



**HAL**  
open science

## Apports du séquençage haut-débit sur la connaissance de l'épigénome aviaire

Marjorie Mersch, Sarah-Anne David, Anais Vitorino Carvalho, Sylvain Foissac, Anne Collin, Frederique Pitel, Vincent Coustham

► **To cite this version:**

Marjorie Mersch, Sarah-Anne David, Anais Vitorino Carvalho, Sylvain Foissac, Anne Collin, et al.. Apports du séquençage haut-débit sur la connaissance de l'épigénome aviaire. INRA Productions Animales, 2018, 31 (4), pp.325-336. 10.20870/productions-animales.2018.31.4.2372 . hal-02154452

**HAL Id: hal-02154452**

**<https://hal.science/hal-02154452>**

Submitted on 12 Jun 2019

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

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



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# Apports du séquençage haut-débit sur la connaissance de l'épigénome aviaire

Marjorie MERSCH<sup>1†</sup>, Sarah-Anne DAVID<sup>2†</sup>, Anaïs VITORINO CARVALHO<sup>2†</sup>, Sylvain FOISSAC<sup>1</sup>, Anne COLLIN<sup>2</sup>, Frédérique PITEL<sup>\*1</sup>, Vincent COUSTHAM<sup>2</sup>

<sup>1</sup>GenPhySE, Université de Toulouse, INRA, INPT, INP-ENVT, 31320, Castanet Tolosan, France

<sup>2</sup>BOA, INRA, 37380, Nouzilly, France

<sup>†</sup>Ces auteurs ont contribué de façon égale à ce travail.

Courriel : [vincent.coustham@inra.fr](mailto:vincent.coustham@inra.fr)

■ Le phénotype résulte de l'expression du génome, de l'environnement et de leur interaction médiée par l'épigénome. Alors que l'oiseau est un modèle largement étudié en biologie du développement, en immunologie et en microbiologie, peu d'informations sont disponibles sur son épigénome. Dans cette revue, nous synthétisons les connaissances concernant l'épigénome aviaire, principalement celui du poulet, et présentons des éléments pour mener à bien son analyse.

## Introduction

L'oiseau, c'est en particulier le cas de l'espèce *Gallus*, est un modèle de choix pour l'étude de la biologie du développement (en particulier l'embryogenèse), l'immunologie et la microbiologie. Il s'agit également d'une source majeure de protéines par sa production de viande et d'œufs. Au cours des cinquante dernières années, la sélection génétique a permis un important accroissement de la production avicole grâce notamment à l'obtention d'animaux présentant une croissance rapide. Cependant, des analyses mettant en regard des caractères phénotypiques et des études génomiques ont montré que la variation génétique n'est pas le seul facteur contribuant à la variabilité phénotypique observée (Manolio *et al.*, 2009).

Dans les années 1950, Waddington a défini un caractère épigénétique comme « un phénotype stable héréditaire résultant de changements dans un chromosome sans altération dans la séquence d'ADN » (Berger *et al.*, 2009). De nos

jours, l'épigénétique est communément définie comme l'ensemble des mécanismes moléculaires réversibles et héréditaires (par mitose et potentiellement par méiose) impliqués dans la régulation de l'expression des gènes sans altération de la séquence d'ADN. Il a été montré que les modifications épigénétiques chez les eucaryotes peuvent affecter l'expression d'un certain nombre de gènes sous l'influence de l'environnement (température, polluants... Feil et Fraga, 2012).

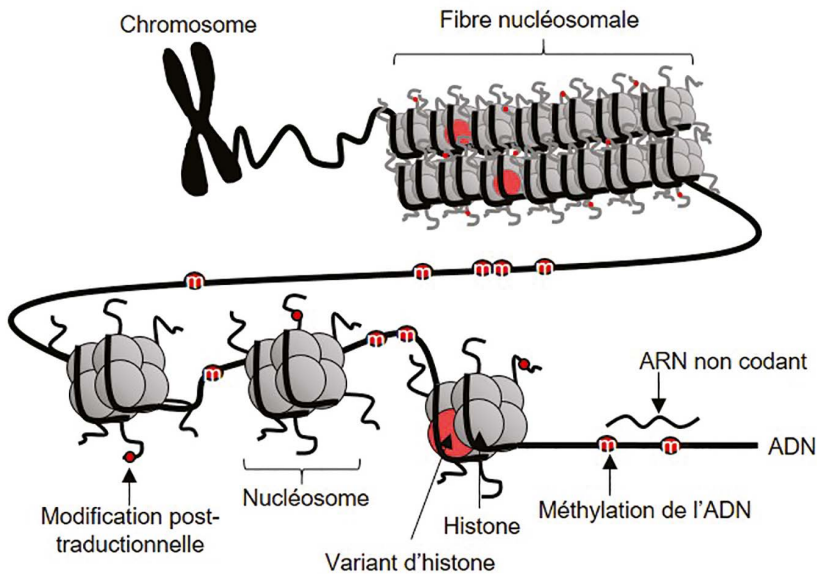
Un nombre croissant d'études montrent que l'épigénétique contribue également à la diversité phénotypique des animaux, en interaction avec leur environnement (Triantaphyllopoulos *et al.*, 2016). Ainsi, il a été suggéré que la modulation épigénétique de l'expression des gènes pouvait avoir des impacts à long-terme sur le phénotype des poulets, notamment sur leurs comportements (Verhulst *et al.*, 2016), leur capacité d'adaptation (Kisliouk et Meiri, 2009) et sur leurs performances (Li *et al.*, 2015).

Dans cette revue adaptée de David *et al.* (2017a), nous présenterons le récent gain de connaissances sur l'épigénome du poulet en nous concentrant sur deux types de marques épigénétiques pour lesquelles de nombreuses informations sont disponibles : la méthylation de l'ADN et certaines Modifications Post-Traductionnelles des Histones (MPTH) très étudiées. Nous nous intéresserons également aux similitudes des profils épigénétiques du poulet avec ceux observés dans d'autres espèces, notamment chez les mammifères. Enfin, nous évoquerons les recommandations à suivre et les pièges à éviter pour la bonne conduite d'études épigénomiques sur les espèces aviaires.

## 1. Les marques épigénétiques

Il existe de nombreux types de facteurs dont la méthylation de l'ADN, les MPTH, les variants d'histones ou

**Figure 1.** Schéma des différents états de compaction de l'ADN au sein du noyau de la cellule eucaryote illustrant les principales modifications épigénétiques.



encore les ARN non-codants (figure 1). Bien que ces deux dernières modifications peuvent également jouer un rôle majeur dans la variabilité phénotypique (Roundtree *et al.*, 2017), peu de connaissances existent à l'heure actuelle chez le poulet et elles ne seront, par conséquent, pas traitées dans cette revue.

### ■ 1.1. La méthylation de l'ADN

La méthylation de l'ADN est actuellement la marque épigénétique la plus étudiée. Chez les vertébrés, cette marque cible principalement les résidus cytosine à la position 5 du carbone (5mC, Gardiner-Garden et Frommer, 1987) dans un contexte de dinucléotides CpG. La méthylation de l'ADN est mise en place par les protéines de la famille DNMT (pour revue voir Lu, 2013). La distribution de la 5mC à travers le génome n'est pas aléatoire (Gardiner-Garden et Frommer, 1987). En effet, les dinucléotides CpG ont tendance à être regroupés dans des régions nommées îlots CpG qui sont souvent hypométhylés. La présence de 5mC dans les promoteurs des gènes est généralement associée à la répression de la transcription (Bird, 1992), alors que sa présence dans le corps des gènes serait plutôt favorable à l'expression génique (Jones, 2012), mais la situation est en général com-

plexe, et il est impossible d'établir des règles universelles. Chez les jumeaux monozygotes humains, la 5mC diffère dans plusieurs tissus, suggérant que cette marque contribue à l'expression différentielle de gènes liés à l'histoire de vie de l'individu (Busche *et al.*, 2015). Il existe une corrélation entre la méthylation de l'ADN et la variation des phénotypes chez plusieurs espèces végétales et animales (Feil et Fraga, 2012). Par exemple, la couleur du pelage des souris agoutis est régulée par l'expression du gène *ASIP*. L'ingestion par des souris agoutis gestantes d'aliments présentant une forte teneur en donneurs de groupes méthyles (tels que l'acide folique, la vitamine B12, la bêtaïne et la choline) induit un assombrissement de la couleur du pelage des petits expliqué par la méthylation différentielle d'un rétro-transposon en amont du gène *ASIP* (Wolff *et al.*, 1998). La 5mC n'est pas la seule modification du cinquième carbone des résidus cytosine potentiellement impliquée dans la régulation de la transcription. En effet, de récentes études ont montré que la 5-hydroxyméthylcytosine (5hmC), influencerait également l'expression de gènes (Ponnaluri *et al.*, 2017). La 5hmC est catalysée par la famille des protéines TET et est proposée comme un intermédiaire de déméthylation de l'ADN.

### ■ 1.2. Les modifications post-traductionnelles des histones

Les MPTH sont une autre catégorie de marques épigénétiques fortement étudiées. Chez les eucaryotes, les protéines histones (H2A, H2B, H3 et H4) sont impliquées dans la compaction de l'ADN via la formation d'octamères enveloppés par 147 paires de bases d'ADN, l'ensemble formant une unité structurale appelé nucléosome (Kornberg et Lorch, 1999). Les histones possèdent deux régions : le cœur et la queue. Ce sont les cœurs des histones qui constituent les nucléosomes. Les queues, elles, restent accessibles en dehors de la structure nucléosomale (figure 1). Les acides aminés principalement situés au niveau de la queue, mais aussi parfois du noyau, peuvent être la cible de modifications covalentes : méthylation, acétylation, ubiquitination, phosphorylation, sumoylation, etc. (Bannister et Kouzarides, 2011). Plusieurs familles d'enzymes, dont les histones méthyl-transférases, sont impliquées dans la modification de ces acides aminés (pour revue voir Lu, 2013). L'une des fonctions des MPTH est la régulation structurale des nucléosomes (dont l'interaction entre les histones et la double hélice de l'ADN) et par conséquent de la conformation de la chromatine, qui va être plus ou moins condensée et donc plus ou moins accessible à la machinerie transcriptionnelle (Tessarz et Kouzarides, 2014). Ces MPTH jouent également un rôle sur le recrutement de protéines sur l'ADN (notamment des facteurs de transcription) et peuvent ainsi moduler l'expression des gènes, que ce soit pour la favoriser ou la réprimer (Kouzarides, 2007). Par exemple, l'acétylation de la lysine 27 sur l'histone H3 (H3K27ac) et la tri-méthylation de la lysine 4 sur l'histone H3 (H3K4me3) sont communément dites « activatrices » car ces marques induisent l'ouverture de la chromatine au niveau des promoteurs et donc favorisent l'expression des gènes (Zhang *et al.*, 2015). En revanche, les triméthylations des lysines 9 et 27 de l'histone H3 (H3K9me3, H3K27me3) sont dites « répressives » car généralement associées à l'absence d'expression du gène. La présence d'un type de MPTH ne prédit pas nécessairement l'état d'expression d'un gène. En effet, des modèles combinatoires de différents

profils de MPTH ont été proposés pour définir des états chromatiniens spécifiques (Baker, 2011).

## 2. L'étude de l'épigénome

Il existe de nombreuses techniques pour analyser la présence de modifications épigénétiques (Kurdyukov et Bullock, 2016), avec des échelles très variables. Dans cette revue, nous nous focaliserons sur les analyses tout-génome nécessitant le recours au SHD.

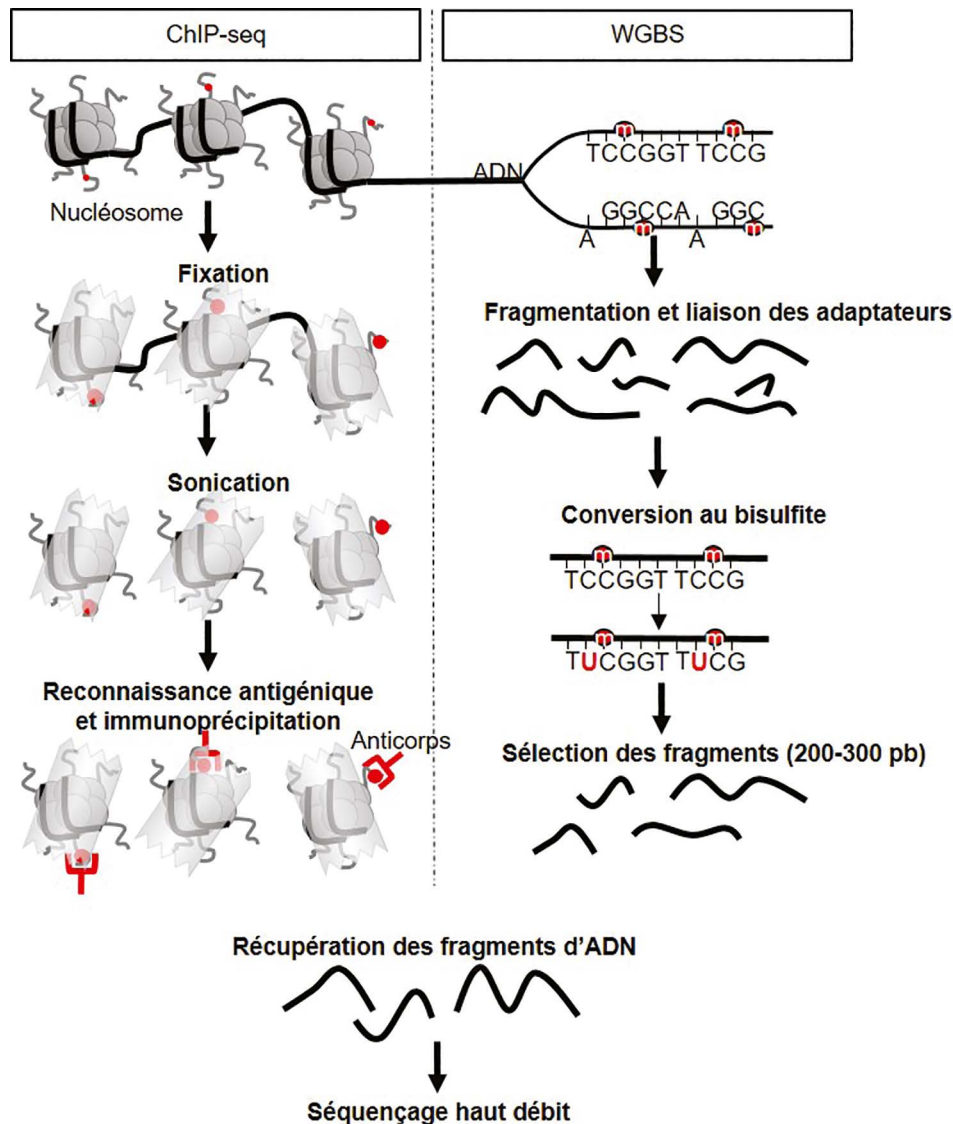
### ■ 2.1. L'étude de la méthylation de l'ADN

À ce jour, la méthode de référence pour l'étude de la 5mC est la conver-

sion de l'ADN génomique par le bisulfite de sodium (figure 2, tableau 1), qui désamine les cytosines en uraciles. Cette transformation est bloquée par la présence d'un groupement méthyle, permettant ainsi d'identifier les cytosines méthylées par comparaison au génome de référence (Frommer *et al.*, 1992 ; Mardis, 2008). En effet, une cytosine non convertie après traitement au bisulfite suggère qu'elle était méthylée dans l'échantillon de départ, alors que sa conversion en thymine suggère qu'elle n'était pas méthylée. En chaque position, la proportion de cytosines non converties donnera le « taux de méthylation » à cette position. Une analyse à l'échelle du génome entier peut ainsi être réalisée quand la conversion au bisulfite est combinée au SHD (Whole

Genome Bisulfite Sequencing – WGBS ou MethylC-seq (Urich *et al.*, 2015), figure 2 et tableau 1) et que le génome de référence est disponible. Une méthode de séquençage d'une fraction du génome a été développée en introduisant une étape de digestion enzymatique grâce à l'enzyme MspI dont le site de restriction est CCGG (Reduced Representation Bisulfite Sequencing – RRBS, Harris *et al.*, 2010). Cette technique est plus économique car elle se focalise principalement sur l'étude des séquences promotrices qui sont généralement riches en îlots CpG, aux dépens des CpG plus isolés souvent exclues de l'analyse (Gu *et al.*, 2011). L'utilisation de différentes combinaisons d'enzymes peut augmenter la portion du génome étudiée, bien que celle-ci reste une

**Figure 2.** Représentation graphique des protocoles d'immunoprécipitation de la chromatine (ChIP-seq) et de Whole Genome Bisulfite Sequencing (WGBS).



**Tableau 1.** Liste des principales technologies d'analyse de la méthylation de l'ADN et des modifications post-traductionnelles des histones.

Modification épigénétique	Nom de la technique	Préparation de l'ADN	Principe d'identification	Fraction ADN étudiée	Résolution
Méthylation de l'ADN	<i>Methylation sensitive Restriction Enzyme sequencing (MRE-seq)</i>	Extraction classique	Enzyme de restriction sensible à la méthylation + séquençage haut-débit	Génome entier	Limitée aux sites de coupure enzymatique
	<i>Methylated DNA ImmunoPrecipitation (MeDIP-seq ou MBD-seq)</i>	Extraction classique	Reconnaissance par anticorps de l'ADN méthylé + séquençage haut-débit	Génome entier	Limitée à la taille des fragments immunoprécipités
	<i>Whole Genome Bisulfite Sequencing (WGBS ou MethylC-seq)</i>	Extraction + conversion au bisulfite de sodium	Séquençage haut-débit	Génome entier	À la base près
	<i>Reduced representation bisulfite sequencing (RRBS)</i>	Extraction + conversion au bisulfite de sodium	Enzyme de restriction + séquençage haut-débit	Sous-représentation du génome enrichie en dinucléotides CpG	À la base près
Modification post-traductionnelle des histones	<i>Immunoprécipitation de la chromatine (ChIP-seq)</i>	Fixation de la chromatine puis sonication ou digestion Mnase	Reconnaissance par anticorps + séquençage haut-débit	Génome entier	Mono/di-nucléosome

sous-représentation du génome. Ainsi, en fonction de la question posée et des moyens financiers disponibles, il peut être conseillé d'utiliser l'une ou l'autre de ces méthodes. Ces deux techniques nécessitent un ADN de très haute qualité pour obtenir une conversion efficace (Ziller *et al.*, 2015). Une autre limitation de ces méthodes est que la conversion au bisulfite ne permet pas de distinguer la 5mC des 5hmC (Globisch *et al.*, 2010). Pour évaluer la part de la 5hmC dans le signal total, il faut effectuer une étape d'oxydation de l'ADN et comparer les résultats à une expérience sans l'étape d'oxydation (correspondant donc aux fractions 5mC et 5hmC, Hahn *et al.*, 2015). Cette étape d'oxydation de l'ADN n'est pas systématiquement effectuée dans les protocoles classiques d'étude de la 5mC, car l'oxydation a tendance à endommager l'ADN et la proportion de 5hmC est souvent résiduelle dans la plupart des tissus (Globisch *et al.*, 2010). D'autres protocoles d'étude de la 5mC n'utilisant pas la conversion au bisulfite et se basant par exemple sur l'utilisation d'enzymes thermosensibles peuvent également être utilisés

(tableau 1). Chaque technique présente ses propres biais qu'il est important de connaître afin de choisir le bon outil et d'interpréter correctement les résultats (Bock, 2012).

Il est probable qu'à l'avenir les analyses de méthylation haut-débit à l'échelle du génome puissent se passer de la conversion au bisulfite de sodium et de l'étape d'amplification de l'ADN grâce par exemple à la technologie Nanopore, en cours de développement, qui pourrait permettre de lire directement, en une seule fois, le génome et le méthylome d'un même échantillon (Simpson *et al.*, 2017).

## ■ 2.2. L'étude des modifications post-traductionnelles des histones

Pour étudier les marques d'histones, la méthode de référence est actuellement l'immunoprécipitation de la chromatine suivie de SHD (ChIP-seq, figure 2, tableau 1). Cette technique vise à identifier les lieux d'association

de l'ADN avec des protéines (facteurs de transcription, histones...) par séquençage de fragments d'ADN immunoprécipités en utilisant des anticorps spécifiques (Barski *et al.*, 2007 ; Johnson *et al.*, 2007 ; Robertson *et al.*, 2007). Ainsi, par l'utilisation d'anticorps spécifiques des MPTH, il est possible d'identifier des séquences d'ADN qui leurs sont associées. Pour ces études, la chromatine extraite peut être soit fixée (protéines liées de façon covalente à l'ADN) et fragmentée par sonication (X-ChIP-seq), soit native (ou non-fixée) et digérée par une enzyme (N-ChIP-seq, David *et al.*, 2017b). Le choix de la méthode dépend de la nature du facteur étudié et du tissu (David *et al.*, 2017b). Une étape de digestion enzymatique par exonucléase peut être ajoutée au protocole de X-ChIP-seq pour réduire la taille des fragments séquencés (ChIP-exo, Rhee et Pugh, 2011) et ainsi améliorer sa résolution. Ainsi, l'état de la chromatine (fixée ou non), la fragmentation et l'affinité de l'anticorps sont des points qui peuvent influencer les résultats obtenus. Il est donc crucial de mettre au point les conditions expérimentales

avec soin et de prendre en compte le choix de la méthode lors des analyses de données (David *et al.*, 2017b). Pour aider au choix de l'anticorps, un portail interactif recense le comportement de différents anticorps commerciaux dirigés contre plusieurs marques d'histones (<http://www.histoneantibodies.com/>; Rothbart *et al.*, 2015).

### 3. La conception d'études épigénomiques

Une particularité essentielle des études basées sur le SHD est qu'en raison du volume de données générées elles nécessitent une forte capacité d'analyse bioinformatique. Ainsi, pour faciliter ces analyses et obtenir des résultats les plus fiables possibles, il est important d'intégrer des spécialistes en bioinformatique et biostatistiques dès la conception du projet (Rigaill *et al.*, 2018). Pour guider les analyses, de nombreux consortiums ont défini des recommandations.

#### ■ 3.1. Les recommandations pour l'analyse du séquençage haut-débit

Au cours des deux dernières décennies, les techniques de SHD se sont démocratisées grâce notamment à la baisse de leur coût, entraînant ainsi une meilleure compréhension des transcriptomes et des épigénomes. Afin d'homogénéiser les résultats, plusieurs consortiums sont apparus pour définir des recommandations sur les protocoles basés sur le SHD. Ces recommandations concernent l'élaboration, l'application et l'analyse de résultats de ces protocoles. Elles ont été émises principalement à partir des données obtenues chez les mammifères, afin d'encadrer toute procédure touchant au SHD et ainsi faciliter la comparaison des résultats entre les études. Ces différentes directives sont issues de projets variés comme le programme ENCODE (ENCyclopedia Of DNA Elements, ENCODE Project Consortium, 2012), le consortium NIH (National Institute of Health) Roadmap Epigenomics Project Mapping Consortium (Bernstein *et al.*, 2010), le consortium IHEC (International

Human Epigenome Consortium) (Bae, 2013) et l'application d'ENCODE aux espèces modèles (modENCODE, Celniker *et al.*, 2009). Depuis 2015, la volonté de mieux comprendre la diversité génétique, épigénétique et phénotypique des espèces d'élevage comme le poulet a conduit à la création d'un consortium appelé FAANG (Functional Annotation of ANimal Genomes, Andersson *et al.*, 2015). Au-delà de son objectif de doter les génomes animaux d'annotations fonctionnelles de haute qualité – épigénomes compris –, FAANG contribue aussi à la recherche médicale et agronomique en normalisant les tests de base et les protocoles expérimentaux et en établissant une infrastructure pour l'analyse des données haut-débit.

Les recommandations émises par ces consortiums concernent notamment le nombre de répétitions biologiques, mais également la profondeur de séquençage nécessaire à l'obtention de résultats de SHD vérifiables. La profondeur de séquençage est définie comme « le nombre moyen de fois qu'un nucléotide particulier (une position définie sur le génome) est représenté dans une collection de séquences brutes aléatoires » (Stein, 2001) et s'exprime en « fois » (X). La première étude tout-génome de la méthylation de l'ADN a été publiée en 2008 sur *Arabidopsis thaliana* avec une profondeur de séquençage d'environ 20 X (Cokus *et al.*, 2008). Rapidement, d'autres études ont adapté cette méthode chez différentes espèces, en particulier chez l'Homme (Lister *et al.*, 2009) et la souris (Smith *et al.*, 2009). Pour le WGBS, il est ainsi recommandé par le Epigenomics Roadmap de séquencer à un minimum de 30 X de profondeur ; pour le RRBS, une profondeur d'au moins 10 X est nécessaire (Doherty et Couldrey, 2014). La quantité de séquences nécessaires à une analyse RRBS dépend de la proportion d'ADN sélectionnée et donc du profil de l'enzyme de restriction utilisée : pour une même quantité totale de lectures, la profondeur de chaque site observé sera d'autant plus élevée que la représentation du génome est réduite. À notre connaissance, aucune étude de RRBS chez le poulet n'a été publiée ; par conséquent, le profil de l'enzyme

de restriction MspI doit être soigneusement examiné en amont pour estimer la quantité de séquences nécessaires chez cette espèce.

La première étude tout-génome de MPTH sur des cellules humaines date de 2007 (Barski *et al.*, 2007). Depuis, le consortium ENCODE a publié les lignes directrices pour obtenir des résultats interprétables d'expériences de CHIP-seq : la qualité de l'anticorps, les contrôles utilisés, le nombre de répétitions biologiques et le nombre de lectures uniques obtenues après le séquençage (Landt *et al.*, 2012). Le consortium ENCODE recommande 20 à 40 millions de lectures uniques pour l'analyse CHIP-seq en fonction du type de marque étudiée (Landt *et al.*, 2012). Pour la marque H3K27me3, présentant des pics larges, une couverture élevée de 40-50 millions de lectures alignées uniques est recommandée pour identifier des pics chez l'humain (Jung *et al.*, 2014 ; David *et al.*, 2017b). Bien que ces recommandations aient été initialement définies pour les espèces de mammifères, elles devraient être suivies pour les études de CHIP-seq chez les oiseaux. En effet, pour H3K27me3, 40 millions de lectures alignées de façon unique semblent être suffisantes chez les oiseaux (David *et al.*, 2017b). Il est important de noter que plus l'immunoprécipitation est efficace (voir partie 2.2 pour les différents leviers d'optimisation), moins il sera nécessaire d'augmenter la profondeur de séquençage (David *et al.*, 2017b).

#### ■ 3.2. Le plan expérimental

Le SHD est maintenant bien documenté chez de nombreuses espèces (dont l'humain, la souris, la drosophile et les plantes), nous permettant ainsi de définir des considérations communes à l'étude d'un épigénome quelle que soit l'espèce, et donc également valables pour les oiseaux. Ainsi, de nombreux facteurs peuvent influencer l'épigénome comme l'âge des individus ou encore leur histoire de vie (conditions d'élevage, nutrition...). Il est également important de connaître l'histoire parentale dont l'âge des reproducteurs et les conditions de transport, de conservation des œufs et d'éclosion.

Enfin, idéalement, il est intéressant de contrôler la variabilité génétique des individus étudiés, pour maximiser les chances d'observer des différences épigénétiques significatives. L'utilisation de lignées fortement consanguines peut permettre d'étudier les modifications épigénétiques dans une situation où la variation génétique est nulle ou presque, et pourrait donc faciliter l'interprétation des résultats. Ainsi, biologiquement on élimine la variabilité épigénétique sous le contrôle de la génétique, et techniquement, pour la méthylation de l'ADN où le traitement au bisulfite entraîne un polymorphisme apparent, on élimine les taux de méthylation artéfactuels dus à un polymorphisme ADN de type C/T (Daxinger et Whitelaw, 2010). En plus des biais techniques liés aux protocoles de biologie moléculaire précédemment évoqués, tous ces aspects doivent donc être pris en compte dans les analyses statistiques à défaut d'être strictement contrôlés.

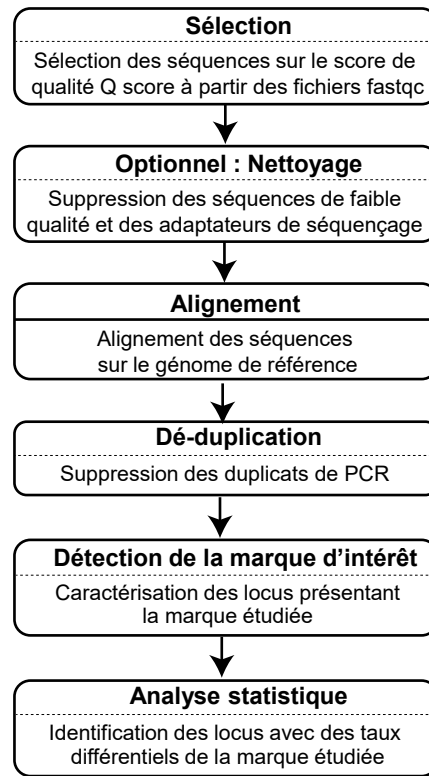
## 4. L'analyse bioinformatique

Souvent sous-estimée, l'analyse des données issues de SHD est un autre élément critique dans les études épigénomiques. Elle nécessite des compétences analytiques particulières ainsi qu'une forte capacité de calcul et de stockage des données.

### ■ 4.1. Les compétences analytiques

Actuellement, les plateformes de SHD proposent généralement l'analyse des données générées. La réalisation d'une analyse complète de données épigénétiques nécessite une version correctement annotée du génome ainsi que des outils informatiques dont les mises à jour sont fréquentes (figure 3). Ainsi, la première étape est de vérifier la disponibilité du génome ainsi que son annotation. Ensuite, il faut prendre en compte l'évolution rapide des outils bioinformatiques qui peut entraîner l'utilisation de nouveaux formats de stockage et des algorithmes de traitement au cours du projet. Par exemple, les approches de référence jusque-là utilisées pour l'ali-

**Figure 3. Exemple d'analyse bioinformatique de données d'immunoprécipitation de la chromatine (ChIP-seq) et de Whole Genome Bisulfite Sequencing (WGBS).**



gnement (identification de la position d'une séquence sur le génome de référence) comme les approches BLAST/BLAT se sont révélées inefficaces pour traiter les positions de milliards de courtes lectures sur le génome entier. Ainsi, des méthodes spécifiques ont dû être développées pour répondre à ce type d'alignement (Langmead *et al.*, 2009 ; Li et Durbin, 2009). Des outils comme Bowtie (Langmead *et al.*, 2009), BWA (Li et Durbin, 2009) et SOAP2 (Li *et al.*, 2009) présentent chacun des paramètres spécifiques (par exemple le nombre de bases non identiques à la référence autorisées pour valider l'alignement) qui impactent significativement l'alignement des lectures et donc l'identification des zones d'intérêt (Han et He, 2016). De plus, pour l'alignement de séquences ayant subi un traitement au bisulfite de sodium on utilise en général des outils dédiés, basés sur les aligneurs précédemment décrits. Ces spécificités ont été intégrées dans plusieurs pipelines mis au point pour étudier les données issues de SHD comme le pipeline d'analyse

de WGBS du consortium ENCODE (The ENCODE Project Consortium) ou le pipeline d'analyse de ChIP-Seq du BLUEPRINT (Epigenome Blueprint). Des normes similaires sont attendues pour les espèces aviaires dans le cadre du consortium FAANG (Andersson *et al.*, 2015).

### ■ 4.2. Les capacités de calcul et de mémoire

Deux autres facteurs importants concernent la capacité de calcul et de mémoire pour exécuter les pipelines d'analyses bioinformatiques, et l'espace disque pour stocker les données et les résultats. Les installations de bioinformatique fournissent généralement des supports informatiques pour mettre à disposition des capacités de calcul suffisantes (plateformes de l'Institut Français de Bioinformatique, <https://www.france-bioinformatique.fr>). Tout comme le protocole expérimental et le budget d'un projet donné, le volume de séquences brutes obtenues doit être défini en amont. Alors qu'on observe une réduction drastique de la taille des données tout au long de l'analyse bioinformatique, les opérations initiales comme le nettoyage des données et l'alignement des lectures issues de SHD produisent des fichiers volumineux. À titre d'exemple, l'ensemble des fichiers générés par une analyse WGBS d'environ 10 X de profondeur chez le poulet (données brutes, intermédiaires et finales) représente près de 35 Go par échantillon. Chez le même modèle, les analyses de ChIP-seq, elles, nécessitent environ 10 Go de stockage, comprenant les séquences des fractions contrôles et immunoprécipitées (fichiers fastq.gz compressée initiaux – environ 1 à 2 Go) ainsi que les fichiers de coordonnées des pics (500 Ko). Il est donc nécessaire de prévoir des capacités de stockage adéquates. Le stockage des données brutes permet leur ré-analyse par de nouveaux outils, élément important étant donné la vitesse de développement de nouveaux programmes. Il permet également d'actualiser les analyses quand les connaissances internationales évoluent, comme les versions de l'assemblage des génomes. Par ailleurs, un intérêt majeur du stockage

de données réside dans la possibilité de réaliser de nouveaux types d'analyses, non existants au moment de leur production, et dans la réalisation de méta-analyses utilisant des données de même type produites dans différents laboratoires à différentes époques. Le stockage au sein des laboratoires qui analysent ces données n'a pas vocation en général à être maintenu sur le long terme : actuellement de nombreux programmes de recherche entrent dans le cadre de « l'open science », et la plupart des données sont mises à disposition de la communauté scientifique sur des bases de données internationales, à long terme, grâce auxquelles de nouvelles analyses peuvent être réalisées par d'autres équipes (voir Korthauer et Irizarry, 2018 pour un exemple récent).

En outre, les besoins de stockage peuvent s'avérer supérieurs dans les cas des espèces de rente car les analyses impliquent souvent de nombreuses optimisations (test de différents paramètres et de logiciels qui ne sont pas développés pour l'étude de ces génomes). En effet, certains outils d'analyse sont optimisés pour l'étude d'un génome constitué de 22 + XY chromosomes humains et non pas des 32 + ZW chromosomes et 524 « scaffolds » que contient l'assemblage actuel du génome du poulet (GRCg6a ; GCA\_000002315.4). L'adaptation de ces protocoles aux organismes non-modèles requiert donc souvent quelques ajustements. Enfin, l'étude des marques épigénétiques sera d'autant plus pertinente et informative que le génome de référence de l'espèce étudiée disposera d'un assemblage de son génome de bonne qualité, et bien annoté.

## 5. L'épigénome du poulet

### ■ 5.1. Le génome du poulet

Le génome du poulet est environ 2,5 fois plus petit que le génome humain (1 200 Mb contre 3 200 Mb). Il est composé de 32 paires d'autosomes et de deux chromosomes sexuels (Z et W). La plupart des autosomes (29 paires au total) sont relativement petits et de taille uniforme, et sont appelés « microchromosomes ».

Le génome du poulet fut le premier génome aviaire, et le premier génome d'une espèce de rente, à avoir été séquencé (International Chicken Genome Sequencing Consortium, 2004). L'assemblage du génome de *Gallus gallus* a été obtenu à partir d'une seule poule de race "Red Jungle Fowl". La dernière version de l'assemblage du génome du poulet (GRCg6a ; GCA\_000002315.4), d'une taille de 1,1 Gb, a été publiée en 2018 et a été obtenue grâce à la technologie de lecture longue (PacBio). Cette sixième version du génome a été améliorée de façon notable, avec une réduction du nombre de « scaffolds non assignés » (larges morceaux de séquence dont la localisation chromosomique est inconnue) qui est passé de 23 870 (version 5) à 525.

Environ 38 % des îlots CpG sont conservés entre l'Homme et le poulet (International Chicken Genome Sequencing Consortium, 2004). Chez cette espèce, les îlots CpG sont principalement localisés sur des microchromosomes présentant une forte densité de gènes (McQueen *et al.*, 1996).

### ■ 5.2. La méthylation de l'ADN chez le poulet

La première analyse de méthylation de l'ADN dans le foie et les tissus musculaires du poulet a été publiée en 2011 (Li *et al.*, 2011). La technique du MeDIP-seq était utilisée. Cette étude a permis de mettre en évidence une diminution du niveau de méthylation de l'ADN dans les régions promotrices, en particulier au niveau du site d'initiation de la transcription (TSS), et une augmentation dans les corps des gènes (Li *et al.*, 2011). Ces observations sont cohérentes avec la localisation des îlots CpG qui sont principalement dans les promoteurs, et confirmées dans une étude ultérieure basée sur la technique de MeDIP-seq (Hu *et al.*, 2013). De légères variations de méthylation ont été observées le long des gènes (Li *et al.*, 2011). La densité de méthylation est plus élevée dans les introns que dans les régions transcrites non traduites (UTRs) et les exons (Hu *et al.*, 2013). Les niveaux de méthylation des régions promotrices sont en général négativement corrélés avec l'expression des gènes associés. On retrouve donc le

même type de corrélation entre niveaux d'expression et méthylation que celle observée chez l'humain (Suzuki et Bird, 2008). Le niveau de méthylation varie entre les tissus, l'ADN hépatique étant moins méthylé que celui du muscle (Li *et al.*, 2011). Malgré une densité d'éléments répétés inférieure chez le poulet (11 %) que chez les mammifères (40-50 %, International Chicken Genome Sequencing Consortium, 2004), les profils de méthylation sont similaires avec environ 60 % des séquences répétées totalement méthylées (Hu *et al.*, 2013). Ainsi, de nombreuses similitudes entre les profils de méthylation des mammifères et du poulet ont pu être observées.

La première analyse du méthylome de poulet à haute-résolution, permise grâce au WGBS, a été publiée en 2015 dans une étude portant sur la résistance à la maladie de Marek dans deux lignées de poulet (Li *et al.*, 2015). Cette étude confirme la présence d'un niveau de méthylation plus important dans les exons que dans le reste du gène (Tian *et al.*, 2013). Ces résultats mettent en évidence une corrélation entre la variation de la 5mC et une résistance à la maladie. En effet, l'étude de gènes liés à l'immunité, y compris le Toll Like Receptor 4 (TLR4), a montré des variations de méthylation et d'expression génique dans les deux lignées de poulets caractérisées pour avoir une résistance à la maladie de Marek différente (Li *et al.*, 2015). Une autre étude WGBS récente a mis en évidence la variation des profils de 5mC en fonction de l'âge dans le tissu musculaire de poules pondeuses (Zhang *et al.*, 2017). Les animaux âgés présentent un niveau de méthylation de l'ADN plus élevé que les jeunes individus. Cette augmentation de la méthylation a également été observée dans les zones promotrices de gènes impliqués dans la qualité de la viande (Zhang *et al.*, 2017 ; Zhang *et al.*, 2018).

À ce jour, une seule analyse à l'échelle de l'épigénome a été effectuée dans une espèce aviaire différente du poulet, la mésange charbonnière (Laine *et al.*, 2016). L'étude de la 5mC par WGBS dans le cerveau et le sang de cet oiseau a mis en évidence des taux réduits de méthylation sur les îlots CpG et autour des TSS.



Par ailleurs, dans le cerveau, une corrélation négative entre les niveaux de méthylation dans le corps des gènes et leurs expressions a été observée (Laine *et al.*, 2016). Enfin, le niveau de 5mC était plus important dans le cerveau que dans le sang, montrant à nouveau des différences dépendantes du tissu étudié (Derks *et al.*, 2016).

### ■ 5.3. Les modifications post-traductionnelles des histones chez le poulet

À ce jour, seuls quelques travaux ont étudié des MPTH chez les oiseaux (Luo *et al.*, 2012 ; Mitra *et al.*, 2012, 2015 ; Jahan *et al.*, 2016). La première étude concerne l'analyse des profils de MPTH dans la maladie de Marek dans deux lignées de poulet présentant des niveaux de résistance différents (Luo *et al.*, 2012 ; Mitra *et al.*, 2012, 2015). Ainsi, les profils de H3K4me3 et H3K27me3 ont été étudiés dans trois tissus impliqués dans la réponse immunitaire : la rate (Luo *et al.*, 2012), le thymus (Mitra *et al.*, 2012) et la bourse de Fabricius (Mitra *et al.*, 2015). La marque H3K4me3 se retrouve principalement autour des TSS et est associée à des gènes actifs, comme précédemment montré chez les mammifères (Barski *et al.*, 2007). Cela a également été confirmé récemment dans les érythrocytes de poulet (Jahan *et al.*, 2016). La marque H3K27me3 est largement enrichie au niveau des TSS et dans le corps des gènes (Luo *et al.*, 2012 ; Mitra *et al.*, 2012, 2015) et associée à des gènes réprimés, comme observé dans d'autres espèces (Barski *et al.*, 2007 ; David *et al.*, 2017b). De façon intéressante, des analyses comparatives ont révélé des régions différentiellement enrichies en ces MPTH entre les lignées de poulet sensibles et résistantes (Mitra *et al.*, 2012, 2015), suggérant l'existence d'une « signature épigénétique » pour la sensibilité à la maladie. De plus, le profil de ces MPTH

est altéré après l'infection au niveau des locus de gènes impliqués dans la réponse du système immunitaire (Luo *et al.*, 2012). Par exemple, le niveau de H3K4me3 augmente au locus du gène CD8a de façon concomitante à l'expression de ce gène dans la rate (Luo *et al.*, 2012). Dans le thymus, une tendance similaire a été observée au locus CTLA-4 (Mitra *et al.*, 2012). Récemment, les profils de H3K27ac et H3K4me3 dans les érythrocytes nucléés ont été publiés chez le poulet (Jahan *et al.*, 2016). Comme attendu, ces MPTH activatrices étaient associées à des gènes transcriptionnellement actifs et leurs distributions à l'échelle du génome entier étaient similaires à celles rapportées dans d'autres espèces (Barski *et al.*, 2007).

En conclusion, ces études tout-génome dans plusieurs lignées de poulet (Luo *et al.*, 2012 ; Mitra *et al.*, 2012, 2015 ; Jahan *et al.*, 2016) montrent des profils de MPTH similaires entre les oiseaux et les mammifères, en accord avec la haute conservation des mécanismes épigénétiques entre espèces allant des plantes aux mammifères (Suzuki et Bird, 2008).

## Conclusion

Avec la démocratisation des techniques de SHD, de nombreuses avancées dans les connaissances des génomes aviaires et de leurs épigénomes ont été réalisées. Cependant, ces développements nécessitent de prendre en compte, dès la conception des projets, les diverses recommandations des consortiums internationaux et les besoins en analyses bioinformatiques et biostatistiques.

L'étude du méthylome et des profils des MPTH met en évidence de nombreuses caractéristiques communes entre l'épigénome du poulet et celui

des mammifères : variations le long du génome liées à des caractéristiques génomiques telles que les TSS, corrélations de certaines marques présentes dans les régions promotrices avec l'expression des gènes, etc. Elles montrent des différences de profils de méthylation et de MPTH en fonction des tissus, de l'âge des individus ainsi que de la sensibilité des poulets aux pathogènes.

L'étude des épigénomes aviaires complétée par la cartographie des sites de chromatine ouverts (Buenrostro *et al.*, 2015) ou l'analyse de la conformation du génome (Belton *et al.*, 2012) devrait apporter une vision beaucoup plus claire des mécanismes de régulation de l'activité du génome liés au fonctionnement des organes, en contexte sain ou malade, ou plus généralement à la variation phénotypique. Ces données devraient permettre de mieux comprendre les mécanismes de réponse de l'épigénome face aux variations de l'environnement et donc faciliter la prise en compte progressive de ces mécanismes, quand c'est pertinent, dans l'amélioration des systèmes d'élevage, contribuant ainsi au développement d'une production avicole durable (Ibeagha-Awemu et Zhao, 2015).

## Remerciements

Ce travail est soutenu par la région Centre-Val de Loire, par le Département de génétique animale et le Département de physiologie animale et systèmes d'élevage de l'Institut National de la Recherche Agronomique (INRA). Il a été financé par l'ANR (programmes « ChickStress » – ANR-13-ADAP-0014 – et « QuailHeatE » – ANR-15-CE02-0009-01) et par le Département de physiologie animale et systèmes d'élevage de l'INRA (crédit incitatif INRA « Stressepimark »)

## Références

Andersson L., Archibald A.L., Bottema C.D., Brauning R., Burgess S.C., Burt D.W., Casas E., Cheng H.H., Clarke L., Couldrey C., Dalrymple B.P., Elsik C.G., Foissac S., Giuffra E., Groenen M.A., Hayes B.J., Huang L.S., Khatib H., Kijas J.W., Kim H., Lunney J.K., McCarthy F.M., McEwan J.C., Moore S., Nanduri B., Notredame C., Palti Y., Plastow

G.S., Reecy J.M., Rohrer G.A., Sarropoulou E., Schmidt C.J., Silverstein J., Tellam R.L., Tixier-Boichard M., Tosser-Klopp G., Tuggle C.K., 2015. Coordinated International action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Gen. Biol.*, 16, 57.

Bae J.B., 2013. Perspectives of International Human Epigenome Consortium. *Gen. Inf.* 11, 7.

Baker M., 2011. Making sense of chromatin states. *Nature Meth.*, 8, 717-722.

- Bannister A.J., Kouzarides T., 2011. Regulation of chromatin by histone modifications. *Cell Res.*, 21, 381-395.
- Barski A., Cuddapah S., Cui K., Roh T.Y., Schones D.E., Wang Z., Wei G., Chepelev I., Zhao K., 2007. High-resolution profiling of histone methylations in the human genome. *Cell*, 129, 823-837.
- Belton J.M., McCord R.P., Gibcus J.H., Naumova N., Zhan Y., Dekker J., 2012. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods (San Diego, Calif.)* 58, 268-276.
- Berger S.L., Kouzarides T., Shiekhattar R., Shilatifard A., 2009. An operational definition of epigenetics. *Genes Dev.*, 23, 781-783.
- Bernstein B.E., Stamatoyannopoulos J.A., Costello J.F., Ren B., Milosavljevic A., Meissner A., Kellis M., Marra M.A., Beaudet A.L., Ecker J.R., Farnham P.J., Hirst M., Lander E.S., Mikkelsen T.S., Thomson J.A., 2010. The NIH Roadmap Epigenomics Mapping Consortium. *Nature Biotechnol.*, 28, 1045-1048.
- Bird A., 1992. The essentials of DNA methylation. *Cell*, 70, 5-8.
- Bock C., 2012. Analysing and interpreting DNA methylation data. *Nature Reviews. Genetics* 13, 705-719.
- Buenrostro J.D., Wu B., Chang H.Y., Greenleaf W.J., 2015. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.*, 109, 21.29.1-21.29.9.
- Busche S., Shao X., Caron M., Kwan T., Allum F., Cheung W.A., Ge B., Westfall S., Simon M.M., The Multiple Tissue Human Expression Resource. Barrett A., Bell J.T., McCarthy M.J., Deloukas P., Blanchette M., Bourque G., Spector T.D., Lathrop M., Pastinen T., Grundberg E., 2015. Population whole-genome bisulfite sequencing across two tissues highlights the environment as the principal source of human methylome variation. *Genome Biol.*, 16, 290.
- Celniker S.E., Dillon L.A.L., Gerstein M.B., Gunsalus K.C., Henikoff S., Karpen G.H., Kellis M., Lai E.C., Lieb J.D., MacAlpine D.M., Micklem G., Piano F., Snyder M., Stein L., White K.P., Waterston R.H., 2009. Unlocking the secrets of the genome. *Nature* 459, 927-930.
- Cokus S.J., Feng S., Zhang X., Chen Z., Merriman B., Haudenschild C.D., Pradhan S., Nelson S.F., Pellegrini M., Jacobsen S.E., 2008. Shotgun bisulfite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, 452, 215-219.
- David S.A., Mersch M., Foissac S., Collin A., Pitel F., Coustham V., 2017a. Genome-wide epigenetic studies in chicken: A Review. *Epigenomes*, 1, 20.
- David S.A., Piégu B., Hennequet-Antier C., Pannetier M., Aguirre-Lavin T., Crochet S., Bordeau T., Couroussé N., Brionne A., Bigot Y., Collin A., Coustham V., 2017b. An assessment of fixed and native chromatin preparation methods to study histone post-translational modifications at a whole genome scale in skeletal muscle tissue. *Biol. Procedures Online*. 19, 10.
- Daxinger L., Whitelaw E., 2010. Transgenerational epigenetic inheritance: more questions than answers. *Genome Res.*, 20, 1623-1628.
- Derks M.F.L., Schachtschneider K.M., Madsen O., Schijlen E., Verhoeven K.J.F., van Oers K., 2016. Gene and transposable element methylation in great tit (*Parus major*) brain and blood. *BMC Genomics*, 17, 332.
- Doherty R., Couldrey C., 2014. Exploring genome wide bisulfite sequencing for DNA methylation analysis in livestock: a technical assessment. *Frontiers in Genetics*, 5, 126.
- ENCODE Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489, 57-74.
- Feil R., Fraga M.F., 2012. Epigenetics and the environment: emerging patterns and implications. *Nature Reviews. Genetics*, 13, 97-109.
- 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. *Proc. Nat. Acad. Sci. United States of America*, 89, 1827-1831.
- Gardiner-Garden M., Frommer M., 1987. CpG islands in vertebrate genomes. *J. Mol. Biol.*, 196, 261-282.
- 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 e15367.
- 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, 468-481.
- Hahn M.A., Li A.X., Wu X., Pfeifer G.P., 2015. Single base resolution analysis of 5-methylcytosine and 5-hydroxymethylcytosine by RRBS and TAB-RRBS. *Methods Mol. Biol.*, (Clifton, N.J.) 1238, 273-287.
- Han Y., He X., 2016. Integrating epigenomics into the understanding of biomedical insight. *Bioinf. Biol. Insights*, 10, 267-289.
- 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., Olshen A., Ballinger T., Zhou X., Forsberg K.J., Gu J., Echipare L., O'Geen H., Lister R., Pelizzola M., Xi Y., Epstein C.B., Bernstein B.E., Hawkins R.D., Ren B., Chung W.Y., Gu H., Bock C., Gnirke A., Zhang M.Q., Haussler D., Ecker J.R., Li W., Farnham P.J., Waterland R.A., Meissner A., Marra M.A., Hirst M., Milosavljevic A., Costello J.F., 2010. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nature Biotechnol.*, 28, 1097-1105.
- Hu Y., Xu H., Li Z., Zheng X., Jia X., Nie Q., Zhang X., 2013. Comparison of the genome-wide DNA methylation profiles between fast-growing and slow-growing broilers. *PLoS One* 8 e56411.
- Ibeagha-Awemu E.M., Zhao X., 2015. Epigenetic marks: Regulators of livestock phenotypes and conceivable sources of missing variation in livestock improvement programs. *Frontiers Genet.*, 6, 1-17.
- International Chicken Genome Sequencing Consortium, 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432, 695-716.
- Jahan S., Xu W., He S., Gonzalez C., Delcuve G.P., Davie J.R., 2016. The chicken erythrocyte epigenome. *Epigenetics Chromatin*, 9, 19.
- Johnson D.S., Mortazavi A., Myers R.M., Wold B., 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science (New York, N.Y.)* 316, 1497-1502.
- Jones P.A., 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Rev. Genet.*, 13, 484-492.
- Jung Y.L., Luquette L.J., Ho J.W.K., Ferrari F., Tolstorukov M., Minoda A., Issner R., Epstein C.B., Karpen G.H., Kuroda M.I., Park P.J., 2014. Impact of sequencing depth in ChIP-seq experiments. *Nucleic Acids Res.*, 42, e74.
- Kisliouk T., Meiri N., 2009. A critical role for dynamic changes in histone H3 methylation at the *Bdnf* promoter during postnatal thermotolerance acquisition. *Eur. J. Neurosci.*, 30, 1909-1922.
- Kornberg R.D., Lorch Y., 1999. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell*, 98, 285-294.
- Korthauer K., Irizarry R.A., 2018. Genome-wide repressive capacity of promoter DNA methylation is revealed through epigenomic manipulation. *BioRxiv* posted online Aug. 1, 2018 <https://doi.org/10.1101/381145>
- Kouzarides T., 2007. Chromatin modifications and their function. *Cell*, 128, 693-705.
- Kurdyukov S., Bullock M., 2016. DNA Methylation Analysis: Choosing the Right Method. *Biology*, 5, 3. <https://doi.org/10.3390/biology5010003>
- Laine V.N., Gossman 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., Corcoran P., The Great Tit HapMap Consortium, 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. *Nature, Communications*, 7, 10474.
- Landt S.G., Marinov G.K., Kundaje A., Kheradpour P., Pauli F., Batzoglou S., Bernstein B.E., Bickel P., Brown J.B., Cayting P., Chen Y., DeSalvo G., Epstein C., Fisher-Aylor K.I., Euskirchen G., Gerstein M., Gertz J., Hartemink A.J., Hoffman M.M., Iyer V.R., Jung Y.L., Karmakar S., Kellis M., Kharchenko P.V., Li Q., Liu T., Liu X.S., Ma L., Milosavljevic A., Myers R., Park J., Pazin M.J., Perry M.D., Raha D., Reddy T.E., Rozowsky J., Shores N., Sidow A., Slattey M., Stamatoyannopoulos J.A., Tolstorukov M.Y., White K.P., Xi S., Farnham P.J., Lieb J.D., Wold B.J., Snyder M.S., 2012. ChIP-seq guidelines and practices of the

- ENCODE and modENCODE consortia. *Genome Res.*, 22, 1813-1831.
- Langmead B., Trapnell C., Pop M., Salzberg S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, 10, R25.
- Li H., Durbin R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*. 25, 1754-1760.
- Li R., Yu C., Li Y., Lam T.W., Yiu S.M., Kristiansen K., Wang J., 2009. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics (Oxford, England)*. 25, 1966-1967.
- Li Q., Li N., Hu X., Li J., Du Z., Chen L., Yin G., Duan J., Zhang H., Zhao Y., Wang J., 2011. Genome-wide mapping of DNA methylation in chicken. *PLoS One*, 6, e19428.
- Li J., Li R., Wang Y., Hu X., Zhao Y., Li L., Feng C., Gu X., Liang F., Lamont S.J., Hu S., Zhou H., Li N., 2015. Genome-wide DNA methylome variation in two genetically distinct chicken lines using MethylC-seq. *BMC Genomics*, 16, 851. <https://doi.org/10.1186/s12864-015-2098-8>
- 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., Harvey Millar A., Thomson J.A., Ren B., Ecker J.R., 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462, 315-322.
- Lu D., 2013. Epigenetic modification enzymes: catalytic mechanisms and inhibitors. *Acta Pharmaceutica Sinica B*, 3, 141-149.
- Luo J., Mitra A., Tian F., Chang S., Zhang H., Cui K., Yu Y., Zhao K., Song J., 2012. Histone methylation analysis and pathway predictions in chickens after MDV infection. *PLoS One* 7 e41849.
- Manolio T.A., Collins F.S., Cox N.J., Goldstein D.B., Hindorf L.A., Hunter D.J., McCarthy M.I., Ramos E.M., Cardon L.R., Chakravarti A., Cho J.H., Guttmacher A.E., Kong A., Kruglyak L., Mardis E., Rotimi C.N., Slatkin M., Valle D., Whittemore A.S., Boehnke M., Clark A.G., Eichler E.E., Gibson G., Haines J.L., Mackay T.F.C., McCarroll S.A., Visscher P.M., 2009. Finding the missing heritability of complex diseases. *Nature*, 461, 747-753.
- Mardis E.R., 2008. The impact of next-generation sequencing technology on genetics. *Trends in Genetics : TIG*, 24, 133-141.
- McQueen H.A., Fantes J., Cross S.H., Clark V.H., Archibald A.L., Bird A.P., 1996. CpG islands of chicken are concentrated on microchromosomes. *Nature Genet.*, 12, 321-324.
- Mitra A., Luo J., Zhang H., Cui K., Zhao K., Song J., 2012. Marek's disease virus infection induces widespread differential chromatin marks in inbred chicken lines. *BMC Genomics*, 13, 557.
- Mitra A., Luo J., He Y., Gu Y., Zhang H., Zhao K., Cui K., Song J., 2015. Histone modifications induced by MDV infection at early cytolytic and latency phases. *BMC Genomics*, 16, 311.
- Ponnaluri V.K.C., Ehrlich K.C., Zhang G., Lacey M., Johnston D., Pradhan S., Ehrlich M., 2017. Association of 5-hydroxymethylation and 5-methylation of DNA cytosine with tissue-specific gene expression. *Epigenetics*, 12, 123-138.
- Rhee H.S., Pugh B.F., 2011. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell*, 147, 1408-1419.
- Rigauil G., Balzergue S., Brunaud V., Blondet E., Rau A., Rogier O., Caius J., Maugis-Rabusseau C., Soubigou-Taconnat L., Aubourg S., Lurin C., Martin-Magniette M.L., Delannoy E., 2018. Synthetic data sets for the identification of key ingredients for RNA-seq differential analysis. *Brief. Bioinform.* 19, 65-76.
- Robertson G., Hirst M., Bainbridge M., Bilenky M., Zhao Y., Zeng T., Euskirchen G., Bernier B., Varhol R., Delaney A., Thiessen N., Griffith O.L., He A., Marra M., Snyder M., Jones S., 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature Methods*, 4, 651-657.
- Rothbart S.B., Dickson B.M., Raab J.R., Grzybowski A.T., Krajewski K., Guo A.H., Shanle E.K., Josefowicz S.Z., Fuchs S.M., Allis C.D., Magnuson T.R., Ruthenburg A.J., Strahl B.D., 2015. An interactive database for the assessment of histone antibody specificity. *Mol. Cell.*, 59, 502-511.
- Roundtree I.A., Evans M.E., Pan T., He C., 2017. Dynamic RNA Modifications in Gene Expression Regulation. *Cell.*, 169, 1187-1200.
- Simpson J.T., Workman R.E., Zuzarte P.C., David M., Dursi L.J., Timp W., 2017. Detecting DNA cytosine methylation using nanopore sequencing. *Nature Methods*, 14, 407-410.
- Smith Z.D., Gu H., Bock C., Gnirke A., Meissner A., 2009. High-throughput bisulfite sequencing in mammalian genomes. *Methods (San Diego, Calif.)* 48, 226-232.
- Stein L., 2001. Genome annotation: from sequence to biology. *Nature Reviews. Genetics*, 2, 493-503.
- Suzuki M.M., Bird A., 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nature Rev. Genet.*, 9, 465-476.
- Tessarz P., Kouzarides T., 2014. Histone core modifications regulating nucleosome structure and dynamics. *Nature Rev. Mol. Cell Biol.*, 15, 703-708.
- The ENCODE Project Consortium Whole-Genome Bisulfite Sequencing Data Standards and Processing Pipeline.
- Tian F., Zhan F., VanderKraats N.D., Hiken J.F., Edwards J.R., Zhang H., Zhao K., Song J., 2013. DNMT gene expression and methylome in Marek's disease resistant and susceptible chickens prior to and following infection by MDV. *Epigenetics*, 8, 431-444.
- Triantaphyllopoulos K.A., Ikononopoulos I., Bannister A.J., 2016. Epigenetics and inheritance of phenotype variation in livestock. *Epigenetics Chromatin*, 9, 31.
- 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, 475-483.
- Verhulst E.C., Mateman A.C., Zwier M.V., Caro S.P., Verhoeven K.J.F., van Oers K., 2016. Evidence from pyrosequencing indicates that natural variation in animal personality is associated with DRD4 DNA methylation. *Mol. Ecol.*, 25, 1801-1811.
- Wolff G.L., Kodell R.L., Moore S.R., Cooney C.A., 1998. Maternal epigenetics and methyl supplements affect agouti gene expression in *Avy/a* mice. *FASEB J. Official Publication Fed. Amer. Soc. Exp. Biol.*, 12, 949-957.
- Zhang T., Cooper S., Brockdorff N., 2015. The interplay of histone modifications – writers that read. *EMBO Reports*, 16, 1467-1481.
- Zhang M., Yan F.B., Li F., Jiang K.R., Li D.H., Han R.L., Li Z.J., Jiang R.R., Liu X.J., Kang X.T., Sun G.R., 2017. Genome-wide DNA methylation profiles reveal novel candidate genes associated with meat quality at different age stages in hens. *Sci. Reports*, 7, 45564.
- Zhang Z., Du H., Bai L., Yang C., Li Q., Li X., Qiu M., Yu C., Jiang Z., Jiang X., Liu L., Hu C., Xia B., Xiong X., Song X., Jiang X., 2018. Whole genome bisulfite sequencing reveals unique adaptations to high-altitude environments in Tibetan chickens. *PLoS ONE* 13.
- Ziller M.J., Hansen K.D., Meissner A., Aryee M.J., 2015. Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. *Nat. Methods*, 12, 230-232.

## Résumé

Comme toutes les espèces de rente, les oiseaux d'élevage ont été sélectionnés génétiquement afin d'augmenter leurs performances. Cependant, de nombreuses études ont montré que le phénotype d'un individu n'est pas le simple résultat de son patrimoine génétique. En effet, il est également sensible aux variations de l'environnement (température, nutrition...) qui peuvent notamment induire des altérations de marques épigénétiques aboutissant à des changements durables des programmes d'expression de gènes. Il est donc essentiel de comprendre comment les épigénomes sont régulés afin d'envisager de futurs leviers ou marqueurs d'adaptation des animaux. Aujourd'hui, le séquençage à haut-débit est devenu relativement abordable et les outils bio-informatiques matures pour envisager l'étude des marques

épigénétiques à l'échelle du génome. Chez les oiseaux, un nombre croissant d'études utilisent ces technologies à haut-débit pour explorer les mécanismes épigénétiques impliqués dans des processus tels que l'immunité ou l'adaptation à l'environnement. Dans cette synthèse, nous décrivons en quoi ces technologies ont contribué à enrichir les connaissances sur l'épigénome aviaire et plus particulièrement celui du poulet, en nous focalisant sur les deux types de modifications épigénétiques les plus étudiées, la méthylation de l'ADN et les modifications post-traductionnelles des histones. Nous présentons également les concepts nécessaires à la conception et la réalisation des analyses haut-débit des épigénomes aviaires. En plus des connaissances fondamentales fortement attendues par la communauté scientifique, une meilleure compréhension des mécanismes épigénétiques à l'origine de la régulation de l'expression génique en réponse aux modifications environnementales chez l'oiseau pourrait contribuer au développement d'une production avicole durable.

## Abstract

### ***New insights into the avian epigenome from high-throughput sequencing experiments***

*Over the years, farmed birds have been genetically selected to improve performance traits. However, many studies have shown that phenotype not only depends on genetics: gene expression is also influenced by the environment, through epigenetic changes that may alter the phenotypes of the animals. Since high-throughput sequencing techniques are extremely developing and become less expensive, a growing number of studies took advantage of them to gain insights into the epigenetic mechanisms of gene regulation in birds. Here we review the current gain of knowledge on the chicken epigenome, focusing on the two most studied epigenetic modifications (DNA methylation and histones post-translational modifications) based on recent data obtained with high-throughput sequencing techniques. We also provide advices about designing and performing analyses to further explore avian epigenomes. A better understanding of the molecular mechanisms underlying the epigenetic involvement in the bird phenotypes may provide new markers that should undoubtedly contribute to improve the sustainability of poultry production.*

MERSCH M., DAVID S.-A., VITORINO CARVALHO A., FOISSAC S., COLLIN A., PITEL F., COUSTHAM V., 2018. Apports du séquençage haut-débit sur la connaissance de l'épigénome aviaire. INRA Prod. Anim., 31, 324-336.

<https://doi.org/10.20870/productions-animales.2018.31.4.2372>

