
ONT	7 537 007	22 325

Table 1. Number of measured CpGs sites and epiline DMCs detected from WGBS and ONT methylation data. A CpG site refers to a dinucleotide CG.

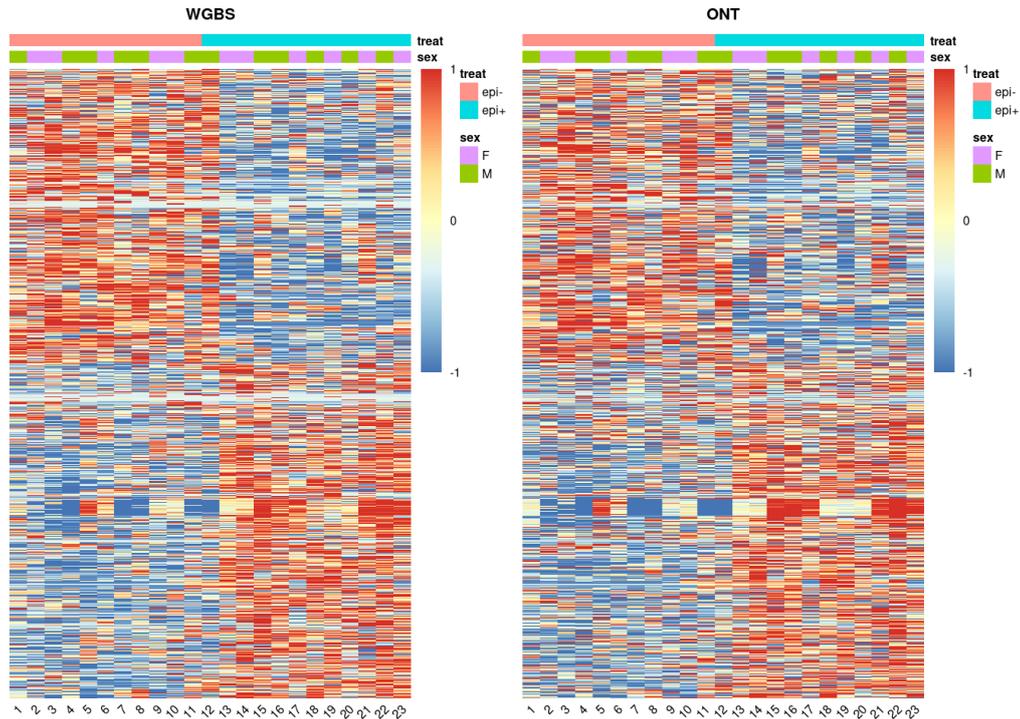


Figure 1. Heatmap of methylation rate of 10,000 common epilines DMCs detected from WGBS and ONT. The red and blue colors refer to the epi- and epi+ groups, respectively. The violet and green colors refer to the female (F) and male (M) quails, respectively. The diverging color gradient refers to a low (blue) or a high (red) methylation rate.

## Acknowledgements

This work was supported by FEDER (European Funding for Regional Economic Development), the PhD scholarship to CC from Région Occitanie and INRAE Animal Genetics Division. We are grateful to the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, <https://doi.org/10.15454/1.5572369328961167E12>) for providing computing and storage resources.

## References

- [1] C. Cerutti, S. Leroux, D. Gourichon, Y. Labrune, I. David, T. Zerjal, V. Coustham, G. Devailly, F. Pitel. *Short communication: Effects of in-ovo injection of endocrine disruptors and methyltransferase unhibitor on quail growth and egg-laying performances*. Vol. 16, pp. :1751-7311, 2022.
- [2] S. Leroux, D. Gourichon, C. Leterrier, Y. Labrune, V. Coustham, S. Rivière, T. Zerjal, J.L. Coville, M. Morisson, F. Minvielle, F. Pitel. *Embryonic environment and transgenerational effects in quail*. Vol. 49, pp. :1297-9686, 2017.

## Transgenerational epigenetics in quail: whole genome DNA methylation analysis

Chloé Cerutti<sup>1</sup>, Sophie Leroux<sup>1</sup>, Paul Terzian<sup>2</sup>, Joanna Lledo<sup>3</sup>, David Gourichon<sup>4</sup>, Jean-Noel Hubert<sup>1</sup>, Rémy-Félix Serre<sup>3</sup>, Céline Vandecasteele<sup>3</sup>, Christophe Klopp<sup>2</sup>, Carole Iampietro<sup>3</sup>, Christine Gaspin<sup>2</sup>, Denis Milan<sup>1</sup>, Cécile Donnadiou<sup>3</sup>, Frédérique Pitel<sup>1</sup> and Guillaume Devailly<sup>1</sup>

<sup>1</sup>GenPhySE, Université de Toulouse, INRAE, ENVT, 31326, Castanet Tolosan, France

<sup>2</sup>MIAT, UR875 INRAE, 31326, Castanet-Tolosan, France

<sup>3</sup>GeT-PlaGe - Génome et Transcriptome - Plateforme Génomique, US1426 GET - INRAE, 31326, Castanet-Tolosan, France

<sup>4</sup>PEAT, UE1295 - INRAE Centre Val de Loire, 37380, Nouzilly, France

Prenatal environment influence on adult phenotype development is partially mediated by epigenetics phenomena. Our previous study highlighted significant effects of the *in-ovo* injection of endocrine disruptors or a DNA methyltransferase inhibitor on quail development, significantly reducing their weight [1]. An increasing number of studies highlighted the transmission of epigenetics marks between generations following environmental exposure. However, there is much debate about their acquired transmission beyond the exposed generation. Recent studies revealed that non-genetics inheritance was probably present in avian species. In one of them [2], fertilized eggs were divided into two groups: one group injected with an endocrine disruptor, Genistein, and a non-injected control group. After three generations without any other injection, several traits were impacted by the ancestor treatment such as reproduction and behavioral traits. This pilot study highlighted the potential existence of transgenerational transmission of environmental effects in quails. To better understand the transgenerational transmission of these environmental effects, DNA methylation data were generated from blood samples of the third generation (WGBS and ONT). To analyse the DNA methylation state, we developed a bioinformatics pipeline in order to detect differential methylated cytosines (DMCs) affected by these transgenerational phenomena. We detected thousands of DMCs between both groups. A preliminary genetic analysis of the data suggests that most of these differentially methylated positions are related to genetic variability between both epilines, even if a mirror crossing was set up to reduce the genetic differences between both lines. In addition to these analyses, these data allowed us to perform a comparative analysis between the WGBS and ONT sequencing technologies for DNA methylation analysis.

### Acknowledgements

This work was funded by the French National Research Agency (ANR; EpiBird grant, ANR-09-GENM-004), FEDER-FSE MIDI-PYRENEES ET GARONNE 2014-2020 (SeqOccIn “Sequençage Occitanie Innovation”), Région Occitanie and INRAE Animal Genetics Division. We are grateful to the genotoul bioinformatics platform Toulouse Occitanie for providing computing and storage resources.

### References

[1] C. Cerutti, S. Leroux, D. Gourichon, Y. Labrune, I. David, T. Zerjal, V. Coustham, G. Devailly, F. Pitel. Short communication: *Effects of in-ovo injection of endocrine disruptors and methyltransferase inhibitor on quail growth and egg-laying performances*. Vol. 16, pp. :1751-7311, 2022.

[2] S. Leroux, D. Gourichon, C. Leterrier, Y. Labrune, V. Coustham, S. Rivière, T. Zerjal, J.L. Coville, M. Morisson, F. Minvielle, F. Pitel. *Embryonic environment and transgenerational effects in quail*. Vol. 49, pp. :1297-9686, 2017.

# Transgenerational epigenetics in quail: whole genome DNA methylation analysis

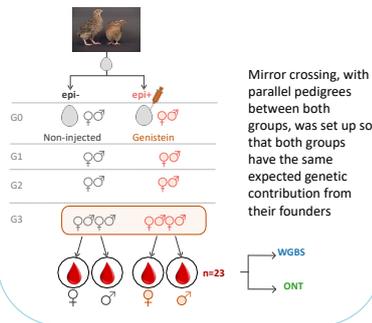
Chloé Cerutti<sup>1</sup>, Sophie Leroux<sup>1</sup>, Paul Terzian<sup>2</sup>, Joanna Lledo<sup>3</sup>, David Gourichon<sup>4</sup>, Jean-Noel Hubert<sup>1</sup>, Yann Labrune<sup>1</sup>, Rémy-Félix Serre<sup>3</sup>, Céline Vandecasteele<sup>3</sup>, Christophe Klopp<sup>2</sup>, Carole Iampietro<sup>3</sup>, Christine Gaspin<sup>2</sup>, Denis Milan<sup>1</sup>, Cécile Donnadieu<sup>3</sup>, Vincent Coustham<sup>5,6</sup>, Frédérique Pitel<sup>1</sup> and Guillaume Devailly<sup>1</sup>

<sup>1</sup> GenPhySE, Université de Toulouse, INRAE, ENVT, 31320 Castanet-Tolosan, France  
<sup>2</sup> MIAI, FF Bioinfo Genotoul, Université de Toulouse, INRAE, Chemin de Borde Rouge, 31320 Castanet-Tolosan, France  
<sup>3</sup> GeP/Race - Génome et Transcription - Plateforme Génétique, GeT - INRAE, 31320 Castanet-Tolosan, France  
<sup>4</sup> INRAE, PEAT, 37380 Nouzilly, France  
<sup>5</sup> INRAE, Université de Tours, B0A, 37380, Nouzilly, France  
<sup>6</sup> INRAE, Université de Pau et Pays de l'Adour, NuMeA, E2S UPPA, Aquaspl@e, 64310, Saint-Pée-sur-Nivelle, France

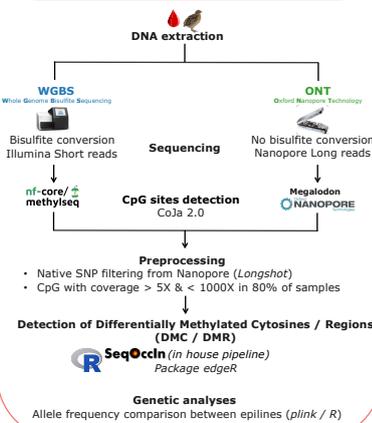
## Introduction

The influence of the prenatal environment on the adult phenotype development is partially mediated by epigenetic phenomena. An increasing number of studies highlight the transmission of epigenetic marks between generations following an environmental exposure. However, there is much debate about their acquired transmission beyond the exposed individuals. Recent studies revealed that non-genetic inheritance was probably present in avian species. In one of them, fertilized eggs were injected with an endocrine disruptor (genistein) and after three generations without any other injection, several traits were impacted by the ancestor treatment such as the reproduction and the behavior. Here we analyse the DNA methylation between control and treated lines of the third generation.

## Experimental design



## Methods



## Conclusion

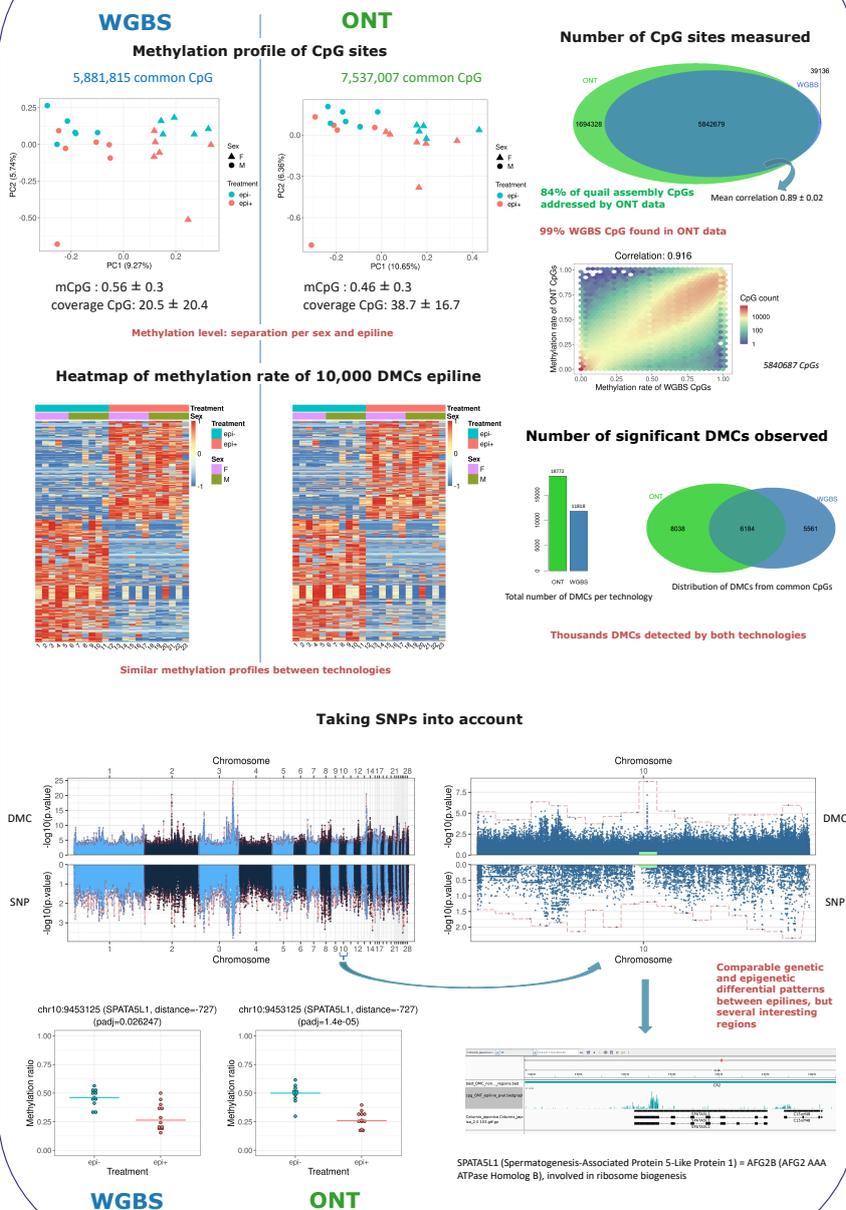
ONT vs WGBS: comparable results but best detection of differential methylation from ONT data

Thousands of DMCs detected between epilines: transgenerational effects of genistein on DNA methylation level

Most of the methylation level variability seems to be correlated with genetic variability

Several candidate regions will be further analysed

## Results





## 20. Continuité du stage de M2 : analyse des circRNA

Poster/Présentation - Cerutti et al., 2020 - *JOBIM* - **Circular RNAs detection and cross-species conservation.** hal-03703612.

Article - (Robic et al., 2021) - *Frontiers in Genetics* - **Comparative analysis of the circular transcriptome in muscle, liver, and testis in three livestock species.** DOI:10.3389/fgene.2021.665153.

Communication orale - Robic et al., 2021 - *EMBL Symposium* - **The lariat-derived circRNA from ATXN2L : an outstanding circRNA in pigs.** hal-03445759.

Poster - Robic et al., 2021 - *ISAG* - **The importance of annotations (reference genome and parent gene) for the study of circRNAs.** hal-03445792.

Article - (Robic et al., 2022) - *BBA Gene Regulatory Mechanisms* - **From the comparative study of a circRNA originating from an mammalian ATXN2L intron to understanding the genesis of intron lariat-derived circRNAs** DOI:10.1016/j.bbagr.2022.194815.



# VIII Références bibliographiques



## Références

- Acuna-Hidalgo, R., Veltman, J. A., & Hoischen, A. (2016, Nov). New insights into the generation and role of de novo mutations in health and disease. *Genome Biol*, *17*(1), 241.
- Albini, A., Rosano, C., Angelini, G., Amaro, A., Esposito, A., Maramotti, S., Noonan, D., & Pfeffer, U. (2014). Exogenous hormonal regulation in breast cancer cells by phytoestrogens and endocrine disruptors. *Current Medicinal Chemistry*, *21*(9), 1129–1145.
- Andersson, L., Archibald, A. L., Bottema, C. D., Brauning, R., Burgess, S. C., Burt, D. W., Casas, E., Cheng, H. H., Clarke, L., Coudrey, C., et al. (2015). Coordinated international action to accelerate genome-to-phenome with faang, the functional annotation of animal genomes project. *Genome biology*, *16*(1), 1–6.
- Andrews, S., et al. (2010). *Fastqc : a quality control tool for high throughput sequence data*. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.
- Anway, M. D., Cupp, A. S., Uzumcu, M., & Skinner, M. K. (2005). Epigenetic transgenerational actions of endocrine disruptors and male fertility. *science*, *308*(5727), 1466–1469.
- Anway, M. D., Leathers, C., & Skinner, M. K. (2006). Endocrine disruptor vinclozolin induced epigenetic transgenerational adult-onset disease. *Endocrinology*, *147*(12), 5515–5523.
- Anway, M. D., & Skinner, M. K. (2006). Epigenetic transgenerational actions of endocrine disruptors. *Endocrinology*, *147*(6), s43–s49.
- Arneson, A., Haghani, A., Thompson, M. J., Pellegrini, M., Kwon, S. B., Vu, H., Maciejewski, E., Yao, M., Li, C. Z., Lu, A. T., et al. (2022). A mammalian methylation array for profiling methylation levels at conserved sequences. *Nature communications*, *13*(1), 783.
- Attig, L., Gabory, A., & Junien, C. (2010). Early nutrition and epigenetic programming : chasing shadows. *Current Opinion in Clinical Nutrition & Metabolic Care*, *13*(3), 284–293.
- Bannister, A. J., & Kouzarides, T. (2011, Mar). Regulation of chromatin by histone modifications. *Cell Res*, *21*(3), 381–395.
- Bardou, P. (2022). *Coturnix japonica - VCF reference dataset*. Portail Data INRAE. doi: 10.15454/TKMI53
- Barlow, D. P. (2011). Genomic imprinting : a mammalian epigenetic discovery model. *Annual review of genetics*, *45*, 379–403.
- Barton, S. C., Surani, M., & Norris, M. (1984). Role of paternal and maternal genomes in mouse development. *Nature*, *311*(5984), 374–376.
- Barturen, G., Rueda, A., Oliver, J. L., & Hackenberg, M. (2013). Methylextract : high-quality methylation maps and snv calling from whole genome bisulfite sequencing data. *F1000Research*, *2*.
- Beaujean, N., Boutinaud, M., Devinoy, È., Jammes, H., Le Guillou, S., Le Provost, F., Leroux, C., Mobuchon, L., Pannetier, M., Sellem, E., et al. (2020). l'épigénétique et la construction du phénotype chez le bovin. *INRAE Productions Animales*, *33*(2), 109–124.
- Beck, D., Ben Maamar, M., & Skinner, M. K. (2021). Integration of sperm ncna-directed dna methylation and dna methylation-directed histone retention in epigenetic transgenerational inheritance. *Epigenetics & chromatin*, *14*(1), 1–14.
- Beck, D., Ben Maamar, M., & Skinner, M. K. (2022). Genome-wide cpg density and dna

- methylation analysis method (medip, rrbs, and wgbs) comparisons. *Epigenetics*, 17(5), 518–530.
- Becot, L., Bedere, N., & Le roy, P. (2021, juin). Sélection sur la ponte des poules en systèmes alternatifs à la cage : (full text available in english). *INRAE Productions Animales*, 34(1), 1–14. doi: 10.20870/productions-animales.2021.34.1.4680
- Bell, J. T., Pai, A. A., Pickrell, J. K., Gaffney, D. J., Pique-Regi, R., Degner, J. F., Gilad, Y., & Pritchard, J. K. (2011). Dna methylation patterns associate with genetic and gene expression variation in hapmap cell lines. *Genome biology*, 12, 1–13.
- Bergeron, L. A., Besenbacher, S., Zheng, J., Li, P., Bertelsen, M. F., Quintard, B., Hoffman, J. I., Li, Z., St Leger, J., Shao, C., Stiller, J., Gilbert, M. T. P., Schierup, M. H., & Zhang, G. (2023, Mar). Evolution of the germline mutation rate across vertebrates. *Nature*, 615(7951), 285–291.
- Bhandari, R., Vom Saal, F., & Tillitt, D. (2015). Transgenerational effects from early developmental exposures to bisphenol a or 17 $\alpha$ -ethinyloestradiol in medaka, oryzias latipes. sci rep 5 : 9303.
- Bhandari, R. K. (2016). Medaka as a model for studying environmentally induced epigenetic transgenerational inheritance of phenotypes. *Environmental epigenetics*, 2(1), dvv010.
- Bhardwaj, V., Heyne, S., Sikora, K., Rabbani, L., Rauer, M., Kilpert, F., Richter, A. S., Ryan, D. P., & Manke, T. (2019). snakepipes : facilitating flexible, scalable and integrative epigenomic analysis. *Bioinformatics*, 35(22), 4757–4759.
- Binder, A. M., & Michels, K. B. (2013, Dec). The causal effect of red blood cell folate on genome-wide methylation in cord blood : a Mendelian randomization approach. *BMC Bioinformatics*, 14, 353.
- Bioinfo, G. (2018). Genotoul bioinformatics facility. Consulté sur <https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html>
- Bird, A. (2002). Dna methylation patterns and epigenetic memory. *Genes & development*, 16(1), 6–21.
- Bird, A. P. (1986). CpG-rich islands and the function of dna methylation. *Nature*, 321(6067), 209–213.
- Bisoni, L., Batlle-Morera, L., Bird, A. P., Suzuki, M., & McQueen, H. A. (2005). Female-specific hyperacetylation of histone h4 in the chicken z chromosome. *Chromosome Research*, 13, 205–214.
- Biswas, S., & Rao, C. M. (2018, Oct). Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. *Eur J Pharmacol*, 837, 8–24.
- Bochtler, M., Kolano, A., & Xu, G. L. (2017, Jan). DNA demethylation pathways : Additional players and regulators. *Bioessays*, 39(1), 1–13.
- Bock, C. (2012). Analysing and interpreting dna methylation data. *Nature Reviews Genetics*, 13(10), 705–719.
- Boichard, D., GROHS, C., BOURGEOIS, F., CERQUEIRA, F., FAUGERAS, R., NEAU, A., MILAN, D., RUPP, R., AMIGUES, Y., BOSCHER, M., & LEVÉZIEL, H. (2000, déc.). La recherche de qtl à l'aide de marqueurs : résultats chez les bovins laitiers. *INRAE Productions*

- Animales, 13(HS), 217–222. doi: 10.20870/productions-animales.2000.13.HS.3841
- Bonefeld-Jorgensen, E. C., Long, M., Hofmeister, M. V., & Vinggaard, A. M. (2007). Endocrine-disrupting potential of bisphenol a, bisphenol a dimethacrylate, 4-n-nonylphenol, and 4-n-octylphenol in vitro : new data and a brief review. Environmental health perspectives, 115(Suppl 1), 69–76.
- Braunschweig, M., Jagannathan, V., Gutzwiller, A., & Bee, G. (2012). Investigations on transgenerational epigenetic response down the male line in f2 pigs. PloS one, 7(2), e30583.
- Breiling, A., & Lyko, F. (2015). Epigenetic regulatory functions of dna modifications : 5-methylcytosine and beyond. Epigenetics & chromatin, 8, 1–9.
- Brionne, A., Juanchich, A., & Hennequet-Antier, C. (2019). Viseago : a bioconductor package for clustering biological functions using gene ontology and semantic similarity. BioData mining, 12(1), 1–13.
- Brun, J.-M., Bernadet, M.-D., Cornuez, A., Leroux, S., Bodin, L., Basso, B., Davail, S., Jaglin, M., Lessire, M., Martin, X., et al. (2015). Influence of grand-mother diet on offspring performances through the male line in muscovy duck. BMC genetics, 16, 1–11.
- Bruner-Tran, K. L., & Osteen, K. G. (2011). Developmental exposure to tcdd reduces fertility and negatively affects pregnancy outcomes across multiple generations. Reproductive toxicology, 31(3), 344–350.
- Bure, I. V., Nemtsova, M. V., & Kuznetsova, E. B. (2022). Histone modifications and non-coding rnas : mutual epigenetic regulation and role in pathogenesis. International Journal of Molecular Sciences, 23(10), 5801.
- Cao-Lei, L., Elgbeili, G., Massart, R., Laplante, D. P., Szyf, M., & King, S. (2015). Pregnant women’s cognitive appraisal of a natural disaster affects dna methylation in their children 13 years later : Project ice storm. Translational psychiatry, 5(2), e515–e515.
- Carvalho, A. V., Hennequet-Antier, C., Rouger, R., Delaveau, J., Bordeau, T., Crochet, S., Couroussé, N., Pitel, F., Collin, A., & Coustham, V. (2023). Thermal conditioning of quail embryos has transgenerational and reversible long-term effects. Journal of Animal Science and Biotechnology, 14, 124.
- Casals-Casas, C., & Desvergne, B. (2011). Endocrine disruptors : from endocrine to metabolic disruption. Annual review of physiology, 73, 135–162.
- Castelli, M. A., Whiteley, S. L., Georges, A., & Holleley, C. E. (2020). Cellular calcium and redox regulation : the mediator of vertebrate environmental sex determination ? Biological Reviews, 95(3), 680–695.
- Cerutti, C., Leroux, S., Gourichon, D., Labrune, Y., David, I., Zerjal, T., Coustham, V., Devailly, G., & Pitel, F. (2022, Mar). Short communication : Effects of in-ovo injection of endocrine disruptors and methyltransferase inhibitor on quail growth and egg-laying performances. Animal, 16(3), 100464.
- Chantalat, S., Depaux, A., Héry, P., Barral, S., Thuret, J.-Y., Dimitrov, S., & Gérard, M. (2011). Histone h3 trimethylation at lysine 36 is associated with constitutive and facultative heterochromatin. Genome research, 21(9), 1426–1437.
- Charrier, M., Lumineau, S., Georgelin, M., Meurisse, M., Palme, R., Angelier, F., Cornilleau, F.,

- Constantin, P., Coustham, V., Nicolle, C., et al. (2022). Prenatal maternal stress is associated with behavioural and epigenetic changes in japanese quail. Psychoneuroendocrinology, 137, 105661.
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp : an ultra-fast all-in-one fastq preprocessor. Bioinformatics, 34(17), i884–i890.
- Chen, Y., Lun, A. T., & Smyth, G. K. (2016). From reads to genes to pathways : differential expression analysis of rna-seq experiments using rsubread and the edger quasi-likelihood pipeline. F1000Research, 5.
- Chen, Y., Pal, B., Visvader, J. E., & Smyth, G. K. (2017). Differential methylation analysis of reduced representation bisulfite sequencing experiments using edger. F1000Research, 6.
- Chen, Z., & Zhang, Y. (2020, Jun). Role of Mammalian DNA Methyltransferases in Development. Annu Rev Biochem, 89, 135–158.
- Cheung, P., & Lau, P. (2005, Mar). Epigenetic regulation by histone methylation and histone variants. Mol Endocrinol, 19(3), 563–573.
- Choudhry, H., & Catto, J. W. (2011). Epigenetic regulation of microrna expression in cancer. MicroRNA and Cancer : Methods and Protocols, 165–184.
- Christman, J. K. (2002). 5-azacytidine and 5-aza-2-deoxycytidine as inhibitors of dna methylation : mechanistic studies and their implications for cancer therapy. Oncogene, 21(35), 5483–5495.
- Chuang, J. C., & Jones, P. A. (2007). Epigenetics and micrnas. Pediatric research, 61(7), 24–29.
- Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M., & Zaret, K. S. (2002). Opening of compacted chromatin by early developmental transcription factors hnf3 (foxa) and gata-4. Molecular cell, 9(2), 279–289.
- Clark, E. L., Archibald, A. L., Daetwyler, H. D., Groenen, M. A., Harrison, P. W., Houston, R. D., Kühn, C., Lien, S., Macqueen, D. J., Reecy, J. M., et al. (2020). From faang to fork : application of highly annotated genomes to improve farmed animal production. Genome Biology, 21, 1–9.
- Class, Q. A., Abel, K. M., Khashan, A. S., Rickert, M. E., Dalman, C., Larsson, H., Hultman, C. M., Långström, N., Lichtenstein, P., & D’Onofrio, B. M. (2014). Offspring psychopathology following preconception, prenatal and postnatal maternal bereavement stress. Psychological medicine, 44(1), 71–84.
- Colborn, T. (1992). Chemically-induced alterations in sexual and functional development : the wildlife/human connection. Advances in modern environmental toxicology, 21.
- Colborn, T., vom Saal, F. S., & Soto, A. M. (1993, Oct). Developmental effects of endocrine-disrupting chemicals in wildlife and humans. Environ Health Perspect, 101(5), 378–384.
- Collins, L. J., Schönfeld, B., & Chen, X. S. (2011). The epigenetics of non-coding rna. Handbook of epigenetics, 49–61.
- Coustham, V., Li, P., Strange, A., Lister, C., Song, J., & Dean, C. (2012). Quantitative modulation of polycomb silencing underlies natural variation in vernalization. Science, 337(6094), 584–587.
- Cridge, A. G., Leask, M. P., Duncan, E. J., & Dearden, P. K. (2015, Mar). What do studies

- of insect polyphenisms tell us about nutritionally-triggered epigenomic changes and their consequences? *Nutrients*, 7(3), 1787–1797.
- Cropley, J. E., Suter, C. M., Beckman, K. B., & Martin, D. I. (2006). Germ-line epigenetic modification of the murine avy allele by nutritional supplementation. *Proceedings of the National Academy of Sciences*, 103(46), 17308–17312.
- Cruz, M. L., Wong, W. W., Mimouni, F., Hachey, D. L., Setchell, K. D., Klein, P. D., & Tsang, R. C. (1994, Feb). Effects of infant nutrition on cholesterol synthesis rates. *Pediatr Res*, 35(2), 135–140.
- Cubas, P., Vincent, C., & Coen, E. (1999, Sep). An epigenetic mutation responsible for natural variation in floral symmetry. *Nature*, 401(6749), 157–161.
- Daley, T., & Smith, A. D. (2013). Predicting the molecular complexity of sequencing libraries. *Nature methods*, 10(4), 325–327.
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., et al. (2021). Twelve years of samtools and bcftools. *Gigascience*, 10(2), giab008.
- David, I., Canario, L., Combes, S., & Demars, J. (2019a). Intergenerational transmission of characters through genetics, epigenetics, microbiota, and learning in livestock. *Frontiers in Genetics*, 10, 1058.
- David, I., Canario, L., Combes, S., & Demars, J. (2019b). Intergenerational Transmission of Characters Through Genetics, Epigenetics, Microbiota, and Learning in Livestock. *Front Genet*, 10, 1058.
- David, I., & Ricard, A. (2019). A unified model for inclusive inheritance in livestock species. *Genetics*, 212(4), 1075–1099.
- Deeb, N., & Cahaner, A. (2002). Genotype-by-environment interaction with broiler genotypes differing in growth rate. 3. growth rate and water consumption of broiler progeny from weight-selected versus nonselected parents under normal and high ambient temperatures. *Poultry Science*, 81(3), 293–301.
- De Koning, L., Corpet, A., Haber, J. E., & Almouzni, G. (2007). Histone chaperones : an escort network regulating histone traffic. *Nature structural & molecular biology*, 14(11), 997–1007.
- Delaney, C., Garg, S. K., & Yung, R. (2015). Analysis of dna methylation by pyrosequencing. *Immunosenescence : Methods and Protocols*, 249–264.
- Dias, B. G., & Ressler, K. J. (2014). Parental olfactory experience influences behavior and neural structure in subsequent generations. *Nature neuroscience*, 17(1), 89–96.
- Disteche, C. M. (2012). Dosage compensation of the sex chromosomes. *Annual review of genetics*, 46, 537–560.
- Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398), 376–380.
- Doherty, R., & Couldrey, C. (2014). Exploring genome wide bisulfite sequencing for dna methylation analysis in livestock : a technical assessment. *Frontiers in genetics*, 5, 126.

- Dolinoy, D. C., Huang, D., & Jirtle, R. L. (2007, Aug). Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. Proc Natl Acad Sci U S A, 104(32), 13056–13061.
- Dolinoy, D. C., Weidman, J. R., Waterland, R. A., & Jirtle, R. L. (2006, Apr). Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome. Environ Health Perspect, 114(4), 567–572.
- Doshi, T., Mehta, S. S., Dighe, V., Balasiner, N., & Vanage, G. (2011, Nov). Hypermethylation of estrogen receptor promoter region in adult testis of rats exposed neonatally to bisphenol A. Toxicology, 289(2-3), 74–82.
- Du, Q., Luu, P. L., Stirzaker, C., & Clark, S. J. (2015). Methyl-CpG-binding domain proteins : readers of the epigenome. Epigenomics, 7(6), 1051–1073.
- Ecsedi, S., Rodríguez-Aguilera, J. R., & Hernandez-Vargas, H. (2018). 5-hydroxymethylcytosine (5hmc), or how to identify your favorite cell. Epigenomes, 2(1). doi: 10.3390/epigenomes2010003
- Edge, P., & Bansal, V. (2019). Longshot enables accurate variant calling in diploid genomes from single-molecule long read sequencing. Nature communications, 10(1), 4660.
- Ellegren, H., Hultin-Rosenberg, L., Brunström, B., Dencker, L., Kultima, K., & Scholz, B. (2007). Faced with inequality : chicken do not have a general dosage compensation of sex-linked genes. BMC biology, 5(1), 1–12.
- Elobeid, M. A., & Allison, D. B. (2008). Putative environmental-endocrine disruptors and obesity : a review. Current opinion in endocrinology, diabetes, and obesity, 15(5), 403.
- Erdmann, R. M., & Picard, C. L. (2020, Oct). RNA-directed DNA Methylation. PLoS Genet, 16(10), e1009034.
- Escher, B. I., Tang, J., Poulsen, A., Leusch, F., Snyder, S., & Jia, A. (2014). Development of bioanalytical techniques to assess the potential human health impacts of recycled water. WateReuse Research Foundation.
- Etier, A., Dumetz, F., reau, S., & Ponts, N. (2022, Apr). Post-Translational Modifications of Histones Are Versatile Regulators of Fungal Development and Secondary Metabolism. Toxins (Basel), 14(5).
- Ewels, P., Hüther, P., Sateesh\_Per, Spix, N., Miller, E., nf-core bot, Hörtenhuber, M., Peltzer, A., F., S., Alneberg, J., Tommaso, P. D., Garcia, M. U., Davenport, C., Ajith, V., Krueger, F., Patel, H., Alessia, Devailly, G., Syme, R., Pearce, S., VIJAY, Menden, K., & Céline, N. (2023, juin). nf-core/methylseq : [2.4.0] gillespie gaia. Zenodo. doi: 10.5281/zenodo.8029942
- Ewels, P., Krueger, F., Käller, M., & Andrews, S. (2016). Cluster flow : A user-friendly bioinformatics workflow tool. F1000Research, 5.
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). Multiqc : summarize analysis results for multiple tools and samples in a single report. Bioinformatics, 32(19), 3047–3048.
- Ewels, P. A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M. U., Di Tommaso, P., & Nahnsen, S. (2020, 03). The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol, 38(3), 276–278.
- Fan, Y., Vilgalys, T. P., Sun, S., Peng, Q., Tung, J., & Zhou, X. (2019, Oct). IMAGE : high-

- powered detection of genetic effects on DNA methylation using integrated methylation QTL mapping and allele-specific analysis. Genome Biol, 20(1), 220.
- Farabollini, F., Porrini, S., & Dessì-Fulgheri, F. (1999). Perinatal exposure to the estrogenic pollutant bisphenol a affects behavior in male and female rats. Pharmacology Biochemistry and Behavior, 64(4), 687–694.
- Feeney, A., Nilsson, E., & Skinner, M. K. (2014). Epigenetics and transgenerational inheritance in domesticated farm animals. Journal of animal science and biotechnology, 5(1), 1–7.
- Feil, R., & Fraga, M. F. (2012, Jan). Epigenetics and the environment : emerging patterns and implications. Nat Rev Genet, 13(2), 97–109.
- Feng, H., Conneely, K. N., & Wu, H. (2014a). A bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. Nucleic acids research, 42(8), e69–e69.
- Feng, H., Conneely, K. N., & Wu, H. (2014b, Apr). A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. Nucleic Acids Res, 42(8), e69.
- Fève, K., Bourrec, T., & Devailly, G. (2022). Transfection adn ou arnm d'organoïdes intestinaux de porc pour une application en édition de l'épigénome. In Organoïdes et recherche agronomique.
- Fitz-James, M. H., & Cavalli, G. (2022). Molecular mechanisms of transgenerational epigenetic inheritance. Nature Reviews Genetics, 23(6), 325–341.
- Franke, A. A., Custer, L. J., & Tanaka, Y. (1998, Dec). Isoflavones in human breast milk and other biological fluids. Am J Clin Nutr, 68(6 Suppl), 1466S-1473S.
- Fresard, L., Leroux, S., Servin, B., Gourichon, D., Dehais, P., Cristobal, M. S., Marsaud, N., Vignoles, F., Bed'Hom, B., Coville, J.-L., et al. (2014). Transcriptome-wide investigation of genomic imprinting in chicken. Nucleic acids research, 42(6), 3768–3782.
- Frésard, L., Morisson, M., Brun, J.-M., Collin, A., Pain, B., Minvielle, F., & Pitel, F. (2013). Epigenetics and phenotypic variability : some interesting insights from birds. Genetics Selection Evolution, 45(1), 1–12.
- Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L., & Paul, C. L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual dna strands. Proceedings of the National Academy of Sciences, 89(5), 1827–1831.
- Frésard, L., Morisson, M., Brun, J. M., Collin, A., Pain, B., Minvielle, F., & Pitel, F. (2013, Jun). Epigenetics and phenotypic variability : some interesting insights from birds. Genet Sel Evol, 45(1), 16.
- Furrow, R. E., Christiansen, F. B., & Feldman, M. W. (2011). Environment-sensitive epigenetics and the heritability of complex diseases. Genetics, 189(4), 1377–1387.
- Gabory. (2008). Etude de la fonction de l'arn non codant h19 et de la régulation de l'empreinte parentale au locus h19/igf2 dans des modèles murins transgéniques.
- Gabory, A., Attig, L., & Junien, C. (2011a). Developmental programming and epigenetics. The American journal of clinical nutrition, 94(suppl\_6), 1943S–1952S.

- Gabory, A., Attig, L., & Junien, C. (2011b). Epigenetic mechanisms involved in developmental nutritional programming. *World journal of Diabetes*, *2*(10), 164.
- Gangaraju, V. K., & Lin, H. (2009). Micromnas : key regulators of stem cells. *Nature reviews Molecular cell biology*, *10*(2), 116–125.
- Gao, S., Zou, D., Mao, L., Liu, H., Song, P., Chen, Y., Zhao, S., Gao, C., Li, X., Gao, Z., et al. (2015). Bs-snp : Snp calling in bisulfite-seq data. *Bioinformatics*, *31*(24), 4006–4008.
- Gapp, K., Parada, G. E., Gross, F., Corcoba, A., Kaur, J., Grau, E., Hemberg, M., Bohacek, J., & Miska, E. A. (2021). Single paternal dexamethasone challenge programs offspring metabolism and reveals multiple candidates in rna-mediated inheritance. *Isience*, *24*(8).
- García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L. M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T. F., & Conesa, A. (2012). Qualimap : evaluating next-generation sequencing alignment data. *Bioinformatics*, *28*(20), 2678–2679.
- Gardiner-Garden, M., & Frommer, M. (1987). CpG islands in vertebrate genomes. *Journal of molecular biology*, *196*(2), 261–282.
- Gaucher, J., Reynoird, N., Montellier, E., Boussouar, F., Rousseaux, S., & Khochbin, S. (2010). From meiosis to postmeiotic events : the secrets of histone disappearance. *The FEBS journal*, *277*(3), 599–604.
- Georges, A., & Holleley, C. E. (2018). How does temperature determine sex ? *Science*, *360*(6389), 601–602.
- Gertz, J., Varley, K. E., Reddy, T. E., Bowling, K. M., Pauli, F., Parker, S. L., Kucera, K. S., Willard, H. F., & Myers, R. M. (2011). Analysis of dna methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation. *PLoS genetics*, *7*(8), e1002228.
- Geyer, H. J., Rimkus, G. G., Scheunert, I., Kaune, A., Schramm, K.-W., Kettrup, A., Zeeman, M., Muir, D. C., Hansen, L. G., & Mackay, D. (2000). Bioaccumulation and occurrence of endocrine-disrupting chemicals (edcs), persistent organic pollutants (pops), and other organic compounds in fish and other organisms including humans. *Bioaccumulation–New Aspects and Developments*, 1–166.
- Globisch, D., Münzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M., & Carell, T. (2010). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PloS one*, *5*(12), e15367.
- Goddard, M. E., & Whitelaw, E. (2014). The use of epigenetic phenomena for the improvement of sheep and cattle. *Frontiers in genetics*, *5*, 247.
- Golding, J., Gregory, S., Northstone, K., Pembrey, M., Watkins, S., Iles-Caven, Y., & Suderman, M. (2022). Human transgenerational observations of regular smoking before puberty on fat mass in grandchildren and great-grandchildren. *Scientific Reports*, *12*(1), 1139.
- Goldman, L. R., Newbold, R., & Swan, S. H. (2001, Nov). Exposure to soy-based formula in infancy. *JAMA*, *286*(19), 2402–2403.
- Goll, M. G., & Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annu Rev Biochem*, *74*, 481–514.
- Gong, T., Borgard, H., Zhang, Z., Chen, S., Gao, Z., & Deng, Y. (2022). Analysis and performance

- assessment of the whole genome bisulfite sequencing data workflow : currently available tools and a practical guide to advance dna methylation studies. *Small Methods*, *6*(3), 2101251.
- Gong, W., Pan, X., Xu, D., Ji, G., Wang, Y., Tian, Y., Cai, J., Li, J., Zhang, Z., & Yuan, X. (2022). Benchmarking dna methylation analysis of 14 alignment algorithms for whole genome bisulfite sequencing in mammals. *Computational and structural biotechnology journal*, *20*, 4704–4716.
- González-Recio, O., López-Catalina, A., Peiró-Pastor, R., Nieto-Valle, A., Castro, M., & Fernández, A. (2023). Evaluating the potential of (epi) genotype-by-low pass nanopore sequencing in dairy cattle : a study on direct genomic value and methylation analysis. *Journal of Animal Science and Biotechnology*, *14*(1), 98.
- González-Recio, O., Toro, M. A., & Bach, A. (2015). Past, present, and future of epigenetics applied to livestock breeding. *Frontiers in genetics*, *6*, 305.
- Gore, A. C. (2008). Neuroendocrine systems as targets for environmental endocrine-disrupting chemicals. *Fertility and sterility*, *89*(2), e101–e102.
- Gouil, Q., & Keniry, A. (2019). Latest techniques to study dna methylation. *Essays in biochemistry*, *63*(6), 639–648.
- Graña, O., López-Fernández, H., Fdez-Riverola, F., González Pisano, D., & Glez-Peña, D. (2018). Bicycle : a bioinformatics pipeline to analyze bisulfite sequencing data. *Bioinformatics*, *34*(8), 1414–1415.
- Grunau, C., Clark, S., & Rosenthal, A. (2001). Bisulfite genomic sequencing : systematic investigation of critical experimental parameters. *Nucleic acids research*, *29*(13), e65–e65.
- Grundberg, E., Meduri, E., Sandling, J. K., Hedman, Å. K., Keildson, S., Buil, A., Busche, S., Yuan, W., Nisbet, J., Sekowska, M., et al. (2013). Global analysis of dna methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *The American Journal of Human Genetics*, *93*(5), 876–890.
- Gu, H., Smith, Z. D., Bock, C., Boyle, P., Gnirke, A., & Meissner, A. (2011). Preparation of reduced representation bisulfite sequencing libraries for genome-scale dna methylation profiling. *Nature protocols*, *6*(4), 468–481.
- Guidotti, A., Ruzicka, W., Grayson, D. R., Veldic, M., Pinna, G., Davis, J. M., & Costa, E. (2007, Jan). S-adenosyl methionine and DNA methyltransferase-1 mRNA overexpression in psychosis. *Neuroreport*, *18*(1), 57–60.
- Guo, W., Zhu, P., Pellegrini, M., Zhang, M. Q., Wang, X., & Ni, Z. (2018). Cgmaptools improves the precision of heterozygous snv calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. *Bioinformatics*, *34*(3), 381–387.
- Halldin, K. (2005, Aug). Impact of endocrine disrupting chemicals on reproduction in Japanese quail. *Domest Anim Endocrinol*, *29*(2), 420–429.
- Halldin, K., Berg, C., Bergman, A., Brandt, I., & m, B. (2001, Dec). Distribution of bisphenol A and tetrabromobisphenol A in quail eggs, embryos and laying birds and studies on reproduction variables in adults following in ovo exposure. *Arch Toxicol*, *75*(10), 597–603.
- Han, Y., Zheleznyakova, G. Y., Marincevic-Zuniga, Y., Kakhki, M. P., Raine, A., Needhamsen, M., & Jagodic, M. (2022). Comparison of em-seq and pbat methylome library methods for

- low-input dna. Epigenetics, 17(10), 1195–1204.
- Hannenhalli, S., & Levy, S. (2001). Promoter prediction in the human genome. Bioinformatics, 17(suppl\_1), S90–S96.
- Hansen, K. D., Langmead, B., & Irizarry, R. A. (2012). Bsmooth : from whole genome bisulfite sequencing reads to differentially methylated regions. Genome biology, 13, 1–10.
- Harris, R. A., Wang, T., Coarfa, C., Nagarajan, R. P., Hong, C., Downey, S. L., Johnson, B. E., Fouse, S. D., Delaney, A., Zhao, Y., et al. (2010). Comparison of sequencing-based methods to profile dna methylation and identification of monoallelic epigenetic modifications. Nature biotechnology, 28(10), 1097–1105.
- Heard, E., & Martienssen, R. A. (2014, Mar). Transgenerational epigenetic inheritance : myths and mechanisms. Cell, 157(1), 95–109.
- Hearnshaw, H., Brown, J. M., Cumming, I. A., Goding, J. R., & Nairn, M. (1972, Jan). Endocrinological and histopathological aspects of the infertility in the ewe caused by oetrogenic clover. J Reprod Fertil, 28(1), 160–161.
- Heijmans, B. T., Tobi, E. W., Stein, A. D., Putter, H., Blauw, G. J., Susser, E. S., Slagboom, P. E., & Lumey, L. (2008). Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proceedings of the National Academy of Sciences, 105(44), 17046–17049.
- Helal, A. A., Saad, B. T., Saad, M. T., Mosaad, G. S., & Aboshanab, K. M. (2022). Evaluation of the available variant calling tools for oxford nanopore sequencing in breast cancer. Genes, 13(9), 1583.
- Hilakivi-Clarke, L., Cho, E., Onojafe, I., Raygada, M., & Clarke, R. (1999). Maternal exposure to genistein during pregnancy increases carcinogen-induced mammary tumorigenesis in female rat offspring. Oncol Rep, 6(5), 1089–1095.
- Hilakivi-Clarke, L., Onojafe, I., Raygada, M., Cho, E., Skaar, T., Russo, I., & Clarke, R. (1999, Aug). Prepubertal exposure to zearalenone or genistein reduces mammary tumorigenesis. Br J Cancer, 80(11), 1682–1688.
- Hiller-Sturmhöfel, S., & Bartke, A. (1998). The endocrine system : an overview. Alcohol health and research world, 22(3), 153.
- Hitchins, M. P. (2010). Inheritance of epigenetic aberrations (constitutional epimutations) in cancer susceptibility. Adv Genet, 70, 201–243.
- Ho, D. H., Reed, W. L., & Burggren, W. W. (2011). Egg yolk environment differentially influences physiological and morphological development of broiler and layer chicken embryos. Journal of Experimental Biology, 214(4), 619–628.
- Ho, S.-M., Tang, W.-Y., Belmonte de Frausto, J., & Prins, G. S. (2006). Developmental exposure to estradiol and bisphenol a increases susceptibility to prostate carcinogenesis and epigenetically regulates phosphodiesterase type 4 variant 4. Cancer research, 66(11), 5624–5632.
- Hollox, E. J., Zuccherato, L. W., & Tucci, S. (2022). Genome structural variation in human evolution. Trends in Genetics, 38(1), 45–58.
- Ho-Shing, O., & Dulac, C. (2019). Influences of genomic imprinting on brain function and

- behavior. *Current opinion in behavioral sciences*, 25, 66–76.
- Hou, Z., Sun, Q., Hu, Y., Yang, S., Zong, Y., & Zhao, R. (2018). Maternal betaine administration modulates hepatic type 1 iodothyronine deiodinase (dio1) expression in chicken offspring through epigenetic modifications. *Comparative Biochemistry and Physiology Part B : Biochemistry and Molecular Biology*, 218, 30–36.
- Howdeshell, K. L., Hotchkiss, A. K., Thayer, K. A., Vandenberg, J. G., & Vom Saal, F. S. (1999). Exposure to bisphenol a advances puberty. *Nature*, 401(6755), 763–764.
- Hu, Y., Feng, Y., Ding, Z., Lv, L., Sui, Y., Sun, Q., Abobaker, H., Cai, D., & Zhao, R. (2020). Maternal betaine supplementation decreases hepatic cholesterol deposition in chicken offspring with epigenetic modulation of sreb2 and cyp7a1 genes. *Poultry science*, 99(6), 3111–3120.
- Hubert, J.-N., & Demars, J. (2022). Genomic imprinting in the new omics era : a model for systems-level approaches. *Frontiers in Genetics*, 13, 838534.
- Hudson, N. O., & Buck-Koehntop, B. A. (2018, Oct). Zinc Finger Readers of Methylated DNA. *Molecules*, 23(10).
- Huo, W., Adams, H. M., Zhang, M. Q., & Palmer, K. L. (2015, Jun). Genome Modification in *Enterococcus faecalis* OG1RF Assessed by Bisulfite Sequencing and Single-Molecule Real-Time Sequencing. *J Bacteriol*, 197(11), 1939–1951.
- Hutchinson, J. N., Raj, T., Fagerness, J., Stahl, E., Vitoria, F. T., Gimelbrant, A., Seddon, J., Daly, M., Chess, A., & Plenge, R. (2014). Allele-specific methylation occurs at genetic variants associated with complex disease. *PLoS One*, 9(6), e98464.
- Illumina. (2022). *Sequencing Technology | Sequencing by synthesis*. Consulté le 2022-09-06, sur <https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html>
- Irwin, D. E. (2018). Sex chromosomes and speciation in birds and other zw systems. *Molecular Ecology*, 27(19), 3831–3851.
- Itoh, Y., Replogle, K., Kim, Y.-H., Wade, J., Clayton, D. F., & Arnold, A. P. (2010). Sex bias and dosage compensation in the zebra finch versus chicken genomes : general and specialized patterns among birds. *Genome research*, 20(4), 512–518.
- Jabbari, K., & Bernardi, G. (2004). Cytosine methylation and cpg, tpg (cpa) and tpa frequencies. *Gene*, 333, 143–149.
- Jahan, S., Xu, W., He, S., Gonzalez, C., Delcuve, G. P., & Davie, J. R. (2016). The chicken erythrocyte epigenome. *Epigenetics & Chromatin*, 9(1), 1–11.
- Janulaitis, A., Klimašauskas, S., Petrušyte, M., & Butkus, V. (1983). Cytosine modification in dna by bcn1 methylase yields n 4-methylcytosine. *FEBS letters*, 161(1), 131–134.
- Jeltsch, A. (2006). Molecular enzymology of mammalian dna methyltransferases. In W. Doerfler & P. Böhm (Eds.), *Dna methylation : Basic mechanisms* (pp. 203–225). Berlin, Heidelberg : Springer Berlin Heidelberg. doi: 10.1007/3-540-31390-7\_7
- Ji, Z., Huo, C., & Yang, P. (2020, Jun). Genistein inhibited the proliferation of kidney cancer cells via CDKN2a hypomethylation : role of abnormal apoptosis. *Int Urol Nephrol*, 52(6), 1049–1055.
- Jin, M. J., Wang, Z. L., Wu, Z. H., He, X. J., Zhang, Y., Huang, Q., Zhang, L. Z., Wu, X. B.,

- Yan, W. Y., & Zeng, Z. J. (2023). Phenotypic dimorphism between honeybee queen and worker is regulated by complicated epigenetic modifications. *Iscience*, *26*(4).
- Johannes, F., & Colomé-Tatché, M. (2011). Concerning epigenetics and inbreeding. *Nature Reviews Genetics*, *12*(5), 376–376.
- Jones, P. A. (2012, May). Functions of DNA methylation : islands, start sites, gene bodies and beyond. *Nat Rev Genet*, *13*(7), 484–492.
- Jung, Y. H., Wang, H.-L. V., Ruiz, D., Bixler, B. J., Linsenbaum, H., Xiang, J.-F., Forestier, S., Shafik, A. M., Jin, P., & Corces, V. G. (2022). Recruitment of ctf to an fto enhancer is responsible for transgenerational inheritance of bpa-induced obesity. *Proceedings of the National Academy of Sciences*, *119*(50), e2214988119.
- Kameswaran, V., & Kaestner, K. H. (2014). The missing lnc (rna) between the pancreatic  $\beta$ -cell and diabetes. *Frontiers in genetics*, *5*, 200.
- Kayang, B. B., Fillon, V., Inoue-Murayama, M., Miwa, M., Leroux, S., ve, K., Monvoisin, J. L., Pitel, F., Vignoles, M., Mouilhayrat, C., Beaumont, C., Ito, S., Minvielle, F., & Vignal, A. (2006, May). Integrated maps in quail (*Coturnix japonica*) confirm the high degree of synteny conservation with chicken (*Gallus gallus*) despite 35 million years of divergence. *BMC Genomics*, *7*, 101.
- Keller, H., & Wahli, W. (1997). Steroid hormone and related receptors. In *Principles of medical biology* (Vol. 10, pp. 255–296). Elsevier.
- Kerkel, K., Spadola, A., Yuan, E., Kosek, J., Jiang, L., Hod, E., Li, K., Murty, V. V., Schupf, N., Vilain, E., et al. (2008). Genomic surveys by methylation-sensitive snp analysis identify sequence-dependent allele-specific dna methylation. *Nature genetics*, *40*(7), 904–908.
- Khanal, J., Tayara, H., Zou, Q., & Chong, K. T. (2021). Identifying DNA N4-methylcytosine sites in the rosaceae genome with a deep learning model relying on distributed feature representation. *Comput Struct Biotechnol J*, *19*, 1612–1619.
- Kiparissis, Y., Balch, G. C., Metcalfe, T. L., & Metcalfe, C. D. (2003, Jul). Effects of the isoflavones genistein and equol on the gonadal development of Japanese medaka *Oryzias latipes*. *Environ Health Perspect*, *111*(9), 1158–1163.
- Knecht, A. L., Truong, L., Marvel, S. W., Reif, D. M., Garcia, A., Lu, C., Simonich, M. T., Teeguarden, J. G., & Tanguay, R. L. (2017). Transgenerational inheritance of neurobehavioral and physiological deficits from developmental exposure to benzo [a] pyrene in zebrafish. *Toxicology and applied pharmacology*, *329*, 148–157.
- Kornberg, R. D., & Lorch, Y. (1999, Aug). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell*, *98*(3), 285–294.
- Kretzmer, H., Otto, C., & Hoffmann, S. (2017). Bat : Bisulfite analysis toolkit : Bat is a toolkit to analyze dna methylation sequencing data accurately and reproducibly. it covers standard processing and analysis steps from raw read mapping up to annotation data integration and calculation of correlating dmrs. *F1000Research*, *6*.
- Krueger, F., & Andrews, S. R. (2011, avril). Bismark : a flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics*, *27*(11), 1571–1572. doi: 10.1093/bioinformatics/btr167

- Krueger, F., James, F., Ewels, P., Afyounian, E., & Schuster-Boeckler, B. (2021). Felixkrueger/-trimgalore : v0.6.7 - doi via zenodo. doi: 10.5281/ZENODO.5127899
- Kucharski, R., Maleszka, J., Foret, S., & Maleszka, R. (2008a). Nutritional control of reproductive status in honeybees via dna methylation. *Science*, *319*(5871), 1827–1830.
- Kucharski, R., Maleszka, J., Foret, S., & Maleszka, R. (2008b, Mar). Nutritional control of reproductive status in honeybees via DNA methylation. *Science*, *319*(5871), 1827–1830.
- Kuiper, G. G., Lemmen, J. G., Carlsson, B., Corton, J. C., Safe, S. H., Van Der Saag, P. T., Van Der Burg, B., & Gustafsson, J.-A. (1998). Interaction of estrogenic chemicals and phytoestrogens with estrogen receptor  $\beta$ . *Endocrinology*, *139*(10), 4252–4263.
- Kurdyukov, S., & Bullock, M. (2016). Dna methylation analysis : choosing the right method. *Biology*, *5*(1), 3.
- Kurosawa, T., Hiroi, H., Tsutsumi, O., Ishikawa, T., Osuga, Y., Fujiwara, T., Inoue, S., Muramatsu, M., Momoeda, M., & Taketani, Y. (2002). The activity of bisphenol a depends on both the estrogen receptor subtype and the cell type. *Endocrine Journal*, *49*(4), 465–471.
- Lamartiniere, C. A., Zhang, J. X., & Cotroneo, M. S. (1998, Dec). Genistein studies in rats : potential for breast cancer prevention and reproductive and developmental toxicity. *Am J Clin Nutr*, *68*(6 Suppl), 1400S-1405S.
- Lamb, H. J., Hayes, B. J., Randhawa, I. A., Nguyen, L. T., & Ross, E. M. (2021). Genomic prediction using low-coverage portable nanopore sequencing. *Plos one*, *16*(12), e0261274.
- Lämke, J., Brzezinka, K., Altmann, S., & Bäurle, I. (2016). A hit-and-run heat shock factor governs sustained histone methylation and transcriptional stress memory. *The EMBO journal*, *35*(2), 162–175.
- Lazar, N. H., Nevonen, K. A., O’Connell, B., McCann, C., O’Neill, R. J., Green, R. E., Meyer, T. J., Okhovat, M., & Carbone, L. (2018). Epigenetic maintenance of topological domains in the highly rearranged gibbon genome. *Genome research*, *28*(7), 983–997.
- Legault, L.-M., Bertrand-Lehouillier, V., & McGraw, S. (2018). Pre-implantation alcohol exposure and developmental programming of fasd : an epigenetic perspective. *Biochemistry and Cell Biology*, *96*(2), 117–130.
- Leroux, S., Gourichon, D., Leterrier, C., Labrune, Y., Coustham, V., Rivière, S., Zerjal, T., Coville, J. L., Morisson, M., Minvielle, F., & Pitel, F. (2017, 01). Embryonic environment and transgenerational effects in quail. *Genet Sel Evol*, *49*(1), 14.
- Li, C., Vagin, V. V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M. D., Syrzycka, M., Honda, B. M., et al. (2009). Collapse of germline pirnas in the absence of argonaute3 reveals somatic pirnas in flies. *Cell*, *137*(3), 509–521.
- Li, D., He, M., Tang, Q., Tian, S., Zhang, J., Li, Y., Wang, D., Jin, L., Ning, C., Zhu, W., Hu, S., Long, K., Ma, J., Liu, J., Zhang, Z., & Li, M. (2022, May). Comparative 3D genome architecture in vertebrates. *BMC Biol*, *20*(1), 99.
- Li, X., Peng, J., & Yi, C. (2021). The epitranscriptome of small non-coding rnas. *Non-coding RNA Research*, *6*(4), 167–173.
- Li, X., Xiong, X., & Yi, C. (2016, 12). Epitranscriptome sequencing technologies : decoding RNA modifications. *Nat Methods*, *14*(1), 23–31.

- Li, X., Xiong, X., & Yi, C. (2017). Epitranscriptome sequencing technologies : decoding rna modifications. *Nature methods*, *14*(1), 23–31.
- Liang, G., Chan, M. F., Tomigahara, Y., Tsai, Y. C., Gonzales, F. A., Li, E., Laird, P. W., & Jones, P. A. (2002). Cooperativity between dna methyltransferases in the maintenance methylation of repetitive elements. *Molecular and cellular biology*, *22*(2), 480–491.
- Lin, A., Du, Y., & Xiao, W. (2020, Aug). Yeast chromatin remodeling complexes and their roles in transcription. *Curr Genet*, *66*(4), 657–670.
- Lindner, M., Gawehns, F., Te Molder, S., Visser, M. E., van Oers, K., & Laine, V. N. (2022). Performance of methods to detect genetic variants from bisulphite sequencing data in a non-model species. *Molecular Ecology Resources*, *22*(2), 834–846.
- Lister, R., Pelizzola, M., Downen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., et al. (2009a). Human dna methylomes at base resolution show widespread epigenomic differences. *nature*, *462*(7271), 315–322.
- Lister, R., Pelizzola, M., Downen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q. M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., & Ecker, J. R. (2009b, Nov). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, *462*(7271), 315–322.
- Liu, D., Diorio, J., Tannenbaum, B., Caldji, C., Francis, D., Freedman, A., Sharma, S., Pearson, D., Plotsky, P. M., & Meaney, M. J. (1997). Maternal care, hippocampal glucocorticoid receptors, and hypothalamic-pituitary-adrenal responses to stress. *Science*, *277*(5332), 1659–1662.
- Liu, X. S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R. A., & Jaenisch, R. (2016). Editing dna methylation in the mammalian genome. *Cell*, *167*(1), 233–247.
- Liu, Y., Rosikiewicz, W., Pan, Z., Jillette, N., Wang, P., Taghbalout, A., Fook, J., Mason, C., Carroll, M., Cheng, A., et al. (2021). Dna methylation-calling tools for oxford nanopore sequencing : a survey and human epigenome-wide evaluation. *Genome biology*, *22*(1), 1–33.
- Liu, Y., Siegmund, K. D., Laird, P. W., & Berman, B. P. (2012). Bis-snp : combined dna methylation and snp calling for bisulfite-seq data. *Genome biology*, *13*(7), 1–14.
- Lokk, K., Modhukur, V., Rajashekar, B., Märtens, K., Mägi, R., Kolde, R., Koltšina, M., Nilsson, T. K., Vilo, J., Salumets, A., et al. (2014). Dna methylome profiling of human tissues identifies global and tissue-specific methylation patterns. *Genome biology*, *15*, 1–14.
- Lu, H., Giordano, F., & Ning, Z. (2016, Oct). Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics Proteomics Bioinformatics*, *14*(5), 265–279.
- Lv, Z., Fan, H., Zhang, B., Ning, C., Xing, K., & Guo, Y. (2018, Aug). Dietary genistein supplementation in laying broiler breeder hens alters the development and metabolism of offspring embryos as revealed by hepatic transcriptome analysis. *FASEB J*, *32*(8), 4214–4228.
- Lv, Z., Fan, H., Zhang, B., Xing, K., & Guo, Y. (2018, Mar). Dietary genistein supplementation for breeders and their offspring improves the growth performance and immune function of

- broilers. *Sci Rep*, *8*(1), 5161.
- Lyko, F. (2018, Feb). The DNA methyltransferase family : a versatile toolkit for epigenetic regulation. *Nat Rev Genet*, *19*(2), 81–92.
- Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., & Maleszka, R. (2011). Correction : The honey bee epigenomes : Differential methylation of brain dna in queens and workers. *PLoS biology*, *9*(1).
- Maffini, M. V., Rubin, B. S., Sonnenschein, C., & Soto, A. M. (2006). Endocrine disruptors and reproductive health : the case of bisphenol-a. *Molecular and cellular endocrinology*, *254*, 179–186.
- Majid, S., Dar, A. A., Shahryari, V., Hirata, H., Ahmad, A., Saini, S., Tanaka, Y., Dahiya, A. V., & Dahiya, R. (2010). Genistein reverses hypermethylation and induces active histone modifications in tumor suppressor gene b-cell translocation gene 3 in prostate cancer. *Cancer : Interdisciplinary International Journal of the American Cancer Society*, *116*(1), 66–76.
- Manders, F., van Boxtel, R., & Middelkamp, S. (2021). The Dynamics of Somatic Mutagenesis During Life in Humans. *Front Aging*, *2*, 802407.
- Mardis, E. R. (2008). The impact of next-generation sequencing technology on genetics. *Trends in genetics*, *24*(3), 133–141.
- Markey, C. M., Wadia, P. R., Rubin, B. S., Sonnenschein, C., & Soto, A. M. (2005). Long-term effects of fetal exposure to low doses of the xenoestrogen bisphenol-a in the female mouse genital tract. *Biology of reproduction*, *72*(6), 1344–1351.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, *17*(1), 10–12.
- Martincorena, I., & Campbell, P. J. (2015). Somatic mutation in cancer and normal cells. *Science*, *349*(6255), 1483–1489.
- Martin-Herranz, D. E., Aref-Eshghi, E., Bonder, M. J., Stubbs, T. M., Choufani, S., Weksberg, R., Stegle, O., Sadikovic, B., Reik, W., & Thornton, J. M. (2019). Screening for genes that accelerate the epigenetic aging clock in humans reveals a role for the h3k36 methyltransferase nsd1. *Genome biology*, *20*, 1–19.
- Mazzio, E. A., & Soliman, K. F. (2012, Feb). Basic concepts of epigenetics : impact of environmental signals on gene expression. *Epigenetics*, *7*(2), 119–130.
- McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor rna-seq experiments with respect to biological variation. *Nucleic acids research*, *40*(10), 4288–4297.
- McGowan, P. O., Sasaki, A., D'alessio, A. C., Dymov, S., Labonté, B., Szyf, M., Turecki, G., & Meaney, M. J. (2009). Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nature neuroscience*, *12*(3), 342–348.
- Meaney, M. J., & Szyf, M. (2022). Environmental programming of stress responses through dna methylation : life at the interface between a dynamic environment and a fixed genome. *Dialogues in clinical neuroscience*.
- Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S., & Jaenisch, R. (2005a).

- Reduced representation bisulfite sequencing for comparative high-resolution dna methylation analysis. Nucleic acids research, 33(18), 5868–5877.
- Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S., & Jaenisch, R. (2005b). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res, 33(18), 5868–5877.
- Merkel, A., Fernández-Callejo, M., Casals, E., Marco-Sola, S., Schuyler, R., Gut, I. G., & Heath, S. C. (2019). gembs : high throughput processing for dna methylation data from bisulfite sequencing. Bioinformatics, 35(5), 737–742.
- Merkel, A., Fernández-Callejo, M., Casals, E., Marco-Sola, S., Schuyler, R., Gut, I. G., & Heath, S. C. (2019, 03). gemBS : high throughput processing for DNA methylation data from bisulfite sequencing. Bioinformatics, 35(5), 737–742.
- Mersch, M., David, S.-A., Carvalho, A. V., Foissac, S., Collin, A., Pitel, F., & Coustham, V. (2018). Apports du séquençage haut-débit sur la connaissance de l'épigénome aviaire. INRA Productions Animales, 31(4), 325–336.
- Mills, A. D., & Faure, J. M. (1991, Mar). Divergent selection for duration of tonic immobility and social reinstatement behavior in Japanese quail (*Coturnix coturnix japonica*) chicks. J Comp Psychol, 105(1), 25–38.
- Minvielle, F. (2004). The future of japanese quail for research and production. World's Poultry Science Journal, 60(4), 500–507.
- Minvielle, M. J. C. J. F. A. M. Y., B.F. (1999, Oct). Changes in heterosis under within-line selection or reciprocal recurrent selection : an experiment on early egg production in Japanese quail. Journal of Animal Breeding and Genetics, 116(5), 363–377.
- Miodini, P., Fioravanti, L., Di Fronzo, G., & Cappelletti, V. (1999, Jun). The two phyto-oestrogens genistein and quercetin exert different effects on oestrogen receptor function. Br J Cancer, 80(8), 1150–1155.
- Miska, E. A., & Ferguson-Smith, A. C. (2016a). Transgenerational inheritance : Models and mechanisms of non-dna sequence-based inheritance. Science, 354(6308), 59–63.
- Miska, E. A., & Ferguson-Smith, A. C. (2016b, Oct). Transgenerational inheritance : Models and mechanisms of non-DNA sequence-based inheritance. Science, 354(6308), 59–63.
- Mizoguchi, B., Topping, N. E., Lavin, A. M., & Valenzuela, N. (2022). Cadmium ecotoxic effects on embryonic dmrt1 and aromatase expression in chrysemys picta turtles may implicate changes in dna methylation. Genes, 13(8), 1318.
- Monk, D. (2015). Germline-derived dna methylation and early embryo epigenetic reprogramming : the selected survival of imprints. The international journal of biochemistry & cell biology, 67, 128–138.
- Moore, L. D., Le, T., & Fan, G. (2013, Jan). DNA methylation and its basic function. Neuropsychopharmacology, 38(1), 23–38.
- Morel, L., Harrison, P., & Devailly, G. (2022). Vizfada : Visualisations of faang data. In Faang workshop 2022.
- Morgan, H. D., Sutherland, H. G., Martin, D. I., & Whitelaw, E. (1999, Nov). Epigenetic inheritance at the agouti locus in the mouse. Nat Genet, 23(3), 314–318.

- Morris, K. M., Hindle, M. M., Boitard, S., Burt, D. W., Danner, A. F., Eory, L., Forrest, H. L., Gourichon, D., Gros, J., Hillier, L. W., et al. (2020). The quail genome : insights into social behaviour, seasonal biology and infectious disease response. *BMC biology*, *18*(1), 1–18.
- Muñoz-de Toro, M., Markey, C. M., Wadia, P. R., Luque, E. H., Rubin, B. S., Sonnenschein, C., & Soto, A. M. (2005). Perinatal exposure to bisphenol-a alters peripubertal mammary gland development in mice. *Endocrinology*, *146*(9), 4138–4147.
- Nakamura, K., Itoh, K., Yaoi, T., Fujiwara, Y., Sugimoto, T., & Fushiki, S. (2006). Murine neocortical histogenesis is perturbed by prenatal exposure to low doses of bisphenol a. *Journal of neuroscience research*, *84*(6), 1197–1205.
- Newbold, R. R., Banks, E. P., Bullock, B., & Jefferson, W. N. (2001, Jun). Uterine adenocarcinoma in mice treated neonatally with genistein. *Cancer Res*, *61*(11), 4325–4328.
- Ni, P., Huang, N., Zhang, Z., Wang, D.-P., Liang, F., Miao, Y., Xiao, C.-L., Luo, F., & Wang, J. (2019). DeepSignal : detecting dna methylation state from nanopore sequencing reads using deep-learning. *Bioinformatics*, *35*(22), 4586–4595.
- Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., Van Berkum, N. L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the x-inactivation centre. *Nature*, *485*(7398), 381–385.
- North, K., & Golding, J. (2000, Jan). A maternal vegetarian diet in pregnancy is associated with hypospadias. The ALSPAC Study Team. Avon Longitudinal Study of Pregnancy and Childhood. *BJU Int*, *85*(1), 107–113.
- Nunn, A., Otto, C., Fasold, M., Stadler, P. F., & Langenberger, D. (2022). Manipulating base quality scores enables variant calling from bisulfite sequencing alignments using conventional bayesian approaches. *BMC genomics*, *23*(1), 477.
- Olova, N., Krueger, F., Andrews, S., Oxley, D., Berrens, R. V., Branco, M. R., & Reik, W. (2018). Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting dna methylation data. *Genome biology*, *19*(1), 1–19.
- Onn Wong, K., Woon Leo, L., & Leng Seah, H. (2005). Dietary exposure assessment of infants to bisphenol a from the use of polycarbonate baby milk bottles. *Food additives and contaminants*, *22*(3), 280–288.
- Osakabe, A., Tachiwana, H., Takaku, M., Hori, T., Obuse, C., Kimura, H., Fukagawa, T., & Kurumizaka, H. (2013). Vertebrate spt2 is a novel nucleolar histone chaperone that assists in ribosomal dna transcription. *Journal of cell science*, *126*(6), 1323–1332.
- Osberg, B. O., Ronen, J. R., Godschan, A. G., Uyar, B. U., Franke, V. F., Akalin, A. A., Wreczycka, K. W., & Wurmus, R. W. (2018). Pigx : reproducible genomics analysis pipelines with gnu guix.
- Osgood, M. P. (1994). X-chromosome inactivation : the case of the calico cat. *American Journal of Pharmaceutical Education*, *58*(2), 204–205.
- Oxford Nanopore Technologies. (2021). *Megalodon*. Consulté le 2021-09-19, sur <https://nanoporetech.github.io/megalodon>.
- Panchenko, P. E., Lemaire, M., Fneich, S., Voisin, S., Jouin, M., Junien, C., & Gabory, A. (2015). Epigénétique et nutrition : impacts de l'alimentation maternelle sur le développement

- placentaire et la santé de la descendance. *Biologie Aujourd'hui*, *209*(2), 175–187.
- Park, Y., & Wu, H. (2016, 05). Differential methylation analysis for BS-seq data under general experimental design. *Bioinformatics*, *32*(10), 1446–1453.
- Patisaul, H. B., Fortino, A. E., & Polston, E. K. (2006). Neonatal genistein or bisphenol-a exposure alters sexual differentiation of the avpv. *Neurotoxicology and teratology*, *28*(1), 111–118.
- Payne, A., Holmes, N., Rakyan, V., & Loose, M. (2019). Bulkvis : a graphical viewer for oxford nanopore bulk fast5 files. *Bioinformatics*, *35*(13), 2193–2198.
- Perez, M. F., & Lehner, B. (2019). Intergenerational and transgenerational epigenetic inheritance in animals. *Nature cell biology*, *21*(2), 143–151.
- Peters, T. J., Buckley, M. J., Chen, Y., Smyth, G. K., Goodnow, C. C., & Clark, S. J. (2021). Calling differentially methylated regions from whole genome bisulphite sequencing with dmrcate. *Nucleic Acids Research*, *49*(19), e109–e109.
- Piao, Y., Xu, W., Park, K. H., Ryu, K. H., & Xiang, R. (2021). Comprehensive evaluation of differential methylation analysis methods for bisulfite sequencing data. *International Journal of Environmental Research and Public Health*, *18*(15), 7975.
- Pidsley, R., Fernandes, C., Viana, J., Paya-Cano, J. L., Liu, L., Smith, R. G., Schalkwyk, L. C., & Mill, J. (2012). Dna methylation at the igf2/h19 imprinting control region is associated with cerebellum mass in outbred mice. *Molecular brain*, *5*, 1–9.
- Pidsley, R., Y Wong, C. C., Volta, M., Lunnon, K., Mill, J., & Schalkwyk, L. C. (2013). A data-driven approach to preprocessing illumina 450k methylation array data. *BMC genomics*, *14*(1), 1–10.
- Pidsley, R., Zotenko, E., Peters, T. J., Lawrence, M. G., Risbridger, G. P., Molloy, P., Van Dijk, S., Muhlhausler, B., Stirzaker, C., & Clark, S. J. (2016). Critical evaluation of the illumina methylationepic beadchip microarray for whole-genome dna methylation profiling. *Genome biology*, *17*(1), 1–17.
- Piestun, Y., Shinder, D., Ruzal, M., Halevy, O., Brake, J., & Yahav, S. (2008). Thermal manipulations during broiler embryogenesis : effect on the acquisition of thermotolerance. *Poultry science*, *87*(8), 1516–1525.
- Pihlajamaa, P., Zhang, F.-P., Saarinen, L., Mikkonen, L., Hautaniemi, S., & Jänne, O. A. (2011). The phytoestrogen genistein is a tissue-specific androgen receptor modulator. *Endocrinology*, *152*(11), 4395–4405.
- Prins, G. S., Tang, W.-Y., Belmonte, J., & Ho, S.-M. (2008). Perinatal exposure to oestradiol and bisphenol a alters the prostate epigenome and increases susceptibility to carcinogenesis. *Basic & clinical pharmacology & toxicology*, *102*(2), 134–138.
- Probst, A. V., Dunleavy, E., & Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. *Nature reviews Molecular cell biology*, *10*(3), 192–206.
- Pukkila, P. J., Peterson, J., Herman, G., Modrich, P., & Meselson, M. (1983, Aug). Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in *Escherichia coli*. *Genetics*, *104*(4), 571–582.
- Qin, T., Zhang, X., Guo, T., Yang, T., Gao, Y., Hao, W., & Xiao, X. (2020). Epigenetic Alteration

- Shaped by the Environmental Chemical Bisphenol A. *Front Genet*, 11, 618966.
- Quina, A., Buschbeck, M., & Di Croce, L. (2006). Chromatin structure and epigenetics. *Biochemical pharmacology*, 72(11), 1563–1569.
- Raddatz, G., Arsenault, R. J., Aylward, B., Whelan, R., hl, F., & Lyko, F. (2021, Jan). A chicken DNA methylation clock for the prediction of broiler health. *Commun Biol*, 4(1), 76.
- Rakyan, V. K., Down, T. A., Balding, D. J., & Beck, S. (2011). Epigenome-wide association studies for common human diseases. *Nature Reviews Genetics*, 12(8), 529–541.
- Rand, A. C., Jain, M., Eizenga, J. M., Musselman-Brown, A., Olsen, H. E., Akeson, M., & Paten, B. (2017). Mapping dna methylation with high-throughput nanopore sequencing. *Nature methods*, 14(4), 411–413.
- Rasmussen, K. D., & Helin, K. (2016, Apr). Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev*, 30(7), 733–750.
- Reed, W. L., & Clark, M. E. (2011). Beyond maternal effects in birds : responses of the embryo to the environment. *Integrative and Comparative Biology*, 51(1), 73–80.
- Reik, W., Dean, W., & Walter, J. (2001). Epigenetic reprogramming in mammalian development. *Science*, 293(5532), 1089–1093.
- Rice, P., Longden, I., & Bleasby, A. (2000). Emboss : the european molecular biology open software suite. *Trends in genetics*, 16(6), 276–277.
- Riggs, M. R. R. V., A.D. (1996, Feb). Epigenetic mechanisms of gene regulation. *Cold Spring Harbor Laboratory Press*, 7(2), 119–130.
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic acids research*, 43(7), e47–e47.
- Rivard, G. E., Momparler, R. L., Demers, J., Benoit, P., Raymond, R., Lin, K.-T., & Momparler, L. F. (1981). Phase i study on 5-aza-2-deoxycytidine in children with acute leukemia. *Leukemia research*, 5(6), 453–462.
- Robic, A., Cerutti, C., Demars, J., & Kühn, C. (2022). From the comparative study of a circrna originating from an mammalian atxn2l intron to understanding the genesis of intron lariat-derived circrnas. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1865(4), 194815.
- Robic, A., Cerutti, C., Kühn, C., & Faraut, T. (2021). Comparative analysis of the circular transcriptome in muscle, liver, and testis in three livestock species. *Frontiers in Genetics*, 12, 665153.
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR : a bioconductor package for differential expression analysis of digital gene expression data. *bioinformatics*, 26(1), 139–140.
- Sacks, D., Baxter, B., Campbell, B. C. V., Carpenter, J. S., Cognard, C., Dippel, D., Eesa, M., Fischer, U., Hausegger, K., Hirsch, J. A., Shazam Hussain, M., Jansen, O., Jayaraman, M. V., Khalessi, A. A., Kluck, B. W., Lavine, S., Meyers, P. M., Ramee, S., fenacht, D. A., Schirmer, C. M., & Vorwerk, D. (2018, Aug). Multisociety Consensus Quality Improvement Revised Consensus Statement for Endovascular Therapy of Acute Ischemic Stroke. *Int J*

- Stroke*, 13(6), 612–632.
- Santé-Publique-France. (2022). *Perturbateurs endocriniens*. Consulté le 2022-02-28, sur <https://www.santepubliquefrance.fr/determinants-de-sante/exposition-a-des-substances-chimiques/perturbateurs-endocriniens/publications>
- Sarda, S., & Hannenhalli, S. (2018). Orphan cpg islands as alternative promoters. *Transcription*, 9(3), 171–176.
- Scanes, C. G., & Dridi, S. (2021). *Sturkie's avian physiology*. Academic Press.
- Schatz, M. C. (2017). Nanopore sequencing meets epigenetics. *Nature methods*, 14(4), 347–348.
- Seralini, G.-E., & Jungers, G. (2021). Endocrine disruptors also function as nervous disruptors and can be renamed endocrine and nervous disruptors (ends). *Toxicology Reports*, 8, 1538–1557.
- Setchell, K. D. (2001). Soy isoflavones—benefits and risks from nature's selective estrogen receptor modulators (serms). *Journal of the American College of Nutrition*, 20(sup5), 354S–362S.
- Setchell, K. D., Gosselin, S. J., Welsh, M. B., Johnston, J. O., Balistreri, W. F., Kramer, L. W., Dresser, B. L., & Tarr, M. J. (1987, Aug). Dietary estrogens—a probable cause of infertility and liver disease in captive cheetahs. *Gastroenterology*, 93(2), 225–233.
- Settar, P., Yalcin, S., Turkmut, L., Ozkan, S., & Cahanar, A. (1999). Season by genotype interaction related to broiler growth rate and heat tolerance. *Poultry science*, 78(10), 1353–1358.
- Shen, L., & Zhang, Y. (2012). Enzymatic analysis of Tet proteins : key enzymes in the metabolism of DNA methylation. *Methods Enzymol*, 512, 93–105.
- Shi, Y. Y., Huang, Z. Y., Zeng, Z. J., Wang, Z. L., Wu, X. B., & Yan, W. Y. (2011). Diet and cell size both affect queen-worker differentiation through dna methylation in honey bees (*apis mellifera*, *apidae*). *PloS one*, 6(4), e18808.
- Shimbo, T., & Wade, P. A. (2016). Proteins That Read DNA Methylation. *Adv Exp Med Biol*, 945, 303–320.
- Shojaei Saadi, H. A., O'Doherty, A. M., Gagné, D., Fournier, É., Grant, J. R., Sirard, M.-A., & Robert, C. (2014). An integrated platform for bovine dna methylome analysis suitable for small samples. *BMC genomics*, 15(1), 1–18.
- Silva, L. B. A. R., Pinheiro-Castro, N., Novaes, G. M., Pascoal, G. F. L., & Ong, T. P. (2019, 11). Bioactive food compounds, epigenetics and chronic disease prevention : Focus on early-life interventions with polyphenols. *Food Res Int*, 125, 108646.
- Simpson, J. T., Workman, R. E., Zuzarte, P., David, M., Dursi, L., & Timp, W. (2017). Detecting dna cytosine methylation using nanopore sequencing. *Nature methods*, 14(4), 407–410.
- Sohoni, P., & Sumpter, J. (1998). Several environmental oestrogens are also anti-androgens. *Journal of endocrinology*, 158(3), 327–340.
- Song, X., Zhou, X., Yang, F., Liang, H., Wang, Z., Li, R., Miao, M., & Yuan, W. (2020, Nov). Association between prenatal bisphenol a exposure and promoter hypermethylation of CAPS2, TNFRSF25, and HKR1 genes in cord blood. *Environ Res*, 190, 109996.
- Stepper, P., Kungulovski, G., Jurkowska, R. Z., Chandra, T., Krueger, F., Reinhardt, R., Reik, W., Jeltsch, A., & Jurkowski, T. P. (2017). Efficient targeted dna methylation with chimeric

- dcas9-dnmt3a-dnmt3l methyltransferase. *Nucleic acids research*, 45(4), 1703–1713.
- Strom, B. L., Schinnar, R., Ziegler, E. E., Barnhart, K. T., Sammel, M. D., Macones, G. A., Stallings, V. A., Drulis, J. M., Nelson, S. E., & Hanson, S. A. (2001, Aug). Exposure to soy-based formula in infancy and endocrinological and reproductive outcomes in young adulthood. *JAMA*, 286(7), 807–814.
- Sun, D., Xi, Y., Rodriguez, B., Park, H. J., Tong, P., Meong, M., Goodell, M. A., & Li, W. (2014). Moabs : model based analysis of bisulfite sequencing data. *Genome biology*, 15(2), 1–12.
- Szyf, M. (2015, Feb). Nongenetic inheritance and transgenerational epigenetics. *Trends Mol Med*, 21(2), 134–144.
- Ségurel, L., Wyman, M. J., & Przeworski, M. (2014). Determinants of mutation rate variation in the human germline. *Annu Rev Genomics Hum Genet*, 15, 47–70.
- Tadić, V., Josipović, G., Zoldoš, V., & Vojta, A. (2019). Crispr/cas9-based epigenome editing : An overview of dcas9-based tools with special emphasis on off-target activity. *Methods*, 164, 109–119.
- Tagu, D., & Jaubert-Possamai, S. (2018). *Principes des techniques de biologie moléculaire et génomique : 3e édition revue et augmentée*. Quae.
- Takeda, K., Kobayashi, E., Akagi, S., Nishino, K., Kaneda, M., & Watanabe, S. (2017). Differentially methylated cpg sites in bull spermatozoa revealed by human dna methylation arrays and bisulfite analysis. *Journal of Reproduction and Development*, 63(3), 279–287.
- Tal, O., Kisdi, E., & Jablonka, E. (2010). Epigenetic contribution to covariance between relatives. *Genetics*, 184(4), 1037–1050.
- Tammen, S. A., Friso, S., & Choi, S.-W. (2013). Epigenetics : the link between nature and nurture. *Molecular aspects of medicine*, 34(4), 753–764.
- Tanaka, K., & Okamoto, A. (2007). Degradation of dna by bisulfite treatment. *Bioorganic & medicinal chemistry letters*, 17(7), 1912–1915.
- Terrazas-Salgado, L., García-Gasca, A., Betancourt-Lozano, M., Llera-Herrera, R., Alvarado-Cruz, I., & Yáñez-Rivera, B. (2022). Epigenetic transgenerational modifications induced by xenobiotic exposure in zebrafish. *Frontiers in Cell and Developmental Biology*, 10, 200.
- Thayer, Z. M., Wilson, M. A., Kim, A. W., & Jaeggi, A. V. (2018, Mar). Impact of prenatal stress on offspring glucocorticoid levels : A phylogenetic meta-analysis across 14 vertebrate species. *Sci Rep*, 8(1), 4942.
- Thirlwell, C., Eymard, M., Feber, A., Teschendorff, A., Pearce, K., Lechner, M., Widschwendter, M., & Beck, S. (2010). Genome-wide dna methylation analysis of archival formalin-fixed paraffin-embedded tissue using the illumina infinium humanmethylation27 beadchip. *Methods*, 52(3), 248–254.
- Tiffon, C. (2018). The impact of nutrition and environmental epigenetics on human health and disease. *International journal of molecular sciences*, 19(11), 3425.
- Timms, B. G., Howdeshell, K. L., Barton, L., Bradley, S., Richter, C. A., & vom Saal, F. S. (2005). Estrogenic chemicals in plastic and oral contraceptives disrupt development of the fetal mouse prostate and urethra. *Proceedings of the National Academy of Sciences*, 102(19), 7014–7019.

- Torres-Perez, J. V., Irfan, J., Febrianto, M. R., Di Giovanni, S., & Nagy, I. (2021, Nov). Histone post-translational modifications as potential therapeutic targets for pain management. Trends Pharmacol Sci, 42(11), 897–911.
- Tremolizzo, L., Carboni, G., Ruzicka, W. B., Mitchell, C. P., Sugaya, I., Tueting, P., Sharma, R., Grayson, D. R., Costa, E., & Guidotti, A. (2002, Dec). An epigenetic mouse model for molecular and behavioral neuropathologies related to schizophrenia vulnerability. Proc Natl Acad Sci U S A, 99(26), 17095–17100.
- Triantaphyllopoulos, K. A., Ikonomopoulos, I., & Bannister, A. J. (2016). Epigenetics and inheritance of phenotype variation in livestock. Epigenetics Chromatin, 9, 31.
- Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021, 26 Aug). Genome-wide association studies. Nature Reviews Methods Primers, 1(1), 59. doi: 10.1038/s43586-021-00056-9
- Urich, M. A., Nery, J. R., Lister, R., Schmitz, R. J., & Ecker, J. R. (2015). Methylc-seq library preparation for base-resolution whole-genome bisulfite sequencing. Nature protocols, 10(3), 475–483.
- Vander Velden, J. W., & Osborne, D. M. (2022). Prolonged diet-induced obesity modifies dna methylation and gene expression in the hippocampus. Neuroscience Letters, 780, 136656.
- Van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018). The third revolution in sequencing technology. Trends in Genetics, 34(9), 666–681.
- Van Eijk, K. R., de Jong, S., Boks, M. P., Langeveld, T., Colas, F., Veldink, J. H., de Kovel, C. G., Janson, E., Strengman, E., Langfelder, P., et al. (2012). Genetic analysis of dna methylation and gene expression levels in whole blood of healthy human subjects. BMC genomics, 13, 1–13.
- Vanyushin, B., Belozersky, A., Kokurina, N., & Kadirova, D. (1968). 5-methylcytosine and 6-methylaminopurine in bacterial dna. Nature, 218(5146), 1066–1067.
- Varona, L., Munilla, S., Mouresan, E., González-Rodríguez, A., Moreno, C., et al. (2015). A bayesian model for the analysis of transgenerational epigenetic variation. g3 (bethesda) 5 : 477–485.
- Verdel, A., Vavasseur, A., Le Gorrec, M., & Touat-Todeschini, L. (2009). Common themes in sirna-mediated epigenetic silencing pathways. The International journal of developmental biology, 53(2-3), 245–257.
- Verhulst, E. C., Mateman, A. C., Zwier, M. V., Caro, S. P., Verhoeven, K. J., & Van Oers, K. (2016). Evidence from pyrosequencing indicates that natural variation in animal personality is associated with drd 4 dna methylation. Molecular ecology, 25(8), 1801–1811.
- Viggiani, M. T., Polimeno, L., Di Leo, A., & Barone, M. (2019). Phytoestrogens : Dietary intake, bioavailability, and protective mechanisms against colorectal neoproliferative lesions. Nutrients, 11(8), 1709.
- Vitorino Carvalho, A., Hennequet-Antier, C., Crochet, S., Bordeau, T., Couroussé, N., Cailleau-Audouin, E., Chartrin, P., Darras, V. M., Zerjal, T., Collin, A., et al. (2020). Embryonic thermal manipulation has short and long-term effects on the development and the physiology of the japanese quail. Plos one, 15(1), e0227700.

- Vitorino Carvalho, A., Hennequet-Antier, C., Rouger, R., Delaveau, J., Bordeau, T., Crochet, S., Courousse, N., Pitel, F., Collin, A., & Coustham, V. (2023). Thermal conditioning of quail embryos has transgenerational and reversible long-term effects. *bioRxiv*, 2023-02.
- Vom Saal, F. S., Akingbemi, B. T., Belcher, S. M., Birnbaum, L. S., Crain, D. A., Eriksen, M., Farabollini, F., Guillette Jr, L. J., Hauser, R., Heindel, J. J., et al. (2007). Chapel hill bisphenol a expert panel consensus statement : integration of mechanisms, effects in animals and potential to impact human health at current levels of exposure. *Reproductive toxicology*, *24*(2), 131–138.
- Vom Saal, F. S., & Welshons, W. V. (2006). Large effects from small exposures. ii. the importance of positive controls in low-dose research on bisphenol a. *Environmental research*, *100*(1), 50–76.
- Waddington, C. (1942, Nov). Canalization of development and the inheritance of acquired characters. *Nature*, *150*(3811), 563–565.
- Wagner, J. R., Busche, S., Ge, B., Kwan, T., Pastinen, T., & Blanchette, M. (2014). The relationship between dna methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. *Genome biology*, *15*, 1–17.
- Wallén, E., Auvinen, P., & Kaminen-Ahola, N. (2021). The effects of early prenatal alcohol exposure on epigenome and embryonic development. *Genes*, *12*(7), 1095.
- Wang, J., Xie, S., Yang, J., Xiong, H., Jia, Y., Zhou, Y., Chen, Y., Ying, X., Chen, C., Ye, C., et al. (2019). The long noncoding rna h19 promotes tamoxifen resistance in breast cancer via autophagy. *Journal of hematology & oncology*, *12*(1), 1–14.
- Wang, J., Zhao, J., Hu, P., Gao, L., Tian, S., & He, Z. (2022). Long non-coding rna hotair in central nervous system disorders : New insights in pathogenesis, diagnosis, and therapeutic potential. *Frontiers in Molecular Neuroscience*, *15*, 949095.
- Watson, J. D., & Crick, F. H. (1953, Apr). Molecular structure of nucleic acids ; a structure for deoxyribose nucleic acid. *Nature*, *171*(4356), 737–738.
- Weaver, I. C., Cervoni, N., Champagne, F. A., D'Alessio, A. C., Sharma, S., Seckl, J. R., Dymov, S., Szyf, M., & Meaney, M. J. (2004). Epigenetic programming by maternal behavior. *Nature neuroscience*, *7*(8), 847–854.
- Webster, T. H., Vannan, A., Pinto, B. J., Denbrock, G., Morales, M., Dolby, G. A., Fiddes, I. T., DeNardo, D. F., & Wilson, M. A. (2023). Incomplete dosage balance and dosage compensation in the zz/zw gila monster (*heloderma suspectum*) revealed by de novo genome assembly. *bioRxiv*, 2023-04.
- Weng, Y.-I., Hsu, P.-Y., Liyanarachchi, S., Liu, J., Deatherage, D. E., Huang, Y.-W., Zuo, T., Rodriguez, B., Lin, C.-H., Cheng, A.-L., et al. (2010). Epigenetic influences of low-dose bisphenol a in primary human breast epithelial cells. *Toxicology and applied pharmacology*, *248*(2), 111–121.
- Wocławek-Potocka, I., Mannelli, C., Boruszewska, D., Kowalczyk-Zieba, I., Waśniewski, T., Skarżyński, D. J., et al. (2013). Diverse effects of phytoestrogens on the reproductive performance : cow as a model. *International journal of endocrinology*, 2013.
- Wolstenholme, J. T., Rissman, E. F., & Connelly, J. J. (2011). The role of bisphenol a in shaping

- the brain, epigenome and behavior. *Hormones and behavior*, 59(3), 296–305.
- Wong, N. C., Ng, J., Hall, N. E., Lunke, S., Salmanidis, M., Brumatti, G., Ekert, P. G., Craig, J. M., & Saffery, R. (2013). Exploring the utility of human dna methylation arrays for profiling mouse genomic dna. *Genomics*, 102(1), 38–46.
- Wöste, M., Leitão, E., Laurentino, S., Horsthemke, B., Rahmann, S., & Schröder, C. (2020). wg-blimp : an end-to-end analysis pipeline for whole genome bisulfite sequencing data. *BMC bioinformatics*, 21, 1–8.
- Wu, K.-J. (2020). The epigenetic roles of dna n6-methyladenine (6ma) modification in eukaryotes. *Cancer Letters*, 494, 40–46.
- Wutz, A. (2011). Gene silencing in x-chromosome inactivation : advances in understanding facultative heterochromatin formation. *Nature Reviews Genetics*, 12(8), 542–553.
- Y. Benjamini, Y. H. (1995). Controlling the false discovery rate : a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 116, 289–300.
- Yehuda, R., Daskalakis, N. P., Bierer, L. M., Bader, H. N., Klengel, T., Holsboer, F., & Binder, E. B. (2016). Holocaust exposure induced intergenerational effects on flkbp5 methylation. *Biological psychiatry*, 80(5), 372–380.
- Yilmaz, B., Terekeci, H., Sandal, S., & Kelestimur, F. (2020). Endocrine disrupting chemicals : exposure, effects on human health, mechanism of action, models for testing and strategies for prevention. *Reviews in endocrine and metabolic disorders*, 21, 127–147.
- Yong, W. S., Hsu, F. M., & Chen, P. Y. (2016). Profiling genome-wide DNA methylation. *Epigenetics Chromatin*, 9, 26.
- Ytournal, F., Gilbert, H., & Boichard, D. (2008, juin). Comment affiner la localisation d'un qtl ? *INRAE Productions Animales*, 21(2), 147–158. doi: 10.20870/productions-animales.2008.21.2.3388
- Yu, G. (2009, juillet). Variance stabilizing transformations of poisson, binomial and negative binomial distributions. *Statistics & Probability Letters*, 79(14), 1621–1629. doi: 10.1016/j.spl.2009.04.010
- Yu, L., Xia, K., Cen, X., Huang, X., Sun, W., Zhao, Z., & Liu, J. (2020). Dna methylation of noncoding rnas : new insights into osteogenesis and common bone diseases. *Stem Cell Research & Therapy*, 11(1), 1–8.
- Zeng, Y., & Chen, T. (2019). Dna methylation reprogramming during mammalian development. *Genes*, 10(4), 257.
- Zhang, T., Cooper, S., & Brockdorff, N. (2015). The interplay of histone modifications–writers that read. *EMBO reports*, 16(11), 1467–1481.
- Zhang, Y., & Reinberg, D. (2001). Transcription regulation by histone methylation : interplay between different covalent modifications of the core histone tails. *Genes & development*, 15(18), 2343–2360.
- Zhao, Y., & Garcia, B. A. (2015). Comprehensive catalog of currently documented histone modifications. *Cold Spring Harbor perspectives in biology*, 7(9), a025064.
- Zhao, Y., Yang, M., Wang, S., Abbas, S. J., Zhang, J., Li, Y., Shao, R., & Liu, Y. (2022). An Overview of Epigenetic Methylation in Pancreatic Cancer Progression. *Front Oncol*, 12,

854773.

Zhuo, Z., Lamont, S. J., & Abasht, B. (2017). Rna-seq analyses identify frequent allele specific expression and no evidence of genomic imprinting in specific embryonic tissues of chicken. Scientific reports, 7(1), 11944.

Ziller, M. J., Hansen, K. D., Meissner, A., & Aryee, M. J. (2015). Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. Nature methods, 12(3), 230–232.





## Epigénétique transgénérationnelle : l'environnement embryonnaire impacte-t-il les performances des générations ultérieures via l'épigénome chez la caille ?

### Résumé

---

Le changement climatique, les contraintes économiques et les préoccupations sociétales vont fortement influencer les systèmes de production animale et nécessitent une caractérisation plus fine de la capacité d'adaptation du génome des animaux aux fluctuations environnementales. Cela repose notamment sur l'étude des marques épigénétiques (méthylation de l'ADN), facilitée par les progrès des technologies de séquençage et des outils bioinformatiques. Nous avons montré, chez la caille, que l'injection de perturbateurs endocriniens dans les œufs impacte à la fois la croissance et la méthylation de l'ADN au niveau de gènes impliqués dans la croissance, le comportement et la reproduction. Des observations similaires ont été identifiées trois générations après injection. Ces résultats suggèrent une transmission épigénétique transgénérationnelle des effets environnementaux. Cette information pourrait être intégrée dans les modèles de sélection, en incluant données génétiques et épigénétiques.

**Mots-clés :** *Coturnix japonica*, méthylation de l'ADN, bisulfite, nanopore, perturbateurs endocriniens, épigénétique transgénérationnelle.

## Transgenerational epigenetics in quail : does the embryonic environment impact the performance of subsequent generations via the epigenome ?

### Abstract

---

Climate change, economic constraints and societal concerns will significantly influence animal production systems and require a more detailed characterization of the adaptability of the animal genome to environmental fluctuations. This relies on the study of epigenetic marks (DNA methylation), made possible by advances in sequencing technologies and the use of bioinformatics tools. Here we showed that the injection of endocrine disruptors into the egg affects both growth and DNA methylation in genes related to growth, behavior and reproduction. Similar observations were identified three generations after injection. These results suggest a transgenerational epigenetic transmission of environmental effects. This information could be integrated into selection programs using models that incorporate genetic and epigenetic data from each individual.

**Keywords :** *Coturnix japonica*, DNA methylation, bisulfite, nanopore, endocrine disruptors, transgenerational epigenetics.