



HAL
open science

Épigénétique chez les poissons d'élevage

Audrey Laurent, Delphine Lallias, Lucie Marandel, Catherine Labbé, Pierre Boudry

► **To cite this version:**

Audrey Laurent, Delphine Lallias, Lucie Marandel, Catherine Labbé, Pierre Boudry. Épigénétique chez les poissons d'élevage. Épigénétique - Mécanismes moléculaires, biologie du développement et réponses à l'environnement, QUAE, pp.163, 2022, 978-2-7592-3769-2. <hal-04692663>

HAL Id: hal-04692663

<https://hal.inrae.fr/hal-04692663v1>

Submitted on 10 Sep 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.



HAL Authorization

Épigénétique

Mécanismes moléculaires,
biologie du développement
et réponses à l'environnement

Hélène Jammes, Pierre Boudry,
Stéphane Maury, coordinateurs

Éditions Quæ

Collection Synthèses

L'antibiorésistance : Un fait social total

C. Harpet, coord.

2022, 168 p.

Zéro pesticide.

*Un nouveau paradigme de recherche
pour une agriculture durable*

F. Jacquet, M.-H. Jeuffroy, J. Jouan,

E. Le Cadre, T. Malausa, X. Reboud,

C. Huyghe, coord.

2022, 244 p.

*Durabilité des systèmes pour la sécurité
alimentaire. Combiner les approches
locales et globales*

A. Thomas, A. Alpha, A. Barczak,

N. Zakhia-Rozis, coord.

2024, 246 p.

*One health, une seule santé :
théorie et pratique des approches
intégrées de la santé*

J. Zinsstag, E. Schelling,

D. Waltner-Toews, M. A. Whittaker,

M. Tanner, coord.

2021, 584 p.

Pour citer cet ouvrage :

Hélène Jammes, Pierre Boudry, Stéphane Maury, 2024. *Épigénétique. Mécanismes moléculaires, biologie du développement et réponses à l'environnement*. Versailles, éditions Quæ, 196 p.

Éditions Quæ

RD 10, 78026 Versailles Cedex

www.quae.com – www.quae-open.com

© Éditions Quæ, 2024

ISBN (papier): 978-2-7592-3769-2

ISBN (ePub): 978-2-7592-3771-5

ISBN (pdf): 978-2-7592-3770-8

ISSN: 1777-4624

The digital versions of these books are distributed under the CC-by-NC-ND 4.0 license.

Sommaire

Remerciements	7
Introduction	9
<i>Hélène Jammes, Stéphane Maury, Pierre Boudry</i>	
Chapitre 1. Épigenétique, un mot dans l'histoire des sciences	13
<i>Gaëlle Pontarotti</i>	
L'épigénétique, un mot qui a une histoire	13
L'épigénétique, une question transversale en biologie.....	16
L'épigénétique, un concept aux multiples enjeux épistémologiques et sociaux.....	18
Conclusion	21
Références bibliographiques	21
Chapitre 2. Épigenétique, une exploration des processus moléculaires	23
<i>Hélène Jammes et David L'Hôte</i>	
Le patrimoine génétique dans tous ses états	23
Le « code des histones ».....	25
La méthylation de l'ADN, marque épigénétique la plus explorée.....	34
Activités des longs et petits ARN non codants dans les régulations épigénétiques.....	40
Les protéines Polycomb et Trithorax.....	42
Conclusion	42
Références bibliographiques	42
Chapitre 3. L'inactivation du chromosome X	45
<i>Clara Roidor, Véronique Duranthon, Maud Borensztein</i>	
La compensation de dose génique	46
Découverte de l'inactivation du chromosome X	48
La puissance des modèles murins.....	52
Le cycle d'inactivation du chromosome X chez les rongeurs, un modèle à part?.....	54
Diversité du processus d'inactivation du chromosome X chez les euthériens.....	57
Avantages et pathologies liés à l'inactivation du chromosome X.....	61
Conclusion et perspectives	63
Références bibliographiques	64
Chapitre 4. Empreinte génomique parentale : découverte et mécanismes de régulation	67
<i>Thierry Forné</i>	
Évolution et empreinte génomique parentale.....	68
Origine moléculaire des mécanismes de l'empreinte génomique	70
Contrôle épigénétique des gènes soumis à l'empreinte parentale	70

Centres d’empreinte et cycle de l’empreinte génomique parentale	73
Exemples de mécanismes moléculaires des empreintes fonctionnelles.....	78
Conclusion	81
Références bibliographiques	82
Chapitre 5. Empreinte génomique parentale chez les animaux d’élevage	83
<i>Julie Demars, Catherine Labbé, Frédérique Pitel</i>	
Empreinte génomique parentale et variabilité phénotypique	83
Caractérisation des mécanismes d’empreinte chez les animaux d’élevage.....	85
Implication potentielle pour l’élevage.....	89
Références bibliographiques	90
Chapitre 6. Empreinte parentale chez les plantes : mécanismes, fonctions et applications	93
<i>Clément Lafon Placette</i>	
Rappel sur la reproduction sexuée chez les plantes.....	93
Mécanismes épigénétiques aboutissant à l’empreinte parentale	94
Mort de la graine hybride : l’empreinte parentale, <i>usual suspect</i>	97
Causes évolutives de l’apparition de l’empreinte parentale chez les plantes	98
Application pour l’amélioration des plantes	101
Références bibliographiques	102
Chapitre 7. La vie <i>in utero</i> : programmation épigénétique des caractères.....	105
<i>Anne Gabory</i>	
Les origines développementales des maladies.....	105
Épigénétique et DOHaD	109
Épigénétique et variation de réponse selon le sexe.....	114
Conclusion : les origines développementales de la santé	115
Références bibliographiques	116
Chapitre 8. Horloge épigénétique	119
<i>Sarah Voisin</i>	
Montre-moi ton épigénome, et je te dirai ton âge.....	119
Une horloge épigénétique, ce n’est pas sorcier!.....	120
Âge biologique <i>versus</i> âge chronologique.....	124
Peut-on avoir simultanément un foie vieux et un muscle jeune?.....	127
Aucune espèce animale n’échappe au vieillissement épigénétique	128
Le vieillissement s’accompagne d’une érosion du méthylome.....	130
Conclusion	131
Références bibliographiques	131
Chapitre 9. Mécanismes épigénétiques des interactions hôte-agent pathogène	133
<i>Christoph Grunau, Isabelle Fudal, Nadia Ponts</i>	
Modification de l’épigénome à la suite d’une infection par des agents pathogènes	133
Mémoire épigénétique et <i>priming</i> des défenses des plantes lors de l’infection.....	135
Rôle de la méthylation de l’ADN dans l’induction de réactions de défense chez les plantes.....	136
Questions futures.....	141
Références bibliographiques	143

Chapitre 10. Mémoires épigénétiques et santé des plantes	145
<i>Philippe Gallusci, Margot M.J. Berger</i>	
La mémoire des plantes : importance des mécanismes épigénétiques.....	146
L'acclimatation des plantes : mémoire somatique et santé des plantes	148
Mémoire épigénétique inter-transgénérationnelle des stress :	
conséquences pour la santé des plantes.....	153
Diversité épigénétique et amélioration des plantes cultivées.....	155
Conclusion	157
Références bibliographiques	159
Chapitre 11. Épigénétique chez les poissons d'élevage	163
<i>Audrey Laurent, Delphine Lallias, Lucie Marandel, Catherine Labbé, Pierre Boudry</i>	
Des génomes dupliqués complexes à l'origine de nouveaux régulateurs épigénétiques	163
Originalité du support génétique du méthylome chez les téléostéens.....	164
Mécanismes épigénétiques et modulation des réponses physiologiques intragénérationnelles.....	165
Transmission intergénérationnelle d'informations environnementales.....	171
Conclusion	173
Références bibliographiques	174
Chapitre 12. Édition de l'épigénome	175
<i>David L'Hôte</i>	
Les outils de l'édition de l'épigénome.....	176
L'édition de l'épigénome au service de la santé	185
Conclusion	189
Références bibliographiques	190
Liste des auteurs	191
Liste des abréviations	193

Remerciements

Nous remercions vivement les auteurs qui ont participé à cette aventure. Ils ont tous, avec leur enthousiasme, su prendre le temps pour expliciter, illustrer et synthétiser leur thématique de recherche, décrire les avancées des connaissances, pas à pas, présenter les arguments scientifiques, définir les limites des savoirs et souligner les questionnements d'aujourd'hui.

Nous avons une pensée pour notre collègue Hélène Bierne, directrice de recherche INRAE, qui nous a quittés trop tôt. Elle aurait su nous faire partager ses connaissances de *Listeria*, bactérie ciblant la machinerie épigénétique de ses hôtes.

Introduction

HÉLÈNE JAMMES, STÉPHANE MAURY, PIERRE BOUDRY

Pourquoi l'épigénétique? À quel public est destiné cet ouvrage? Avec qui se lancer dans cette aventure?

Répondre à la première question, c'était faire la constatation que l'épigénétique est à la fois un champ de recherche scientifique qui mobilise de très nombreux chercheurs à l'échelle mondiale et qui génère une littérature foisonnante¹); une discipline qui fait la une du magazine américain *Time*, l'objet de chroniques sur France Inter ou dans le journal *Le Monde*, et d'une série d'émissions sur France Culture; un argument publicitaire pour l'industrie cosmétique ou qui suscite un intérêt pour les vendeurs de bien-être acquis par un contrôle de ses gènes, pour tous ceux en quête de qualité de vie. Il y avait donc clairement utilité et intérêt d'exprimer, dans ce paysage plus ou moins médiatisé, le point de vue des scientifiques, afin de mettre à disposition de la curiosité des lecteurs des faits expérimentaux, des hypothèses validées et les dernières connaissances acquises.

Dans cet ouvrage, l'épigénétique est entendue, sous sa définition contemporaine, comme la science des molécules et des mécanismes impliqués dans l'architecture de la chromatine et responsables de la régulation de l'expression des gènes sans modification de la séquence de l'ADN génomique nucléaire (Cavalli et Heard, 2019)². Autrement dit, est présenté ici un ensemble de connaissances et de résultats d'études focalisées sur des molécules et des mécanismes intervenant au sein du noyau des organismes eucaryotes, pour piloter des états alternatifs de l'expression des gènes dans un contexte de séquences d'ADN identiques. Informer sur l'épigénétique, c'est décrire la complexité et la diversité des mécanismes en jeu au sein du monde vivant dans différents processus biologiques, avec autant de rigueur que possible, en s'affranchissant de toute simplification abusive mais en essayant de les rendre compréhensibles à un large lectorat. Nous avons pensé en priorité aux étudiants, mais aussi aux enseignants, aux chercheurs dans différents domaines et à un public plus large volontaire pour apprendre encore et toujours.

1. Plus de 135685 publications scientifiques référencées dans la base données PubMed du National Institutes of Health depuis 1958.

2. Advances in epigenetics link genetics to the environment and disease, *Nature*. 2019, 571(7766):489-499. doi: 10.1038/s41586-019-1411-0. En anglais, la définition énoncée par Giacomo Cavalli et Edith Heard est : "The study of molecules and mechanisms that can perpetuate alternative gene activity states in the context of the same DNA sequence".

La coordination de cet ouvrage est issue de la rencontre entre des biologistes d'horizons divers : une biologiste de la reproduction animale et épigénétique, un spécialiste de la physiologie et de l'épigénétique des plantes, et un généticien de l'aquaculture. Tous trois sont animés par une vision commune et fédératrice de l'épigénétique : moléculaire, dynamique, héritable, impliquée dans le maintien de l'identité cellulaire et la mémorisation de l'environnement à court et à long termes, après disparition du *stimulus* initial. Au vu de l'importance de l'épigénétique dans de multiples domaines de la biologie, être exhaustif était une tâche impossible : nous avons donc fait des choix. Au gré de la construction des chapitres, nous avons balayé la diversité du vivant avec quelques exemples chez l'homme, tout en favorisant des exemples chez des animaux d'élevage, terrestres et aquatiques, chez des invertébrés et chez des plantes. Dans cet ouvrage, la place de l'épigénétique dans la compréhension des mécanismes moléculaires à l'origine de pathologies humaines (cancers, diabète, l'obésité, etc.) n'est donc évoquée que ponctuellement.

Il a fallu aussi des rencontres : de celles que chacun a pu faire au cours de sa carrière, de celles qui construisent, qui aident à réfléchir, à prendre du recul, qui sortent du rang. Nous aurions aimé citer ici quelques noms, les plus emblématiques pour chacun de nous. Il nous est néanmoins trop difficile de faire un choix, ils resteront anonymes. Que chacun soit remercié ici.

Contactés par les éditions Quæ pour réaliser cette aventure, nous avons construit cet ouvrage en associant diverses histoires, à la façon d'un recueil de nouvelles, rédigées indépendamment mais coordonnées et harmonisées. Chaque chapitre reste néanmoins sous la responsabilité des auteurs qui ont tenu à illustrer les multiples façons d'appréhender le champ d'investigation offert par l'épigénétique. « Définir l'épigénétique » nous a semblé être la porte d'entrée. Les deux premiers chapitres s'y attellent. Dans le premier, le point de vue d'une philosophe et historienne des sciences évoque l'émergence des concepts au cours du temps, leurs oppositions et leurs complémentarités. Le second chapitre retrace l'histoire des découvertes et des preuves expérimentales qui ont jeté les bases moléculaires des connaissances d'aujourd'hui et abouti aux mécanismes d'action communément acceptés. Les cinq chapitres suivants retracent la compréhension que nous avons, après de nombreuses années de recherche expérimentale, des processus fondamentaux, nécessaires pour assurer le développement des individus dès les premières phases de leur vie. La diversité des mécanismes y est décrite, les processus se déclinant aussi bien chez les animaux que chez les plantes. Le huitième chapitre mêle modèles mathématiques et données moléculaires épigénétiques et présente la notion d'« horloge épigénétique », redéfinissant le vieillissement chez de nombreuses espèces animales. Les deux chapitres suivants détaillent la contribution des mécanismes épigénétiques dans la compréhension des relations entre hôte et agents pathogènes et dans la réponse à long terme qu'elles peuvent générer, que ce soit chez des invertébrés ou chez les plantes. Y sont aussi décrits les développements possibles pour le futur. Le onzième chapitre évoque la contribution de l'épigénétique chez les poissons, organismes particulièrement exposés aux modifications de leur milieu de vie (qualité de l'eau, nutrition, température, etc.). Enfin, le dernier chapitre présente les possibilités de modifications de l'épigénome de manière ciblée et contrôlée; loin de clore cet ouvrage, il ouvre l'horizon vers le champ des possibles.

Les recherches en épigénétique, comme pour toute la biologie, avancent à grands pas, les technologies développées étant de plus en plus performantes. De plus en plus de données sont générées et intégrées, de nouvelles hypothèses et de nouveaux modèles sont proposés, testés, validés ou infirmés. Les connaissances en épigénétique restent un gigantesque puzzle dont les pièces minuscules, petites briques de savoir de plus en plus nombreuses, s'assemblent progressivement. L'épigénétique n'est qu'une des multiples facettes de la biologie, dont les mécanismes ont été explorés ces dernières décennies. Là réside tout le défi de cet ouvrage : présenter, expliquer, illustrer, donner des clés de compréhension. C'est avec cette ambition que nous nous sommes mis à l'ouvrage et espérons avoir atteint notre objectif.

Chapitre 1

Épigénétique, un mot dans l'histoire des sciences

GAËLLE PONTAROTTI

Renvoyant à ce qui se situe autour ou sur le gène, le concept d'épigénétique fait l'objet de nombreuses définitions dans différents champs de recherche. Les travaux conduits en épigénétique annonceraient par ailleurs la disgrâce des approches relevant du « tout génétique », et induiraient une véritable révolution dans nos conceptions sur le développement, l'hérédité et l'évolution. Dans ce contexte, on peut se demander si l'épigénétique est un mot comme un autre dans l'histoire des sciences, un concept pouvant désigner un ensemble circonscrit de phénomènes, ou si ce terme s'apparente plutôt à une question ouverte, à l'image de l'activité scientifique elle-même.

Dans un premier temps, il s'agira de revenir sur l'histoire du concept d'épigénétique et sur l'évolution des définitions qui en ont été proposées. Ce bref aperçu historique permettra de mettre en évidence que l'épigénétique s'est fondamentalement présentée, au fil du temps, comme une discipline complémentaire de la génétique. À la lumière de ce constat, il conviendra dans un second temps de dresser une cartographie des domaines dans lesquels l'épigénétique complète la génétique, et de revenir sur la façon dont cette discipline met à l'honneur le rôle de l'environnement dans les différents phénomènes biologiques. Enfin, la dernière partie du chapitre proposera des éléments de réflexion sur la caractérisation du changement scientifique induit par l'épigénétique, mais aussi sur les usages sociopolitiques des données.

► L'épigénétique, un mot qui a une histoire

Petite histoire de l'épigénétique

Les historiens des sciences s'accordent à dire que c'est dans les années 1940 que le terme « épigénétique » a fait son apparition dans le discours scientifique. Le mot aurait plus précisément été estampillé par Conrad Waddington (1942) pour désigner l'étude de l'ensemble des mécanismes développementaux permettant à un génotype de produire des phénotypes. Pour le biologiste, il s'agit alors de penser l'articulation entre deux disciplines séparées depuis le début du ^{xx}e siècle, à savoir l'embryologie – science du développement – et la génétique – science de l'hérédité. Pour illustrer ses idées, Waddington (1957) élabore la désormais célèbre métaphore du paysage épigénétique.

Celle-ci suggère que les différentes trajectoires développementales (représentées par les vallées du paysage) sont façonnées par les interactions à l'œuvre entre les gènes (représentés par des piquets sous-jacents) et leurs produits. Il convient ici de noter que le mot « épigénétique » fait écho à la vieille théorie de l'épigénèse qui, aux XVII^e et XVIII^e siècles, s'oppose à celle de la préformation. La première propose en substance que le développement d'un organisme repose sur une différenciation progressive de ses parties, alors que la seconde considère que chaque organisme est déjà préformé dans les germes parentaux.

À la fin des années 1950, David Nanney (1958) associe épigénétique et différenciation cellulaire. Il s'agit pour lui de comprendre comment des cellules abritant les mêmes gènes peuvent se différencier au fil du temps, devenir par exemple des neurones, des hépatocytes, des lymphocytes ou encore des fibroblastes. Nanney décrit deux systèmes de contrôle cellulaire à l'œuvre dans le développement : le système génétique et le système épigénétique. Le système génétique désigne les mécanismes de réplication basés sur un modèle et permettant le maintien d'une bibliothèque de spécificités. Le système épigénétique renvoie quant à lui aux mécanismes auxiliaires déterminant quelles spécificités doivent être exprimées dans les cellules. D'autres biologistes participent de ce tournant cellulaire de l'épigénétique. Ils s'intéressent par exemple à l'hérédité extrachromosomique (et plus précisément à la transmission des mitochondries présentes dans le cytoplasme), ainsi qu'à la transmission des états fonctionnels du noyau à travers les divisions cellulaires (pour une analyse plus complète, voir Haig, 2004).

Dans les années 1970, l'épigénétique tend à devenir moléculaire (voir chapitre 2). Des scientifiques enquêtent alors sur l'implication des histones (protéines constitutives de la chromatine) dans la régulation de la synthèse de l'ARN et commencent à décrire les modifications post-traductionnelles de ces histones (Allfrey et Mirsky, 1964). D'autres étudient le rôle de la méthylation de l'ADN dans la régulation de l'expression des gènes, en particulier dans le cadre des premiers travaux sur l'inactivation du chromosome X chez les mammifères femelles (Riggs, 1975 ; voir chapitre 3). Dans la continuité de ces recherches, les biologistes contemporains étudient aussi l'activité des petits ARN non codants (voir chapitre 2). Ils tentent plus généralement d'appréhender l'ensemble des marques épigénétiques et leurs interactions à l'échelle du génome entier. On parle alors d'épigénomique.

Peut-on distinguer deux épigénétiques ?

Ce bref parcours historique permet d'apercevoir que les études relevant de l'épigénétique se sont intéressées à différentes échelles d'analyse, allant des organismes aux molécules, en passant par les cellules. La question est alors de savoir s'il existe un lien entre ces travaux, ou si les spécialistes ont, au fil du temps, employé un même terme pour désigner plusieurs domaines de recherche. En d'autres termes, l'épigénétique a-t-elle historiquement constitué un champ d'étude unifié ?

Au premier abord, toutes les études mentionnées, même si elles se focalisent sur des échelles diverses, interrogent le même processus, à savoir l'ontogenèse, c'est-à-dire le développement progressif d'un organisme depuis sa conception jusqu'à sa forme mature. De ce point de vue, tous les travaux relevant de l'épigénétique auraient affronté le problème du développement, mais à différents niveaux, sachant que le processus par lequel un organisme adulte est progressivement produit, à partir de l'œuf fécondé,

implique un phénomène de différenciation cellulaire par la régulation de l'expression des gènes (Nicoglou et Merlin, 2017). En d'autres termes, l'épigénétique aurait toujours désigné « l'étude analytique du développement individuel (ontogenèse) avec son problème central de la différenciation cellulaire » (Huxley, 1957, cité par Haig, 2004).

Toutefois, il est également possible de penser une séparation plus nette entre deux conceptions de l'épigénétique. Il s'agit alors de distinguer « deux usages historiques [...] correspondant à des contextes scientifiques très différents » (Morange 2005), ou encore deux origines de l'épigénétique (Haig, 2004). La première conception de l'épigénétique, qui peut être rattachée à la figure tutélaire de Waddington, s'intéresse au développement en général; la seconde, pouvant être associée au nom de Nanney, traite avant tout le problème de la différenciation cellulaire, du contrôle de l'expression des gènes. Dans une veine similaire, quoique légèrement différente, une distinction pourrait être établie entre une épigénétique systémique d'une part, et une épigénétique moléculaire et réductionniste, d'autre part (Baedke, 2018). La première, étudiant des réseaux de gènes, serait notamment à l'œuvre dans la biologie des cellules souches. La seconde, s'intéressant aux facteurs non génétiques impliqués dans les modifications héréditaires de la chromatine et dans la régulation génétique, partagerait des traits majeurs avec la génétique moléculaire.

L'épigénétique mobilisée en biologie de l'évolution peut être d'inspiration développementale, lorsqu'il s'agit de comprendre le rôle des mécanismes développementaux dans la transformation des espèces. Mais elle peut aussi être cellulaire et moléculaire, si les biologistes entendent surtout interroger l'existence d'une variation héréditable non génétique qui pourrait être scrutée par la sélection naturelle (dans le contexte de la synthèse évolutive étendue³) (Nicoglou et Merlin, 2017).

Ainsi, même si la question commune à tous les travaux relevant de l'épigénétique semble être celle du développement, les échelles considérées introduisent de la diversité dans la perception du concept qui nous intéresse et des mécanismes auxquels il renvoie.

L'épigénétique, un concept aux nombreuses définitions

Comment, dès lors, définir l'épigénétique? Dans quelle mesure est-il possible d'en esquisser des contours précis? Un rapide examen de la littérature permet de constater que les définitions sont plurielles, même si elles semblent pouvoir être rattachées, pour la plupart, aux deux traditions présentées ci-dessus.

Inaugurant la tradition développementale, Waddington (1942) définit l'épigénétique comme l'étude des processus développementaux qui permettent de produire le phénotype à partir du génotype. Dans la même veine, Jablonka et Raz (2009) soutiennent que l'épigénétique est « l'étude des processus qui sous-tendent la plasticité développementale et la canalisation et qui entraînent des effets développementaux persistants à la fois chez les procaryotes et les eucaryotes ». Dans ce contexte, l'épigénotype désigne des processus développementaux (Waddington, 1942).

3. La synthèse étendue entend dépasser et compléter la théorie synthétique de l'évolution élaborée dans les années 1920-1950. Alors que la seconde fait la synthèse entre un groupe restreint de disciplines, notamment la génétique mendélienne et la biologie de l'évolution néo-darwinienne, la première entend inclure d'autres disciplines, comme la biologie développementale, dans les recherches des biologistes de l'évolution.

D'autres définitions relèvent davantage de la tradition cellulaire et moléculaire. Elles présentent l'épigénétique comme « l'hérédité nucléaire qui n'est pas basée sur des différences dans la séquence d'ADN » (Holliday, 2006) ou encore comme « l'étude des changements héréditaires dans les fonctions des gènes à travers la mitose et/ou la méiose qui ne peuvent pas être expliqués par des changements dans la séquence d'ADN »⁴ (Riggs et al., 1996). Autrement dit, l'épigénétique renvoie ici au « contrôle de l'activité des gènes par méthylation de l'ADN ou modification des composants de la chromatine » (Morange, 2005). Dans ce contexte, l'épigénotype est envisagé comme un état fonctionnel de la cellule (Holliday, 2006).

Il est toutefois à noter qu'une articulation est possible entre l'échelle de l'organisme et celle des molécules. Ainsi, Nicoglou et Merlin (2017) décrivent l'épigénétique comme « l'étude de plusieurs facteurs intracellulaires qui ont un effet sur la stabilité des processus développementaux à travers leur action sur les potentialités du génome (c'est-à-dire, la susceptibilité du génome d'être exprimé de façon différentielle) ». À travers cette définition synthétique, les philosophes explicitent le rapport implicite qui existe dès le départ entre la question du développement et celle de la différenciation cellulaire.

Notons enfin que d'autres définitions, peu rigoureuses, s'écartent des cadres développementaux et moléculaires mentionnés ci-dessus. Ainsi, l'épigénétique peut renvoyer, de façon assez vague, à tout ce qui n'est pas génétique. D'aucuns ont même affirmé qu'elle introduit l'idée de libre arbitre dans nos conceptions de la génétique (Jirtle, cité par E. Watters, 2006, dans la revue *Discover*) !

Faut-il voir dans la diversité des définitions de l'épigénétique un obstacle au progrès des connaissances ? Une première option consiste à soutenir que la polysémie du concept d'épigénétique signale l'aspect lacunaire de nos connaissances, en cohérence avec l'idée selon laquelle l'épigénétique, au cours de l'histoire, est toujours venue combler les insuffisances de la génétique (Morange 2005 ; voir comme exemple l'empreinte parentale, chapitre 4). Cependant, selon un tout autre point de vue, l'imprécision définitionnelle de l'épigénétique pourrait être liée au fait que le concept ne désigne pas tant des phénomènes particuliers qu'un « champ actif de questions ouvertes », dépassant le cadre de la biologie et nourrissant notamment les analyses des sociologues (Landecker et Panofsky, 2013).

Dans ce contexte, les différentes définitions de l'épigénétique ne seraient pas de nature à obérer le progrès scientifique, mais elles permettraient au contraire de le stimuler. Il est d'ailleurs tout à fait possible de considérer que le flou d'un concept va de pair avec son potentiel explicatif. Enfin, la polysémie du concept d'épigénétique ne semble pas constituer un frein au dialogue entre les disciplines comme la biologie moléculaire, la biologie développementale et la sociologie. Les spécialistes appartenant à ces domaines semblent en effet s'accorder sur une définition minimale selon laquelle l'épigénétique est ce qui permettrait d'aller « au-delà de la génétique » (Arimondo *et al.*, 2019).

►► L'épigénétique, une question transversale en biologie

L'exercice de définition, réalisé dans la partie précédente, ouvre la voie à une cartographie des différents usages de l'épigénétique. Se calquant sur ceux de la génétique,

4. Citation en langue originale : « the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence ».

ces derniers sont extrêmement divers. Dans chaque cas, il s'agit cependant pour les biologistes d'aller au-delà de l'idée selon laquelle les gènes⁵ et leurs activités suffisent à expliquer les différents phénomènes qui se situent au cœur de l'investigation biologique : développement, hérédité et évolution.

Dépasser le génocentrisme dans différents domaines de la biologie

Les recherches en épigénétique se déploient tout d'abord en biologie du développement. Dans ce domaine, la question est de savoir comment le génotype, l'ensemble des facteurs héréditaires, permet de produire le phénotype, l'ensemble des traits observables. Si la génétique (mendélienne puis moléculaire) a conduit à l'élaboration de la métaphore du « programme », qui semble incarner une version moderne de la doctrine de la préformation, l'épigénétique pourrait quant à elle fonder une nouvelle épigénèse dans laquelle l'environnement jouerait un rôle plus important que précédemment envisagé au cours de l'ontogenèse. On trouve d'ailleurs dans la littérature l'idée de « programmation » ou de « reprogrammation » épigénétique. Faisant clairement écho à la métaphore du programme génétique, celle-ci suggère que l'information développementale ne se situe pas exclusivement dans l'ADN. Il s'agit donc, en biologie du développement, de déterminer dans quelle mesure les marques épigénétiques participent de la construction des trajectoires individuelles des organismes. La recherche se joue ici principalement à l'échelle moléculaire, l'objectif étant de mieux comprendre les chaînes causales impliquées dans la régulation des gènes. Plusieurs exemples connus de variations phénotypiques sous-tendues par des marques épigénétiques sont présentés dans la littérature : symétrie de la linaire commune, couleur de la tomate, couleur du pelage de la souris, etc. Il est en outre établi que les marques épigénétiques sont impliquées dans le développement de certains cancers et dans le vieillissement.

L'épigénétique modifie aussi le champ des travaux sur l'hérédité biologique. Alors que la génétique a régné sur les recherches consacrées à la transmission des traits pendant une large partie du xx^e siècle, l'idée d'une hérédité strictement génétique est fortement contestée depuis une quarantaine d'années. La transmission épigénétique transgénérationnelle, ou l'hérédité épigénétique, constitue un ingrédient majeur de ce qu'on appelle parfois « hérédité étendue ». La thèse d'une hérédité épigénétique trouve manifestement son origine dans les définitions apparues lors de la molécularisation de l'épigénétique, et selon lesquelles l'épigénétique désigne l'étude des changements héréditaires dans les états fonctionnels des cellules. En effet, ces changements peuvent être hérités à travers la mitose, soit à l'échelle intragénérationnelle, mais aussi *via* la méiose, autrement dit à l'échelle inter- voire transgénérationnelle (Holliday, 2006). Une littérature de plus en plus fournie recense ainsi des cas de transmissions transgénérationnelles chez les plantes, les levures et chez différentes espèces animales, notamment chez les vers (Cavalli et Heard, 2019). Chez les mammifères, l'hérédité épigénétique est plus controversée même si quelques publications récentes en font état (Boscardin *et al.*, 2022; Van de Pette *et al.*, 2022).

Enfin, l'épigénétique constitue un champ d'étude pertinent en biologie de l'évolution. Dans ce domaine, la théorie néo-darwinienne soutient que la transformation des espèces résulte de la sélection naturelle de petites variations génétiques, apparues de

5. Ici, le gène sera défini comme une portion d'ADN qui est utilisée pour fabriquer une protéine.

manière aléatoire et conférant un avantage sélectif dans un environnement donné. Or la littérature montre que les marques épigénétiques, acquises au cours du développement et potentiellement adaptatives, constituent une source additionnelle (non-génétique) de variation héritable. Aussi laisse-t-elle songer que ces marques pourraient avoir une incidence sur le destin des espèces.

Le tournant environnemental des années 2010

La prise en compte de l'épigénétique dans les travaux des biologistes semble souvent associée à celle de l'environnement. Il s'agit plus précisément, au moins depuis quelques années, de repenser à travers l'épigénétique le rôle de l'environnement dans les phénomènes que sont le développement, l'hérédité et l'évolution, et d'envisager les mécanismes épigénétiques comme les médiateurs des effets de l'environnement sur les organismes. De ce point de vue, l'épigénétique participe de ce que d'aucuns appellent un « tournant environnemental » dans les sciences du vivant (Landecker et Panofsky, 2013).

Le rôle de l'environnement sur le développement à travers des mécanismes épigénétiques a surtout été bien documenté chez les plantes et chez quelques animaux comme la drosophile (Cavalli et Heard, 2019). Depuis quelque temps, la discipline appelée « épigénétique environnementale » présente par ailleurs les mécanismes épigénétiques comme des moyens d'incorporer des expériences environnementales et sociales chez les êtres humains. Aussi l'objectif de ce domaine de recherche est-il de traquer « les mécanismes par lesquels les forces sociales – pollution, nutrition, soins maternels, expérience traumatique – sont incorporées à l'échelle moléculaire, affectent l'expression des gènes et induisent des changements durables dans le comportement et la santé » (Landecker et Panofsky, 2013).

Plusieurs sous-champs disciplinaires partagent globalement cet objectif. Par exemple, « l'épigénétique comportementale » étudie la façon dont l'expérience précoce laisse des traces ayant une incidence sur le comportement et la santé de l'adulte. On trouve également dans la littérature l'idée d'une « épigénétique des traumatismes », traumatismes notamment liés à l'holocauste et à différentes formes de discriminations (Dubois *et al.*, 2018). Ces travaux sont cependant encore souvent spéculatifs et manquent d'assise empirique précise.

» L'épigénétique, un concept aux multiples enjeux épistémologiques et sociaux

Enfin, l'épigénétique est associée à des enjeux épistémologiques et sociaux majeurs. Les premiers concernent la question du changement scientifique. Les seconds portent sur l'exploitation des données à des fins sociales et politiques.

Enjeux épistémologiques : la question du changement scientifique

Si les découvertes réalisées en épigénétique semblent s'inscrire dans la continuité de celles de la génétique (Gayon, 2016), dès lors que l'épigénétique étudie de nouveaux mécanismes de régulation génétique, l'apparition de l'épigénétique est parfois envisagée comme un point de rupture dans l'étude des êtres vivants. L'épigénétique constitue-t-elle dès lors une discipline à part entière, ou n'est-elle que le prolongement de la génétique, voire l'une de ses branches ? S'inscrit-elle dans la continuité de

la génétique ou marque-t-elle l'avènement d'un changement radical dans l'étude et la compréhension des phénomènes biologiques? Quel type de changement scientifique l'épigénétique induit-elle? Aller au-delà de la génétique revient-il à approfondir les connaissances constituées dans ce domaine, ou bien à s'affranchir des limites théoriques qui le caractérisent?

L'analyse du changement scientifique développée par Thomas Kuhn dans *La Structure des révolutions scientifiques* (1983) offre quelques outils pour répondre à ces questions, sachant que plusieurs acteurs du débat mobilisent une terminologie kuhnienne. Pour le dire brièvement, Kuhn considère que le changement scientifique s'opère à travers des révolutions, qui sont assimilées à des changements de paradigmes. Les paradigmes peuvent être définis comme des travaux exemplaires (par exemple les lois de Newton) ou comme des matrices disciplinaires (cadres théoriques, concepts, méthodes) qui fondent des traditions de recherche, et qui norment l'activité scientifique du quotidien. Dans le cadre de la « science normale », la connaissance progresse par accumulation. Mais lorsque le paradigme se heurte à des phénomènes non prévus ou non compatibles avec ses principes (des « anomalies »), il est fragilisé. S'amorce alors une phase de « crise », qui ne s'achève que lorsque le paradigme défaillant se trouve remplacé par un autre. Le passage d'un paradigme à un autre (e.g. de l'astronomie ptolémaïque à l'astronomie copernicienne) constitue une révolution scientifique.

Dans le cas présent, la question est de savoir si l'épigénétique contribue au développement du paradigme de la génétique, et se contente de combler ses insuffisances, ou si elle constitue un nouveau paradigme de nature à fonder une révolution scientifique et une nouvelle tradition de recherche. En réalité, il semble difficile de répondre globalement à cette question tant les usages de l'épigénétique sont divers, et tant les définitions en sont variées. Aussi paraît-il plus prudent de décliner l'analyse dans les différents domaines concernés. Il s'agira alors d'interroger, dans chaque cas, les éléments théoriques mis en cause – concepts, méthodes, questions de recherche, hypothèses fondamentales, présupposés théoriques, etc. – afin de mieux évaluer l'ampleur du changement en cours.

En biologie moléculaire, on pourra notamment se demander si l'épigénétique vient modifier le concept fondamental d'information biologique, qui correspond selon Francis Crick à un « code » qui s'incarne dans les séquences de bases de l'ADN, ou encore si elle ébranle le présupposé théorique du dogme central, également formulé par Crick, selon lequel l'information génétique est unidirectionnelle, allant des séquences d'ADN vers les protéines. Dans la continuité des travaux sur la régulation génétique, les recherches conduites en épigénétique soulignent en réalité que l'information biologique, y compris génétique, ne peut se résumer à la séquence d'ADN dès lors qu'elle nécessite, pour être construite, le concours de toute une machinerie cellulaire. Par ailleurs, les récents travaux sur l'assimilation génétique suggèrent que les mécanismes épigénétiques peuvent affecter le taux de mutation de l'ADN, et avoir une incidence sur le contenu informationnel des gènes (Danchin *et al.*, 2019)⁶. Ainsi, l'épigénétique semble signaler l'obsolescence des analyses de Crick.

6. Cette assimilation génétique est à distinguer de celle qui est associée au nom de Waddington. Pour Waddington, l'assimilation génétique désigne la canalisation du développement sous l'effet de la sélection naturelle.

En biologie de l'évolution, l'épigénétique semble en mesure de modifier les perspectives sur la nature de la variation héritable, et sur le rôle du développement dans l'histoire du vivant. En effet, la théorie néo-darwinienne présente l'évolution comme le résultat de la sélection naturelle de petites variations héréditaires génétiques, aléatoires et imperméables aux demandes de l'environnement. Les études conduites en épigénétique suggèrent quant à elles que la variation héritable n'est pas seulement génétique, qu'elle peut être acquise au cours du développement et adaptative. Ainsi, l'épigénétique questionne une hypothèse fondamentale concernant le moteur de la transformation des espèces. Elle peut également nourrir de nouvelles questions de recherche, notamment sur le rôle du développement dans la détermination des trajectoires évolutives. D'aucuns évoquent alors avec fracas le retour du lamarckisme, souvent réduit à la thèse de l'hérédité des caractères acquis (Jablónka et Lamb, 2005). De fait, la dimension la plus révolutionnaire de l'épigénétique résiderait dans sa capacité à réhabiliter cette thèse qui, après avoir été largement admise jusqu'à la fin du XIX^e siècle, est tombée en désuétude suite aux travaux de Weismann et à l'émergence de la génétique. Cependant, plusieurs auteurs ont montré les limites d'un rapprochement entre épigénétique et lamarckisme, et, comme mentionné plus haut, la transmission transgénérationnelle des marques épigénétiques n'est fermement établie que chez les plantes et chez quelques autres organismes eucaryotes (Cavalli et Heard, 2019).

Enfin, le véritable changement scientifique induit par l'épigénétique pourrait être celui du rapprochement disciplinaire : entre génétique et embryologie, entre biologie du développement et biologie de l'évolution (Nicoglou et Merlin, 2017). Ce changement pourrait aussi concerner les pratiques explicatives (Baedke, 2018), et le passage d'une ontologie mécaniste, atomiste et réductionniste (dans laquelle le comportement d'un système s'explique par les propriétés de ses parties) à une ontologie systémique des processus (dans laquelle il est impossible d'isoler des parties pour comprendre le comportement du tout, qui évolue en permanence).

Enjeux sociaux : la libération par l'épigénétique ?

L'épigénétique, soulève des enjeux sociaux majeurs. Cette discipline, fondée sur l'hypothèse selon laquelle les marques épigénétiques permettent d'incorporer et de mémoriser biologiquement certains effets environnementaux, ouvre en effet une fenêtre pour évaluer le poids causal des diverses expositions environnementales dans le développement des individus. Les données de l'épigénétique peuvent par exemple être exploitées en santé publique, ou en épidémiologie sociale, pour mieux comprendre l'origine des inégalités sociales de santé, et pour mieux identifier les responsabilités des acteurs publics dans l'existence et le maintien d'environnements pathogènes.

Pour autant, il importe de conserver une certaine prudence face à des exploitations abusives et caricaturales des données. Il convient plus précisément de rester circonspect face aux discours qui affirment dans la presse grand public que l'ADN n'est pas notre destin – comme si nous ne le savions pas déjà! – et qui présentent l'épigénétique comme la discipline qui pourra enfin nous affranchir de la tyrannie de nos gènes. Cette promesse d'émancipation fait de l'épigénétique un sujet en vogue, presque un argument marketing. On trouve en effet des shampoings épigénétiques, mais aussi, étonnamment, des coaches qui délivrent des certificats...

Face à ces analyses peu sérieuses, il est indispensable de se souvenir que les données de l'épigénétique sont en cours de construction, que le caractère généralement adaptatif des marques épigénétiques n'est pas établi, et que la transmission transgénérationnelle de ces marques, chez les êtres humains, est encore mal comprise.

► Conclusion

Le bref parcours historique et l'état des lieux critique proposés dans ce chapitre montrent que l'épigénétique n'est pas qu'un mot dans l'histoire des sciences. Concept en perpétuelle évolution, l'épigénétique désigne des éléments hétérogènes : une discipline, des processus développementaux, des mécanismes moléculaires, des outils permettant de médiatiser les effets de l'environnement sur les organismes à différentes échelles, une promesse d'émancipation, etc.

Quelle que soit la définition retenue, l'épigénétique semble toujours se présenter comme ce qui va plus loin que la génétique, soit en complétant les données des chercheurs, soit en contestant leurs présupposés théoriques. Plus qu'un mot dans l'histoire des sciences, le concept d'épigénétique apparaît finalement comme un défi. Il incarne peut-être à lui seul l'esprit de la science, notamment expérimentale, qui consiste à compléter, à questionner et à dépasser sans cesse un savoir, certes établi avec méthode et ayant fait la preuve de sa validité, mais toujours provisoire, et toujours perfectible.

Remerciements

L'auteure remercie l'ANR Envirobiosoc.

► Références bibliographiques

- Allfrey V.G., Faulkner R., Mirsky A.E., 1964. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc Natl Acad Sci USA*, 51, 786-794.
- Arimondo P.B., Barberousse A., Pontarotti G., 2019. The many faces of epigenetics Oxford, December 2017. *Epigenetics*, 14 (6), 623-631. <https://doi.org/10.1080/15592294.2019.1595298>
- Baedke J., 2018. *Above the Gene, Beyond Biology: Toward a Philosophy of Epigenetics*, University of Pittsburgh Press. <http://www.jstor.org/stable/j.ctv14h5kr>
- Boscardin C., Manuella F., Mansuy I.M., 2022. Paternal transmission of behavioural and metabolic traits induced by postnatal stress to the 5th generation in mice. *Environ Epigenet.*, 8 (1), dvac024. <https://doi.org/10.1093/eep/dvac024>
- Cavalli G., Heard E., 2019. Advances in epigenetics link genetics to the environment and disease. *Nature*, 571 (7766), 489-499. <https://doi.org/10.1038/s41586-019-14110>
- 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. Reviews*, 94, 289-282. <https://doi.org/10.1111/brv.12453>
- Dubois M., Guaspere C., Louvel S., 2018. De la génétique à l'épigénétique : une révolution « post-génomique » à l'usage des sociologues. *Revue française de sociologie*, 59 (1), 71-98.
- Gayon J., 2016. From Mendel to epigenetics: History of genetics. *C R Biol.*, 339 (7-8), 225-230. <https://doi.org/10.1016/j.crvi.2016.05.009>
- Haig D., 2004. The (dual) origin of epigenetics. *Cold Spring Harb Symp Quant Biol.*, 69, 67-70. <https://doi.org/10.1101/sqb.2004.69.67>
- Holliday R., 2006. Epigenetics: A historical overview. *Epigenetics.*, 1(2), 76-80.
- Jablonska E., Lamb M., 2005. *Evolution in Four Dimensions*, Cambridge, MIT Press.

- Jablonka E., Raz G., 2009. Transgenerational epigenetic inheritance: Prevalence, mechanisms, and implications for the study of heredity and evolution. *Q Rev Biol.*, 84 (2), 131-176. <https://doi.org/10.1086/598822>
- Kuhn T., 1983 [1962]. *La Structure des révolutions scientifiques*, Flammarion, 284 p.
- Landecker H., Panovski A., 2013. Regulation and back: A critical introduction to environmental epigenetics for sociology. *Annual Review of Sociology*, 39, 333-357.
- Morange M., 2005. Quelle place pour l'épigénétique? *Med Sci (Paris)*, 21 (4), 367-369. <https://doi.org/10.1051/medsci/2005214367>
- Nanney D.L., 1958. Epigenetic control systems. *Proc. Natl. Acad. Sci.*, 44, 712.
- Nicoglou A., Merlin F., 2017. Epigenetics: A way to bridge the gap between biological fields. *Stud Hist Philos Biol Biomed Sci.*, 66, 73-82. <https://doi.org/10.1016/j.shpsc.2017.10.002>
- Riggs A.D., 1975. X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet.*, 14, 9-25.
- Riggs A.D., Martienssen R.A., Russo V.E.A., 1996. Introduction. In: *Epigenetic Mechanisms of Gene Regulation* (V.E.A. Russo *et al.*, eds), New York, Cold Spring Harbor Laboratory Press, p. 1.
- Van de Pette M., Dimond A., Galvão A.M., Millership S.J., To W. *et al.*, 2022. Epigenetic changes induced by in utero dietary challenge result in phenotypic variability in successive generations of mice. *Nat Commun.*, 13(1), 2464. <https://doi.org/10.1038/s41467-022-30022-2>
- Waddington C., 1942. The epigenotype. *Endeavour*, 1, 18-20.
- Waddington C. H., 1957. *The Strategy of the Genes*, London, George Allen & Unwin.

Chapitre 2

Épigénétique, une exploration des processus moléculaires

HÉLÈNE JAMMES ET DAVID L'HÔTE

Comme indiqué en introduction à cet ouvrage, la définition contemporaine de l'épigénétique fait référence aux molécules et mécanismes responsables de la régulation de l'expression des gènes sans modification de la séquence de l'ADN (Cavalli et Heard, 2019). La contribution de l'épigénétique dans la différenciation cellulaire au cours du développement des individus et dans la réponse à l'environnement (à différentes échelles : cellulaire, tissulaire, systémique) est explicitée à travers de nombreux exemples déclinés au fil des chapitres de l'ouvrage, soulignant la dynamique, la complexité et la diversité des interactions moléculaires nécessaires. Ce chapitre a pour objectif d'illustrer comment, au fil du temps, les études scientifiques ont identifié des molécules, proposé et validé des mécanismes (tels que les modifications post-traductionnelles des histones, la méthylation de l'ADN, les petits et longs ARN non codants...), aujourd'hui analysés dans de nombreux contextes biologiques (biologie du développement, réponses adaptatives du génome à l'environnement, survenue de pathologies). Ce chapitre a par ailleurs l'ambition de donner au lecteur des définitions et des bases en biologie moléculaire et cellulaire, véritable boîte à outils pouvant être ouverte pour le plaisir de la connaissance ou pour faciliter la lecture des chapitres suivants. Le lecteur aura ainsi en main les clés de compréhension et pourra partir à la découverte des champs du possible offerts par l'épigénétique, une des multiples facettes de la science du vivant.

► Le patrimoine génétique dans tous ses états

Chez les eucaryotes, le patrimoine génétique, ou ADN génomique nucléaire (ADNg), se présente sous différentes formes au gré de la vie de la cellule : sous forme de chromatine plus ou moins compactée ou sous forme de chromosomes individualisés (stade métaphasique au cours de la division cellulaire). La chromatine est décrite pour la première fois par Walther Flemming (~1880) comme une substance présente dans le noyau cellulaire, répartie en régions claires, peu condensées, laissant passer la lumière du microscope – l'euchromatine –, et en régions foncées, opaques à la lumière du fait d'une extrême condensation – l'hétérochromatine (figure 2.1).

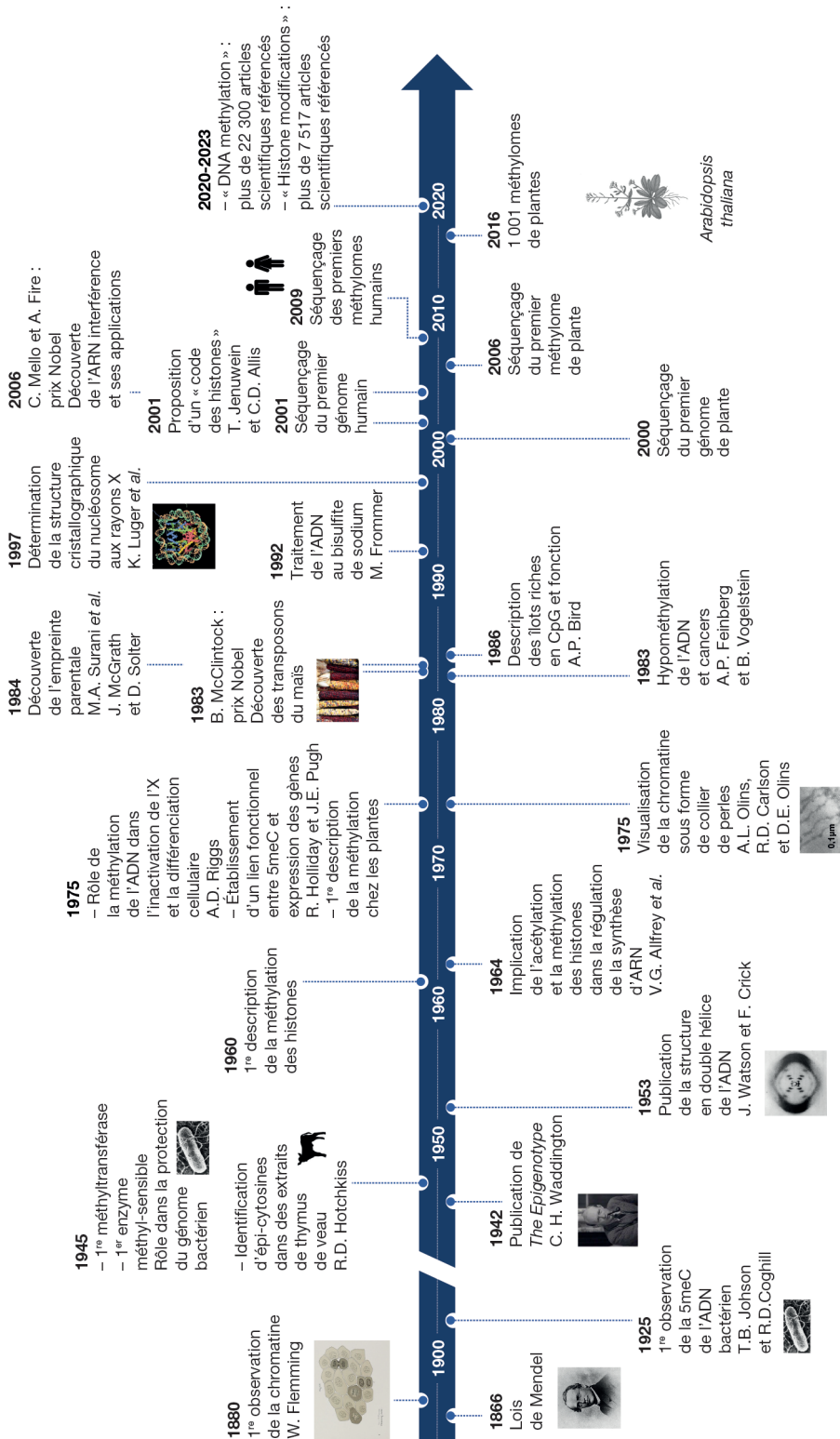


Figure 2.1. Fresque des principales avancées et découvertes en épigénétique.

La condensation, ou compaction de l'ADNg, est nécessaire, d'une part, pour être empaqueté dans le noyau cellulaire – chez l'homme, cet ensemble fait environ 2 m et est contenu dans une sphère d'à peine 10 µm de diamètre –, et, d'autre part, pour être protégé des agressions mécaniques et chimiques. Dans un type cellulaire donné, la répartition entre euchromatine et hétérochromatine participe à une organisation tridimensionnelle qui régit les différentes fonctions de l'ADNg telles que la réplication lors des divisions cellulaires, la transcription en ARN messagers (ARNm), la réparation et la recombinaison. Dès 1974, Olins *et al.* précisent la structure de la chromatine en la visualisant par microscopie électronique sous forme de particules sphériques reliées entre elles, tel un collier de perles (Olins *et al.*, 1975). Il faut attendre 1997 pour percer complètement le mystère de cette structure par des approches de cristallographie aux rayons X (Luger *et al.*, 1997) : chaque perle, appelée « nucléosome », est un octamère protéique composé de deux dimères d'histones H2A-H2B et de deux dimères d'histones H3.1/H3.2-H4 (figure 2.2). S'ensuivent l'extraction et la purification des protéines des nucléosomes à partir de divers organismes et de divers tissus ou organes et la caractérisation des histones. Les histones canoniques (H2A-H2B et H3.1/H3.2-H4) sont de petites protéines (de l'ordre de 100 à 220 acides aminés ; masse moléculaire entre 10 et 24 kDa). La nature basique des histones, due à leur richesse en lysine et en arginine, leur confère une charge positive et permet une forte interaction avec l'ADNg (chargé négativement par la présence d'un groupe phosphate), qui s'enroule approximativement deux fois autour du nucléosome (soit environ 147 paires de bases). La participation des histones canoniques à la structure du nucléosome est décrite pour un très large éventail d'organismes (homme, souris, poule, drosophile, riz, levure, etc.), révélant une très forte conservation pour l'ensemble des eucaryotes. En plus des histones canoniques citées ci-dessus, l'existence de formes variantes aux caractéristiques structurales et fonctionnelles différentes est démontrée. Chez l'homme, une vingtaine de variants est décrite, alors que chez la levure, seuls deux variants sont mis en évidence. La présence des variants d'histones est observée dans des types cellulaires particuliers ou au cours d'une phase du cycle cellulaire. Leur distribution génomique n'est pas aléatoire ; certains d'entre eux sont plutôt associés à l'hétérochromatine.

►► Le « code des histones »

Un contrôle de l'intensité des interactions histones-ADNg est exercé par l'apposition de groupements chimiques sur certains acides aminés (ou résidus) positionnés en partie N-terminale des histones, s'échappant du nucléosome et donc accessibles. De nombreuses modifications covalentes post-traductionnelles sont possibles (tableau 2.1).

En prenant en compte la position et le nombre d'acides aminés modifiés, plus d'une centaine de modifications post-traductionnelles sont répertoriées, constituant une combinatoire de marques épigénétiques connue sous le nom de « code des histones » (Strahl et Allis, 2000 ; Zhao et Garcia, 2015 ; figure 2.2). La complexité de ce code est loin d'être complètement élucidée, malgré les efforts intenses de la communauté scientifique depuis de nombreuses années. Ces marques jouent un rôle déterminant dans l'architecture de la chromatine. Leur distribution génomique définit de grands domaines dont l'activité transcriptionnelle est variable (domaines permissif/répressif), contrôlant l'expression des gènes (encadré 2.1).

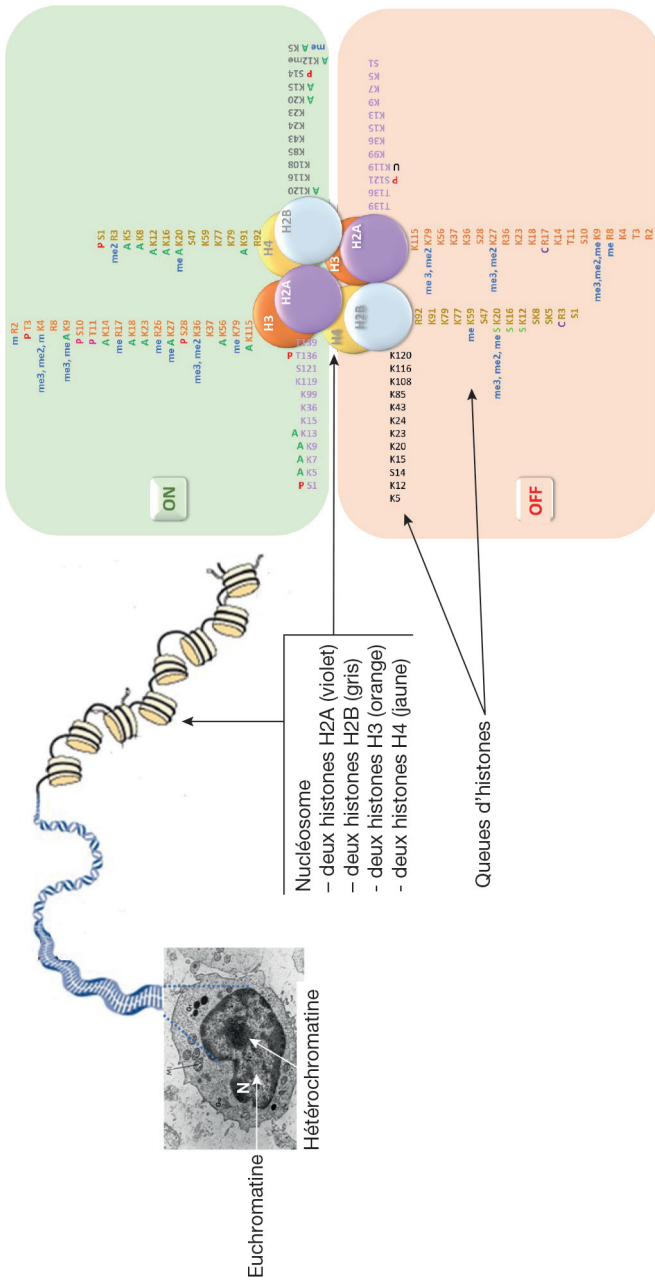


Figure 2.2. Représentation de la chromatine et de la structure moléculaire du nucléosome.

L'observation microscopique d'une cellule révèle son noyau (N), contenant le patrimoine génétique, ou ADN génomique (ADNg), le plus souvent sous forme de chromatine. La chromatine est organisée en hétérochromatine et en euchromatine. La chromatine déboulinée se présente sous forme de nucléosomes autour desquels l'ADNg s'enroule (147 bases/nucléosome) et ressemble à une structure en collier de perles observée par Orlins *et al.* en 1975. Chaque nucléosome est composé de : deux histones H2A (violet), deux histones H2B (gris), deux histones H3 (orange) et deux histones H4 (jaune). Les acides aminés, accessibles et listés en fonction de leur position, forment les « queues d'histones » porteuses de modifications chimiques (gris), elles-mêmes classées en fonction de leur rôle permissif (ON) ou répressif (OFF) sur la transcription. Quelques exemples de modifications chimiques post-traductionnelles sont représentées : la méthylation (bleu) avec mono- (me), di- (me2), ou tri-méthylation (me3) ; l'acétylation (A, vert) ; la citrullination (C, violet) ; la sumoylation (S, noir) ; la phosphorylation (P, rouge) ; l'ubiquitination (U, noir).

Tableau 2.1. Liste non exhaustive des modifications post-traductionnelles portées par des acides aminés des histones.

Acides aminés	Modification chimique
Lysine (K)	Acétylation, ubiquitylation, sumoylation, biotinylation, butyrylation
Lysine (K) acétylée	2-hydroxyisobutyrylation, N-formylation, glutathionylation, malonylation, hydroxylation, oxydation et crotonylation
Lysine (K) et arginine (R)	Méthylation (mono, di et triméthylation);
Sérines (S) et thréonines (T)	Phosphorylation
Arginines (R)	O-GlcNAcylation, citrullination
Acide glutamique (E)	ADP-polyribosylation
Arginine (R)	Désamination
Proline (P)	Isomérisation

Encadré 2.1. Gènes, régions régulatrices et marques épigénétiques.

Le gène est l'unité codante du génome. Il est transcrit sous la forme d'un ARN, et cette transcription est régulée par un ensemble de séquences d'ADN présentes à plus ou moins grande proximité du gène. Les données récentes de génomique fonctionnelle démontrent que ces séquences régulatrices sont très nombreuses. En effet, on considère que les gènes ne représentent qu'environ 5% du génome (donc 95% du génome n'est pas transcrit en ARNm pour fabriquer une protéine). À quoi sert donc tout cet ADN en plus ? On a longtemps pensé qu'il s'agissait majoritairement d'ADN inutile, que l'on appelait élogiquement ADN « poubelle » (*Junk DNA*).

Depuis les années 2010, cette notion d'ADN poubelle a peu à peu disparu, alors que nous nous rendons compte que cet ADN non codant est en réalité le manuel d'utilisation du génome : c'est dans ces 95% qu'est cachée l'information de « à quel moment », « avec quelle intensité » et « dans quelles cellules » les gènes doivent s'exprimer. Nous autres mammifères, sommes composés d'environ 200 types cellulaires différents (neurones moteurs, sensoriels, hépatocytes, cellules musculaires striées lisses et cardiaques, fibroblastes, etc.). Chacun de ces types cellulaires exprime une combinaison de gènes parmi les ~30 000 gènes du génome humain. Au delà de l'information génique, le génome doit donc contenir les instructions nécessaires pour que chaque type cellulaire puisse savoir quels sont les gènes à exprimer en conditions physiologiques ou en réponse à un challenge : par exemple, lors d'une coupure de l'épiderme (peau), les gènes impliqués dans la production de protéines du système de réparation cellulaire, normalement silencieux, sont exprimés spontanément.

Le génome est donc le regroupement de plus 200 manuels à l'usage des cellules pour définir leur identité et leurs fonctions tout au long de notre vie ! D'un point de vue moléculaire, ces fameuses instructions sont contenues dans de petites régions appelées « régions régulatrices » de l'expression des gènes. Il faut noter que les régions régulatrices peuvent virtuellement se trouver n'importe où par rapport aux gènes dont elles régulent l'expression. Les cellules ont donc besoin d'un système de reconnaissance de ces régions et elles utilisent pour cela les marques épigénétiques.

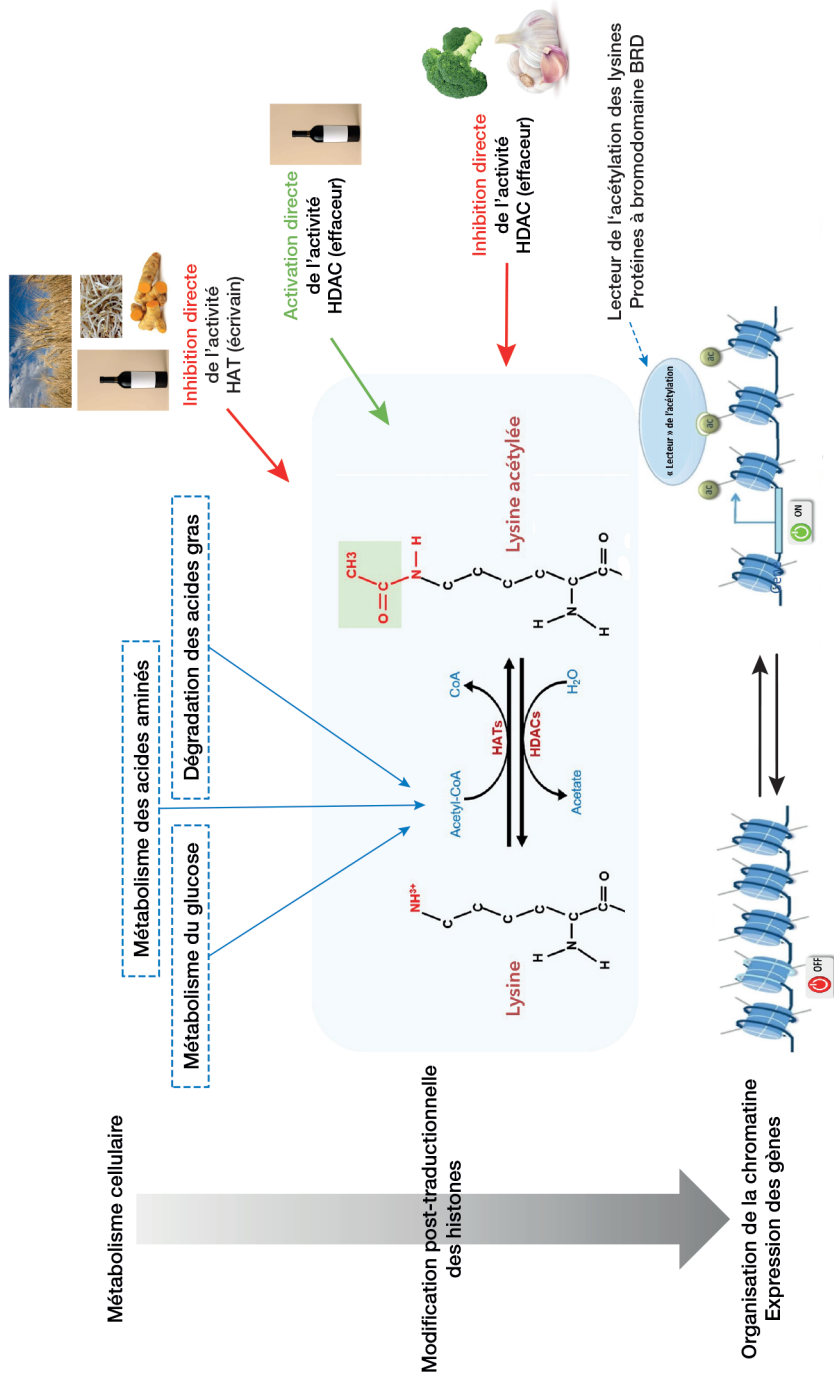


Figure 2.3. Acétylation /déacétylation des lysines : rôle de l'alimentation.

Dans chaque noyau cellulaire et à chaque instant, le code doit être écrit, lu – déchiffré –, interprété, voire modifié ou écrasé (réversion), nécessitant l'existence de nombreux acteurs protéiques, classés, pour chaque type de modification, en fonction de leur activité enzymatique, en « écrivains », « lecteurs » et « effaceurs » (*writers, readers and erasers*; Strahl et Allis, 2000).

Parmi les modifications post-traductionnelles possibles, l'acétylation/désacétylation et la méthylation des histones ont fait l'objet de nombreuses études illustrant l'ampleur des processus régulateurs mis en jeu.

Les acteurs de l'acétylation/désacétylation des lysines

Les « écrivains » créent la modification. L'acétylation consiste en l'ajout d'un groupe acétyle ($-\text{CO}-\text{CH}_3$) sur certaines lysines (K) – notée « H3K4ac », c'est-à-dire acétylation de la lysine 4 de l'histone H3 – sous l'action d'enzymes nommées « histones acétyltransférases » (HAT; la première fut découverte en 1996) en utilisant l'acétyl-CoA⁷ comme substrat (figure 2.3). La lysine possède un groupement ammonium (NH_3^+) chargé positivement à pH neutre (qui est le pH cellulaire) qui augmente l'affinité entre l'histone et l'ADN. L'acétylation du groupement conduit à une neutralisation de la charge positive et diminue fortement les contacts avec l'ADN enroulé sur le nucléosome : la chromatine devient plus flexible et se décondense. L'acétylation des lysines est souvent associée à l'euchromatine, ainsi plus accessible à l'ensemble des complexes protéiques régulateurs; cette modification est associée à la transcription des gènes, à la réplication au cours du cycle cellulaire, à la recombinaison et à la réparation de l'ADNg.

Les « lecteurs » de l'acétylation sont des protéines nucléaires présentant un bromodomaine (succession d'un peu plus d'une centaine d'acides aminés, décrite pour la première fois chez la drosophile au cours de l'étude de la régulation du gène *BRAHMA* en 1992). En se liant aux nucléosomes dont les histones présentent une acétylation, ces protéines utilisent l'énergie libre d'hydrolyse de l'ATP, modifient les positions des nucléosomes tout le long de l'ADNg et induisent des changements conformationnels dans la chromatine. La partie de l'ADN ainsi dépliée sera alors exposée aux différents facteurs de la transcription comme l'ARN polymérase II, nécessaires à la transcription.

Les « effaceurs » de l'acétylation sont des protéines appelées « histones désacétylases » (HDAC). Chez les mammifères, il existe 18 HDAC réparties en deux grandes familles. La première famille utilise l'ion zinc comme co-activateur. La seconde famille a besoin pour fonctionner de nicotinamide adénine dinucléotide (NAD^+), composé présent dans toutes les cellules (impliqué dans des réactions d'oxydo-réduction du métabolisme cellulaire). De manière générale, les HDAC sont peu spécifiques et peuvent agir sur de nombreux substrats.

Parmi les facteurs pouvant influencer les modifications post-traductionnelles des histones, le métabolisme de l'individu et son alimentation jouent un rôle très important (Janke et al., 2015). En effet, l'acétyl-CoA, fournisseur du groupe acétyle, provient de voies métaboliques de l'organisme telles que le cycle de Krebs, la glycolyse, la β -oxydation des lipides et le métabolisme des acides aminés, soulignant un lien majeur

7. L'acétyl-CoA est un composé cellulaire issu à la fois de la dégradation des glucides via la glycolyse et de la dégradation des acides gras constituant les lipides via la β -oxydation et intervenant dans le cycle de Krebs au sein des mitochondries.

entre le métabolisme général, le contrôle de l'architecture de la chromatine et l'expression des gènes (figure 2.3; Pietrocola *et al.*, 2015). Toute augmentation de l'acétyl-CoA peut être reliée à une augmentation de l'acétylation des histones, à une ouverture de la chromatine et à une expression génique. L'activité acétyltransférase peut être directement inhibée par le polyphénol (vin), la curcumine (racines de curcuma), la lunasine (soja, blé et orge), le garcinol (zeste de garcinia) ou l'acide anacardique (coque de noix de cajou). Certains composés alimentaires sont des inhibiteurs des HDAC : le sulforaphane du brocoli ou les organosulfures de l'ail. D'autres composés comme le resvératrol, contenu dans le vin rouge, augmentent l'activité des HDAC (figure 2.3). Les effets bénéfiques ou délétères restent à identifier, au cas par cas, puisqu'ils sont déterminés par l'amplitude et la spécificité de l'activité enzymatique : acétylation générale de toutes les histones du génome ou spécifique de certaines liées à des régions génomiques particulières ?

Les acteurs de la méthylation/déméthylation des histones

La méthylation est une modification post-traductionnelle des histones qui consiste en l'apport d'un groupe méthyle (-CH₃ ou Me) sur certaines lysines (K), arginines (R), glutamines (Q) ou sur l'acide glutamique (E). Cette modification a été décrite dès 1960 (figure 2.1). Le rôle répressif ou activateur de la méthylation sur l'expression génique dépend du résidu touché et du niveau de méthylation. Par exemple, la méthylation de la lysine 4 de l'histone H3 (H3K4me) engendre une conformation de chromatine transcriptionnellement active (figure 2.4), alors que la méthylation de la lysine 9 de l'histone H3 (H3K9me) est associée à un état répressif.

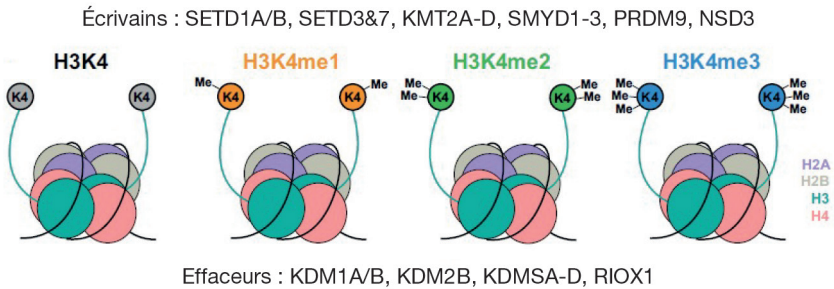


Figure 2.4. Variations de l'état de méthylation de la lysine 4 des histones H3.

Les « écrivains », ou méthyltransférases, apposent un groupe méthyle provenant de la S-adenosylméthionine (SAM). Chez les mammifères, la SAM est le donneur universel du groupe méthyle pour les histones mais aussi pour l'ADN (voir la section « La méthylation de l'ADN, marque épigénétique la plus explorée »). La biodisponibilité de la SAM résulte du métabolisme monocarboné mettant en jeu des composés directement fournis par l'alimentation comme le folate ou vitamine B9, la méthionine, la choline et la bétaine (voir figure 2.6). Les concentrations de ces composés sont variables selon le type d'aliments (apport en folate : noix, pois chiche, épinards, fromages affinés (100-200 µg/100 g) > pommes de terre, poisson, fruits frais (10-20 µg/100 g)⁸.

8. <https://www.anses.fr/fr/content/tout-savoir-sur-la-vitamine-b9>

SUV39H1, première méthyltransférase identifiée chez la drosophile en 1994, est spécialisée dans la méthylation des lysines (K) et contient un domaine catalytique SET (pour Su var3-9, Enhancer of Zeste and Trithorax) de 130 à 140 acides aminés. En sélectionnant toutes les protéines dont la séquence présente ce domaine fonctionnel, il a été possible d'identifier d'autres méthyltransférases chez la levure, chez l'homme, mais aussi chez certaines bactéries et virus.

La méthylation des arginines (R) est assurée, quant à elle, par des «protéines arginine méthyltransférases» (PRMT). Chez les mammifères, les 9 PRMT identifiées sont capables de mono-méthyliser l'arginine (fixation de CH₃ sur un des azotes (N) du groupe guanidine) Elles sont ensuite classées en fonction de leur capacité à di-méthyliser. Le type I (PRMT1– 4, 6, 8) catalyse l'addition d'un second groupe méthyl sur le même azote générant une diméthylation asymétrique. Le type II (PRMT5 et 9) catalyse l'addition des groupes méthyl de façon symétrique sur chacun des azotes du groupe guanidine Le type III (PRMT7) n'a qu'une activité de mono-méthylation.

Les «lecteurs» sont des protéines possédant des domaines fonctionnels particuliers, capables de reconnaître le taux de méthylation (mono-, di- ou triméthylée; symétrique ou asymétrique) et de se fixer aux lysines et aux arginines.

Les «effaceurs» n'ont été identifiés qu'en 2004, ce qui a valu à la méthylation des lysines et des arginines, la renommée d'être une marque extrêmement stable, faute de preuve de l'existence d'une activité enzymatique de déméthylation. La découverte de la première déméthylase a bouleversé cette idée : La lysine déméthylase 1A (notée KDM1A/LSD1), identifiée dans de nombreuses espèces (de la levure à l'homme), est capable de déméthyliser H3K4. Une autre lysine déméthylase a ensuite été découverte, capable de déméthyliser H3K4me1 et H3K4me2 (notée LSD2/KDM1B) (figure 2.4).

Une autre famille de protéines, JARID1 (jumonji AT-rich interactive domain 1), est aussi capable de déméthyliser H3K4. À ce jour, sont identifiées deux classes de lysines déméthylases : l'une est dépendante du fer et s'attaque aux trois états de méthylation (me1, me2 et me3; flavin-dépendent KDM1 enzymes) et l'autre dépend de l'oxygène (JmjC KDM) et s'attaque plutôt aux di- et mono-méthylations (Hyun *et al.*, 2017).

De manière générale, acétylation et méthylation sont des marques antagonistes; la désacétylation des lysines doit précéder leur méthylation. Cet antagonisme entraîne une modification dynamique de l'état chromatidien avec, pour une région donnée, le passage d'un état d'hétérochromatine (où les histones H3 sont désacétylées et méthylées et les gènes sont généralement non exprimés) à un état d'euchromatine (où les histones H3 sont acétylées et déméthylées et les gènes sont en général exprimés) et *vice versa* (encadré 2.2).

Au cours des dernières décennies, la spectrométrie de masse (MS) s'est avérée être une méthode performante, facilitant la découverte et l'étude d'un large spectre de modifications post-traductionnelles des histones en une seule analyse, mais ne renseignant pas sur la distribution génomique des modifications. Les modifications des histones sont aussi étudiées par des approches biochimiques faisant appel à la production d'anticorps spécifiques contre une modification particulière, et la révélation de la fixation de l'anticorps sur l'histone modifiée s'effectue par *Western blot*, ou immunoprécipitation de la chromatine (ChIP). La séquence d'ADNg impliquée dans l'interaction mise ainsi en évidence peut être recherchée par amplification

spécifique (si on s'intéresse à une région du génome particulière) ou identifiée par séquençage haut débit (si la stratégie est sans *a priori*). La mobilisation internationale mettant en œuvre ce type d'études et la mise en commun des données au sein de larges consortiums permettent aujourd'hui de produire une cartographie du génome en identifiant des domaines aux fonctionnalités variables en lien avec le type cellulaire ou l'organe⁹, chez l'homme et la souris. Il est ainsi possible de définir un épigénome cellulaire, correspondant à l'ensemble des marques épigénétiques (modifications post-traductionnelles des histones et méthylation de l'ADN) portées par et distribuées dans le génome d'une cellule.

Il est aussi à noter que le secteur de la recherche-développement des industries pharmaceutiques s'est investi dans l'identification de molécules pouvant intervenir sur le code des histones, en tant qu'écrivain, lecteur ou effaceur, avec des objectifs de thérapie pour divers cancers, pathologies de l'immunité ou pathologies neurologiques (Kringel *et al.*, 2021). Comme une des limites de l'utilisation de ces molécules est leur spécificité d'action (modifications ciblées, séquence d'ADNg/gène ciblé), l'exploration d'interventions moléculaires ajustées (*epigenome editing*) est évoquée dans le chapitre 12 de cet ouvrage.

Encadré 2.2. Activation et inhibition des régions régulatrices des gènes.

On distingue deux types de régions régulatrices (figure 2.5) : les promoteurs et les *enhancers*. Les promoteurs sont des séquences d'ADN (de tailles variables) toujours situées juste devant le gène qu'elles régulent. Ces promoteurs sont très souvent associés avec la marque H3K4me3 (histone H3 avec trois groupements méthyles, me3, sur la quatrième lysine, K4; figure 2.5A). Si les promoteurs sont toujours devant leur gène, les *enhancers* sont des régions régulatrices qui peuvent se situer très loin des gènes qu'ils régulent (devant, derrière, dedans même, quand ce n'est pas sur un autre chromosome). Le nom d'*enhancer* (« exhausteur ») vient à l'origine de leur capacité à augmenter l'effet des promoteurs. Un gène peut être régulé par plusieurs *enhancers* et un *enhancer* peut réguler plusieurs gènes. Les *enhancers* sont très souvent associés à la marque H3K4me1 (histone 3 avec un seul groupement méthyle, CH3, sur la quatrième lysine). Ces deux marques définissent les régions régulatrices à utiliser. Bien sûr, la cellule ne va pas utiliser tout le temps toutes les régions régulatrices disponibles : il y a donc d'autres marques qui indiquent si ces régions sont vraiment actives (et donc stimulent l'expression des gènes qu'elles régulent) ou inactives (et restent sans effet). Pour simplifier, la présence de la marque H3K27ac (l'acétylation de l'histone 3 sur la lysine 27; figure 2.5B) est le signe que la région régulatrice est active et que le gène associé est probablement exprimé. Alors que la présence de la marque H3K27me3 (la triméthylation de l'histone 3 sur la lysine 27) est souvent le signe d'une inhibition de la région régulatrice (figure 2.5C). La méthylation de l'ADN sur les cytosines dans les nucléotides CpG (voir la section « La méthylation de l'ADN, marque épigénétique la plus explorée ») est elle aussi une marque dite « répressive » des régions régulatrices et donc de l'expression génique. À noter que de nombreuses autres marques activatrices (H3K9ac par exemple) ou inhibitrices (H3K9me3) sont connues pour réguler l'expression des gènes.

9. <https://www.encodeproject.org/chip-seq/histone-encode4>

Encadré 2.2. (suite)

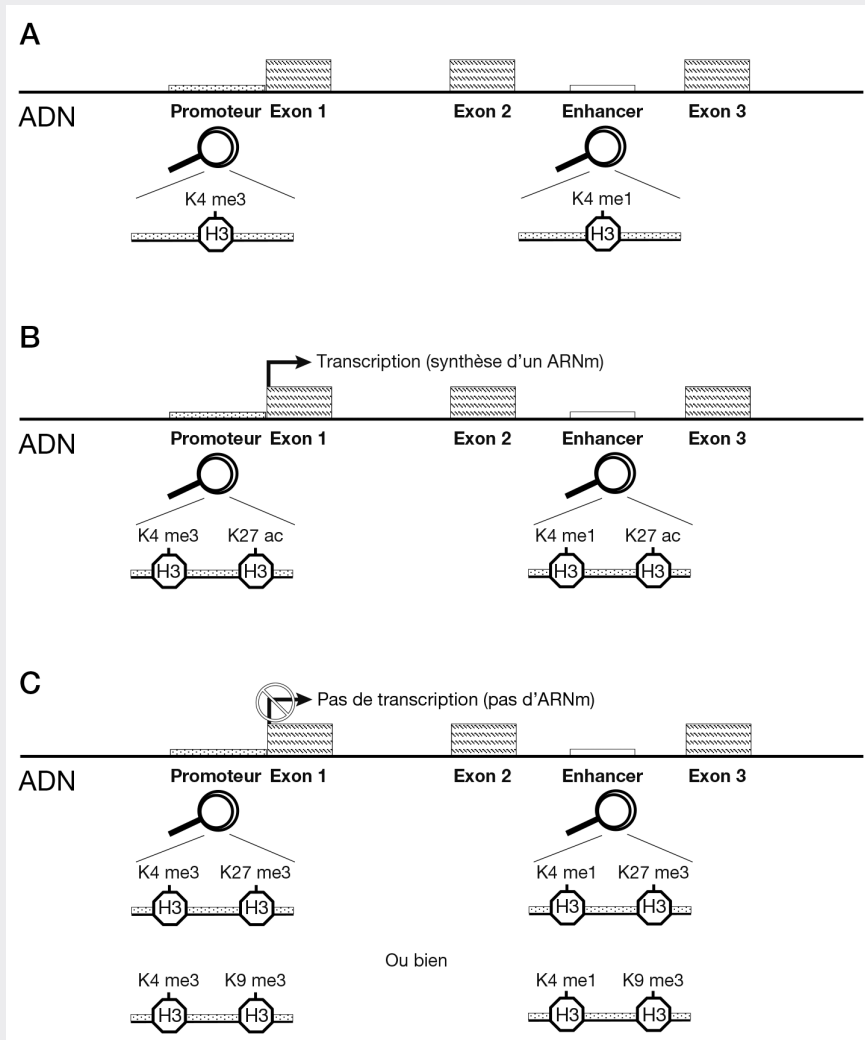


Figure 2.5. Représentation schématique des modifications post-traductionnelles des histones associées aux promoteurs et aux *enhancers* pilotant l'état transcriptionnel d'un gène.

A) Structure simplifiée d'un gène modèle; B) Décorations épigénétiques d'un gène actif (transcrit); C) Décorations épigénétiques d'un gène inactif (non-transcrit).

En résumé, si les régions régulatrices d'un gène (*enhancers* avec H4K4me1 et promoteurs avec H3K4me3) sont activées (donc avec du H3K27ac), le gène s'exprime, c'est-à-dire qu'un ARNm est produit. Inversement, si les régions régulatrices sont associées avec soit la marque H3K27me3 soit la méthylation de l'ADN sur les CpG, le gène reste silencieux.

» La méthylation de l'ADN, marque épigénétique la plus explorée

En 1925, Treat B. Johnson et Robert D. Coghil purifient les acides nucléiques de la bactérie *Mycobacterium tuberculosis*, les cristallisent, les hydrolysent en présence d'acide sulfurique et les observent sous lumière polarisée par microscopie (figure 2.1). Ils distinguent et nomment différentes structures : cytosine (C) et 5-méthylcytosine (5mC) sont ainsi identifiées. Quelques années plus tard, les acides nucléiques nucléaires sont extraits à partir de nombreux organes et analysés par chromatographie sur couche, permettant de séparer par migration les quatre bases, adénine (A), cytosine (C), guanine (G) et thymine (T). En 1945, Rollin D. Hotchkiss observe dans des extraits de thymus de veau l'existence d'une bande supplémentaire en plus de celle de la cytosine, qu'il nomme alors « épi-cytosine ». La présence de 5mC sera rapidement confirmée pour différents tissus de différentes espèces. La même année, l'intérêt fonctionnel de la 5mC est décrit chez la bactérie en démontrant que différentes souches bactériennes possèdent une activité « méthyltransférase » souche-spécifique. Il est alors proposé que la 5mC pourrait avoir un rôle dans la défense des bactéries contre les phages par un mécanisme dit de « système de restriction-modification » : lors d'une attaque virale, la bactérie synthétise des enzymes capables d'éliminer l'ADN viral par clivage. L'ADN bactérien étant méthylé, il est protégé de cette activité enzymatique. Ainsi, l'existence d'enzymes méthyl-sensibles est mise en évidence.

Dans les années 1960-1970, il est démontré que l'activité méthyltransférase est présente dans tous les tissus testés de différentes espèces animales, en quantité variable en fonction du type cellulaire, suggérant que le taux de 5mC peut être différent. Au cours du développement de l'œuf d'oursin, Scarano *et al.* démontrent que la distribution de la 5mC n'est pas aléatoire dans le génome et qu'elle apparaît dans un contexte CG, appelé aussi CpG (correspondant à une cytosine liée à une guanine *via* un groupement phosphate). La méthylation de l'ADN est l'addition covalente d'un groupe méthyl (CH₃) fourni par la S-adénosylméthionine (SAM), sur le 5^e carbone de la cytosine (Figure 2.6).

Il faut attendre 1975 pour que la 5mC soit décrite dans le génome des plantes, à différents taux en fonction de l'espèce et dans des contextes génomiques CG, CHG et CHH (où H représente les bases A, C ou T). Sur le plan fonctionnel, il est montré que le taux de 5mC dépend des phases de division pour les bactéries, tout comme pour les fibroblastes en culture, suggérant la nécessité d'une activité enzymatique capable de « recopier » la distribution génomique de la 5mC lors de la réplication de l'ADN. Cette propriété est fondamentale, puisqu'elle souligne la notion de mémoire des marques épigénétiques transmises à travers la mitose, pour le maintien de l'identité cellulaire. En 1975, Robin Holliday et John E. Pugh suggèrent qu'il est possible d'établir un lien fonctionnel entre le taux de méthylation et l'expression des gènes, en particulier au cours du développement des individus. Depuis, les études de la 5mC se sont intensifiées (Mattei *et al.*, 2022) et ont permis de décrypter son implication dans de très nombreux processus biologiques : quelques exemples sont donnés dans cet ouvrage, en particulier dans le contrôle de l'inactivation du chromosome X (chapitre 3) ou de l'empreinte parentale (chapitres 4, 5 et 6) et illustrent en détail ces avancées de la connaissance fondamentale.

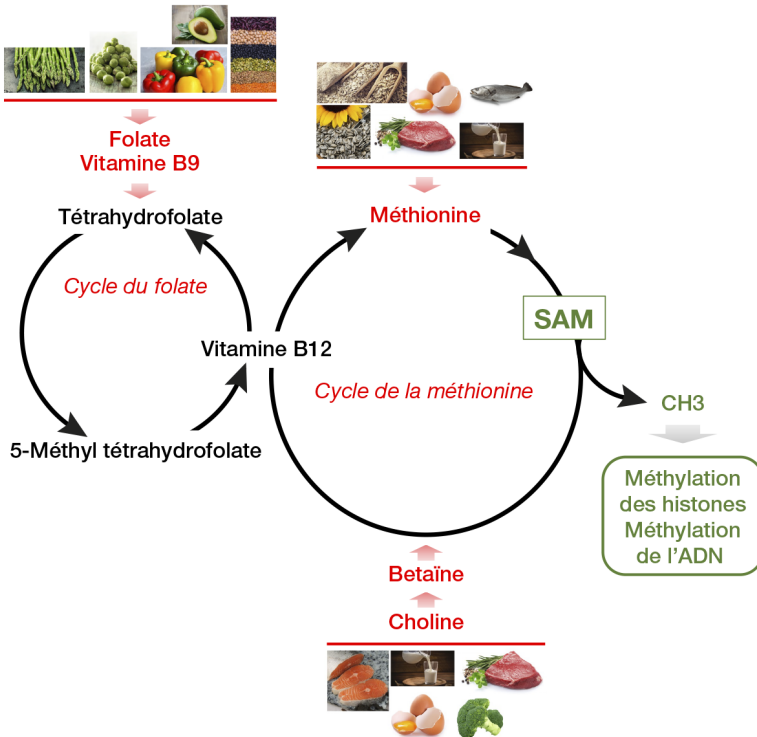


Figure 2.6. Biodisponibilité de la S-adenosylméthionine (SAM), donneur universel du groupe méthyl (CH₃) pour la méthylation des histones et de l'ADN : rôle de l'alimentation.

La SAM est issue du métabolisme des monocarbones. Le cycle métabolique du folate ou vitamine B9 aboutit à la formation de la vitamine B12 ou cobalamine, elle-même impliquée dans le cycle de la méthionine produisant la SAM. Le folate, la méthionine, la bétaine et la choline sont directement fournis par l'alimentation.

Aujourd'hui, il est admis que la 5mC est présente chez un grand nombre d'eucaryotes. Pourtant, chez certains invertébrés, comme le nématode *Caenorhabditis elegans* ou le diptère *Drosophila melanogaster*, il n'a pas été détecté de 5mC génomique. Chez d'autres invertébrés comme l'abeille (*Apis mellifera*), le nombre de cytosines génomiques méthylées reste modeste, moins de 1%, mais contrôle des gènes majeurs impliqués dans le dimorphisme entre reine et ouvrières (taille, fertilité et comportement). Chez les mammifères, seules 5 à 10% des cytosines du génome sont méthylées, dont une très large majorité (98%) intervient dans un contexte CpG de manière symétrique (sur chaque brin de l'ADN); les 2% restants apparaissent dans des contextes de type CHG ou CHH. Seuls certains types cellulaires, tels que les cellules souches embryonnaires ou les neurones, présentent un taux de 5mC dans un contexte génomique CHH dit non CpG plus important. Chez les plantes, la méthylation peut concerner jusqu'à 50% des cytosines du génome et intervenir dans les différents contextes (CpG et/ou non CpG), de manière symétrique ou asymétrique, suggérant une implication forte de cette marque épigénétique dans divers processus biologiques du monde végétal (plus détaillés dans les chapitres 9 et 10) ou chez les poissons (chapitre 11).

Chez les mammifères, les sites CpG ne sont pas répartis de manière uniforme le long du génome. Des régions denses en sites CpG, nommées «îlots CpG» (séquences de 100 à 1 000 bases, avec plus de 60% de CG; Illingworth and Bird, 2009), sont souvent localisées au niveau de séquences répétées du génome. Parmi elles citons les séquences satellites centromériques et péri-centromériques et les rétrotransposons, éléments mobiles du génome, d'origine virale, accumulés au cours de l'évolution. Dans ce cas, les îlots CpG sont méthylés et permettent une compaction de la chromatine. Au niveau des rétrotransposons, la méthylation/compaction de la chromatine assure une protection du génome en bloquant toute réplication et donc toute invasion par ces séquences. L'hétérochromatine constitutive est riche en séquences satellites péri-centromériques méthylées, afin de limiter les recombinaisons et les ségrégations chromosomiques indésirables. Il est à noter que le taux de méthylation de l'ADN au niveau des régions flanquantes des îlots CpG, est particulièrement variable en fonction du type cellulaire, du stade de développement, de la physiologie de l'animal ou de l'environnement. Localisée sur des éléments régulateurs ou dans des promoteurs en amont des gènes, la méthylation de l'ADN est généralement inhibitrice de l'expression. Localisée en intragénique, elle a plutôt un rôle activateur en limitant les démarrages de transcription illégitime (encadré 2.2). La méthylation de l'ADN intervient dans de très nombreux processus biologiques.

Les «écrivains» de la 5mC sont des ADN méthyltransférases (DNMT; figure 2.7). La famille des ADN méthyltransférases est caractérisée par des domaines fonctionnels extrêmement conservés au cours de la phylogénèse, dont un domaine de liaison à l'ADN et un domaine catalytique responsable du transfert de groupe CH₃. La plupart d'entre elles sont réparties selon deux classes :

- les ADN méthyltransférases *de novo*, capables d'apporter un groupe CH₃ sur des cytosines non méthylées et d'établir de nouveaux profils de méthylation propres à chaque type cellulaire (DNMT3A, DNMT3B, DNMT3C dans le règne animal; DRM pour *domains rearranged methylase* dans le règne végétal). La DNMT3L ne possède pas de domaine catalytique mais joue un rôle primordial comme cofacteur et guide de la méthylation *de novo* dans les cellules germinales souches;
- les ADN méthyltransférases, capables de reconnaître les sites CpG méthylés sur un seul brin d'ADN à l'issue de la réplication et d'assurer la propagation des patrons de méthylation à travers la division cellulaire (DNMT1 dans le règne animal; MET1 (methyltransférases 1) et CMT (chromomethylase) dans le règne végétal). Un cofacteur de DNMT1 important est le facteur URF1 (ou NP95), qui est capable de se lier au brin d'ADN néosynthétisé et hémi-méthylé et de recruter DNMT1 pour le maintien du profil de méthylation initial.

Une grande question est de savoir comment les DNMT *de novo* ciblent les séquences génomiques à méthyler (Hervouet et al., 2018). Les résultats de la littérature scientifique suggèrent que différents mécanismes mettent en jeu la présence de certaines modifications post-traductionnelles des histones et/ou celle de certaines séquences répétées et/ou l'intervention d'une famille de petits ARN non codants (les ARN Piwi (piARN), voir section «Activités des longs et petits ARN non codants dans les régulations épigénétiques»). La fixation de certains facteurs de transcription peut aussi façonner la distribution de la 5mC en masquant certaines cibles potentielles.

Les effets sur la transcription des gènes ne sont pas directement dus à la présence de 5mC, mais nécessitent la fixation de protéines nucléaires capables de reconnaître les

régions d'ADN méthylé (domaine de liaison à l'ADN méthylé ou « *methyl CpG-binding domain* », MBD). Ces protéines, appelées « Methyl Binding Protein » (MeCP1, MeCP2, MBD2, MBD3, etc.) sont les « lecteurs » de la 5mC. Ces protéines sont aussi capables de recruter d'autres protéines, des répresseurs transcriptionnels (mSin3A) ou des enzymes de modification des marques d'histones, en particulier des histones désacétylases. Les progrès de séquençage des deux dernières décennies ont été notables et profitables à l'analyse de la 5mC. En complément de la mise en œuvre d'une conversion de l'ADNg par le bisulfite de sodium (transformation de toutes les cytosines non méthylées en uracile par désamination et conservation des cytosines méthylées résistantes à ce traitement), des techniques d'analyse de la conformation de la chromatine ont permis de mettre en évidence l'implication de la 5mC dans l'architecture de la chromatine et l'apparition de boucles de régulation autour des gènes ou des clusters de gènes.

Parmi les lecteurs, il faut citer le facteur protéique en doigt de zinc CTCF (*CCCTC-binding factor*), ubiquitaire et très conservé chez les eucaryotes (Kim *et al.*, 2015). Sa fixation est méthyl-sensible. CTCF reconnaît et se fixe sur des sites riches en CpG non méthylés. Chez les mammifères, on dénombre 55 000 à 65 000 sites, distribués dans le génome dans les régions intergéniques (50 %), dans le corps des gènes (35 %) et au sein des régions proximales des promoteurs de gènes (15 %). La fixation du CTCF bloque les interactions entre *enhancer* et promoteur et l'activité transcriptionnelle. Il est montré aussi que CTCF aide à l'attachement de la chromatine à la matrice nucléaire, interagit avec la cohésine, qui stabilise les boucles de chromatine entre *enhancers* et promoteurs, et gère les interactions entre domaines intra et interchromosomiques, aboutissant à une structure chromatinienne supra-ordonnée. La méthylation de ces sites de fixation a pour conséquence une déstabilisation génomique ou une modification de la conformation de l'ADN et du statut transcriptionnel du domaine. Il est aussi montré que la fixation du CTCF limite la méthylation *de novo*.

Les « effaceurs » de la 5mC assurent une déméthylation active. Chez les mammifères, au cours du développement précoce de l'embryon, il est observé une grande vague de reprogrammation épigénétique associée à un effacement quasi total de la méthylation (seules quelques régions génomiques ne sont pas affectées par cette reprogrammation; parmi elles les centres d'empreinte parentale – chapitre 4). Par immunocytochimie (suivi de la 5mC à l'aide d'un anticorps anti-5méthylcytosine), il est possible de suivre, au cours des premiers clivages cellulaires de l'œuf en post-fécondation chez la souris, le devenir de la 5mC associée aux génomes maternel et paternel. Il est alors observé que la 5mC s'efface brutalement dès le stade 2 cellules au niveau du génome paternel, mais progressivement au niveau du génome maternel en suivant la division cellulaire. Ces observations ont conduit à postuler, d'une part, à une perte de méthylation par non-conservation des profils au cours de la réplication (pas d'activité DNMT1 nucléaire à ces stades) et, d'autre part, à la possibilité d'avoir un processus actif assuré par des activités enzymatiques.

En 2009, les enzymes TET (*ten-eleven translocation*) sont identifiées (figure 2.7) : elles permettent une conversion des 5mC en dérivés oxydés, en 5-hydroxyméthylcytosine (5hmC), elle-même transformée en 5-formylcytosine (5fC), puis en 5-carboxylcytosine (5caC). D'autres enzymes sont aussi capables d'intervenir pour exciser et réparer, induisant une perte de méthylation. Les différentes voies de déméthylation sont activables en fonction des *stimuli* extérieurs et conduisent à de nouveaux profils de méthylation pangénomique (Kriaucionis *et al.*, 2009).

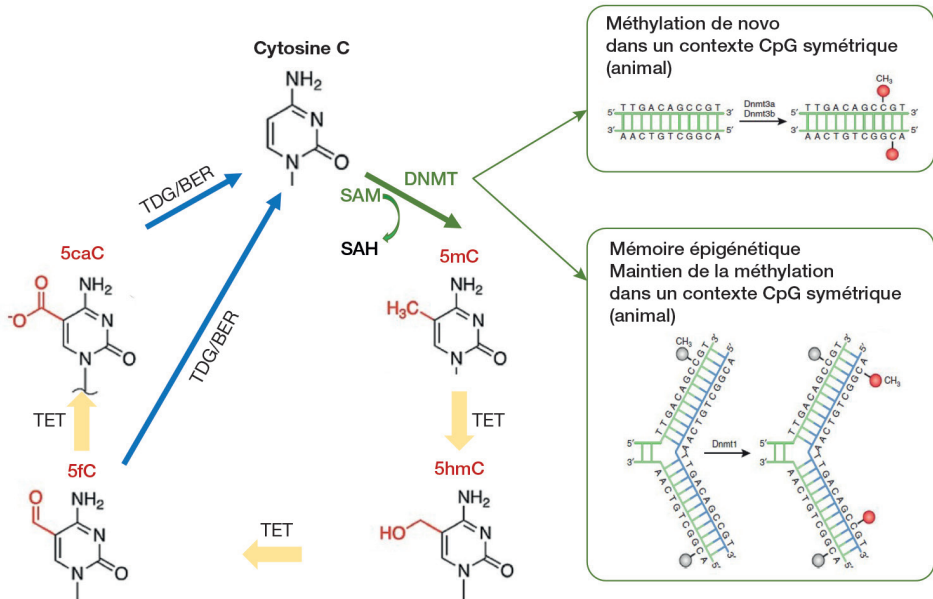


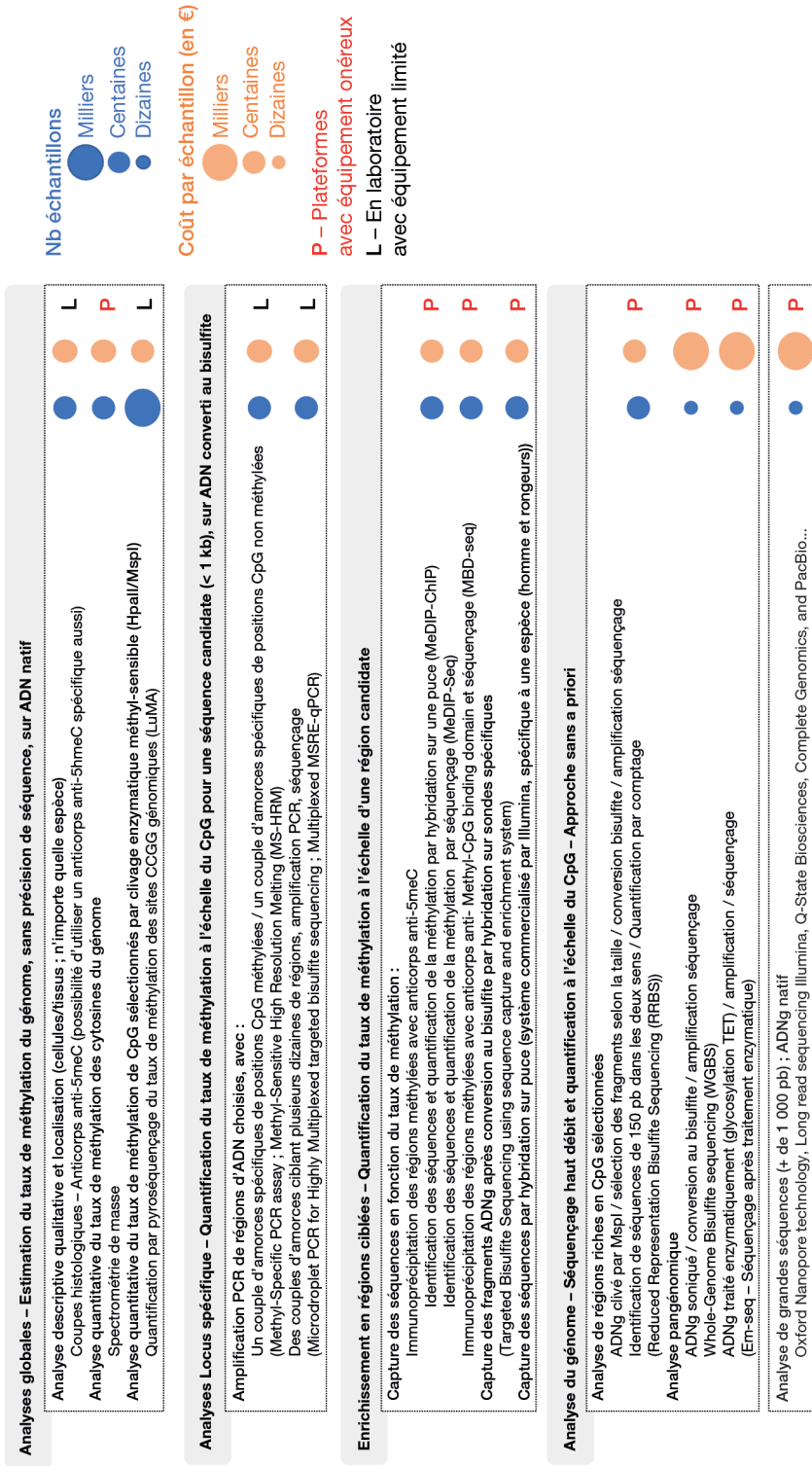
Figure 2.7. Cycle de méthylation et de déméthylation de l'ADN.

L'addition du groupe CH₃ sur la cinquième cytosine s'effectue sous le contrôle d'ADN méthyltransférases (DNMT). Les DNMT sont capables soit de méthyliser *de novo* l'ADN (DNMT3A et DNMT3B), soit de maintenir l'état de méthylation du brin d'ADN néosynthétisé lors de la réplication (DNMT1). Cette propriété confère à cette marque épigénétique son héritabilité au cours de la mitose et constitue une véritable mémoire (de l'identité cellulaire par exemple). La déméthylation active est sous le contrôle d'une famille d'enzymes, les TET pour « *ten-eleven translocation* » : elles convertissent la 5mC en 5-hydroxyméthylcytosine (5hmC), puis en 5-formylcytosine (5fC) et en 5-carboxylcytosine (5caC). Finalement, la forme canonique de la cytosine non méthylée est obtenue sous l'action d'une thymine ADN glycosylase (TDG) et du système de réparation par excision (*Base Excision Repair*, BER).

Les dernières décennies ont été très riches en obtention de données illustrant l'importance de la 5mC dans le contrôle des profils d'expression de gènes en lien avec la différenciation cellulaire, la physiopathologie humaine – cancers, maladies métaboliques, maladies neurodégénératives ou psychiatriques, vieillissement (voir chapitre 8 qui décrit l'horloge épigénétique), effets environnementaux – l'exposition à divers toxiques, l'hygiène de vie, le stress chez l'homme ou des modèles animaux, mais aussi en lien avec l'adaptation des plantes à leur environnement (voir chapitre 10) ou les interactions hôtes-agents pathogènes (voir chapitre 9). Ces progrès de la connaissance ont suscité le développement d'un large éventail de méthodologies (figure 2.8). Des progrès notables ont été obtenus par l'exploitation de toutes les nouvelles technologies liées au séquençage haut débit.

Comme pour les modifications post-traductionnelles des histones, l'écriture, la lecture et l'effacement de la méthylation de l'ADN sont sources de recherche de nouvelles molécules pouvant modifier l'état de méthylation et avoir des applications en thérapie. Les profils de méthylation de l'ADN sont aussi utilisés aujourd'hui en routine pour classer les tumeurs cancéreuses et adapter les traitements¹⁰.

10. <https://cordis.europa.eu/article/id/86034-dna-methylation-for-cancer-diagnosis/fr>



Nb échantillons



Coût par échantillon (en €)



P – Plateformes avec équipement onéreux
L – En laboratoire avec équipement limité

Figure 2.8. Présentation de quelques méthodes d'étude de la méthylation de l'ADN classées en fonction de l'échelle d'analyse.

► Activités des longs et petits ARN non codants dans les régulations épigénétiques

À partir des travaux de J. Watson et F. Crick, il a été considéré que l'information génétique était portée par les gènes, transcrits en ARNm, eux-mêmes traduits en protéines. Aujourd'hui, le concept ADN → ARN → protéine est bousculé. Il est bien montré qu'une part importante du génome des eucaryotes est non codante : chez l'homme, 51,8% du génome sont transcrits et seuls 1,2% (les exons) codent pour une protéine (GENCODE v25 annotation). Le progrès technique du séquençage appliqué à l'exploration du transcriptome (ensemble des ARN synthétisés à partir du génome) a aussi permis de découvrir l'existence de plus de 600 familles d'ARN dits « non codants » (leur séquence ne permet la synthèse d'aucune protéine répertoriée dans des bases de données).

Les familles d'ARN non codants se différencient les unes des autres par leur structure (linéaire, en tige-boucle ou circulaire) et par leur longueur, de quelques bases à plusieurs centaines; ainsi trouve-t-on les « petits » (< 200 nucléotides, *small non-coding RNAs* ou petits ARN non codants) et les « longs » (> 200 nucléotides, *long non-coding RNAs* ou longs ARN non codants). Certains participent aux processus épigénétiques.

Les longs ARN non codants jouent un rôle majeur dans le contrôle de la transcription de certaines régions chromosomiques. Certains d'entre eux sont localisés dans des régions intergéniques et d'autres dans le corps des gènes (généralement ils couvrent les exons et introns d'un gène). En général, une fois transcrits, ils s'accrochent sur des domaines chromosomiques cibles, bloquant ainsi la transcription d'un ensemble de gènes du domaine. L'exemple le plus connu est celui de *XIST*, dont l'histoire est intimement liée à celle de l'inactivation du chromosome X détaillée dans le chapitre 3.

Les petits ARN non codants se subdivisent en deux catégories selon leur fonction. Les structurants, généralement exprimés de manière constitutive, sont indispensables au fonctionnement normal de la cellule. Il s'agit des ARN ribosomiques (ARNr constituant les ribosomes) ou des ARN de transfert (ARNt apportant les acides aminés lors de la traduction). Les petits ARN non codants régulateurs, incluant les micro-ARN (miARN), les ARN interagissant avec PIWI (piARN) ou encore les ARN circulaires (circARN), sont exprimés de manière spécifique en réponse à un *stimulus* particulier au cours du développement de la différenciation cellulaire ou induits par des changements de l'environnement.

Les ARN de transfert ou ribosomiques sont étudiés depuis de longues années (1958) tandis que d'autres, comme les miARN, le sont depuis seulement une trentaine d'années. Leur biogenèse, leurs rôles ou leurs mécanismes d'action ont été décrits chez de nombreux eucaryotes.

Les miARN représentent la famille la plus étudiée. Au début des années 1990, le premier miARN identifié a été *lin-4* chez le nématode *Caenorhabditis elegans*. *Lin-4* n'existait que sous la forme d'une courte séquence de nucléotides dépourvue de phase codante. Il a été montré que cette courte séquence pouvait se lier à un ARNm, et que cette liaison empêchait toute production de protéine. Depuis, plusieurs milliers de miARN ont été identifiés dans 271 organismes différents¹¹. Si certains comme *lin-4* sont spécifiques d'un organisme, d'autres apparaissent plus conservés entre espèces.

11. Voir la base de données *miRBase102* : <http://www.mirbase.org>

Près de la moitié des miARN est localisée dans une région intergénique, la seconde moitié étant distribuée entre les régions codantes (avec une localisation préférentielle dans les introns) et les régions régulatrices non codantes des gènes. La majorité des miARN sont « isolés », mais certains sont organisés en clusters et transcrits en même temps. La description de la biogenèse des miARN et des régulations de leur activité ont fait l'objet de nombreuses études. Chez les mammifères, les miARN cibleraient environ 60% des gènes et auraient donc un rôle crucial dans la plupart des processus biologiques (prolifération, développement, différenciation, réponse au stress, adhésion cellulaire, mobilité, apoptose, etc.).

La découverte des piARN a suivi un chemin détourné. Vers la fin des années 1990, des protéines, les PIWI (*P-element induced wimpy testis*), ont été identifiées dans les cellules souches germinales de *Drosophila melanogaster*, avant d'être également décrites chez d'autres espèces. Ces protéines nucléaires appartiennent à la famille des Argonautes. Chez la souris, il existe trois PIWI protéines : MILI, MIWI1 et MIWI2, exprimées différemment selon le sexe de l'animal (MILI existe dans les 2 sexes, alors que MIWI1 et 2 ne sont exprimées que dans la lignée germinale mâle). Les protéines PIWI comportent 4 domaines : N-terminal, PAZ, MID et PIWI. Quelques années plus tard, plusieurs publications font état de petits ARN qui se lient à ces protéines. C'est la découverte des piARN, de petits ARN non codants de 26-33 nucléotides, présentant une méthylation de leur extrémité 3' ainsi qu'une base uridine à leur extrémité 5'. Leur biogenèse est aussi particulière, faisant intervenir plusieurs voies de régulation. Les piARN ayant été initialement identifiés dans la lignée germinale des animaux, il a été proposé qu'ils aient un rôle essentiel dans le développement des gamètes. Le premier rôle attribué à cette famille de petits ARN non codants est de lutter contre l'instabilité du génome, par la répression des transposons. Les éléments transposables peuvent en effet se dupliquer et se transloquer tout au long du génome, provoquant des insertions dans le corps des gènes, conduisant à une traduction en protéine non fonctionnelle ou dans les régions régulatrices en 3', dérégulant l'expression du gène. Ils engendrent également des cassures de l'ADN. Une fois captés par les protéines PIWI dans le cytoplasme et retournés dans le noyau, ces complexes sont capables à la fois de reconnaître les séquences transposables et de recruter les DNMT3L, 3A et 3B, afin d'induire localement une méthylation de l'ADN et de bloquer ainsi toute activité de ces séquences (LINE1 ou LTR). Il semblerait également que les complexes protéines PIWI-piARN répriment l'expression des transposons par des modifications locales de la condensation de la chromatine en recrutant des facteurs capables d'apposer des modifications post-traductionnelles des histones répressives (histone méthyltransférase). Ces actions conduisent à la fermeture locale de la chromatine, réprimant donc l'expression du transposon.

Mais depuis quelques années, l'existence d'une faible quantité de piARN et/ou de séquences dites *piRNA-like* présentant les mêmes caractéristiques que des piARN a également été montrée dans les cellules somatiques. Cette découverte s'accompagne également d'un nouveau rôle fonctionnel associé aux piARN. En effet, certains piARN identifiés chez plusieurs espèces ne sont pas complémentaires d'éléments transposables et seraient plutôt impliqués dans la régulation de l'expression génique, *via* l'activité catalytique médiée par l'un des domaines des protéines PIWI.

► Les protéines Polycomb et Trithorax

Les protéines du groupe Polycomb (PcG pour *Polycomb group*) et les protéines du groupe Trithorax (TrxG) ont été initialement identifiées chez la drosophile comme régulateurs des gènes homéotiques *HOX*, un groupe de gènes codant pour des facteurs de transcription spécifiant l'identité cellulaire tout au long de l'axe antéro-postérieur (Schuettengruber *et al.*, 2017). Des études ont permis d'identifier les séquences consensus permettant la fixation de ces protéines. Se fixant sur des éléments de réponse spécifiques (*Polycomb/Trithorax Response Elements, PRE et TRE*), médiateurs d'un état silencieux pour les PcG (voir chapitre 3) ou actif pour les TrxG de la chromatine, ces complexes protéiques constituent la mémoire de ces régulations au cours des divisions cellulaires, sous forme d'héritabilité épigénétique (Schwartz et Pirrotta, 2007). Chez les vertébrés, ces protéines régulent aussi les gènes *HOX*, mais sont en plus impliquées à la fois dans la régulation de la prolifération cellulaire, dans l'identité des cellules souches, dans les processus de l'empreinte parentale et dans l'inactivation du chromosome X chez la femelle.

► Conclusion

Au cours des dernières décennies, les mécanismes épigénétiques ont fait l'objet d'intenses recherches. Il est cependant important de garder à l'esprit que leur complexité n'est certainement pas complètement élucidée. En particulier, la cinétique des interactions entre mécanismes est toujours difficile à appréhender, puisque la plupart des méthodes d'analyse permettent de faire un état des lieux à un temps *t*, et non un suivi spatio-temporel des modifications moléculaires. Au vu du rôle de ces mécanismes dans de très nombreux processus biologiques tels qu'ils seront présentés au cours des chapitres suivants, un vaste champ de recherche s'offre aux chercheurs d'aujourd'hui et aux futures générations.

► Références bibliographiques

- Allfrey V.G., Faulkner R., Mirsky A.E., 1964. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc Natl Acad Sci U S A.*, 51(5), 786-794. <https://doi.org/10.1073/pnas.51.5.786>
- Cavalli G., Heard E., 2019. Advances in epigenetics link genetics to the environment and disease. *Nature*, 571, 489-499. <https://doi.org/10.1038/s41586-019-1411-0>
- Feinberg A., Vogelstein B., 1983. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature*, 301, 89-92. <https://doi.org/10.1038/301089a0>
- Hervouet E., Peixoto P., Delage-Mourroux R., Boyer-Guittaut M., Cartron P.F., 2018. Specific or not specific recruitment of DNMTs for DNA methylation, an epigenetic dilemma. *Clin Epigenetics.*, 10:17. <https://doi.org/10.1186/s13148-018-0450-y>
- Holliday R., Pugh J.E., 1975. DNA Modification Mechanisms and Gene Activity during Development. *Science*, 187, 226-232. <https://doi.org/10.1126/science.1111098>
- Hotchkiss R.D., 1948. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J. Biol. Chem.*, 175, 315-332.
- Hyun K., Jeon J., Park K., Kim J. 2017. Writing, erasing and reading histone lysine methylations. *J. Exp Mol Med.*, 49(4), e324. <https://doi.org/10.1038/emm.2017.11>
- Janke R., Dodson A.E., Rine J., 2015. Metabolism and epigenetics. *J. Annu Rev Cell Dev Biol.*, 31, 473-496. <https://doi.org/10.1146/annurev-cellbio-100814-125544>

- Johnson T.B., Coghill R.D., 1925. Researches on pyrimidines C111. The discovery of 5-Methyl-Cytosine in Tuberculinic Acid, the nucleic acid of the tubercle Bacillus1. *J Amer. Chem. Soc.*, 47, 2838-2844.
- Illingworth R.S., Bird A.P., 2009. CpG islands—'a rough guide'. *FEBS Lett*, 583(11), 1713-1720. [https://doi: 10.1016/j.febslet.2009.04.012](https://doi.org/10.1016/j.febslet.2009.04.012)
- Kim S., Yu N.K., Kaang B.K. 2015. CTCF as a multifunctional protein in genome regulation and gene expression. *Exp Mol Med.*, 47(6):e166. [https://doi: 10.1038/emm.2015.33](https://doi.org/10.1038/emm.2015.33)
- Kriaucionis S., Heintz N., 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, 324(5929), 929-930. [https://doi: 10.1126/science.1169786](https://doi.org/10.1126/science.1169786)
- Kringel D., Malkusch S., Lötsch J., 2021. Drugs and Epigenetic Molecular Functions. A Pharmacological Data Scientometric Analysis. *J. Int J Mol Sci.*, 22(14), 7250. [https://doi: 10.3390/ijms22147250](https://doi.org/10.3390/ijms22147250)
- Lang Z., Xie S., Zhu J.K., 2016. The 1001 Arabidopsis DNA Methylomes: An Important Resource for Studying Natural Genetic, Epigenetic, and Phenotypic Variation. *Trends Plant Sci.*, 21(11), 906-908. [https://doi: 10.1016/j.tplants.2016.09.001](https://doi.org/10.1016/j.tplants.2016.09.001)
- Lister R., Pelizzola M., Dowen R.H., Hawkins R.D., Hon G., Tonti-Filippini J., *et al.*, 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271), 315-22. [https://doi: 10.1038/nature08514](https://doi.org/10.1038/nature08514)
- Luger K., Mäder A., Richmond R., *et al.*, 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389, 251-260. <https://doi.org/10.1038/38444>
- Mattei A.L., Bailly N., Meissner A., 2022. DNA methylation: a historical perspective. *Trends Genet.*, 38(7), 676-707. [https://doi: 10.1016/j.tig.2022.03.010](https://doi.org/10.1016/j.tig.2022.03.010)
- McGrath J., Solter, D. 1984. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell*, 37, 179–183. [https://doi: 10.1016/0092-8674\(84\)90313-1](https://doi.org/10.1016/0092-8674(84)90313-1)
- Olins A.L., Carlson D., Olins D.E., 1975. Visualization of chromatin substructure: epsilon bodies. *J Cell Biology*; 64, 528-537.
- Pietrocola P., Galluzzi L., Bravo-San Pedro J.M., Frank Madeo F., Kroemer G., 2015. Acetyl coenzyme A: a central metabolite and second messenger. *Cell Metab.*, 21(6), 805-821. [https://doi: 10.1016/j.cmet.2015.05.014](https://doi.org/10.1016/j.cmet.2015.05.014)
- Scarano E., Iaccarino M., Grippo P., Winckelmans D., 1965. On methylation of DNA during development of the sea urchin embryo. *J. Mol Biol.*, 14(2), 603-607. [https://doi: 10.1016/s0022-2836\(65\)80211-x](https://doi.org/10.1016/s0022-2836(65)80211-x)
- Schuettengruber B., Bourbon M.H., Di Croce L., Cavalli G., 2017. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell.*, 171(1), 34-57. [https://doi: 10.1016/j.cell.2017.08.002](https://doi.org/10.1016/j.cell.2017.08.002)
- Schwartz Y.B., Pirrotta V., 2007. Polycomb silencing mechanisms and the management of genomic programmes. *Nat. Rev. Genet.*, 8(1), 9-22. [https://doi: 10.1038/nrg1981](https://doi.org/10.1038/nrg1981)
- Surani M.A., Barton, S.C., Norris M.L., 1984. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature*, 308, 548-550.
- Zhao Y., Garcia B.A., 2015. Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb. Perspect. Biol.*, 7(9), a025064. [https://doi: 10.1101/cshperspect.a025064](https://doi.org/10.1101/cshperspect.a025064)

Chapitre 3

L'inactivation du chromosome X

CLARA ROIDOR, VÉRONIQUE DURANTHON,
MAUD BORENSZTEIN

Les processus épigénétiques existent chez tous les organismes vivants avec une grande variété d'acteurs. Chez les mammifères, ce sont des mécanismes vitaux qui permettent le contrôle de l'expression génique et la mémoire de l'identité cellulaire conduisant au développement correct d'un individu. Ce contrôle subtil et héritable de l'état transcriptionnel au cours du temps et de l'espace se fait notamment *via* l'apposition de marques sur la chromatine, telles que la méthylation de l'ADN et les modifications post-transcriptionnelles des histones (voir chapitre 2). Dans ce chapitre, nous traiterons des connaissances actuelles concernant les mécanismes épigénétiques responsables de l'inactivation du chromosome X (ICX) chez la femelle de mammifères euthériens (mammifères pour lesquels le développement embryonnaire et foetal s'effectue dans l'utérus maternel), en détaillant certaines caractéristiques d'espèces.

Chez les mammifères euthériens, le déterminisme du sexe se fait *via* la transmission de chromosomes sexuels hétérogamétiques. Les mâles reçoivent un chromosome X et un chromosome Y, et les femelles deux chromosomes X. Parce que le chromosome X porte plus de 1 300 gènes importants pour de nombreux processus cellulaires, contre moins de 100 gènes pour le chromosome Y, majoritairement importants pour la détermination et la gamétogenèse du sexe masculin, la composition en chromosomes sexuels crée un déséquilibre de dose des produits du chromosome X (ARN et protéines). L'inactivation transcriptionnelle d'un des deux chromosomes X chez les mammifères femelles permet de compenser ce déséquilibre entre les sexes. Ce mécanisme épigénétique, mis en place très tôt durant l'embryogenèse, est absolument crucial. Son absence chez les femelles ou sa mise en place aberrante chez les mâles conduisent à une létalité embryonnaire précoce. L'ICX a pour finalité l'expression monoallélique de la majorité des gènes portés par le chromosome X dans chacune des cellules somatiques femelles. En effet, une fois établie de façon aléatoire, l'ICX est extrêmement stable et héritable à travers les divisions cellulaires. Les individus femelles sont donc mosaïques, toutes leurs cellules n'expriment pas le même chromosome X : la moitié des cellules somatiques contiennent le chromosome X inactif hérité de leur père, et l'autre moitié le chromosome X inactif hérité de leur mère. L'ICX est un paradigme de la régulation épigénétique des gènes, à l'échelle d'un chromosome entier, qui conduit à la formation d'un complexe hétérochromatinien répressif (Morey et Avner, 2011).

Un exemple parlant du mosaïcisme et de la clonalité de l'ICX est celui des chattes au pelage calico (Jaraud-Darnault et Abitbol, 2017). Chez le chat domestique, un des gènes qui code pour les pigments du pelage est situé sur le chromosome X. Deux versions de ce gène existent : un allèle codant le pigment orange et un allèle pour le noir. Ainsi, parce que les femelles peuvent porter les deux allèles mais que seul un des deux est exprimé de façon aléatoire, dans chacune des cellules la fourrure des femelles peut être orange et noire (figure 3.1A). Les régions de fourrure noire représentent des zones dont les cellules progénitrices ont inactivé le chromosome portant l'allèle orange. À l'inverse, les zones de fourrure orange sont dues à une inactivation du chromosome portant l'allèle noir. Les chats calico sont donc toutes des femelles XX. À l'inverse, les chats mâles XY sont porteurs soit de l'allèle noir, soit de l'allèle orange : ils ne montrent aucun mosaïcisme, jamais de fourrure bicolore noire et orange. Il s'agit d'un des exemples les plus connus de mosaïcisme de la fourrure, ici dû à l'inactivation aléatoire du chromosome X chez les femelles.

►► La compensation de dose génique

L'ICX permet d'obtenir l'expression d'une dose équivalente de produits des gènes du chromosome X chez les femelles et les mâles (figure 3.1B). Cependant, le chromosome Y porte une centaine de gènes dont une grande partie est conservée sur le chromosome X, en particulier au niveau de la région pseudo-autosomale (*pseudo-autosomal region*, PAR). Cette région permet la ségrégation correcte des chromosomes sexuels lors de la méiose chez les mâles et mime le comportement biallélique des autosomes. Elle n'est pas soumise à l'inactivation du chromosome X et les gènes qui y sont localisés font donc partie des quelques gènes qui « échappent » à l'inactivation chez les femelles (figure 3.1C). Parmi les gènes qui ne sont pas éteints, certains restent toujours exprimés à partir du X inactif dans toutes les cellules au cours de la vie. Ils échappent de façon constitutive à l'ICX, et les gènes localisés dans la PAR en sont l'exemple parfait. L'échappement peut aussi être facultatif, c'est-à-dire uniquement dans certaines cellules, certains tissus ou à certaines périodes du développement de l'individu ; il peut aussi être différent entre individus. De manière générale, les gènes qui échappent à l'inactivation représentent environ 7 % du total des gènes du chromosome X chez la souris et plus de 15 % chez la femme.

L'inactivation de l'un des deux chromosomes X, chez les femelles mammifères, permet une compensation fine de la quantité de produits transcrits à partir de ce chromosome chez les femelles par rapport aux mâles. Cependant, cela crée un déséquilibre de dose génique entre le seul X actif et les paires d'autosomes homologues. Il existe alors un second niveau de compensation de dosage génique : la surexpression des gènes du seul chromosome X actif permet de rétablir l'équilibre avec la transcription biallélique des gènes autosomiaux (figure 3.1C). Bien que ce mécanisme reste encore peu connu, il semble important que les gènes du chromosome X, codant pour des composants de complexes protéiques, soient transcrits à des niveaux similaires à ceux de leurs partenaires, codés par des autosomes. Pour avoir un niveau d'expression global similaire à celui des paires d'autosomes homologues, le chromosome X actif est surexprimé, d'environ 1,5 à 2 fois, grâce à la présence accrue de la machinerie de transcription et à des *pulses* de transcription, plus fréquents le long du chromosome actif. Cette surexpression a été mise en évidence à la fois dans les cellules mâles et dans les cellules femelles

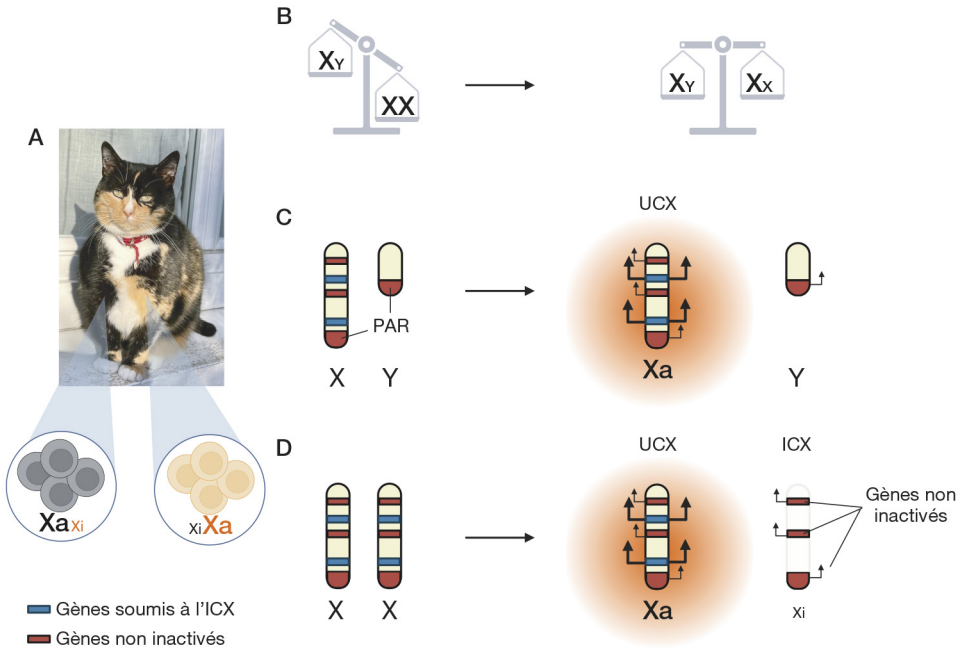


Figure 3.1. Compensation de dose et inactivation du chromosome X.

A. Chouchou, la chatte calico ou tricolore (blanc, orange et noir). Les secteurs de pelage orange et noir sont une conséquence de l'inactivation du chromosome X. **B.** L'inactivation du chromosome X permet d'équilibrer la dose de produits des gènes du chromosome X entre les femelles et les mâles. **C et D.** Seuls les gènes qui échappent à l'inactivation, tels que la région pseudo-autosomale (PAR), sont exprimés sur le chromosome inactif. Pour pallier le déséquilibre de dose génique entre le seul chromosome X actif et les paires d'autosomes homologues, un second niveau de compensation est mis en place : la surexpression globale des gènes du chromosome X actif. L'équilibre est rétabli avec la transcription biallélique des gènes autosomaux. Les gènes non inactivés ne semblent pas subir cette surexpression, car ils sont exprimés à partir des deux chromosomes X. Le seul chromosome X du mâle subit aussi cette surexpression par rapport aux paires d'autosomes homologues.

Xa : X actif; Xi : X inactif; ICX : inactivation du chromosome X; UCX : surexpression (*upregulation*) du chromosome X; PAR : région pseudo-autosomale. Figure créée sous BioRender.com.

(figure 3.1D). De façon intéressante, cette surexpression n'est observée qu'en présence d'un seul chromosome X actif dans des cellules diploïdes. Dans certaines situations de cellules femelles dont les deux chromosomes X sont actifs, tels que le zygote et les cellules souches embryonnaires murines, cette surexpression n'est pas observée.

Les aneuploïdies chromosomiques constituent des exemples de déséquilibre de dosage génique. Elles entraînent des malformations au cours du développement et sont pour la plupart non viables. Mais, du fait des compensations de dosage génique mises en place en fonction de la quantité de chromosomes X, les aneuploïdies des chromosomes sexuels sont quant à elles souvent viables et représentent la majorité des aneuploïdies détectées chez l'enfant. Les plus connues sont le syndrome de Klinefelter et de Turner. Ces syndromes sont à l'origine d'altérations du développement, de troubles hormonaux et de stérilité. Les individus « Klinefelter » ont trois chromosomes sexuels (XXY), soit 47 chromosomes; ce sont des hommes stériles présentant un blocage de

la spermatogenèse avant la méiose. Les individus « Turner » n'ont qu'un chromosome sexuel (X0), soit 45 chromosomes; ce sont des femmes stériles n'ayant peu ou pas de gamètes matures (perte des ovocytes au cours de la méiose; Heard et Turner, 2011).

► Découverte de l'inactivation du chromosome X

En 1948, Murray Barr et son étudiant en thèse Edwart G. Bertram (université Western Ontario, Canada) étudient au microscope des noyaux de neurones de chats par coloration de Nissl (mise en évidence des structures basophiles telles que l'ARN et l'ADN). Ils constatent une différence entre les femelles et les mâles dans le noyau de ces cellules (Barr et Bertram, 1949; Morey et Avner, 2011). Chez les femelles, une petite zone de chromatine densément colorée est observable à proximité du nucléole. Murray Barr émet l'hypothèse que cette région serait de l'hétérochromatine (condensation de la chromatine où les gènes ne sont pas transcrits) des chromosomes sexuels. Pour la première fois, ces observations montrent qu'il est possible de sexer des cellules différenciées d'euthériens grâce à une simple coloration et un microscope. Cette zone dense et hétérochromatinienne, présente uniquement chez les femelles, est depuis connue sous le nom de corpuscule de Barr.

En 1960, le généticien Susumu Ohno (Duarte, États-Unis) compare la quantité de matériel génique présent dans le noyau avec la quantité d'hétérochromatine visible par coloration. Il montre que des cellules cancéreuses diploïdes femelles et des cellules cancéreuses tétraploïdes mâles de souris possèdent deux chromosomes X et un corpuscule de Barr. Cependant, dans des cellules tumorales tétraploïdes de femelles contenant $4n$ chromosomes et donc 4 chromosomes X, 2 corpuscules de Barr sont observés, ce qui sous-entend la présence de 2 X actifs et de 2 X inactifs. Susumu Ohno remarque que cette hétérochromatine est similaire en taille et en contenu au chromosome X. Il fait le lien entre le chromosome X surnuméraire chez les femelles et le corpuscule de Barr. Le généticien propose également pour la première fois l'existence d'une compensation de dose entre le nombre de chromosomes X et la quantité d'autosomes dans chacune des cellules de mammifères (Morey et Avner, 2011).

Au même moment, ce sont les hypothèses de la généticienne Mary Lyon (Harwell, Royaume-Uni) qui sont à l'origine de la compréhension du mécanisme de l'inactivation aléatoire du chromosome X chez les femelles de mammifères. Mary Lyon fait le postulat suivant : le corpuscule de Barr peut être le chromosome X hérité de la mère ou bien celui du père dans différentes cellules d'une même souris. En effet, en étudiant le phénotype de souris portant différents allèles d'un gène du chromosome X codant pour le pelage, la généticienne observe un mosaïcisme de la fourrure tachetée ou mouchetée des femelles rongeurs. Ce mosaïcisme de la fourrure est similaire à celui de la chatte calico décrit plus haut. Mary Lyon propose ainsi pour la première fois qu'un des deux chromosomes X soit inactivé, de façon aléatoire, dans toutes les cellules d'un individu femelle, et cela dès le développement embryonnaire précoce. Au milieu des années 1960, Mary Lyon va plus loin et suggère qu'un centre de contrôle de l'ICX doit se trouver sur le chromosome X lui-même, permettant de mesurer le nombre de chromosomes dans chaque cellule et d'initier l'inactivation du chromosome X. Ceci sera effectivement démontré plus tard par la découverte du centre d'inactivation (*Xic*, pour *X inactivation center*). En hommage à cette chercheuse, le mécanisme d'inactivation du chromosome X a été nommé « lyonisation » (Lyon, 1961; Morey et Avner, 2011).

À la même période, des observations similaires sont faites par Ernest Beutler (États-Unis) chez la femme. Il remarque la présence de deux types différents de globules rouges dans le sang de patientes porteuses d'une mutation hétérozygote pour le gène *G6PDX* (glucose 6-phosphate déshydrogénase, gène porté par le chromosome X). Ce phénotype s'explique par l'expression de la protéine G6PDX dans les hématies normales et l'expression de la protéine déficiente dans les hématies pathologiques.

Ayant acquis la certitude que l'un des chromosomes X s'inactivait chez les femelles de mammifères, plusieurs laboratoires ont cherché à identifier la région du chromosome qui pourrait en être responsable. Bruce Cattanach (Cambridge, Royaume-Uni), dans les années 1970, propose l'existence d'un élément de contrôle du choix du chromosome inactivé, le Xce (pour *X-choice element*). Alors qu'il est attendu une inactivation aléatoire de l'un ou l'autre chromosome X, il apparaît que le choix du chromosome X inactivé est parfois biaisé. En effet, Bruce Cattanach croise plusieurs lignées différentes de souris et réalise que certaines portent un chromosome X plus résistant à l'inactivation que d'autres (Morey et Avner, 2011). Ce biais dans le choix suggère une influence de la séquence nucléotidique précise du X sur le processus de choix. Aujourd'hui encore, la nature exacte du Xce reste inconnue, plusieurs loci sont impliqués. Par ailleurs, le Xce n'a pas pu être mis en évidence chez l'humain, probablement du fait de la difficulté de ce type d'analyse chez cette espèce.

Peu de temps après, plusieurs équipes vont s'intéresser à la mécanistique de l'inactivation et du choix du chromosome X inactivé. Chez certains patients, il est possible de mettre en évidence des translocations réciproques entre régions du chromosome X et régions de chromosomes autosomiques. De façon surprenante, seule une partie de la translocation du chromosome X est inactivée, suggérant qu'il existe une région du X qui est nécessaire à l'inactivation. Différents laboratoires ont ensuite réalisé des translocations chez la souris, pour déterminer la région nécessaire et suffisante du chromosome X qui permet l'inactivation transcriptionnelle de l'autosome receveur. Sohaila Rastan (Clinical Research Center, Royaume-Uni) réduit cet élément de contrôle à un intervalle d'environ 1 000 kilobases (Rastan, 1983; Morey et Avner, 2011). Cet intervalle, suffisant pour initier l'ICX, sera réduit en taille et mieux caractérisé par l'étude de transgènes, et plus récemment par des approches mettant en évidence la conformation tridimensionnelle de cette région, réalisées par l'équipe d'Edith Heard (Institut Curie, France). L'inactivation du chromosome X dépend donc d'une petite région, le Xic (figure 3.2). Quand le Xic est délété, le chromosome X porteur de la délétion n'est pas inactivé. Une autre fonction du Xic est de permettre le comptage du nombre de chromosomes X dans la cellule et de n'initier l'inactivation que dans les cellules diploïdes contenant deux chromosomes X.

Suite à ces découvertes fondamentales pour la compréhension de l'inactivation du chromosome X, plusieurs laboratoires ont analysé en détail le Xic. En 1991, le gène *XIST* (*X-inactive specific transcript*) y est découvert, à la fois chez l'homme et la souris (Brown *et al.*, 1991; Morey et Avner, 2011; figure 3.2).

Le gène *XIST* est transcrit uniquement à partir du chromosome X inactif dans les cellules somatiques femelles. Il s'avère être l'acteur majeur et le premier élément moléculaire de contrôle de l'inactivation découvert. En 1996, l'intérêt se porte sur l'ARN du gène *XIST*. Des études moléculaires montrent que l'ARN *Xist* n'a pas de cadre ouvert de lecture et ne produit donc pas de protéine. Il s'agit d'un long ARN non codant, de près de 17 kilobases, localisé uniquement dans le noyau (Morey et Avner, 2011).

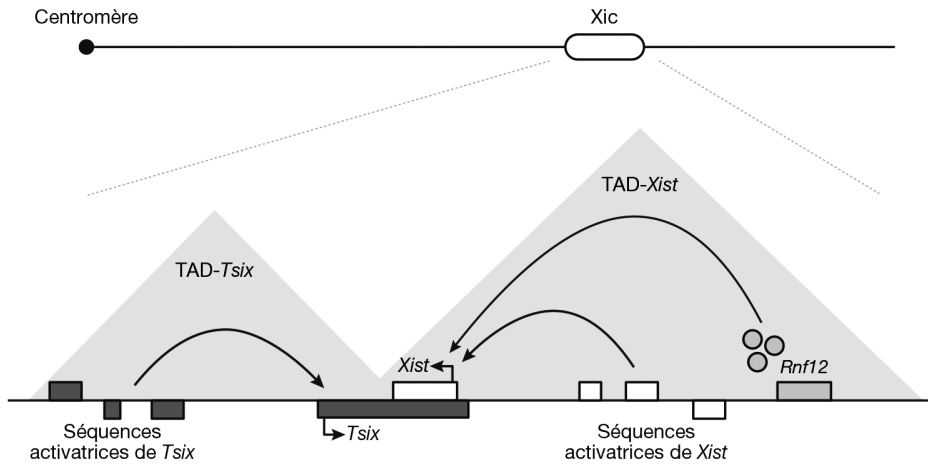


Figure 3.2. Le centre d'inactivation du chromosome X (Xic).

Chez la souris, le Xic se définit comme la région minimale et suffisante pour permettre l'initiation de l'ICX. Cette région comprend plusieurs gènes codant des protéines et des longs ARN non codants, dont RNF12, *XIST* et *TSIX*. Ces derniers sont des acteurs majeurs de l'ICX chez la souris. Les promoteurs des gènes *XIST* et *TSIX* se situent dans deux domaines topologiques (TAD) différents, dans lesquels résident leurs séquences régulatrices. Les TAD de *XIST* et *TSIX* s'étendent respectivement sur environ 550 kb et 330 kb. TAD : *topologically associating domain*; Xic : *X inactivation center*; kb : kilobases. Figure créée sous BioRender.com.

Par des techniques d'hybridation *in situ* par fluorescence de l'ARN (FISH), *XIST* est visualisé dans les noyaux de cellules somatiques femelles, au niveau du chromosome X inactif, qu'il recouvre complètement (figure 3.3). La création de mutants de souris, délétés pour le gène *XIST*, par l'équipe de Rudolph Jaenisch (Massachusetts Institute of Technology, États-Unis), permet ensuite d'observer que l'ICX ne peut être mise en place en absence de *XIST*, ce qui entraîne une létalité de tous les embryons femelles. Peu de temps après, le laboratoire de Jeannie Lee (Harvard, États-Unis) découvre la présence d'un ARN non codant, transcrit en antisens de *XIST*, et le nomme *TSIX* (Lee *et al.*, 1999). Chez la souris, l'ARN *TSIX* réprime *XIST* en *cis*, c'est-à-dire sur le même chromosome. Il est donc logiquement exprimé uniquement à partir du chromosome actif. Sa délétion sur un des deux chromosomes X entraîne l'inactivation préférentielle de ce dernier, et dans des cellules mâles, la perte de *TSIX* s'accompagne d'une expression aberrante de *XIST*. Ces expériences démontrent la nécessité de *XIST* et de sa bonne régulation par *TSIX* pour le développement des femelles chez les mammifères.

Entre les années 2000 et 2020, un grand nombre d'acteurs ont été découverts en lien avec l'inactivation du chromosome X : son initiation, son maintien, la régulation du gène *XIST*, mais aussi le choix du futur chromosome X inactif et le comptage du nombre de X par rapport aux autosomes. L'étude des boucles de la chromatine, c'est-à-dire des régions d'interactions à distance de l'ADN, à l'échelle du génome, a permis une meilleure compréhension de la régulation génique. En 2012, trois laboratoires¹² ont découvert l'existence de compartiments génomiques, tout le long du génome, conservés

12. Dirigés par Giacomo Cavalli (Institut de génétique humaine, France), Edith Heard (Institut Curie, France), et Bin Ren (université de Californie, San Diego, États-Unis).

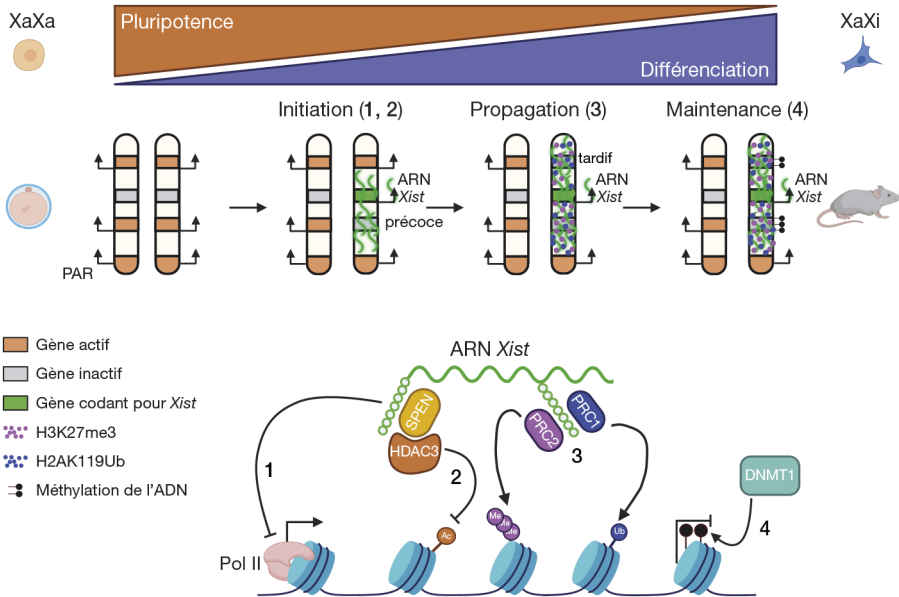


Figure 3.3. Mécanismes moléculaires de l'inactivation du chromosome X chez la souris.

La mise en place de l'inactivation du chromosome X se fait au cours du développement/différenciation cellulaire, lorsque la cellule sort de son état de pluripotence. Il existe une boucle de rétrocontrôle entre l'expression des facteurs de pluripotence et l'expression de l'ARN non codant *XIST*. Chez la souris, la transcription de *XIST* à partir du futur chromosome X inactif initie ce processus. L'ARN *XIST* recouvre complètement ce chromosome et agit comme une plateforme d'accueil pour les acteurs moléculaires de l'inactivation. Il y a tout d'abord exclusion de l'ARN polymérase II (1). Puis *XIST* recrute l'histone désacétylase HDAC3 par l'intermédiaire de la protéine SPEN. Cela permet d'initier l'inactivation transcriptionnelle des gènes précoces (2). Puis les complexes répressifs Polycomb 1 et 2 sont recrutés et apposent les marques d'histone répressive H2AK119Ub et H3K27me3 respectivement (3). Les gènes situés le long du chromosome X continuent à s'inactiver progressivement, jusqu'aux plus tardifs, les derniers gènes éteints. En revanche, certains gènes échappent à l'inactivation du chromosome X, les « escapees », dont font partie les gènes de la région pseudo-autosomale (PAR) ainsi que d'autres gènes répartis le long du chromosome X. Une fois l'inactivation du chromosome mise en place, elle est maintenue de façon stable au cours des divisions cellulaires. L'apposition de méthylation au niveau des cytosines des promoteurs des gènes inactivés par l'ADN méthyltransférase 1, DNMT1, permet cette maintenance (4). ac : groupement acétyl; me : groupement méthyl; ub : ubiquitination; PAR : région pseudo-autosomale; pol II : ARN polymérase II; PRC1 : complexe répressif Polycomb 1; PRC2 : complexe répressif Polycomb 2; Xa : chromosome X actif; Xi : chromosome X inactif. Figure créée sous BioRender.com.

entre les espèces. Ces domaines représentent des régions génomiques contenant des boucles de chromatine qui interagissent préférentiellement entre elles, par exemple une région *enhancer* et son promoteur. Ces domaines d'interactions préférentielles ont été nommés domaines topologiques (TAD, *topologically associating domains*). De façon très intéressante, Edith Heard et son étudiant en thèse Elphège Nora (Institut Curie, France) ont montré que le Xic est partagé entre deux TAD : l'un comprenant le promoteur de *XIST* et ses séquences régulatrices et le second comprenant le promoteur de *TSIX* et ses séquences régulatrices (Nora *et al.*, 2012) (figure 3.2). Nous allons détailler dans les sections suivantes les mécanismes moléculaires de l'inactivation du chromosome X, et sa mise en place au cours du développement, en soulignant les différences de modalités identifiées entre les espèces.

► La puissance des modèles murins

Ces dernières années, le domaine de l'épigénétique a vécu une révolution technologique qui a permis de faire grandement avancer la compréhension mécanistique de l'ICX. L'hétérochromatinisation du chromosome inactif et les modifications chromatiniennes associées ont pu être observées grâce au développement et à la production d'anticorps spécifiques contre les différentes marques d'histone, ce qui a contribué à l'étude de la méthylation de l'ADN et à la mise en évidence de la conformation tridimensionnelle du génome. L'utilisation des cellules souches embryonnaires de souris (mES) a ouvert la voie des études mécanistiques de l'ICX, grâce à la puissance de ce modèle. En effet, les cellules mES femelles ont deux chromosomes X actifs en culture. En modifiant ces conditions de culture, il est possible d'induire une voie de différenciation *in vitro*; l'ICX se met alors en place aléatoirement et récapitule ce qui est observé dans l'embryon post-implantatoire (voir section « Le cycle d'inactivation du chromosome X chez les rongeurs, un modèle à part ? »). La modification de l'expression des gènes du futur X inactif, les changements de paysages chromatiniens et le recrutement des facteurs associés à *Xist* ont été caractérisés par différentes techniques telles que l'immunofluorescence combinée au RNA-FISH (*RNA-fluorescence in situ hybridization*) et l'immunoprécipitation de la chromatine. Récemment, le développement des techniques de séquençage à haut débit et de spectrométrie de masse a permis l'étude de la transcription allèle-spécifique, de la conformation du génome par techniques de type 3C (*chromosome conformation capture*), mais aussi de toutes les protéines associées à l'ARN non codant *XIST* au cours de l'inactivation (immunoprécipitation de l'ARN suivi par spectrométrie de masse).

Ces études ont pu être conduites de façon dynamique, au cours de l'inactivation du chromosome X chez la souris, et ont permis de détailler plus précisément l'initiation, la propagation, puis le maintien de l'état silencieux au cours du développement (figure 3.3). Cette section met en exergue des mécanismes découverts chez les rongeurs. Leur identification pour d'autres espèces eutheriennes nécessite des études complémentaires.

L'initiation de l'inactivation

La pluripotence définit la faculté d'une cellule à se différencier en cellule du mésoderme, de l'endoderme ou de l'ectoderme, c'est-à-dire en une cellule de l'un des trois feuilletts embryonnaires. Les cellules pluripotentes sont maintenues dans cet état par l'expression de gènes codant des facteurs dits « de pluripotence ». Les cellules pluripotentes issues d'organismes femelles possèdent deux chromosomes X actifs. Lorsque ces cellules sortent de leur état de pluripotence et s'engagent dans une voie de différenciation, elles initient l'inactivation de l'un des deux chromosomes X (figure 3.3).

Ainsi, il existe un rétrocontrôle négatif entre les facteurs de pluripotence et l'initiation de l'ICX. Des chercheurs ont découvert que les facteurs de transcription liés à la pluripotence NANOG (codé par le gène *NANOG homeobox*), OCT3/4 (codé par le gène *POU5F1, POU class 5 homeobox 1*) et SOX2 (codé par le gène *sex determining region Y-box 2*) se fixent au locus de *XIST* réprimant son expression dans les cellules pluripotentes. De plus, les facteurs de transcription de pluripotence REX1 (codé par le gène *RNA exonuclease 1*), KLF4 (codé par le gène *Kruppel-like transcription factor 4*) et C-MYC (codé par le gène *myelocytomatosis oncogene*) activent l'expression de *TSIX* en se liant à son promoteur. Au cours des premières phases du développement

embryonnaire et dans les cellules souches embryonnaires pluripotentes, ces facteurs répriment donc l'expression de *Xist* soit directement, soit en activant *TSIX* (Morey et Avner, 2011). Mais lorsque les cellules se différencient, l'expression de ces facteurs de pluripotence décroît, permettant l'initiation de l'expression de *XIST* et donc la mise en place de l'ICX. À l'inverse, il a été montré, dans des embryons délétés pour *XIST*, que l'absence de compensation de dose empêchait la répression transcriptionnelle des gènes de pluripotence et l'entrée en différenciation des cellules. En effet, il semblerait que l'ICX doive être initiée dans les cellules femelles pour que ces dernières puissent se différencier correctement. Ainsi, les cellules mES femelles présentent un léger retard de différenciation par rapport aux mES mâles *in vitro*. La présence de 2 chromosomes actifs pourrait bloquer la sortie de pluripotence et la mise en place de l'inactivation spécifique aux femelles et ainsi expliquer ce retard. Si les cellules femelles perdent un chromosome X et deviennent X0, elles se comporteront comme des cellules mâles.

Dans ce processus d'initiation de l'ICX, la protéine RNF12, pour *Ring Finger Protein 12*, aussi appelée RLIM, joue également un rôle prépondérant pour activer l'expression de *XIST*. Cette ubiquitine ligase E3 est impliquée dans la dégradation de protéines. L'action de RNF12 est dose-dépendante et cette protéine est codée à partir du chromosome X (Morey et Avner, 2011). Chez les individus XX femelles, la double dose de RNF12 est suffisante pour réprimer REX1 et activer l'expression de *XIST*. Mais chez le mâle XY, une seule dose n'est pas suffisante pour dégrader de manière ciblée REX1, qui continue d'activer *TSIX* et donc de réprimer *XIST*.

Une fois le gène *XIST* activé à partir d'un des chromosomes X, son long ARN non codant recouvre progressivement le chromosome à partir duquel il est transcrit et participe à son inactivation. L'ARN *XIST* est un très long transcrit, dont la séquence est très peu conservée au cours de l'évolution des euthériens, à l'exception de plusieurs régions répétées en tandem. Ces régions sont importantes et permettent le recrutement de facteurs indispensables à l'inactivation transcriptionnelle du chromosome X. La délétion de ces régions répétées influe directement sur la mise en place de l'inactivation, le recrutement de *XIST* sur le chromosome et l'extinction génique.

L'ARN *XIST* agit donc comme une plateforme de recrutement de cofacteurs. Son ancrage à des régions spécifiques de la chromatine permet le recrutement du co-represseur SPEN (*Spn family transcriptional repressor*) au niveau des promoteurs et des séquences activatrices. Cette protéine est impliquée dans la répression transcriptionnelle. SPEN interagit à son tour avec l'ARN polymérase II et l'histone désacétylase HDAC3 (figure 3.3). Il y a alors exclusion de l'ARN polymérase II et hypoacétylation du futur X inactif. Cela conduit à l'extinction transcriptionnelle d'un premier groupe de gènes du chromosome X (dits « gènes précoces ») situés au niveau des sites d'ancrage de l'ARN *XIST*.

La propagation de l'inactivation

Le recouvrement de la totalité du chromosome X par les molécules d'ARN *XIST* va se poursuivre et entraîner l'extinction transcriptionnelle des gènes dits « tardifs » sur le chromosome X (figure 3.3). Après la perte des marques chromatiniennes activatrices sur le futur chromosome X inactif, il y a apposition de marques d'histones répressives. Les complexes répressifs Polycomb (PRC1 et PRC2) sont recrutés par les séquences répétées de *XIST* et conduisent respectivement à l'ubiquitination de la lysine 119

de l'histone 2 (H2AK119ub) et à la triméthylation de la lysine 27 de l'histone 3 (H3K27me3) (Morey et Avner, 2011). À ce jour, seules les marques chromatiniennes les mieux caractérisées ont été recensées, mais d'autres marques d'hétérochromatine interviennent, comme certains variants d'histones et d'autres modificateurs épigénétiques. À ce stade, tous les gènes du chromosome X inactif sont transcriptionnellement réprimés, à l'exception des gènes qui échappent à l'inactivation notamment dans la région pseudo-autosomale.

La maintenance

Une fois mise en place, l'inactivation du chromosome X se propage de manière clonale, des cellules mères aux cellules filles, à travers des centaines de divisions cellulaires grâce à des mécanismes épigénétiques différents. Lorsque l'inactivation a eu lieu, les gènes éteints sont méthylés au niveau des cytosines de leur promoteur. Cet état est maintenu au cours des divisions cellulaires sous l'action de l'ADN méthyltransférase 1 (DNMT1), spécialisée dans le maintien et la mémoire de l'information épigénétique (Morey et Avner, 2011 ; figure 3.3). L'utilisation de drogue induisant une hypométhylation de l'ADN a conduit à la réactivation de nombreux gènes sur le chromosome inactif. La méthylation de l'ADN est donc importante pour le maintien de l'inactivation, contrairement à *XIST*. En effet, une délétion du gène *XIST*, une fois l'ICX mise en place, a un effet mineur sur quelques gènes, suggérant un rôle dans l'initiation, mais pas dans le maintien global de l'inactivation.

L'étude de la conformation tridimensionnelle du génome a, quant à elle, permis l'observation d'une structure spécifique du chromosome X inactif. Lors de l'ICX, les TAD disparaissent et font place à seulement deux « mégadomaines » le long du chromosome inactif. La délimitation des domaines d'interaction, les TAD, et les régulations de l'ADN à courtes et moyennes distances sont perdues, sauf dans les régions où sont localisés les gènes qui échappent à l'inactivation. La perte des TAD, pour l'ensemble des gènes inactivés, semble être précédée par l'extinction transcriptionnelle des gènes et pourrait jouer un rôle dans le maintien de l'inactivation au cours du temps.

Enfin, il est connu depuis les années 1990 que les deux chromosomes X des mammifères se répliquent de manière asynchrone. En utilisant des cellules hybrides souris-humains en culture et en étudiant leur ADN nouvellement synthétisé, des chercheuses ont découvert que le gène *HPRT* (*Hypoxanthine phosphoribosyltransferase 1*) est répliqué au début de la phase S sur le chromosome X actif, et en fin de phase S sur le chromosome X inactif. De plus, lorsque le gène *HPRT* est réactivé sur le chromosome X inactif, il se réplique plus rapidement. Cela supporte l'idée d'une relation entre répllication et transcription. Cette répllication asynchrone pourrait aussi jouer un rôle dans le maintien de l'inactivation du chromosome X au cours du temps.

» Le cycle d'inactivation du chromosome X chez les rongeurs, un modèle à part ?

L'inactivation du chromosome X est mise en place très tôt au cours du développement embryonnaire. Comme les mécanismes de l'ICX, le cycle d'inactivation/réactivation du chromosome X au cours du développement a été particulièrement étudié chez la souris. Ce modèle mammifère est en effet facilement accessible et a permis

le développement et l'utilisation d'outils génétiques nombreux et puissants. Plusieurs laboratoires ont étudié l'ICX dans des embryons murins mâles et femelles, au départ par FISH et immunofluorescence, puis plus récemment par transcriptomique allèle-spécifique à l'échelle de la cellule unique (Borensztein *et al.*, 2017). Toutes ces données permettent aujourd'hui d'avoir une idée précise de la dynamique d'inactivation et de réactivation des chromosomes X dans les tissus embryonnaires et extra-embryonnaires de cette espèce (figure 3.4).

Chez la souris, une première vague d'inactivation s'effectue très rapidement après la fécondation. Lors de l'activation du génome zygotique, les deux chromosomes X sont activement transcrits, et l'ARN *XIST* est exprimé uniquement à partir du chromosome paternel. En effet, l'embryon hérite de la lignée germinale maternelle un allèle de *XIST* réprimé par la présence d'un domaine répressif, caractérisé par la triméthylation de la lysine 27 de l'histone H3 (H3K27me3) au niveau du promoteur de *XIST*. Il s'agit d'un mécanisme d'empreinte génomique non canonique qui conduit à une inactivation de *XIST* sur le chromosome X maternel; seul le chromosome X paternel exprime *Xist* et initie l'inactivation. L'activation transcriptionnelle monoallélique de *XIST* est donc non aléatoire, puisque soumise à l'empreinte génomique parentale (voir chapitre 4), et observable dans l'embryon préimplantatoire et les tissus extra-embryonnaires. Dès le stade 4 cellules, les premiers gènes le long du chromosome X paternel sont éteints, jusqu'à une extinction presque complète de ce chromosome au stade blastocyste. Une fois éteint, le chromosome X paternel reste inactivé dans les tissus extra-embryonnaires jusqu'à la fin de la gestation (figure 3.4). En cas de délétion du gène *XIST* situé sur le X paternel, l'allèle maternel ne peut pas prendre le relais. L'allèle maternel *XIST* ne peut être transcrit avant le stade blastocyste, stade auquel l'empreinte non canonique portée par H3K27me3 est effacée. Cela conduit à une dérégulation majeure des gènes du X, principalement dans les tissus extra-embryonnaires. Le placenta ne se développe pas de façon appropriée et perd son rôle nutritionnel dans le développement du fœtus, conduisant à une létalité.

Au stade blastocyste, la première différenciation cellulaire contribue à la mise en place du trophoctoderme (à l'origine du placenta) et de la masse cellulaire interne (MCI), qui elle-même donnera l'endoderme primitif (à l'origine du sac vitellin) et les cellules de l'épiblaste (futur fœtus). Au stade blastocyste, le statut du chromosome X est extrêmement dynamique et transitoire. En quelques heures, le chromosome X paternel va être réactivé uniquement dans les cellules pré-épiblastiques. Ces cellules formeront l'individu en lui-même et sont à l'origine des cellules souches embryonnaires murines (mES), ce qui explique la présence de deux chromosome X actifs dans ces dernières (figure 3.4). L'inactivation soumise à l'empreinte parentale est maintenue dans les tissus extra-embryonnaires, mais perdue dans les cellules du futur embryon.

Rapidement après l'implantation de l'embryon, une seconde vague d'inactivation se produit dans les cellules de l'épiblaste autour de 6-6,5 jours de développement. Cette fois-ci, l'ICX se fait de manière aléatoire entre le chromosome X maternel et le chromosome X paternel (figure 3.4). L'inactivation est ensuite stable, maintenue et transmissible au cours des divisions cellulaires dans les différents tissus. La moitié des cellules somatiques exprime donc le chromosome paternel et l'autre moitié le maternel. Ainsi, toutes les souris femelles XX sont mosaïques dans l'expression des gènes localisés sur le chromosome X.

Cependant, un événement spécifique d'effacement de l'état silencieux du chromosome a lieu dans les cellules germinales primordiales. En effet, ces cellules, à l'origine des futurs gamètes, subissent une reprogrammation épigénétique globale et drastique de leur génome, se caractérisant par un profond remodelage de la chromatine, une déméthylation globale du génome, un effacement des marques de l'empreinte parentale (voir chapitre 4), accompagnés d'une réactivation du X inactif pour les femelles (figure 3.4). Dans tous les cas, la reprogrammation germinale est essentielle pour la spécialisation des cellules reproductrices et la perpétuation de l'espèce. Chez les femelles, avant transmission à la future génération, le chromosome X inactif est réactivé dans les cellules primordiales germinales, l'ARN *Xist* est perdu, les marques répressives de la chromatine effacées et les gènes réexprimés. Il existe donc un véritable cycle d'inactivation puis de réactivation du chromosome X au cours de l'embryogenèse.

Bien que l'ICX aléatoire ainsi que la réactivation du chromosome X dans la lignée germinale semblent conservées chez les euthériens, la présence de deux vagues d'inactivation successives est une particularité des rongeurs, acquise au cours de l'évolution. Une hypothèse pouvant expliquer cette singularité est un développement préimplantatoire rapide chez la souris, avec une activation transcriptionnelle du génome embryonnaire et une implantation précoces, pouvant nécessiter la mise en place rapide de la compensation de dose génique.

► Diversité du processus d'inactivation du chromosome X chez les euthériens

La grande majorité des études portant sur l'inactivation du chromosome X chez les mammifères a été menée chez la souris, comme décrit précédemment. Pendant de nombreuses années, le cycle et les mécanismes d'inactivation du chromosome X étudiés chez cette espèce ont été tenus pour représentatifs du processus chez l'ensemble des mammifères euthériens.

Ce n'est que dans les années 2010 que des chercheurs spécialistes du développement des embryons de mammifères euthériens non murins, dont Jean-Paul Renard et Véronique Duranthon (INRAE, France), conscients des différences importantes de régulation de l'expression des gènes au cours des premières étapes du développement des différentes espèces, posent la question de la conservation des mécanismes chez les euthériens. Comme chez la souris, ils utilisent des techniques de RNA-FISH pour mettre en évidence dans chaque cellule l'expression et l'accumulation de *XIST*. En collaboration avec l'équipe d'Edith Heard, ils étudient l'ICX dans des embryons de lapin et dans des embryons humains, dont l'activation transcriptionnelle du génome est beaucoup plus progressive et tardive que chez la souris (stade 8 cellules chez le lapin et 4-8 cellules chez l'humain) (Morey et Avner, 2011 ; Okamoto *et al.*, 2011). Ainsi, ils mettent en évidence pour la première fois une grande diversité dans la cinétique et la régulation de l'initiation de l'inactivation du chromosome X parmi les mammifères euthériens (figure 3.5).

En effet, les embryons de lapin et humains ne suivent pas du tout le cycle d'inactivation précédemment décrit. Contrairement à la souris, l'expression de *XIST* dans ces embryons n'est pas soumise à l'empreinte parentale. Dès l'activation transcriptionnelle du génome (stade 8 cellules chez le lapin et 4-8 cellules chez l'humain), l'ARN

XIST est exprimé de façon biallélique dans les embryons femelles, mais également de façon transitoire à partir du chromosome X maternel dans les embryons mâles. Une telle expression de *XIST* maternel entre l'activation du génome et le stade blastocyste n'avait jamais été observée chez la souris.

Au stade blastocyste, l'accumulation du transcrit *XIST* le long du chromosome X est détectée uniquement chez les embryons de lapins femelles, et toujours de façon biallélique dans près d'un quart des cellules du trophoctoderme. Ce n'est que plus tard, après environ 24 heures, que l'accumulation de l'ARN *XIST* sur un seul chromosome est observée dans ces cellules trophoblastiques de lapin, qui ont alors acquis le comportement habituel d'une cellule somatique. Dans la masse cellulaire interne du blastocyste de lapin, c'est aussi plus tardivement (120 h *post coitum*) que l'accumulation de *XIST* est détectable sur un ou sur les deux chromosomes X (Okamoto *et al.*, 2011). Il n'y a donc pas de réactivation d'un X paternel préalablement inactivé avant une inactivation aléatoire, comme on l'observe chez la souris dans ce compartiment embryonnaire, mais plutôt une mise en place progressive de l'inactivation d'un chromosome X. Enfin, la répression transcriptionnelle des gènes du chromosome X chez les embryons femelles de lapin ne commence qu'au stade blastocyste et peut se produire sur les deux chromosomes lorsque deux domaines d'accumulation de *XIST* sont en place. Chez cette espèce, l'enrichissement en H3K27me3 semble étroitement associé à l'accumulation de *XIST* sur les chromosomes X (figure 3.5).

L'analyse de blastocystes humains issus de procréation médicalement assistée montre également une accumulation biallélique de *XIST* chez les embryons femelles et monoallélique chez les mâles (Okamoto *et al.*, 2011). Mais, de façon différente à la fois du lapin et de la souris, cette accumulation s'est avérée être découplée de la répression transcriptionnelle des gènes du chromosome X. En effet, les gènes chez l'humain ne semblent pas montrer d'expression monoallélique au stade blastocyste. C'est la première mise en évidence d'une accumulation de *XIST* découplée de l'inactivation du chromosome X correspondant au cours du développement. Une étude récente des groupes de Fredrik Lanner et Rickard Sandberg (Institut Karolinska, Suède; Petropoulos *et al.*, 2016) se basant sur des méthodes d'études du transcriptome allèle-spécifique (mettant à profit des polymorphismes au niveau d'un nucléotide unique, le *single nucleotide polymorphism*, ou SNP, portés par le chromosome X pour distinguer l'origine parentale des allèles exprimés) et couplées à des analyses en cellule unique, confirme l'expression biallélique de la majorité des gènes du chromosome X dans le blastocyste humain. Les auteurs suggèrent également l'existence transitoire d'une compensation de dose génique, avant l'ICX, appelée « *dampening* » sur les deux chromosomes.

Chez les animaux d'élevage, l'inactivation du chromosome X et l'expression de *XIST* ont été étudiées depuis les années 2000, mais les conclusions restaient incertaines car variables selon les techniques utilisées : analyse par RT-PCR (*Reverse Transcription-Polymerase Chain Reaction*) de l'expression du transcrit *XIST* au cours du développement préimplantatoire, comparaison de l'expression de *XIST* et de gènes du chromosome X entre embryons mâles et femelles, entre embryons fécondés, parthénotes et androgénotes (embryons issus d'une reproduction uniquement à partir de gamètes femelles ou mâles respectivement), utilisation de SNP pour détecter l'expression mono ou biallélique de *XIST* et son origine parentale. Finalement, les études les plus récentes, parce qu'elles utilisent des méthodes plus appropriées de

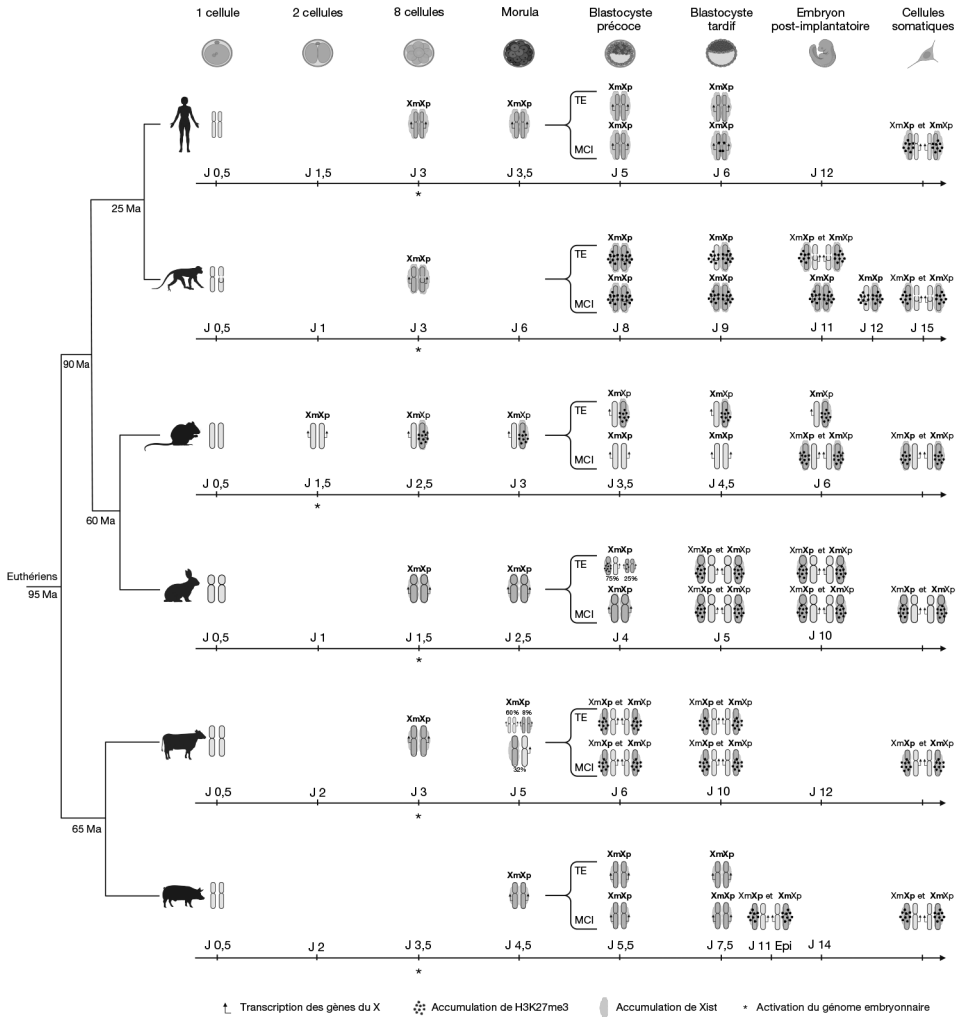


Figure 3.5. Diversité du processus d'inactivation du chromosome X chez les euthériens.

Au sein du clade des mammifères euthériens, il existe une très grande variété dans le processus d'initiation de l'inactivation du chromosome X. Les deux vagues d'inactivation observées chez la souris semblent être une exception au cours de l'évolution. Chez l'homme, la transcription de l'ARN *XIST* débute dès l'activation du génome embryonnaire, à partir des deux chromosomes X. L'inactivation transcriptionnelle des gènes pourrait commencer très tardivement dans l'embryon post-implantatoire humain, de façon aléatoire. Chez le macaque, l'inactivation est uniquement aléatoire également. Tout d'abord, les chromosomes X de macaques sont actifs, recouverts par l'ARN *XIST* et enrichis en marques d'histones répressives H3K27me3. Au cours du développement, l'inactivation d'un des deux chromosomes a lieu, et seul ce chromosome conserve l'accumulation de l'ARN *XIST* et de H3K27me3 après l'implantation. Ce même processus a lieu dans les cellules du trophoblaste et plus tardivement dans la masse cellulaire interne. Chez le lapin, l'inactivation est également aléatoire et semble plus précoce dans le trophoblaste que dans la masse cellulaire interne. Chez le bovin, l'initiation de l'inactivation est observée dès le stade blastocyste précoce, bien que la compensation de dosage génique semble bien plus tardive. Enfin, chez le porc, l'inactivation est seulement observée dans l'épiblaste après implantation. Epi : épiblaste mature; MCI : masse cellulaire interne; TE : trophoblaste; J : jour de développement; Ma : millions d'années; Xm : chromosome X maternel; Xp : chromosome X paternel; X : chromosome X inactif. Figure créée sous BioRender.com.

transcriptomes sur cellules uniques, de façon allèle-spécifique, apportent des résultats moins controversés (Ramos-Ibeas *et al.*, 2019). Ainsi, dans l'embryon de porc, les analyses utilisant des SNP portés par les gènes du chromosome X ont d'abord permis d'écarter l'hypothèse d'une ICX soumise à empreinte parentale. Les analyses en RNAseq (*RNA sequencing*) sur cellules uniques montrent que la compensation de dosage génique pour les gènes du chromosome X entre embryons femelles et mâles n'est atteinte que tardivement. En effet, elle n'est observée dans l'épiblaste des embryons sphériques que 10-11 jours après fécondation, c'est-à-dire juste avant le début de l'élongation des annexes extra-embryonnaires précédant, chez les ongulés, l'implantation très tardive (jour 14 chez le porc). Cette compensation de dosage génique est concomitante de l'apparition de domaines de H3K27me3 sur l'un des chromosomes X de chaque cellule épiblastique (figure 3.5).

Chez le cheval, *XIST* est surexprimé chez les embryons femelles aux stades 8 cellules, morula et blastocyte. L'inactivation du chromosome X ne semble pas soumise à l'empreinte parentale et la compensation de dosage génique entre chromosomes X et autosomes ne débute pas avant le stade blastocyste.

Dans l'embryon bovin, à partir de l'activation transcriptionnelle du génome embryonnaire (8-16 cellules dans cette espèce), l'expression de *XIST* est plus forte dans les embryons femelles que dans les embryons mâles. Les premiers signes d'initiation de l'ICX chez le bovin apparaissent dès le stade morula avec la mise en évidence par RNA-FISH des premières accumulations de *XIST* dans les noyaux des embryons femelles. Comme chez le lapin, il y a coexistence au sein des embryons de cellules présentant l'accumulation de *XIST* sur 0, 1 ou 2 chromosomes X, mais le pourcentage de cellules présentant l'accumulation de *XIST* sur un seul chromosome X va croissant au fur et à mesure du développement femelle. La colocalisation de *XIST* avec H3K27me3 n'est visible qu'à partir du stade blastocyste à 7 jours de développement. Pourtant, chez cette espèce, le dimorphisme sexuel dans l'expression génique est encore très marqué à ce stade, et la compensation de dosage génique entre embryons mâles et femelles ne semble apparaître qu'entre 11 et 14 jours, soit, comme chez le porc, juste avant ou au tout début de l'élongation des annexes extra-embryonnaires, qui précède l'implantation au 20^e jour (figure 3.5).

La découverte d'exceptions au modèle murin d'inactivation du chromosome X a connu un rebondissement majeur en 2021, avec la description détaillée de la cinétique d'ICX au cours du cycle de vie du macaque, par Ikuhiro Okamoto et Mitinori Saitou (Institut Riken, Japon; Okamoto *et al.*, 2021). Chez le macaque, l'inactivation du chromosome X se fait seulement après l'implantation, au 9^e jour, dans les tissus extra-embryonnaires et, beaucoup plus tard, au 17^e jour, dans les lignages embryonnaires dérivés de la masse cellulaire interne (figure 3.5). Les processus moléculaires et leur lien avec l'inactivation diffèrent aussi de ce qui est connu chez les autres espèces. En effet, l'accumulation de *XIST* et des marques d'histones H2AK119Ub et H3K27me3 se fait d'abord sur les deux chromosomes X actifs. Ensuite, *XIST* puis les histones répressives sont séquentiellement éliminées du chromosome X qui restera actif. À l'inverse, le futur X inactif conserve ces marques, et l'inactivation transcriptionnelle commence à peine. Il y a donc ici un découplage pendant une longue période (jusqu'à 14 jours) entre l'accumulation de *XIST* autour du chromosome X et le recrutement des histones répressives d'une part, et l'inactivation transcriptionnelle d'autre part. Dans les cellules

germinales primordiales de macaques, la réactivation du chromosome X a lieu très tôt au cours du développement, ce qui conduit à une expression biallélique des gènes du chromosome X dans la lignée germinale.

Finalement, l'étude fine des cinétiques d'accumulation de *XIST* et de l'apposition des marques d'histones répressives dans différentes espèces d'euthériens montre clairement que le modèle murin, bien que très étudié, est insuffisant pour rendre compte des processus chez l'ensemble des euthériens. Au contraire, de nombreuses différences existent entre espèces et sont importantes à explorer.

► Avantages et pathologies liés à l'inactivation du chromosome X

La reproduction sexuée, qui implique l'intervention de deux génomes parentaux, permet à l'espèce un brassage des allèles dans la population, ainsi qu'une meilleure adaptabilité face à l'apparition de mutations. En effet, la présence de deux copies de chaque gène est avantageuse lorsqu'une des copies présente une mutation délétère. Ceci reste vrai pour le chromosome X chez les femelles. De plus, dans la lignée germinale femelle, les deux chromosomes X peuvent recombinaison lors de la méiose et éviter l'accumulation de mutations décrite pour le chromosome Y.

Dans le cas des chromosomes hétérogamétiques mâles, la mutation d'un gène du chromosome X (mutation liée à l'X) ne pourra pas être « secourue » par l'autre allèle. Les mutations chez le mâle se manifestent donc car ils sont hémizygotés, c'est-à-dire que les gènes du chromosome X chez le mâle sont présents en une seule copie dans un organisme pourtant composé de cellules diploïdes. Chez la femelle, l'hétérogénéité de l'expression des gènes diploïdes du chromosome X a pour conséquence des différences d'expression entre les cellules à l'échelle du tissu, de l'organe et de l'individu. Deux cellules pourtant côte à côte peuvent ainsi exprimer différents répertoires géniques. La possibilité d'hétérozygotie est unique chez les individus XX pour les gènes portés par le chromosome X (à l'exception de la PAR). Ceci donne donc un avantage adaptatif crucial aux femelles, dans le cas de mutations délétères récessives. En effet, l'inactivation aléatoire du chromosome X introduit une diversité phénotypique. Chez la femme, la moitié seulement des cellules exprimerait l'allèle muté dans le cas d'une mutation sur le chromosome X.

Par exemple, la dysplasie ectodermique hypohidrotique est une maladie qui touche les cellules dérivées du feuillet embryonnaire de l'ectoderme (poils, muqueuse, dents). Dans la forme liée à l'X, les femmes sont asymptomatiques ou présentent un phénotype léger à modéré. Chez les hommes en revanche, cette maladie est plus grave. Elle entraîne une diminution, voire une absence de sudation, une intolérance à la chaleur, une absence partielle ou totale de denture, une dystrophie faciale, des atteintes digestives et/ou pulmonaires. Chez la femme, plus l'inactivation se fera sur le chromosome X portant la mutation, plus nombreuses seront les cellules normales et plus atténué sera le phénotype.

Dans le cas de certaines mutations liées à l'X, les cellules exprimant la copie non mutée auront un avantage développemental, une meilleure aptitude à la survie, à la division et à la différenciation cellulaire. La patiente pourra alors présenter une inactivation biaisée, préférentiellement vers le chromosome portant la copie mutée.

Dans ce cas, la mutation n'influe pas sur le choix du chromosome éteint, mais plus tard sur la survie cellulaire. Cela permet de diminuer l'impact de certaines mutations du chromosome X chez les femmes.

Enfin, de nombreux gènes impliqués dans le système immunitaire sont localisés sur le chromosome X, ainsi que des gènes « suppresseurs de tumeur ». Plusieurs de ces gènes font partie des gènes dont l'échappement est facultatif. Ces gènes échappent à l'ICX dans certaines cellules, et ici spécifiquement dans certaines cellules immunitaires. Cela entraîne un dimorphisme sexuel dans la réponse immunitaire, avec des différences dans l'expression des gènes liés à l'hématopoïèse (formation des cellules sanguines), la production des cellules immunitaires et la recombinaison de l'ADN V(D)J (V pour variable, D pour diversité et J pour jonction) lors de la maturation des lymphocytes B et T. Ainsi, l'hétérozygotie conférée aux femmes pour ces gènes et la production d'une plus grande diversité d'anticorps induisent chez les femmes une réponse du système immunitaire meilleure que celle des hommes, ce qui permet de mieux lutter contre les agents pathogènes dans le cas d'infections virales ou bactériennes, ou lors d'un développement tumoral.

Les maladies auto-immunes

Si les femmes ont plus de facilité à combattre certaines infections, elles présentent un plus fort taux d'incidence des maladies auto-immunes, incluant la sclérose en plaques, l'arthrite rhumatoïde et le lupus systémique érythémateux. Ces anomalies du système immunitaire conduisent à une autodestruction des cellules saines de l'organisme (Credendino *et al.*, 2020). Cette tendance est en partie expliquée par le grand nombre de gènes impliqués dans la réponse immunitaire qui échappent à l'inactivation, comme par exemple *TLR7* gène de la famille des récepteurs de type Toll-like, dans le cas du lupus.

Le lupus touche plus d'une personne sur 1 000. Son incidence est 9 fois supérieure chez la femme adulte que chez l'homme. Il s'agit d'une maladie chronique auto-immune qui se traduit par des symptômes cutanés et/ou articulaires et peut dégénérer en dommages permanents du rein et du cœur. La maladie se caractérise par la présence d'anticorps antinucléaires qui se retournent contre l'ADN des patients.

Plusieurs études ont mis le doigt sur un lien entre nombre de chromosomes X et lupus chez la souris et l'humain. Les hommes Klinefelter [47 chromosomes, XXY] ont une prévalence au lupus augmentée par rapport aux femmes Turner [45, X0]. Il a même été montré que les femmes [47, XXX] ont une incidence au lupus trois fois supérieure à celle des femmes [46, XX]. Ces résultats suggèrent un rôle prédominant des gènes du chromosome X qui échappent à l'inactivation. Parmi ceux-ci, *TLR7*, échappe à l'inactivation dans certaines cellules immunitaires, expliquant la plus grande prévalence du lupus chez les femmes. Ainsi, chez les patientes atteintes de lupus le déséquilibre de dosage de *TLR7* est particulièrement important (Brown *et al.*, 2022).

Le cancer

La présence de deux chromosomes X et leur inactivation aléatoire chez la femme la rendent moins vulnérable que l'homme à certains cancers (Credendino *et al.*, 2022). Cependant, certaines dérégulations du processus de l'ICX pourraient à leur tour conférer différentes vulnérabilités face au cancer et réciproquement. La délétion du

suppresseur de tumeur p53 chez la souris entraîne une perte d'expression des gènes du Xic, dont *Xist*, et une réexpression aberrante des gènes le long du chromosome X inactif. Chez la femme, dans le cas de tumeurs provoquées par une mutation de p53, il peut y avoir une ICX déficiente, qui augmente alors la pathogénicité cancéreuse. Lors de cancers du sein et de l'ovaire, plus les cellules cancéreuses réactivent leur chromosome X inactif, plus le pronostic est mauvais pour la patiente, dû à une perte du dosage génique. Dans certains cas, une perte physique du chromosome X inactif peut avoir lieu. Cela conduit à la perte de la double expression de certains gènes « protecteurs » qui échappent à l'inactivation, tels que le suppresseur de tumeur *PR70*, codant pour une protéine de la sous-unité B de la phosphatase 2. Enfin, des études chez des souris femelles adultes dont le gène *XIST* est délété ont mis en évidence une tendance au développement accru de cancers hématopoïétiques, suggérant un nouveau rôle pour *XIST* comme suppresseur de tumeur.

► Conclusion et perspectives

La découverte de l'inactivation du chromosome X chez la femelle a fêté ses 60 ans en 2021. Les chercheurs continuent d'affiner les connaissances sur ce processus biologique crucial et emblématique qui affecte toutes les femelles mammifères. Dans ce chapitre, nous avons abordé la notion de compensation de dose entre les sexes et entre autosomes et chromosomes X. Nous avons retracé la découverte de l'ICX, et montré les avancées technologiques qui ont aidé à comprendre ce processus. L'acteur majeur de la mise en place de l'inactivation est le long ARN non codant *XIST*, dont l'accumulation joue un rôle de plateforme moléculaire. Le cycle d'inactivation du chromosome X chez la souris a été finement décrit. Cependant, la souris reste une exception, car les deux vagues d'inactivation n'existent pas chez les autres mammifères. Les données disponibles chez les autres espèces montrent la diversité des stratégies de mise en place de l'inactivation au cours de l'évolution. Ceci interroge le biologiste à plusieurs titres. Pourquoi la souris a-t-elle mis en place une première vague d'inactivation soumise à l'empreinte parentale? Ces différences de stratégies sont-elles liées à l'évolution chromosomique de la région Xic? Aux variations de séquences de *XIST*? À l'évolution de transcrits tels que *TSIX* faiblement conservé entre espèces? À l'apparition d'acteurs nouveaux chez certaines espèces, comme le transcrit non codant *XACT* (*X Active specific transcript*) présent chez l'homme uniquement et qui s'accumule sur le chromosome X actif?

Les études ultérieures devront approfondir nos connaissances sur ce locus dans différentes espèces pour répondre à ces questions. Dans tous les cas, l'exemple de l'inactivation du chromosome X met en lumière la richesse de la biologie comparée pour comprendre les mécanismes épigénétiques du développement.

Il est à noter que lorsque nous nous éloignons du clade des euthériens, des mécanismes encore plus différents se sont mis en place pour la même finalité : inactiver un chromosome X. Pourquoi les marsupiaux subissent-ils une inactivation du chromosome X soumise à l'empreinte dans tous les tissus et tout au long de la vie de l'individu? Un autre long ARN non codant, *RSX* (*RNA-on-the-silent X*), y régule l'inactivation du chromosome X de façon très similaire à *XIST*.

Par ailleurs, nous ne savons pas exactement quel est l'impact potentiel des gènes qui échappent à l'inactivation dans les divergences entre les sexes. Certains gènes qui échappent, jouent-ils également un rôle dans le dimorphisme entre les femelles et

les mâles? Quel est le rôle de la réactivation du chromosome X dans la lignée germinale? Quelle est son implication dans la fertilité de la femelle? Pourquoi, durant l'embryogenèse, une réactivation du chromosome X inactif peut-elle être physiologique dans la lignée germinale, alors que ces réactivations aberrantes durant la vie adulte sont pathologiques dans les cancers et les maladies auto-immunes? Enfin, jouer sur la réactivation du chromosome inactif sain (non porteur de mutation) dans le cadre d'une pathologie liée au chromosome X chez la femme pourrait-il constituer une stratégie thérapeutique? Des laboratoires se penchent déjà sur la question dans le cas du syndrome de Rett (syndrome neurologique grave) qui touche les petites filles porteuses d'un allèle muté du gène *MECP2* (*methyl CpG binding protein 2*), porté par le chromosome X. Ils étudient la possibilité de réactiver la copie non mutée lorsqu'elle est éteinte, dans des modèles murins, avec des résultats prometteurs. À terme, cela apportera des éléments clés sur les causes de diverses pathologies et pourrait permettre le développement de nouvelles stratégies thérapeutiques.

L'étude de l'inactivation du chromosome X est un domaine en perpétuelle progression. Une meilleure compréhension de l'évolution de ce mécanisme épigénétique et des acteurs impliqués permettra un approfondissement de nos connaissances sur la régulation fine de l'expression des gènes, et le maintien de l'état silencieux hétérochromatinien. Enfin, quelles relations lient inactivation du chromosome X, pluripotence et différenciation des premiers lignages? Les études en cellules ES murines et la cinétique d'inactivation dans l'embryon de cette espèce semblent lier très étroitement ces trois aspects très importants pour le développement. Ainsi, l'étude de l'ICX et de ses conséquences montre à quel point les processus épigénétiques sont cruciaux pour le développement normal d'un individu.

Remerciements

Nous remercions Thierry Forné pour la relecture de notre chapitre. Le travail de l'équipe de Maud Borensztein est soutenu par la FRM AJE202005011598, l'Agence nationale de la recherche au titre du programme d'Investissements d'avenir portant la référence ANR-16IDEX-0006. Nous remercions la Ligue nationale contre le cancer pour le financement de thèse de Clara Roidor.

► Références bibliographiques

- Barr M.L., Bertram E., 1949. A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. *Nature*, 4148, 676-677.
- Borensztein M. *et al.*, 2017. Xist-dependent imprinted X inactivation and the early developmental consequences of its failure. *Nat Struct Mol Biol*, 24, 226-233.
- Brown C.J. *et al.*, 1991. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*, 349, 38-44.
- Brown G.J. *et al.*, 2022. TLR7 gain-of-function genetic variation causes human lupus. *Nature*, 605, 349-356.
- Credendino S.C., Neumayer C., Cantone I., 2020. Genetics and epigenetics of sex bias: Insights from human cancer and autoimmunity. *Trends in Genetics*, 36, 650-663.
- Heard E., Turner J., 2011. Function of the sex chromosomes in Mammalian fertility. *Cold Spring Harb Perspect Biol*, 1-18.
- Jarraud-Darnault A., Abitbol M., 2017. Le mosaïcisme chez les carnivores domestiques. *Bulletin de l'Académie vétérinaire de France*, 170, 70-76.

- Lee J., Davidow L.S., Warshawsky D., 1999. Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat Genet*, 21, 400-404.
- Lyon M., 1961. Gene action in the X-chromosome of the Mouse (*Mus musculus* L.). *Nature*, 190, 373-374.
- Morey C., Avner P., 2011. The demoiselle of X-inactivation: 50 years old and as trendy and mesmerising as ever. *PLoS genetics*, 7, e1002212.
- Nora E.P. *et al.*, 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485, 381-385.
- Okamoto I. *et al.*, 2011. Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature*, 472, 370-374.
- Okamoto I. *et al.*, 2021. The X chromosome dosage compensation program during the development of cynomolgus monkeys. *Science*, 374, eabd8887.
- Petropoulos S., Edsgård D., Reinius B., Deng Q., Panula S.P., Codeluppi S., *et al.*, 2016. Single Cell RNA-Seq reveals lineage and X chromosome dynamics in Human preimplantation embryos. *Cell*, 165 (4), 1012-1026.
- Ramos-Ibeas P. *et al.*, 2019. Pluripotency and X chromosome dynamics revealed in pig pre-gastrulating embryos by single cell analysis. *Nat Commun*, 10, 500.
- Rastan S., 1983. Non-random X-chromosome inactivation in mouse X-autosome translocation embryos-location of the inactivation centre. *J Embryol Exp Morphol.*, 78, 1-22.

Chapitre 4

Empreinte génomique parentale : découverte et mécanismes de régulation

THIERRY FORNÉ

Du point de vue classique de la génétique mendélienne, l'héritage de copies parentales, ou allèles, génétiquement identiques d'un gène autosomique dans une cellule diploïde est considéré comme équivalent. Cependant, la découverte de l'empreinte génomique parentale chez les mammifères, au milieu des années 1980, a radicalement changé cette conception.

En effet, les expériences des équipes de Davor Solter et d'Azim Surani ont montré que les deux génomes parentaux ne contribuent pas de manière égale au développement de l'embryon (McGrath et Solter, 1984; Surani *et al.*, 1984). Les embryons de souris issus d'œufs contenant deux pro-noyaux femelles (gynogénotes) ou deux pro-noyaux mâles (androgénotes) présentent des anomalies de développement et meurent avant la naissance. Les gynogénotes présentent des anomalies placentaires, tandis que les androgénotes sont caractérisés par une surcroissance des tissus extra-embryonnaires. Ces phénotypes sont très similaires à ceux observés chez l'homme en cas de fécondation d'un ovocyte énucléé par un ou deux spermatozoïdes, ou d'activation spontanée d'un ovocyte après duplication du génome maternel. Ainsi, les génomes paternel et maternel ont des rôles complémentaires et sont tous deux nécessaires au développement normal de l'embryon.

La notion d'empreinte génomique avait également été suggérée en 1985 par l'équipe de Bruce Cattanach grâce à des études sur des embryons de souris ayant hérité de deux copies identiques d'une partie d'un chromosome d'un de leurs parents : la disomie uniparentale. Ces embryons présentaient d'importants défauts de développement, et il est intéressant de noter que ces disomies avaient des effets opposés sur le développement embryonnaire. Initialement, dix régions distinctes sujettes à l'empreinte et réparties sur six chromosomes avaient été identifiées dans le génome de la souris, certaines d'entre elles présentant des phénotypes similaires à ceux observés dans les disomies uniparentales humaines.

Ces observations soutenaient l'idée que certains gènes de mammifères pouvaient se comporter différemment selon l'origine parentale de l'allèle. Cette hypothèse a été démontrée quelques années plus tard avec, dès 1991, l'identification des premiers gènes dont l'expression dépend de l'origine parentale des allèles : le gène *H19* qui produit un

ARN non traduit, *IGF2* (*insulin-like growth factor 2*) qui produit un facteur de croissance embryonnaire, et *IGF2R* (*insulin-like growth factor 2 receptor*) qui code pour l'un des récepteurs d'*IGF2*. Ces gènes sont dits « soumis à l'empreinte génomique parentale », car ils sont réprimés de façon stable sur l'un des deux allèles selon son origine parentale, et ils sont donc exprimés de manière monoallélique. Plus d'une centaine de gènes soumis à l'empreinte (GSE) ont aujourd'hui été décrits chez les mammifères. L'empreinte génomique a progressivement acquis une place importante dans le domaine de la biologie du développement, car elle s'est avérée essentielle au développement embryonnaire normal (Tucci *et al.*, 2019) et est altérée dans de nombreuses pathologies humaines (Monk *et al.*, 2019).

» Évolution et empreinte génomique parentale

Le terme d'empreinte a été utilisé dès les années 1960 pour décrire un phénomène observé chez les insectes qui conduit à l'inactivation du génome paternel. Cependant, le phénomène d'empreinte génomique parentale (l'expression monoallélique d'un gène résultant d'une répression stable d'un allèle en fonction de son origine parentale) n'a été observé que chez les mammifères et les plantes. La plupart des recherches sur les GSE dans le règne animal ont été réalisées chez les mammifères placentaires. Pour mieux comprendre l'évolution de l'empreinte, des études ont aussi été menées chez les marsupiaux, les monotrèmes et les oiseaux. Ces analyses ont révélé que l'empreinte génomique est conservée chez les marsupiaux, mais pas chez les monotrèmes ni chez les oiseaux, ce qui indique qu'elle est apparue il y a un peu plus de 150 millions d'années, après la divergence des monotrèmes et des mammifères thériens (placentaires et marsupiaux) (figure 4.1).

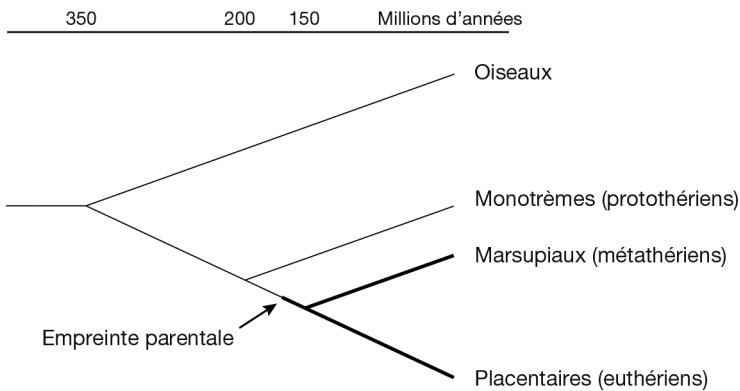


Figure 4.1. Empreinte génomique parentale et évolution chez les mammifères thériens.

L'empreinte génomique est apparue il y a environ 150-200 millions d'années dans la branche menant aux mammifères thériens (placentaires et marsupiaux).

En réprimant un seul allèle de plus d'une centaine de gènes, l'empreinte génomique parentale semble être en contradiction avec l'avantage évolutif que constitue la diploïdie. Par conséquent, comprendre la signification fonctionnelle de l'empreinte est rapidement devenu un défi fondamental pour la biologie du développement chez les mammifères.

Une des clés pour mieux comprendre l'émergence de l'empreinte génomique est de souligner que la plupart des GSE sont impliqués dans le contrôle du développement embryonnaire ou des tissus extra-embryonnaires. En effet, chez les mammifères, plusieurs GSE codent pour des facteurs impliqués dans la voie de l'insuline. D'autres codent pour des facteurs régulant le développement placentaire ou sont impliqués dans le développement cérébral et le comportement de la mère ou du nouveau-né. Il est frappant de constater que les gènes exprimés sur l'allèle paternel ont tendance à favoriser le développement placentaire et/ou la croissance foetale, alors que les gènes exprimés sur l'allèle maternel présentent généralement un effet opposé. Ceci est cohérent avec le phénotype des androgénotes, qui possèdent des annexes embryonnaires plus grandes que la normale, et celui des gynogénotes, dont les annexes présentent un retard de croissance. Ces observations, ainsi que la distribution phylogénique de l'empreinte génomique, soutiennent la théorie dite du « conflit parental » émise dès 1991 par Tom Moore et David Haig (1991). Selon cette théorie, l'empreinte génomique résulte d'un conflit entre les génomes paternels et maternels sur la quantité de nutriments allouée à l'embryon. L'intérêt du père, en matière de transmission de ses gènes, serait de favoriser la croissance des embryons afin de maximiser leur taux de survie, même au détriment des futures progénitures de la mère. En effet, ces derniers ont peu de chance de lui appartenir puisque les mammifères sont essentiellement polygames. À l'inverse, l'intérêt de la mère serait de limiter la croissance du fœtus afin d'économiser ses propres ressources et de garantir la possibilité de futures progénitures. Par conséquent, dans l'embryon, les gènes hérités paternellement sont sélectionnés pour stimuler l'extraction des ressources de la mère, tandis que les gènes transmis maternellement sont sélectionnés pour limiter cette extraction.

Ce modèle prédit que l'empreinte génomique ne devrait pas apparaître chez les espèces ovipares (monotrèmes et oiseaux), puisque la quantité de nutriments allouée à l'embryon est plus étroitement déterminée par la mère. Cette théorie implique également que les GSE ne sont pas nécessairement les mêmes dans chaque espèce et, à cet égard, il a par exemple été montré que les gènes *IGF2R* et *TSSC4* (*tumor suppressing subtransferable candidate 4*) sont soumis à l'empreinte chez la souris mais pas chez l'homme, alors que par exemple le gène *TRPM5* (*transient receptor potential cation channel, subfamily M, member 5*) est soumis à l'empreinte chez l'homme mais pas chez la souris.

En suivant cette logique, on peut argumenter que l'empreinte génomique ne présente aucun rôle intrinsèque dans le développement chez les mammifères, mais reflète seulement un « conflit parental » conduisant à une expression monoallélique. De ce point de vue, les phénotypes associés aux défauts d'empreinte chez les mammifères sont simplement dus à la perturbation du dosage de l'expression des GSE, conduisant à des niveaux aberrants des ARNm et des protéines correspondantes.

Plusieurs autres théories ont été proposées pour expliquer l'évolution de l'empreinte génomique parentale chez les mammifères. Parmi celles-ci, l'une des plus anciennes suggère que l'empreinte génomique a évolué pour empêcher la parthénogenèse, c'est-à-dire le développement d'un embryon chez la femelle sans fécondation par le mâle. Cependant, ces théories n'expliquent pas pourquoi l'empreinte n'a évolué que chez les mammifères thériens ou pourquoi elle affecte principalement les gènes impliqués dans le développement embryonnaire.

► Origine moléculaire des mécanismes de l'empreinte génomique

Chez les mammifères, la plupart des GSE sont regroupés dans de grands domaines chromosomiques englobant des centaines de paires de kilobases (kb). Ces domaines contiennent aussi bien des gènes exprimés au niveau des allèles paternels que maternels, et ils sont régulés par un centre d'empreinte commun. Ils peuvent également contenir des gènes qui échappent à l'empreinte, et qui sont donc exprimés de manière biallélique grâce à des mécanismes subtils de régulation insensibles à l'empreinte. Cette organisation implique que l'empreinte génomique a souvent évolué au niveau de grands domaines chromatiniens plutôt qu'au niveau de gènes individuels.

Il y a pratiquement autant de gènes exprimés sur l'allèle maternel que sur l'allèle paternel, et aucune séquence ou structure qui spécifierait toujours et uniquement l'empreinte n'a été identifiée. Toutefois, de nombreux GSE contiennent des éléments répétés. Ces répétitions ne partagent aucune homologie entre elles, mais peuvent être différentiellement méthylées dans les lignées germinales mâles et femelles, et sont fréquemment situées à proximité de régions de méthylation différentielle des allèles, ce qui avait très tôt suggéré qu'elles puissent être impliquées dans l'établissement d'empreintes de méthylation au niveau des loci soumis à l'empreinte. Ainsi cette observation avait-elle conduit Denise Barlow à émettre dès 1993 l'hypothèse que les mécanismes moléculaires de l'empreinte ont évolué à partir d'un mécanisme de défense de l'hôte contre les séquences répétées parasites (Barlow, 1993). Depuis cette époque, de nombreuses découvertes sont venues étayer cette hypothèse, démontrant en particulier que certains aspects de l'empreinte génomique sont liés à la répression de longues répétitions terminales (*long terminal repeats*, LTR) dérivées de rétrovirus endogènes (*endogenous retrovirus*, ERV) (Kaneko-Ishino et Ishino, 2022). Les ERV sont des éléments répétés des génomes des mammifères correspondant à d'anciennes insertions de rétrovirus et qui constituent jusqu'à environ 10% du génome chez la souris. Les ERV intacts contiennent les gènes nécessaires à la réplication virale et sont entourés, en amont et en aval, par des LTR. Cependant, l'immense majorité des ERV génomiques ont perdu leurs gènes viraux et n'existent plus que sous la forme de LTR isolées appelées « solo-LTR ». Les ERV intacts sont maintenus silencieux dans les génomes grâce à un complexe répresseur constitué de protéines de la famille KRAB (*Krüppel-associated box*) et de nombreuses protéines « à doigt de zinc » (*zinc finger protein*, ZFP), qui apposent des marques épigénétiques répressives, telle la triméthylation de la lysine 9 de l'histone H3 (H3K9me3) et la méthylation de l'ADN. Cependant, les « solo-LTR » échappent souvent à ce contrôle et peuvent parfois être cooptées pour la régulation de certains gènes, et en particulier ceux soumis à l'empreinte génomique. Ainsi, dans le cas du gène *RASGRF1* (*Ras protein specific Guanine nucleotide Releasing Factor 1*), il a été démontré qu'une LTR située en amont du gène est nécessaire pour l'établissement d'une méthylation paternelle de l'ADN durant la spermatogenèse et pour l'expression monoallélique du gène. De façon remarquable, ces répétitions sont absentes chez les espèces de rongeurs, chez qui ce gène n'est pas soumis à l'empreinte.

► Contrôle épigénétique des gènes soumis à l'empreinte parentale

Comment les deux allèles d'un GSE peuvent-ils être exprimés de manière différentielle dans le même contexte cellulaire, alors que leur séquence nucléotidique est identique ?

Comme nous l'avons évoqué pour le gène *RASGRF1*, la réponse à cette question est que l'empreinte génomique parentale implique des modifications épigénétiques, c'est-à-dire des modifications héréditaires qui affectent l'expression des gènes sans changement de la séquence d'ADN (voir chapitre 2). Ici, ces marques dépendent de l'origine parentale et elles sont apposées différenciellement sur les deux centres d'empreinte dans les lignées germinales produisant les gamètes (spermatozoïdes ou ovocytes), puis sont conservées pendant le développement embryonnaire. Elles permettent donc de maintenir une mémoire de l'origine parentale des allèles et sont qualifiées de marques « primaires » d'empreinte. Ces marques sont de deux types principaux : la méthylation de l'ADN dans le cas de l'empreinte dite « canonique », et les modifications post-traductionnelles des histones dans le cas de l'empreinte dite « non canonique ».

Méthylation de l'ADN et empreinte génomique canonique

La méthylation de l'ADN a été la première marque épigénétique à être identifiée comme jouant un rôle essentiel dans l'empreinte génomique chez les mammifères. Dès 1987, plusieurs études avaient montré que l'expression d'un transgène pouvait être régulée par l'établissement de profils de méthylation distincts dans les gamètes parentaux. Par la suite, la fonction de la méthylation de l'ADN a été renforcée par la découverte que les gènes *H19*, *IGF2* et *IGF2R* contiennent des régions qui sont méthylées de façon différentielle sur les allèles parentaux (*differentially methylated regions*, DMR). Finalement, des DMR résultant de différences de méthylation des allèles, établies soit au cours de la gamétogenèse (DMR gamétique, gDMR) soit après la fécondation (DMR secondaire, sDMR), ont été décrites dans tous les GSE.

Cependant, le principal argument en faveur d'un rôle de la méthylation de l'ADN dans l'empreinte génomique est venu de l'observation que l'expression monoallélique de nombreux GSE est perdue dans les embryons de souris *Dnmt1* *-/-* (délétion homozygote du gène *DNMT1*). De plus, la surexpression de DNMT1 induit également une perte d'empreinte de certains gènes en raison d'une méthylation bi-allélique de leurs séquences régulatrices. Enfin, l'utilisation de la 5-azacytidine, un inhibiteur de la méthylation de l'ADN, induit une perte d'empreinte de nombreux gènes aussi bien *ex vivo* dans des cellules en culture qu'*in vivo* chez la souris.

Ces travaux soulignent le rôle fondamental que joue la méthylation de l'ADN dans l'apposition et le maintien de l'empreinte génomique canonique chez les mammifères (SanMiguel et Bartolomei, 2018).

Modifications post-traductionnelles des histones et empreinte génomique non canonique

D'autres modifications épigénétiques sont associées à l'empreinte génomique chez les mammifères. Des expériences de sensibilité à la désoxyribonucléase (DNase) et des essais d'immunoprécipitation de la chromatine (ChIP) ont démontré que les GSE présentent des structures chromatiniennes et des profils d'acétylation des histones différents sur chaque allèle parental. En règle générale, l'allèle réprimé porteur d'une méthylation de l'ADN est caractérisé par une chromatine compacte et une désacétylation des histones, tandis que l'allèle exprimé, dont l'ADN reste généralement non méthylé, est caractérisé par une sensibilité aux nucléases et une acétylation des histones

(voir chapitre 2). Des profils alléliques de méthylation des histones ont également été décrits. Ainsi, la marque H3K9me3 est associée à l'allèle méthylé réprimé, alors que la méthylation de la lysine 4 de l'histone H3 (H3K4me3) est associée à l'allèle actif non méthylé. Au niveau du locus *SNRPN* (*small nuclear ribonucleoprotein polypeptide N*) les modifications chromatiniennes semblent représenter des marques d'empreinte secondaires indépendantes de la méthylation de l'ADN. Chez l'homme, le centre d'empreinte AS-SRO (*Angelman Syndrome-smallest region of deletion overlap*) maintient une structure chromatinienne allélique différentielle en l'absence de méthylation de l'ADN. De même, chez la souris, la marque H3K9me3 au niveau de la séquence maternelle du centre d'empreinte PWS-IC (*Prader-Willi syndrome-imprinting center*) est maintenue en l'absence de méthylation de l'ADN et semble être suffisante pour maintenir la répression du gène *SNRPN*.

Si l'importance des modifications post-traductionnelles des histones dans le maintien de l'empreinte dans les lignées germinales est reconnue de longue date, son rôle dans l'apposition des empreintes est longtemps resté controversé. La question fondamentale est en effet de savoir si les modifications des histones peuvent agir comme des marques primaires d'empreinte apposées et héritées à partir des gamètes, ou si elles sont établies plus tardivement, comme des marques secondaires, après la méthylation des CpG (cytosine liée à une guanine *via* un groupement phosphate), pour maintenir et renforcer l'état réprimé de la chromatine. Des exemples de loci soumis à l'empreinte qui ne semblent pas être contrôlés par la méthylation de l'ADN sont connus depuis longtemps, comme les gènes *GAB1* (*GRB2-associated-binding protein 1*) et *SFMBT2* (*Scm-like with four MBT domains 2*) dont l'expression monoallélique paternelle persiste dans le placenta, alors qu'aucune DMR gamétique n'est présente. L'existence de ces cas particuliers est aujourd'hui mieux comprise suite à la découverte récente d'un mécanisme d'empreinte génomique, l'empreinte non canonique, dont l'apposition est indépendante de la méthylation de l'ADN dans les gamètes (Hanna et Kelsey, 2021). Ce mécanisme, qui semble restreint aux seules lignées extra-embryonnaires et au placenta, dépend de la triméthylation de la lysine 27 de l'histone H3 (H3K27me3), qui est apposée sur les centres d'empreinte de certains loci dans l'ovocyte et qui est ensuite transmise à l'embryon indépendamment de toute marque de méthylation de l'ADN, constituant ainsi une véritable marque épigénétique primaire de l'empreinte génomique. Cependant, cette marque ne résiste pas à la reprogrammation épigénétique qui prend place dans l'embryon préimplantatoire, où elle est alors perdue, y compris dans les tissus extra-embryonnaires où pourtant l'expression monoallélique persiste. Le maintien de l'empreinte non canonique dépend en effet d'une méthylation de l'ADN au niveau de DMR secondaires spécifiques des tissus extra-embryonnaires. À ce jour, la seule exception connue est le gène *SFMBT2*, qui semble parvenir à retenir un enrichissement de la marque H3K27me3 sur l'allèle maternel et une expression monoallélique paternelle en l'absence de DMR secondaire. L'importance physiologique de l'empreinte non canonique a été bien soulignée chez la souris, où au moins sept gènes autosomiques et un gène lié au chromosome X sont soumis à ce phénomène. Il a été montré que l'empreinte non canonique du gène *XIST*, qui produit un long ARN non codant impliqué dans l'inactivation du chromosome X chez les femelles (voir chapitre 3), contribue au développement embryonnaire, tandis que l'empreinte non canonique des gènes autosomiques restreint la croissance placentaire.

► Centres d’empreinte et cycle de l’empreinte génomique parentale

Centres d’empreinte génomique et pathologies liées à l’empreinte génomique

Les marques épigénétiques primaires responsables de l’empreinte génomique parentale sont apposées sur une séquence de quelques kilobases, spécifique de chaque locus, appelée « centre d’empreinte » ou « région de contrôle de l’empreinte » (*imprinting-control region*, ICR).

L’existence d’un centre d’empreinte a été proposée pour la première fois au locus humain 15q11-13 après avoir observé que les microdélétions chez les patients atteints des syndromes d’Angelman et de Prader-Willi sont associées à une perturbation de l’empreinte de plusieurs gènes dispersés sur 2 Mb du locus. Ce centre d’empreinte est un élément régulateur bipartite (figure 4.2A). L’élément AS-SRO, situé 35 kb en amont du gène *SNRPN*, est supprimé chez les patients atteints du syndrome d’Angelman et régule l’établissement des empreintes maternelles dans les cellules germinales. L’élément PWS-SRO, qui comprend l’exon 1 du gène *SNRPN*, est altéré chez les patients atteints du syndrome de Prader-Willi. Cet élément est nécessaire pour contrôler l’établissement des empreintes paternelles et également pour maintenir l’empreinte sur le chromosome paternel pendant le développement embryonnaire.

Le centre d’empreinte du locus *IGF2R*, situé sur le chromosome 17 chez la souris, a également été l’un des premiers à avoir été identifié (figure 4.2B). Cet élément intro- nique, méthylé sur l’allèle maternel, a tout d’abord été décrit comme étant requis pour l’empreinte d’*IGF2R*, et il s’est avéré qu’il agit sur l’allèle paternel comme promoteur pour un ARN non traduit (*AIRN*) qui agit comme un répresseur bidirectionnel pour plusieurs gènes dispersés dans un domaine de 400 kb.

Enfin, deux centres d’empreinte identifiés dans la partie distale du chromosome 7 de la souris, orthologue du locus 11p15.5 humain, comprennent les gènes *CDKN1C*, *IGF2* et *H19*, dont les dérégulations de l’empreinte conduisent au syndrome de Beckwith-Wiedemann. La région KvDMR1, située dans l’intron 10 du gène *KCNQ1*, est spécifiquement méthylée sur l’allèle maternel, et produit sur l’allèle paternel un ARN non codant (*Kcnq1ot1*) qui induit la répression de plusieurs gènes du locus (figure 4.2C). L’empreinte des gènes *IGF2* et *H19* dépend d’un second centre d’empreinte situé 2 kb en amont du gène *H19*. Cette séquence a une activité d’isolateur (*insulator*) qui est régulée par la méthylation de l’ADN. Sur l’allèle maternel, la protéine CTCF (*CCCTC binding factor*) se lie au centre d’empreinte non méthylé et bloque l’accès des promoteurs d’*IGF2* aux activateurs transcriptionnels (*enhancers*) situés en aval d’*H19*. Sur l’allèle paternel, la liaison de CTCF est inhibée par la méthylation de l’ADN.

Les marques épigénétiques responsables de l’empreinte parentale subissent un cycle au cours de la vie de l’organisme qui permet leur reprogrammation à chaque génération. Les marques primaires d’empreinte sont héritées des gamètes parentaux, puis sont maintenues et « lues » dans les cellules somatiques de l’individu avant d’être effacées dans la lignée germinale, puis rétablies en fonction du sexe de l’individu pour la génération suivante (figure 4.3).

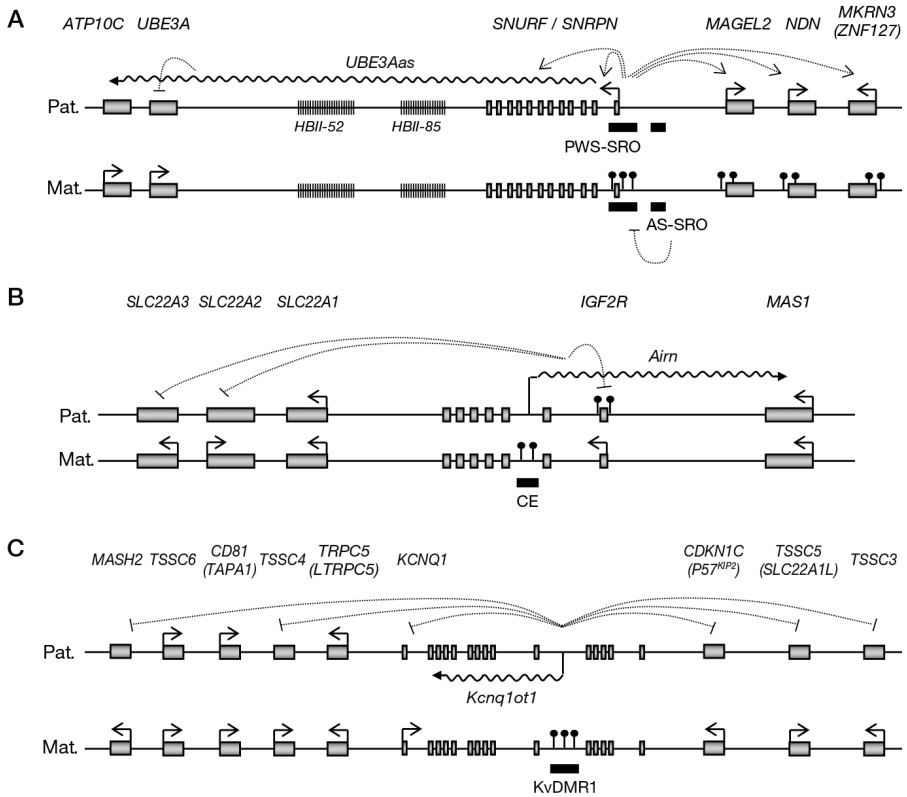


Figure 4.2. Centres d'empreinte et ARN non codants.

A. Le locus humain 15q11-13, impliqué dans les syndromes d'Angelman (AS) et de Prader-Willi (PWS). Le syndrome de Prader-Willi est associé à une déficience des gènes exprimés sur l'allèle paternel, tandis que le syndrome d'Angelman résulte d'une déficience des gènes exprimés sur l'allèle maternel. *UBE3A* est le principal gène impliqué dans ce syndrome. L'élément PWS-SRO, qui comprend le promoteur et l'exon 1 du gène *SNRPN*, contient une DMR dont la méthylation maternelle est héritée des ovocytes chez la souris, mais qui s'établit après la fécondation chez l'homme. L'élément PWS-SRO est nécessaire sur l'allèle paternel pour activer les gènes exprimés paternellement et pour réprimer *UBE3A*, alors que l'élément AS-SRO est nécessaire sur l'allèle maternel pour réprimer les gènes exprimés paternellement et activer *UBE3A*. L'élément AS-SRO ne présente pas de méthylation différentielle, mais présente une structure chromatinienne spécifique sur l'allèle maternel qui inactive l'élément PWS-SRO et empêche l'activation des gènes *in cis*. Le gène *UBE3A* est soumis à l'empreinte uniquement dans le cerveau. L'élément PWS-SRO augmente l'expression d'un ARN antisens (*UBE3Aas*) dans le cerveau qui réprime *UBE3A* sur l'allèle paternel. Cet ARN comprend des petits ARN nucléolaires (snoRNAs) spécifiques du cerveau qui sont répétés en de nombreuses copies (*HBII-85* et *52*).

B. Le locus *IGF2R/AIRN* sur le chromosome 17 de la souris. Le gène *IGF2R* exprimé par la mère contient deux DMR : une dans la région du promoteur, qui porte une méthylation paternelle établie après la fécondation, et une située dans l'intron 2 qui est méthylée dans les ovocytes et correspond au centre d'empreinte (CE) de ce locus. Plus en aval, les gènes *SLC22A2* et *SLC22A3* ne sont exprimés que sur l'allèle maternel dans le placenta. Le CE non méthylé sur l'allèle paternel correspond au promoteur actif du transcrit non codant *AIRN* (Antisense Igf2r RNA) de 108 kb de long. La délétion du CE, ou l'expression d'un ARN *AIRN* tronqué sur l'allèle paternel, lève la répression des gènes *IGF2R*, *SLC22A2* et *SLC22A3*, indiquant un rôle direct de l'ARN *AIRN* dans un mécanisme bidirectionnel de répression sur le chromosome paternel.

C. Le locus *KCNQT1* chez la souris. La région *KvDMR1* intronique est méthylée sur l'allèle maternel dans les cellules somatiques et germinales, et elle correspond au promoteur de l'ARN non codant *Kcnqt1ot1* transcrit sur l'allèle paternel. La délétion de *KvDMR1* sur l'allèle paternel induit la réactivation de nombreux gènes *in cis*, dont *TSSC3*, *TSSC5*, *CDKN1C*, *KCNQ1*, *TSSC4* et *MASH2*.

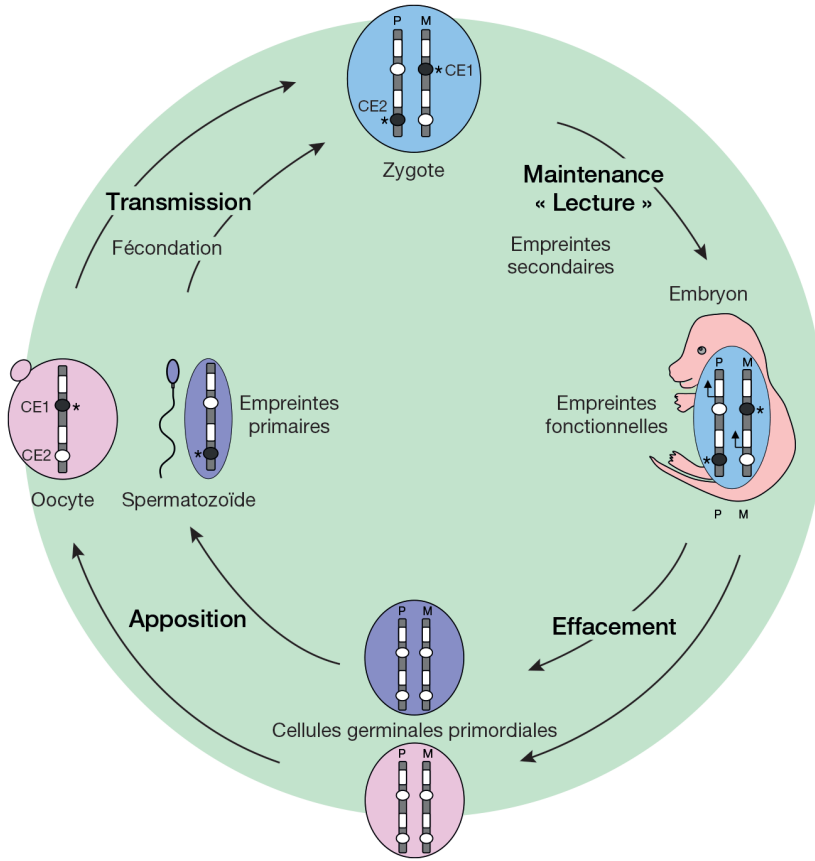


Figure 4.3. Cycle de l'empreinte génomique parentale.

L'empreinte génomique résulte de modifications épigénétiques apposées dans les lignées germinales mâles et femelles (empreintes primaires), maintenues et « lues » au cours des divisions cellulaires de l'embryon pour constituer des empreintes secondaires, puis des empreintes fonctionnelles qui inactivent de manière stable l'un des allèles parentaux, conduisant ainsi à une expression monoallélique. Le schéma montre l'exemple d'un gène réprimé paternellement et d'un gène réprimé maternellement suite à la transmission par l'ovocyte ou le spermatozoïde de marques épigénétiques (étoiles noires) apposées au niveau des centres d'empreinte (CE1 et CE2). Les empreintes parentales sont effacées au cours de la gamétogenèse précoce, puis rétablies selon le sexe de l'individu pour être transmises à la génération suivante. M : chromosome maternel ; P : chromosome paternel.

Effacement

Les empreintes épigénétiques issues des parents sont effacées dans les cellules germinales primordiales de l'embryon avant d'être reprogrammées en fonction du type de gamète. Le génome des cellules germinales primordiales subit une déméthylation globale conduisant à l'effacement de la méthylation de l'ADN au niveau des loci soumis à l'empreinte entre le 10^e et le 13^e jour du développement embryonnaire chez la souris. Les expériences de transplantation nucléaire ont confirmé que les empreintes de méthylation des DMR sont complètement effacées à ce stade. Cependant, chaque centre d'empreinte possède un rythme de déméthylation qui lui est propre : l'effacement de

la méthylation de *NNAT* (*neuronatin*) est le plus précoce, celui d'*H19* est « intermédiaire », tandis que celui qui résiste le plus longtemps à la déméthylation est le *PEG10* (*paternally expressed gene 10*).

Des expériences d'analyse des niveaux de méthylation de l'ADN réalisées grâce à des souris déficientes pour les protéines de déméthylation *ten-eleven translocation* TET1 et TET2 ont montré que TET1 est un médiateur essentiel à l'effacement des empreintes de méthylation dans les cellules germinales primordiales.

Il est intéressant de noter que l'allèle hérité du parent du même sexe est reprogrammé dans les cellules germinales de l'embryon plutôt que transmis sans modifications à la génération suivante.

Apposition

Les marques d'empreinte primaires, parfois aussi dites « germinales » ou « gamétiques », sont donc des modifications épigénétiques qui sont apposées dans les gamètes et qui conservent la mémoire épigénétique de l'origine parentale des chromosomes dans les cellules somatiques tout au long du développement. L'apposition de l'empreinte est définie comme l'acquisition de ces marques d'empreinte primaires dans les lignées germinales entre l'étape d'effacement et la fécondation (figure 4.3).

La méthylation de l'ADN constitue une telle marque d'empreinte primaire pour la plupart des GSE identifiés jusqu'à présent. En effet, de nombreux GSE acquièrent des profils de méthylation différentiels dans les gamètes mâles et femelles (appelés « empreintes primaires de méthylation ») qui sont ensuite maintenus dans l'embryon. L'établissement des empreintes primaires de méthylation se produit pendant la différenciation tardive des gamètes. Dans les cellules de la lignée germinale mâle, l'apposition de la méthylation de l'ADN sur les centres d'empreinte a lieu dès que la déméthylation cesse, à partir du jour 13,5 du développement chez la souris, et elle est totalement terminée à la naissance. En revanche, dans les cellules de la lignée germinale femelle, l'établissement des empreintes de méthylation sur les centres d'empreinte s'effectue essentiellement après la naissance, pendant la phase de croissance des ovocytes. Il existe un biais dans la contribution parentale à ces empreintes primaires de méthylation, puisque la grande majorité d'entre elles sont acquises dans la lignée germinale femelle (ICR maternelles), alors que seulement trois sont connues pour être héritées paternellement (ICR paternelles).

Contrairement aux autres gènes autosomiques, le passage dans la lignée germinale est nécessaire pour établir les profils de méthylation des loci soumis à l'empreinte, ce qui implique qu'ils sont ciblés par des activités de méthylation *de novo* spécifiques aux cellules germinales. Une percée majeure a été réalisée grâce à l'identification du facteur *DNMT3L* (*DNMT3-like*), une protéine appartenant à la famille DNMT3. La *DNMT3L* a été le premier facteur connu pour être impliqué dans l'apposition d'empreintes primaires au cours de la gamétogenèse. *DNMT3L* est exprimé dans les lignées germinales mâles et femelles précisément au moment où les empreintes parentales sont établies. La délétion homozygote de *DNMT3L* chez la souris inhibe la différenciation des spermatozoïdes chez le mâle, tandis que les embryons issus de femelles *DNMT3L* *-/-* présentent une dérégulation de nombreux GSE conduisant à une mort prématurée. En effet, la délétion de *DNMT3L* induit une absence de

méthylation au niveau des séquences soumises à l’empreinte qui sont habituellement méthylées dans les ovocytes sans affecter les profils de méthylation dans le reste du génome. Par conséquent, le gène *DNMT3L* est spécifiquement requis pour établir les empreintes de méthylation maternelles pendant l’ovogenèse. De manière surprenante, la protéine DNMT3L ne possède pas le domaine catalytique caractéristique des ADN méthyltransférases, mais elle se lie à la lysine 4 de l’histone H3 (H3K4) non méthylée et elle interagit avec les deux ADN méthyltransférases de la même famille, DNMT3A et DNMT3B, agissant comme un cofacteur stimulant leur activité. En revanche, l’établissement d’empreintes de méthylation ne nécessite pas DNMT10 (l’isoforme de DNMT1 exprimée pendant l’ovogenèse).

Hormis le cas du gène murin *RASGRF1*, pour lequel nous avons vu qu’une LTR joue un rôle essentiel dans l’apposition d’une méthylation de l’ADN dans la lignée germinale mâle, on ne sait pas précisément comment les cellules germinales mâles ou femelles peuvent établir des profils de méthylation différents, ni comment les centres d’empreinte sont ciblés pour l’apposition de ces empreintes. Cela pourrait résulter de l’existence d’une organisation différente de la chromatine dans les cellules germinales mâles et femelles ou de l’expression de facteurs d’apposition des empreintes spécifiques de chacune des lignées germinales.

Maintenance

La méthyltransférase DNMT1 est nécessaire à la maintenance des patrons de méthylation alléliques des DMR dans les domaines soumis à l’empreinte durant le développement embryonnaire. Cependant, immédiatement après la fécondation, une vague de déméthylation globale affecte le génome. Le génome paternel est activement déméthylé dans le zygote, alors que le génome maternel reste protégé de cette déméthylation active, mais subit une déméthylation passive au cours des divisions cellulaires en raison de l’absence de la méthyltransférase de maintenance jusqu’au stade du blastocyste. De façon remarquable, les séquences portant des empreintes primaires de méthylation ont la propriété unique de résister à cette vague de déméthylation et de maintenir leurs patrons de méthylation gamétique tout au long du développement. Cette propriété est cruciale pour maintenir des marques distinctes entre les génomes parentaux, car les empreintes de méthylation ne pourraient pas être rétablies dans les cellules somatiques de l’embryon. Les mécanismes par lesquels les domaines soumis à l’empreinte résistent à la déméthylation après la fécondation sont encore mal compris. Il a toutefois été montré que la méthylation de certains centres d’empreinte dépend de protéines protectrices telles que DPPA3 (*developmental pluripotency-associated 3*) pour *H19* et ZFP57 pour *IGF2R*.

Le maintien des empreintes de méthylation dans l’embryon implique également que les séquences non méthylées puissent résister à la méthylation *de novo*. Cette propriété est illustrée par le fait que les régions soumises à l’empreinte peuvent être méthylées *de novo* dans les cellules germinales, mais pas dans les cellules somatiques de l’embryon. Les bases moléculaires de cette résistance restent encore méconnues, mais elles impliquent probablement des facteurs qui protègent les séquences non méthylées de la machinerie de méthylation. Tel est le cas de la protéine TET3, dont le rôle dans le zygote ne semble pas être lié à la déméthylation, mais servirait plutôt à protéger certaines régions non méthylées du génome de l’acquisition inappropriée de méthylation de l’ADN.

Lecture de l'empreinte et expression monoallélique

Les marques primaires d'empreinte apposées dans les gamètes conduisent à une répression stable, c'est-à-dire héritable, au cours des divisions cellulaires, d'un des allèles parentaux des GSE dans l'embryon, conduisant à une expression monoallélique de ces gènes. Cette expression monoallélique est la conséquence de cascades distinctes d'événements épigénétiques se produisant sur chacun des allèles parentaux au cours du développement. L'allèle actif subit toutes les réorganisations chromatinienne nécessaires à l'expression du gène. En revanche, l'allèle soumis à l'empreinte acquiert une « empreinte fonctionnelle » qui compromet la programmation épigénétique de l'expression génique sur cet allèle et maintient un état silencieux stable. L'acquisition d'empreintes fonctionnelles, qui s'apparente à une « lecture de l'empreinte » (figure 4.3), implique l'établissement de « marques secondaires d'empreinte », c'est-à-dire de modifications épigénétiques établies en *cis* après la fécondation à partir des marques d'empreinte primaires. Ces marques secondaires comprennent les DMR secondaires qui sont parfois spécifiques du type de cellule, et cette hiérarchie dans l'établissement des DMR est cruciale pour les mécanismes d'empreinte génomique.

Sur l'allèle qui n'est pas soumis à l'empreinte, la programmation épigénétique conduit à une régulation spatio-temporelle appropriée de l'expression génique au cours du développement. Cette programmation épigénétique consiste en une combinaison de réorganisations à la fois locales et d'ordre supérieur dans la chromatine, permettant l'activation et la modulation de la transcription au cours du développement. Elle nécessite des éléments régulateurs spécifiques. Bien que n'étant pas directement impliqués dans le processus d'empreinte proprement dit, la compréhension du rôle de ces éléments régulateurs est cruciale puisque, sur l'allèle réprimé, des empreintes fonctionnelles sont constituées pour bloquer leur activité. Un exemple est fourni par l'élément régulateur DMR2, dans le gène *IGF2*, spécifique de l'endoderme chez la souris. La région DMR2 n'est pas nécessaire pour maintenir l'empreinte d'*IGF2* sur l'allèle maternel, mais une DMR2 méthylée permet l'interaction du gène avec les activateurs transcriptionnels, et elle est donc nécessaire pour avoir des niveaux d'expression élevés d'*IGF2* à partir de l'allèle paternel. Sur l'allèle réprimé, l'empreinte fonctionnelle maintient la DMR2 non méthylée et empêche ainsi son interaction avec les activateurs transcriptionnels, contribuant à empêcher l'expression d'*IGF2* à partir de l'allèle maternel. Cet exemple révèle comment certaines marques d'empreinte secondaires peuvent refléter l'existence d'empreintes fonctionnelles sans être nécessairement impliquées dans le processus d'empreinte, c'est-à-dire la répression de l'allèle soumis à l'empreinte.

» Exemples de mécanismes moléculaires des empreintes fonctionnelles

Il n'existe pas de modèle unique de répression monoallélique qui serait commun à tous les GSE connus. En effet, comme la plupart des GSE sont regroupés et partagent des éléments régulateurs qui peuvent être spécifiques d'un tissu, l'acquisition d'empreintes fonctionnelles est soumise à diverses contraintes spatio-temporelles au cours du développement. Ainsi, les mécanismes d'expression monoalléliques peuvent différer non seulement d'un GSE à l'autre, mais aussi d'un type cellulaire à l'autre.

Les empreintes fonctionnelles dans l'embryon résultent donc de multiples mécanismes qui impliquent les centres d'empreinte, des DMR secondaires et la liaison allèle-spécifique de certains facteurs protéiques ou d'ARN non codants. L'intégration de tous ces facteurs génomiques et épigénétiques conduit à l'acquisition d'une architecture tridimensionnelle (3D) de la chromatine différente sur les deux allèles parentaux qui se traduit par une expression monoallélique finement régulée selon le tissu et au cours du développement.

Méthylation du promoteur

Un mécanisme courant qui donne lieu à une répression monoallélique stable est la méthylation allélique différentielle des promoteurs (figure 4.4A). Cette méthylation spécifique de l'allèle peut être soit apposée directement dans les gamètes, soit, comme c'est le cas pour la plupart des GSE, être acquise secondairement dans l'embryon à partir d'un centre d'empreinte en *cis*. C'est le cas du gène *H19*, où le centre d'empreinte situé 2 kb en amont induit la méthylation du promoteur sur l'allèle paternel pendant le développement précoce.

La méthylation de l'ADN peut aussi conduire à une répression transcriptionnelle en empêchant la liaison de facteurs de transcription, ou en favorisant l'accès aux protéines possédant un domaine qui se lie à l'ADN méthylé (*methyl-CpG-binding domain*, MBD). Ces facteurs induisent en effet une compaction locale de la chromatine en recrutant des complexes d'histones désacétylases ou d'histones méthyltransférases.

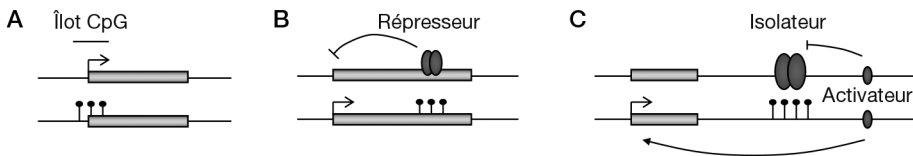


Figure 4.4. Mécanismes de régulation de l'expression des gènes soumis à l'empreinte (GSE).

A. Le promoteur du gène est réprimé par la méthylation de l'ADN (tiges-boucles noires). B. La liaison d'un répresseur agissant en *cis* est régulée par la méthylation de l'ADN. C. Une activité d'isolateur, qui bloque l'accès du promoteur aux activateurs transcriptionnels, est régulée par la méthylation de l'ADN.

Modulation des séquences régulatrices

L'observation que certains GSE abritent des séquences méthylées sur l'allèle actif a conduit à l'hypothèse que ces régions ont des activités de répression qui sont inhibées par la méthylation de l'ADN (figure 4.4B). Ce modèle a été par exemple démontré dans le cas du gène *IGF2*, où la région DMR1, une DMR intragénique portant une marque d'empreinte secondaire, agit comme un répresseur sur l'allèle maternel et est inactivée par la méthylation de l'ADN sur l'allèle paternel dans les tissus mésodermiques. Un modèle similaire a été proposé pour le locus *DLK1/MEG3* (*delta like non-canonical Notch ligand 1/maternally Expressed gene 3*), où la répression de l'allèle maternel *DLK1* dépend d'une DMR germinale intergénique non méthylée (Ig-DMR).

Les empreintes fonctionnelles peuvent également impliquer des mécanismes indirects, comme la modulation de séquences activant ou réprimant la transcription qui sont situées loin du gène. Par exemple, comme nous l'avons vu, l'empreinte fonctionnelle du gène *IGF2* implique la protéine CTCF qui se lie au centre d'empreinte non méthylé,

où elle joue un rôle d'isolateur empêchant l'accès des promoteurs aux activateurs transcriptionnels (figure 4.4C).

Transcrits non codants

De nombreux loci soumis à l'empreinte contiennent des gènes produisant des ARN non codants. Ces ARN jouent un rôle important dans l'apposition et/ou le maintien de l'empreinte génomique, mais leurs mécanismes d'action restent encore largement à clarifier (Llères *et al.*, 2021). Il s'agit d'ARN très longs et souvent exprimés de manière paternelle à partir de centres d'empreinte non méthylés. L'empreinte fonctionnelle établie grâce à ces ARN non codants dépend du recrutement de complexes protéiques répresseurs du groupe Polycomb grâce à des interactions directes entre certaines protéines de ces complexes et les ARN non codants. Les exemples les mieux caractérisés sont ceux des loci *IGF2R/AIRN*, *UBE3A/Ube3Aas* et *KCNQT1/Kcnq1ot1* (figures 4.2A-C).

La première démonstration de l'implication directe d'un ARN non codant dans la régulation de l'empreinte génomique a été celui de l'ARN *AIRN*. Le promoteur *AIRN* est situé dans le centre d'empreinte intronique identifié dans le gène *IGF2R*, et il a été démontré que le transcrit *AIRN* est nécessaire pour la répression paternelle de plusieurs gènes du locus. Une observation déroutante est que, dans les tissus extra-embryonnaires, l'ARN *AIRN* contrôle non seulement l'empreinte de gènes proches situés dans le même locus, mais aussi celle de plusieurs autres gènes situés à plusieurs mégabases de distance. Comme le fait le long ARN non codant *Xist* sur le chromosome X (voir chapitre 3), on peut imaginer que l'ARN *AIRN* initie tout d'abord une répression locale au niveau du gène *IGF2R* qui se propage ensuite aux gènes voisins, puis à une partie du chromosome. Il est possible que le transcrit *Kcnq1ot1* joue un rôle similaire dans la répression de l'allèle paternel au locus *Kcnqt1*.

Architecture de la chromatine

Au cours des vingt dernières années, il est peu à peu devenu évident que la plupart des GSE ne sont pas régulés individuellement, mais de façon coordonnée au niveau de domaines chromatinien structurés. Les centres d'empreinte (CE) régulent généralement l'empreinte de plusieurs gènes sur de longues distances, et l'expression monoallélique qu'ils contrôlent dépend donc de la formation d'empreintes fonctionnelles qui induisent une organisation tridimensionnelle (3D) de la chromatine différente sur les deux allèles parentaux.

L'existence d'une organisation chromatinienne différentielle au niveau des domaines chromosomiques soumis à l'empreinte avait été suggérée dès le milieu des années 1990 par l'observation que les loci humains *PWS/AS* et *IGF2/H19* présentent des fréquences de recombinaison méiotique différentes dans les lignées germinales mâle et femelle. Le développement des méthodes de capture de conformation de chromosome (3C) a ensuite permis de démontrer que les domaines soumis à l'empreinte possèdent effectivement une organisation tridimensionnelle de la chromatine différente selon l'origine parentale du chromosome. Tel est, par exemple, le cas du locus des gènes *IGF2* et *H19*, où les études de 3C réalisées dès la fin des années 2000 ont confirmé les modèles qui avaient été proposés auparavant grâce aux délétions génomiques réalisées chez la souris. Sur le chromosome hérité du père, la méthylation du centre de contrôle de

l’empreinte, localisé en amont du gène *H19*, empêche la liaison de la protéine CTCF, un facteur impliqué dans la fonction d’isolateur (*insulator*) sur la chromatine chez les vertébrés. Son absence permet aux séquences activatrices de la transcription (*enhancers*) situées en aval du gène *H19* de former une boucle de chromatine avec les promoteurs du gène *IGF2* situés plus de 90 kb en amont. Cette interaction favorise ainsi la transcription de ce gène. Sur le chromosome maternel, le centre d’empreinte n’étant pas méthylé, la liaison de la protéine CTCF s’effectue, empêchant la formation de la boucle chromatinienne entre les promoteurs du gène *IGF2* et les séquences régulatrices. Ces dernières interagissent alors avec le promoteur du gène *H19* dont l’expression est ainsi favorisée.

Finalement, les travaux les plus récents ont démontré que les boucles chromatiniennes liées aux empreintes fonctionnelles et à la régulation de l’expression monoallélique des gènes prennent place à l’intérieur de grandes régions génomiques appelées « domaines topologiques » (*topologically-associating domain*, TAD, voir chapitre 3).

» Conclusion

L’empreinte génomique parentale se définit comme un phénomène épigénétique conduisant à une répression allélique stable au cours des divisions cellulaires et dépendante de l’origine parentale. Au niveau moléculaire, le terme « empreinte » peut être utilisé pour qualifier des empreintes « primaires », qui gardent la mémoire des origines parentales, des empreintes « secondaires », ces modifications épigénétiques spécifiques de l’allèle établies après la fécondation, ou des empreintes « fonctionnelles », qui intègrent l’ensemble des caractéristiques génomiques et épigénétiques impliquées dans la répression stable de l’allèle soumis à l’empreinte.

Il est aujourd’hui clairement établi que, chez les mammifères, les gènes subissent au cours du développement une reprogrammation épigénétique qui permet leur expression spatio-temporelle adéquate, et que l’empreinte génomique interfère avec ce processus pour obtenir une répression stable sur l’allèle soumis à l’empreinte. Cette répression stable implique la constitution d’empreintes fonctionnelles qui dépendent du contexte génomique et de contraintes spécifiques au type cellulaire, conduisant à des architectures tridimensionnelles de la chromatine différentes sur les deux allèles parentaux et à une expression monoallélique du gène. Les mécanismes moléculaires de l’empreinte sont donc très différents d’un gène à l’autre. Enfin, les éléments régulateurs impliqués dans l’empreinte, ou dans l’expression des gènes, sont parfois partagés entre plusieurs gènes au niveau des loci soumis à l’empreinte. Un élément régulateur, impliqué dans l’empreinte d’un gène, peut donc être requis pour l’expression d’un autre gène. De ce point de vue, l’empreinte génomique illustre remarquablement la complexité des mécanismes épigénétiques impliqués dans le contrôle de l’expression des gènes chez les mammifères, dont les dérégulations sont à l’origine de nombreux syndromes et d’anomalies de croissance chez l’homme, tels les syndromes de Beckwith-Wiedemann et de Prader-Willi. L’identification complète des domaines soumis à l’empreinte chez l’homme continue de poser des problèmes du fait de la difficulté à obtenir des informations sur l’origine parentale des allèles ou à cause de l’hétérogénéité cellulaire des échantillons disponibles. Cela constitue autant de défis importants à relever pour poursuivre les avancées dans ce domaine.

Remerciements

Nous remercions Maud Borensztein, Annick Lesne, Cosette Rebouissou et Nathan Lecouvreur pour la relecture de ce chapitre. Le travail de l'équipe de Thierry Forné est soutenu par la Mission pour les initiatives transverses et interdisciplinaires (MITI) du CNRS (projet 80|Prime Argecor), la Ligue contre le cancer (comité de l'Hérault, contrat 269152), la Fondation ARC (ARCPJA2023080006958) et la Fondation pour la recherche médicale (MND202310017917).

► Références bibliographiques

- Barlow D.P., 1993. Methylation and imprinting: From host defense to gene regulation? *Science*, 260, 309-310.
- Cattanach B.M., Kirk M., 1985. Differential activity of maternally and paternally derived chromosome regions in mice. *Nature* 315, 496-498. <https://api.istex.fr/ark:/67375/GT4-8HG60FDJ-Z/fulltext.pdf?sid=clickandread>
- Hanna C., Kelsey G., 2021. Features and mechanisms of canonical and noncanonical genomic imprinting. *Genes and Development*, 35, 821-834. <http://www.genesdev.org/cgi/doi/10.1101/gad.348422.121>
- Kaneko-Ishino T., Ishino E., 2022. The Evolutionary Advantage in Mammals of the Complementary Monoallelic Expression Mechanism of genomic imprinting and its emergence from a defense against the insertion into the host genome. *Frontiers in Genetics*, 13, 832983. <https://doi.org/10.3389/fgene.2022.832983>
- Llères D., Imaizumi Y., Feil R., 2021. Exploring chromatin structural roles of non-coding RNAs at imprinted domains. *Biochemical Society Transactions*, 49, 1867-1879. <https://doi.org/10.1042/BST20210758>
- McGrath J., Solter D., 1984. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell*, 37, 179-183. [https://doi.org/10.1016/0092-8674\(84\)90313-1](https://doi.org/10.1016/0092-8674(84)90313-1)
- Monk D., Mackay D., Eggermann T., Maher E., Riccio A., 2019. Genomic imprinting disorders: Lessons on how genome, epigenome and environment interact. *Nature Review Genetics*, 20, 235-248. <https://doi.org/10.1038/s41436-018-0416-7>
- Moore T., Haig D., 1991. Genomic imprinting in mammalian development: A parental tug-of-war. *Trends Genet*, 7, 45-49. [https://doi.org/10.1016/0168-9525\(91\)90230-N](https://doi.org/10.1016/0168-9525(91)90230-N)
- SanMiguel J.M., Bartolomei M.S., 2018. DNA methylation dynamics of genomic imprinting in mouse development. *Biology of Reproduction*, 99, 252-262. <https://doi.org/10.1093/biolre/iyo036>
- Surani M.A., Barton S.C., Norris M.L., 1984. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature*, 308, 548-550. <https://doi.org/10.1038/308548a0>
- Tucci V., Isles A., Kelsey G., Ferguson-Smith A., Erice Imprinting Group, 2019. Genomic Imprinting and Physiological Processes in Mammals. *Cell*, 176, 952-965. <https://doi.org/10.1016/j.cell.2019.01.043>

Chapitre 5

Empreinte génomique parentale chez les animaux d'élevage

JULIE DEMARS, CATHERINE LABBÉ, FRÉDÉRIQUE PITEL

Chez les espèces d'élevage, le rôle essentiel joué par les gènes soumis à empreinte a été démontré par l'étude de l'architecture génétique de certains caractères de production (O'Doherty *et al.*, 2015). Le développement d'outils moléculaires de génomique au cours des deux dernières décennies a entraîné un changement important dans le domaine de l'amélioration génétique animale. L'utilisation de marqueurs moléculaires en grand nombre et répartis de façon homogène sur le génome a alors permis la mise en place et le déploiement d'études d'associations génétiques pour de nombreux caractères d'intérêt agronomique. Ces analyses visent à identifier les régions chromosomiques qui contribuent à un phénotype spécifique en testant les associations entre les performances des animaux et leurs génotypes aux marqueurs. Chez les espèces porcine et ovine, deux mutations dans des gènes soumis à empreinte affectant un caractère de production (l'hypertrophie musculaire) ont ainsi été identifiées, et constituent – avec un syndrome lié à des perturbations de l'empreinte chez les bovins – des exemples majeurs de variabilité phénotypique liée au phénomène d'empreinte génomique parentale chez les animaux d'élevage. Après avoir présenté ces exemples, les méthodes actuelles de caractérisation de ce mécanisme seront décrites, et l'importance non négligeable de l'empreinte génomique parentale dans les productions animales sera développée.

► Empreinte génomique parentale et variabilité phénotypique

Le *large offspring syndrome*, un modèle bovin de pathologie d'empreinte

La principale caractéristique du *large offspring syndrome* (LOS, également appelé AOS, *abnormal offspring syndrome*), observé chez les fœtus au cours de la gestation et chez les nouveau-nés de ruminants, est l'hypercroissance, accompagnée notamment de macroglossie, d'une hernie ombilicale et d'anomalies des membres et de la moelle épinière (Rivera *et al.*, 2022). Les premiers cas ont été rapportés dans les années 1990 après utilisation de biotechnologies de reproduction comme le transfert nucléaire ou la production d'embryons *in vitro*. On sait depuis que ce syndrome peut également être d'origine naturelle, mais que son incidence augmente avec l'utilisation des biotechnologies de reproduction. Chez l'homme, le syndrome de Beckwith-Wiedemann présente

une étiologie similaire (voir chapitre 4) : il peut être d'origine génétique ou épigénétique et aboutit à une altération de l'empreinte génomique parentale (Filippi et Mckusick, 1970). Il est à noter qu'une prévalence anormalement élevée d'enfants conçus par FIV a été observée dans les séries d'enfants présentant un syndrome de croissance excessive tel que le syndrome de Beckwith-Wiedemann (BWS), son miroir clinique qu'est le syndrome de Silver-Russell (SRS, marqué par un retard de croissance intra-utérin), ou le syndrome d'Angelman (AS, présentant un désordre développemental).

De façon intéressante, il a été montré des similitudes dans les mécanismes épigénétiques mis en cause dans le LOS bovin et le syndrome de Beckwith-Wiedemann humain : ils conduiraient à une perte d'empreinte affectant des gènes orthologues entre les deux espèces (Mangiavacchi *et al.*, 2021). En effet, cette dérégulation de l'empreinte génomique parentale concerne les gènes *H19-IGF2* et *KCNQ1OT1-CDKN1C*, et leurs centres de contrôle d'empreinte respectifs appelés CE1 et CE2 (voir chapitre 4). Ces dysfonctionnements passent par des modifications de méthylation de l'ADN de CE1 (gain de méthylation de CE1 entraînant une augmentation de l'expression d'IGF2 (*insulin growth factor 2*), important facteur de croissance) ou CE2 (perte de méthylation de CE2 entraînant la dérégulation de gènes soumis à empreinte situés dans la région). Des études récentes ont également démontré l'implication de miARN, d'ARNt et de modifications dans la conformation de la chromatine dans les manifestations du syndrome.

Les similitudes phénotypiques et moléculaires entre les deux syndromes, humain et bovin, font du LOS bovin un excellent modèle pour l'étude des défauts d'empreinte chez l'homme.

La mutation *IGF2* associée à l'hypertrophie musculaire chez le porc

Un locus à effet quantitatif (ou QTL, *quantitative trait locus*) ayant un effet majeur sur la masse musculaire et le dépôt de graisse chez le porc a été identifié sur le chromosome 2 à partir de différents croisements. Ce QTL est responsable de 30% de la variance observée pour la viande maigre, de 15-30% de la variance pour la masse musculaire et de 10-20% de la variance pour l'augmentation de la teneur en graisse dorsale. La synténie conservée entre la région porcine du chromosome 2 et la région orthologue humaine située sur le chromosome 11 a suggéré le gène *IGF2* à expression paternelle comme gène candidat, étant donné son rôle majeur chez l'homme dans la croissance fœtale et postnatale. Afin de tester cette hypothèse, un modèle statistique adapté faisant l'hypothèse de la présence d'un effet d'empreinte parentale a été développé. Un effet paternel de ce QTL a été mis en évidence, renforçant l'intérêt du gène soumis à empreinte *IGF2* (figure 5.1A).

La mutation causale de ce QTL responsable du phénotype d'hypermuscularité chez le porc correspond à la substitution d'un nucléotide dans l'intron 3 du gène *IGF2* (Van Laere *et al.*, 2003). Seuls les porcs homozygotes pour l'allèle muté ou hétérozygotes ayant reçu l'allèle muté de leur père présentent un phénotype hypermusclé. Sur le plan moléculaire, les animaux porteurs de l'allèle muté sur leur chromosome paternel produisent trois fois plus de messagers issus du gène *IGF2* dans le muscle squelettique après la naissance (Van Laere *et al.*, 2003). La mutation, bien que localisée dans un îlot de dinucléotides CpG (cytosine liée à une guanine *via* un groupement phosphate) conservé au cours de l'évolution et qui est hypométhylé dans le muscle squelettique, n'affecte pas le profil de méthylation. En revanche, elle abroge la liaison

avec le répresseur transcriptionnel ZBED6 (*Zinc finger BED domain-containing protein 6*), ce qui aboutit à une expression accrue des ARNm issus du gène *IGF2* et donc à une hypertrophie musculaire chez les animaux ayant reçu l'allèle muté de leur père (Markljung *et al.*, 2009).

La mutation *DLK1-MEG3* associée au phénotype callipyge chez le mouton

Le phénotype callipyge, en référence au qualificatif des statues de la déesse grecque Aphrodite, se caractérise chez le mouton par une augmentation de ~ 30 % des muscles de l'arrière-train associée à une réduction de ~ 8 % de la teneur en graisse et à une amélioration de l'efficacité alimentaire (Cockett *et al.*, 1996). Le phénotype d'hyper-muscularité est observé uniquement chez les individus hétérozygotes porteurs de l'allèle muté sur leur chromosome paternel. Ce mode atypique de transmission non mendélienne représente un cas particulier d'empreinte génomique parentale appelé « surdominance polaire ».

La mutation responsable du phénotype callipyge chez le mouton correspond à la substitution d'un nucléotide. Elle est localisée sur le chromosome 18 ovin au sein d'un groupe de gènes soumis à empreinte, entre le gène à expression paternelle codant pour la protéine *DLK1* (*delta like non-canonical Notch ligand 1*) et le gène à expression maternelle *MEG3* (*maternally expressed 3*), qui est un long ARN non codant (Cockett *et al.*, 1996) (figure 5.1B). Les animaux callipyges, porteurs de la mutation sur le chromosome paternel, présentent, en comparaison des individus non callipyges : une surexpression des transcrits issus des gènes *DLK1* et *PEG11* (*retrotransposon-like gene 1*) codant pour des protéines du muscle squelettique; une réduction de la méthylation dans le muscle tout le long de la région du groupe de gènes soumis à empreinte; et une diminution de l'expression de l'histone désacétylase *HDAC9* (*histone deacetylase 9*), une enzyme associée à la condensation de la chromatine. Bien que deux modèles transgéniques murins d'expression ectopique des gènes *DLK1* et *PEG11* aient suggéré une action synergique des deux gènes pour induire le phénotype callipyge, il a été montré récemment que *DLK1* serait l'effecteur primaire de l'hypertrophie musculaire.

► Caractérisation des mécanismes d'empreinte chez les animaux d'élevage

Des approches ciblées aux approches tout-génome

Chez les animaux d'élevage, quelques dizaines de gènes soumis à empreinte ont été identifiées et validées expérimentalement à ce jour, alors qu'environ 200 gènes sont documentés chez l'homme et la souris. Plusieurs études ont suggéré que les gènes soumis à empreinte étaient moins conservés chez les mammifères qu'on ne le pensait initialement (Monk *et al.*, 2006; Khatib *et al.*, 2007). À titre d'exemple, l'expression maternelle de *DLX5* (*distal-less homeobox 5*) est conservée entre l'homme et le porc, alors que ce gène présente une expression biallélique chez la souris. Ces résultats soulèvent des questions sur la conservation des mécanismes d'empreinte génomique parentale entre les espèces au cours de l'évolution. Bien que certaines données sur l'empreinte génomique existent pour les espèces d'élevage, tous les gènes soumis à empreinte identifiés



Figure 5.1. Effets moléculaires des mutations affectant les loci *IGF2* chez le porc et *DLK1-MEG3* chez le mouton.

Le gène *IGF2* est exprimé uniquement à partir de l'allèle paternel. La mutation dans un intron du gène empêche la fixation du facteur ZBED6, ce qui a pour conséquence d'amplifier l'expression paternelle du gène *IGF2*. Les gènes *DLK1*, *CLPG* (callipyge) et *MEG3* présentent respectivement une expression paternelle, biallélique et maternelle. La mutation dans le gène *CLPG* affecte la méthylation tout le long du locus, ce qui dérégule l'expression de l'ensemble des gènes en *cis* et en *trans*.

jusqu'à présent ont été mis en évidence par des approches dites «gènes candidats» *via* la comparaison avec des régions orthologues connues dans les espèces murines et humaines. Aujourd'hui, il est possible de tirer parti des nouvelles technologies de séquençage (Hubert et Demars, 2022) et de la mise en place de croisements réciproques à façon pour identifier et cartographier l'atlas des gènes soumis à empreinte dans une espèce ou un tissu d'intérêt (figure 5.2). Cette stratégie tout-génome sans *a priori* a d'ores et déjà été appliquée chez la souris ou le porc. C'est également cette approche qui a été mise en place pour suggérer l'absence d'empreinte génomique parentale chez les oiseaux, dont un focus est présenté ci-dessous (Frésard *et al.*, 2014).

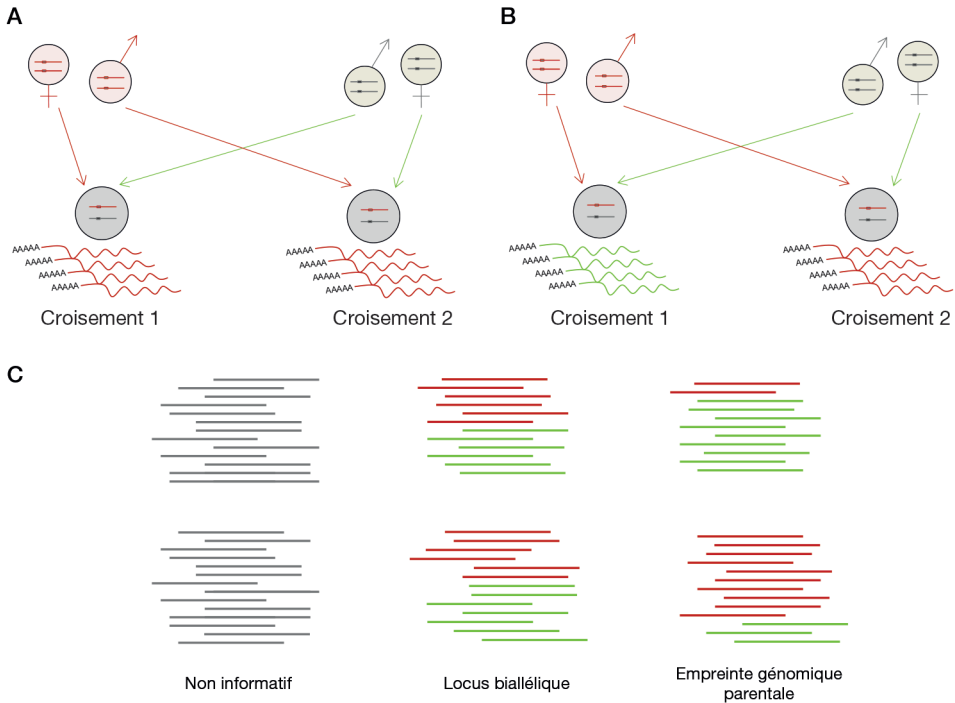


Figure 5.2. Mise en évidence de l'empreinte génomique parentale par des approches tout-génome.

L'utilisation de croisements réciproques permet de discerner l'expression allèle-spécifique (A) de l'empreinte génomique parentale (B). La détection d'empreinte génomique parentale nécessite que le locus présente au moins un SNP permettant de discerner les allèles parentaux (C).

Absence d'empreinte génomique parentale chez les oiseaux

L'empreinte génomique est apparue tant chez les plantes angiospermes que chez les mammifères placentaires (Feil et Berger, 2007). Elle est également observée chez les insectes. Il a même été démontré qu'elle est apparue plusieurs fois au cours de l'évolution des mammifères en raison d'une pression sélective différente sur plusieurs loci. Mais ce phénomène n'a jamais été observé chez les oiseaux (Das *et al.*, 2021). La principale théorie – même si elle n'est pas unique – proposée pour expliquer l'empreinte génomique parentale est l'hypothèse du conflit parental, faisant référence à une stratégie reproductive différente entre mâles et femelles (voir chapitre 4).

Elle suggère que les gènes contrôlant l'allocation des ressources à l'embryon devraient être particulièrement affectés. Le génome maternel limiterait l'utilisation des ressources afin de préserver la mère et sa future progéniture, alors que le génome paternel contrôlerait positivement l'allocation des nutriments au fœtus pour favoriser la croissance de sa descendance.

L'empreinte génomique a été étudiée chez les oiseaux, mais jusqu'à récemment seule l'expression des gènes connus pour être soumis à empreinte chez les mammifères avait été analysée chez la poule, la plupart d'entre eux montrant une expression biallélique et suggérant l'absence d'empreinte dans cette espèce. Cette idée est renforcée par l'absence de DNMT3L (cofacteur clé de l'établissement de la méthylation de l'ADN, voir chapitre 2) dans le génome de la poule. Mais les effets réciproques, nombreux en production avicole, peuvent s'expliquer en partie par l'empreinte, et quelques études semblaient mettre en évidence des QTL à origine parentale. Ces observations contradictoires ont incité la communauté scientifique à répondre définitivement à la question (Frésard *et al.*, 2013), et plusieurs études tout-génome ont été réalisées grâce aux progrès des technologies de séquençage, qui permettent désormais d'étudier l'empreinte génomique par le séquençage du transcriptome entier. Les analyses réalisées sur embryons entiers, cerveaux adultes ou tissus embryonnaires, n'ont jusqu'alors jamais observé d'empreinte génomique parentale chez la poule.

Existence d'une empreinte génomique parentale chez les poissons ?

Chez les poissons, la question de l'empreinte parentale a été discutée d'un point de vue évolutif il y a plus de vingt ans (McGowan et Martin, 1997). Depuis, il a été recherché par des approches ciblées (gènes candidats) si certains gènes connus pour être soumis à empreinte parentale chez l'homme pourraient aussi l'être chez les poissons. Par exemple, une région du gène *igf2* du poisson rouge s'avère différentiellement méthylée entre le sperme et l'ovocyte, mais ce différentiel spécifique aux gamètes parentaux n'est pas maintenu dans l'embryon. Dans le même sens, l'orthologue chez le poisson zèbre du gène *peg1/mest* soumis à empreinte chez les mammifères a en fait une expression biallélique dans les larves de poisson zèbre, réfutant ainsi la présence d'un dosage d'origine parentale pour cet autre candidat.

Paulsen *et al.* (2005) ont également démontré que certains éléments de contrôle de l'empreinte identifiés chez les mammifères n'étaient pas présents chez le fugu ou le poisson zèbre, et que les groupes de gènes soumis à empreinte parentale chez les mammifères étaient peu conservés chez les poissons. Comme chez les oiseaux, l'absence d'un orthologue de *Dnmt3l*, ce cofacteur clé dans la méthylation des gènes soumis à empreinte chez les mammifères, a été considérée comme un indice supplémentaire de l'absence de mécanisme d'empreinte chez le poisson zèbre. Au contraire, Ma *et al.* (2001) ont montré que le gène *ntl* (*no tail*) du poisson rouge est différentiellement méthylé entre le spermatozoïde et les ovules, et que l'allèle paternel est transcrit plus tôt que l'allèle maternel dans l'embryon. Des indices d'un possible mécanisme d'empreinte ont également été fournis dans une étude portant sur l'expression d'un transgène chez le poisson zèbre. À partir de ces études divergentes, il est encore impossible de se prononcer définitivement sur l'existence ou non de l'empreinte parentale chez les poissons, et des études à l'échelle du génome sont encore nécessaires.

► Implication potentielle pour l'élevage

Le nombre de gènes soumis à empreinte représente environ 1 à 2 % des gènes annotés dans le génome des mammifères. Comparé aux quelque 40 000 gènes – codants ou non – présents dans le génome, ce chiffre laisse penser que si l'empreinte a bien un effet sur les phénotypes, celui-ci reste d'une ampleur limitée. Mais, comme nous l'avons vu plus haut, certains loci soumis à empreinte sont à l'origine d'une part importante de la variance phénotypique, comme *IGF2* pour la masse musculaire ou la quantité de graisse dorsale, ou *DLK1* pour la quantité de muscles de l'arrière-train. Même si peu nombreux, les effets des gènes soumis à empreinte sur les phénotypes peuvent être importants et représentent donc un domaine d'étude à ne pas négliger. La communauté scientifique continue d'ailleurs de fournir des preuves de l'implication de l'empreinte génomique parentale dans la croissance et le développement des espèces de rente (Ibeagha-Awemu et Zhao, 2015).

Pourtant, comme souligné par Michael Goddard et Emma Whitelaw (2014), l'empreinte n'est pas intégrée en routine dans les modèles d'évaluation des valeurs génétiques, tout comme la plupart des phénomènes non mendéliens (hérédité cytoplasmique, épigénétique transgénérationnelle, épistasie, etc.). Adapter les modèles en tenant compte de ces phénomènes permettrait d'améliorer l'estimation de la valeur des reproducteurs, et donc l'efficacité de la sélection, même sans rechercher d'effet cumulatif entre générations. Dans les espèces utilisées en croisement, prendre en compte le phénomène d'empreinte permettrait également de réaliser des plans d'accouplement plus pertinents. De plus, des exemples comme celui du LOS montrent que des dysfonctionnements des mécanismes d'empreinte génomique parentale, en particulier les problèmes potentiels liés à l'utilisation des biotechnologies de reproduction, peuvent avoir des conséquences sérieuses sur la survie ou la santé/les performances des individus. Améliorer les connaissances dans ce domaine permettrait de diminuer l'incidence de ces problèmes dans les systèmes de production faisant appel à ces méthodes.

Plusieurs points importants doivent être gardés à l'esprit lorsqu'il est question d'épigénétique et d'implication pour l'élevage. L'empreinte génomique parentale est un exemple atypique de régulation de l'expression des gènes par des mécanismes épigénétiques. Le rôle des marques épigénétiques dans la variabilité des caractères est aujourd'hui reconnu. Les récentes avancées technologiques de séquençage permettent d'analyser l'ensemble de l'épigénome soit de façon unique en accédant au méthylome de l'ADN, soit de façon combinée en utilisant aussi le séquençage du génome en longues lectures. Ces analyses comportent cependant des contraintes majeures : leur coût sur génome entier, même s'il diminue, reste très élevé pour une application à un grand nombre d'individus. De plus, alors que le génotype est le même pour toutes les cellules d'un animal, l'épigénotype dépend de l'âge (ou du stade de développement) de l'individu et, surtout, est spécifique du tissu. Il faut donc choisir avec soin les échantillons à analyser, notamment le tissu sélectionné, afin d'obtenir des données pertinentes.

Malgré tout, avec le développement de nouvelles analyses génomiques et de nouvelles méthodes de laboratoire pour l'épigénétique, il est possible qu'à l'avenir les éleveurs ne s'appuient plus seulement sur les informations génétiques, mais utilisent également l'épigénome des candidats à la sélection.

Remerciements

Le travail de Julie Demars est soutenu par l'Agence nationale de la recherche au titre du programme «Jeunes chercheuses et jeunes chercheurs» portant la référence ANR-18-CE20-0018.

► Références bibliographiques

- Babak T. *et al.*, 2008. Global survey of genomic imprinting by transcriptome sequencing. *Curr. Biol.*, 18 (22), 1735-1741. <https://doi.org/10.1016/j.cub.2008.09.044>
- Chen Z. *et al.*, 2015. Characterization of global loss of imprinting in fetal overgrowth syndrome induced by assisted reproduction. *Proc. Natl. Acad. Sci. USA*, 112 (15), 4618-4623. <https://doi.org/10.1073/pnas.1422088112>
- Cockett N.E. *et al.*, 1996. Polar overdominance at the ovine callipyge locus. *Science*, 273 (5272), 236-238. <https://doi.org/10.1126/science.273.5272.236>
- Das R. *et al.*, 2012. Convergent and divergent evolution of genomic imprinting in the marsupial *Monodelphis domestica*. *BMC Genomics*, 13, 394. <https://doi.org/10.1186/1471-2164-13-394>
- Feil R., Berger F., 2007. Convergent evolution of genomic imprinting in plants and mammals. *Trends Genet. TIG*, 23 (4), 192-199. <https://doi.org/10.1016/j.tig.2007.02.004>
- Filippi G., Mckusick V.A., 1970. The Beckwith-Wiedemann syndrome. *Medicine (Baltimore)*, 49 (4), 279-298. <https://doi.org/10.1097/00005792-197007000-00002>
- Freking B.A. *et al.*, 2002. Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. *Genome Res.*, 12 (10), 1496-1506. <https://doi.org/10.1101/gr.571002>
- Frésard L. *et al.*, 2013. Epigenetics and phenotypic variability: some interesting insights from birds. *Genet. Sel. Evol. GSE*, 45 (1), 16. <https://doi.org/10.1186/1297-9686-45-16>
- Frésard L. *et al.*, 2014. Transcriptome-wide investigation of genomic imprinting in chicken. *Nucleic Acids Res.*, 42 (6), 3768-3782. <https://doi.org/10.1093/nar/gkt1390>
- Goddard M.E., Whitelaw E., 2014. The use of epigenetic phenomena for the improvement of sheep and cattle. *Front. Genet.*, 5. <https://doi.org/10.3389/fgene.2014.00247>
- Haig D., 2014. Coadaptation and conflict, misconception and muddle, in the evolution of genomic imprinting. *Heredity*, 113 (2), 96-103. <https://doi.org/10.1038/hdy.2013.97>
- Hubert J.-N., Demars J., 2022. Genomic imprinting in the new omics era: A model for systems-level approaches. *Front. Genet.*, 13. <https://doi.org/10.3389/fgene.2022.838534>
- Ibeagha-Awemu E.M., Zhao X., 2015. Epigenetic marks: Regulators of livestock phenotypes and conceivable sources of missing variation in livestock improvement programs. *Front. Genet.*, 6. <https://doi.org/10.3389/fgene.2015.00302>
- Jeon J.-T. *et al.*, 1999. A paternally expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the IGF2 locus. *Nat. Genet.*, 21 (2). <https://doi.org/10.1038/5938>
- Khatib H., Zaitoun I., Kim E.S., 2007. Comparative analysis of sequence characteristics of imprinted genes in human, mouse, and cattle. *Mamm Genome*, 18, 538-547. <https://doi.org/10.1007/s00335-007-9039-z>
- Ma C., Fan L., Ganassin R., Bols N., Collodi P., 2001. Production of zebrafish germ-line chimeras from embryo cell cultures. *Proc. Natl. Acad. Sci. USA*, 98 (5), 2461-2466. <https://doi.org/10.1073/pnas.041449398>
- Mangiavacchi P.M., Caldas-Bussiere M.C., da Mendonça S.M., Dias A.J.B., Rios Á.F.L., 2021. Multi-locus imprinting disturbances of Beckwith-Wiedemann and Large offspring syndrome/Abnormal offspring syndrome: A brief review. *Theriogenology*, 173, 193-201. <https://doi.org/10.1016/j.theriogenology.2021.08.005>
- Markljung E. *et al.*, 2009. ZBED6, a novel transcription factor derived from a domesticated DNA transposon regulates IGF2 expression and muscle growth. *PLoS Biol.*, 7 (12), e1000256. <https://doi.org/10.1371/journal.pbio.1000256>

- McGowan R.A., Martin C.C., 1997. DNA methylation and genome imprinting in the zebrafish, *Danio rerio*: some evolutionary ramifications. *Biochem. Cell Biol. Biochim. Biol. Cell.*, 75 (5), 499-506.
- Monk D. *et al.*, 2006. Limited evolutionary conservation of imprinting in the human placenta. *Proc. Natl. Acad. Sci.*, 103 (17), 6623-6628. <https://doi.org/10.1073/pnas.0511031103>
- Murphy S.K. *et al.*, 2006. Callipyge mutation affects gene expression in cis: A potential role for chromatin structure. *Genome Res.*, 16 (3), 340-346. <https://doi.org/10.1101/gr.4389306>
- Nezer C. *et al.*, 1999. An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs. *Nat. Genet.*, 21 (2). <https://doi.org/10.1038/5935>
- O'Doherty A.M., MacHugh D.E., Spillane C., Magee D.A., 2015. Genomic imprinting effects on complex traits in domesticated animal species. *Front. Genet.*, 6. <https://doi.org/10.3389/fgene.2015.00156>
- Paulsen M., Khare T., Burgard C., Tierling S., Walter J., 2005. Evolution of the Beckwith-Wiedemann syndrome region in vertebrates. *Genome Res.*, 15 (1), 146-153. <https://doi.org/10.1101/gr.2689805>
- Rivera R.M. *et al.*, 2022. Identification of large offspring syndrome during pregnancy through ultrasonography and maternal blood transcriptome analyses. *Sci. Rep.*, 12, 10540. <https://doi.org/10.1038/s41598-022-14597-w>
- Smit M. *et al.*, 2003. Mosaicism of Solid Gold Supports the Causality of a Noncoding A-to-G Transition in the Determinism of the Callipyge Phenotype. *Genetics*, 163 (1), 453-456. <https://doi.org/10.1093/genetics/163.1.453>
- Smith T.H.L., Collins T.M., McGowan R.A., 2011. Expression of the dnmt3 genes in zebrafish development: similarity to Dnmt3a and Dnmt3b. *Dev. Genes Evol.*, 220 (11-12), 347-353. <https://doi.org/10.1007/s00427-010-0347-z>
- Takeda H. *et al.*, 2006. The callipyge mutation enhances bidirectional long-range DLK1-GTL2 intergenic transcription in cis. *Proc. Natl. Acad. Sci.*, 103 (21), 8119-8124. <https://doi.org/10.1073/pnas.0602844103>
- Van Laere A.-S. *et al.*, 2003. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature*, 425 (6960). <https://doi.org/10.1038/nature02064>

Chapitre 6

Empreinte parentale chez les plantes : mécanismes, fonctions et applications

CLÉMENT LAFON PLACETTE

Comme expliqué dans les chapitres 4 et 5, on définit l’empreinte parentale comme le phénomène épigénétique aboutissant à l’expression d’un allèle dépendant de son origine parentale : certains gènes sont exprimés uniquement à partir de l’allèle provenant de la mère, alors que pour d’autres gènes c’est l’allèle paternel qui est exprimé. L’empreinte parentale est observée chez les plantes à fleur (angiospermes, ce qui exclut conifères, fougères et mousses), essentiellement dans la graine, et plus précisément dans l’albumen, le tissu nourricier de l’embryon, l’analogie du placenta des mammifères.

La première observation de l’empreinte parentale chez les plantes a été faite sur le maïs dans les années 1970. Les grains de maïs peuvent être jaunes ou violets. Si l’on croise deux plantes, l’une en tant que mâle et l’autre en tant que femelle (la plupart des espèces sont hermaphrodites, on peut donc utiliser la même plante comme mâle et femelle), on obtient une couleur des grains, par exemple jaune. Mais si le croisement inverse est réalisé, on obtient une couleur violette. Ceci ne s’explique pas facilement par la génétique classique, où le sens du croisement ne devrait pas changer le résultat de la descendance, puisqu’une plante donnée a la même séquence d’ADN, qu’elle transmette ses allèles en tant que mâle ou en tant que femelle. L’empreinte parentale permet d’expliquer ce phénomène : dans l’albumen, les chromosomes hérités du mâle expriment des gènes différents de ceux hérités de la femelle, les génomes paternel et maternel ont donc un effet différent sur la couleur de la graine. Depuis cette première observation, l’empreinte parentale a été retrouvée chez toutes les espèces d’angiospermes étudiées : *Arabidopsis*, ricin, riz, blé, tomate, nénuphar, fleur de singe, etc. Et on sait aujourd’hui qu’elle affecte d’autres gènes que ceux responsables de la coloration du grain de maïs. On peut alors se demander : l’empreinte parentale joue-t-elle un rôle biologique important et pourquoi est-elle apparue chez les plantes ? Comment fonctionne-t-elle chez les plantes ? Peut-on utiliser notre savoir de l’empreinte parentale à des fins agronomiques ? Ce sont les questions sur lesquelles nous allons nous pencher dans ce chapitre.

► Rappel sur la reproduction sexuée chez les plantes

Les plantes, comme les animaux, ont une sexualité. Ceci paraît peut-être trivial, mais jusqu’en 1717, on ignorait complètement que le sexe existait aussi chez les plantes.

Cette date fait référence à la production par Thomas Fairchild d'œillettes hybrides en mettant en contact du pollen, qu'il pensait être la partie mâle des plantes, avec un pistil, qu'il pensait être la partie femelle de la fleur. Son intuition fut bonne, puisqu'il réussit ainsi à produire des plantes à partir des graines obtenues.

Bien que la reproduction sexuée chez les plantes présente certaines analogies avec celle des animaux, elle montre aussi plusieurs particularités. L'une d'elles est l'existence de la « double fécondation » : chaque grain de pollen (il y en a des milliers dans chaque fleur) ne contient pas une, mais deux cellules spermatiques. Une autre particularité est que ce que l'on appelle « ovule » chez les plantes est un organe complexe, plutôt analogue à l'ovaire chez nous, qui contient non seulement des tissus maternels, mais aussi, à maturité, deux cellules gamétiques. Lors de la fécondation, une cellule spermatique mâle va fusionner avec une des cellules femelles pour donner le zygote puis l'embryon (figure 6.1). La deuxième cellule mâle va fusionner avec l'autre gamète femelle, la cellule centrale, pour donner naissance à un tissu appelé « albumen ». Ce tissu a pour fonction de nourrir l'embryon jusqu'à la maturité de la graine, voire parfois même pendant la germination de celle-ci. D'un point de vue anthropocentrique, l'albumen est d'une importance agronomique majeure puisqu'il compose la plus grande partie des graines de céréales comme le riz, le maïs ou le blé. Par ailleurs, une analogie importante avec les mammifères placentaires est que les plantes à fleurs sont aussi vivipares (l'embryon grandit au sein de la mère, qui apporte les ressources nécessaires pour le bon développement de l'embryon), et, tout comme le placenta, les cellules de l'albumen sont biparentales, composées des génomes à la fois maternels et paternels. Il est probable que ces deux caractéristiques soient la condition nécessaire pour l'émergence et l'évolution de l'empreinte parentale au sein du vivant, plantes comprises.

Il est important de noter que normalement les gamètes sont haploïdes, c'est-à-dire que, du fait de la méiose, les gamètes possèdent un nombre de chromosomes divisé de moitié par rapport aux cellules somatiques. Ainsi, quand les gamètes mâles et femelles fusionnent pendant la fécondation, ils redonnent une cellule avec un nombre de chromosomes équivalent à celui des cellules somatiques. C'est bien le cas des cellules spermatiques dans le pollen, et de l'oosphère chez la femelle chez la plupart des espèces de plantes. En revanche, la cellule centrale, l'autre cellule gamétique femelle, est en fait une fusion de deux gamètes. Elle possède donc un nombre de chromosomes double comparé aux autres gamètes. La fusion entre une cellule spermatique et la cellule centrale aboutit donc à des cellules de l'albumen ayant deux copies de chromosomes maternels et une copie de chromosomes paternels. Le ratio entre les chromosomes maternels et paternels n'est donc pas de 1 pour 1 dans l'albumen, mais de 2 pour 1.

► Mécanismes épigénétiques aboutissant à l'empreinte parentale

Comment l'empreinte parentale se met-elle en place au niveau moléculaire chez les plantes, et par quels processus épigénétiques ? Pour répondre à cette question, il faut s'intéresser à ce qu'il se passe dans les gamètes, après la méiose et avant la fécondation.

À maturité, le pollen est composé de deux types de cellules : la cellule végétative et les cellules spermatiques (figure 6.1). Dans le noyau végétatif, une forte décondensation de la chromatine s'effectue, notamment à cause de l'absence d'une enzyme responsable

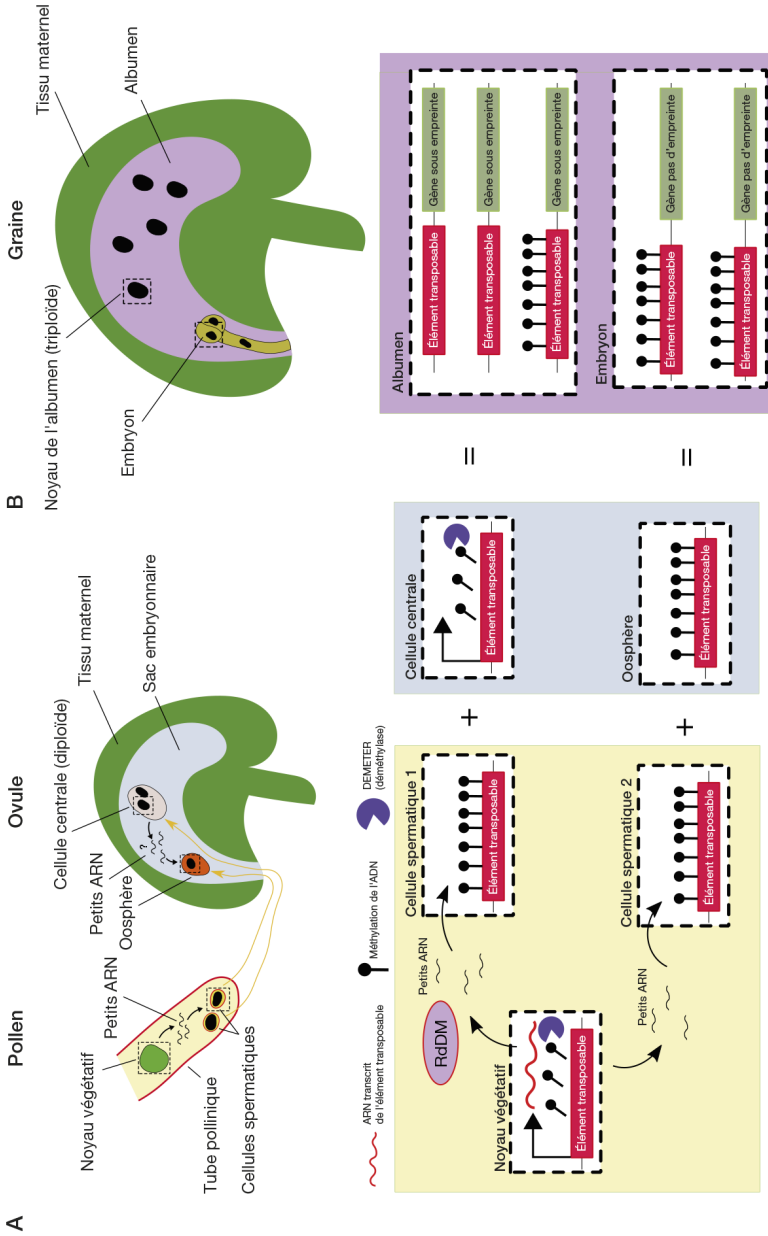


Figure 6.1. La reproduction sexuée chez les plantes et les processus épigénétiques l'entourant.

A. Avant la double fécondation. Dans le pollen, les petits ARN sont produits dans le noyau végétatif, suite notamment à l'action déméthylante de DEMETER sur les éléments transposables. Les petits ARN migrent dans les cellules spermatisques. Dans le sac embryonnaire, un processus similaire se déroule entre cellule centrale et oosphère. B. Après la double fécondation. L'action de déméthylation dans l'oosphère aboutit à une asymétrie épigénétique entre génomes paternel et maternel dans l'albumen, à la base de l'empreinte parentale. Cette asymétrie n'est pas présente dans l'embryon.

de la formation d'hétérochromatine, DDM1 (*decrease in DNA methylation 1*), habituellement présente dans toutes les cellules. En revanche, une enzyme appelée DEMETER, qui a une action déméthylante (remplacement des cytosines méthylées par des cytosines non méthylées), est elle active dans le noyau végétatif (Ibarra *et al.*, 2012). Cette enzyme cible particulièrement les éléments transposables. Ceci aboutit dans le noyau végétatif à une réexpression des éléments transposables, normalement réprimés par la machinerie épigénétique. Cette réactivation des éléments transposables va être un signal pour l'activation d'une voie moléculaire particulière, la RdDM (*RNA-directed DNA methylation*; voir chapitre 10). En reconnaissant les transcrits d'éléments transposables, cette voie va entraîner le clivage de ceux-ci en petits ARN, qui vont à leur tour servir à la répression d'autres éléments transposables reconnus par homologie de séquence avec ces petits ARN. Il est à noter que ces petits ARN ne vont pas avoir une action dans le noyau végétatif, mais vont migrer de la cellule végétative vers les cellules spermatiques où ils vont s'accumuler (Martínez *et al.*, 2016). Le noyau végétatif apparaît donc comme « usine » à petits ARN, les destinataires étant les cellules spermatiques.

Les gamètes femelles sont plus difficiles à étudier du fait de leur faible nombre (comparé aux grains de pollen). Néanmoins, les données actuelles montrent un scénario très similaire à celui du pollen, notamment, dans la cellule centrale : l'enzyme DEMETER est aussi active, ce qui aboutit à la réduction de la méthylation de l'ADN, spécialement sur les éléments transposables. Par manipulation génétique, il a été montré que des petits ARN artificiels produits dans la cellule centrale sont capables de réprimer des gènes exprimés dans l'oosphère (Ibarra *et al.*, 2012). Des petits ARN artificiels peuvent donc migrer entre cellule centrale et oosphère. Il paraît donc probable que les petits ARN naturels en soient capables aussi, mais cela reste à démontrer.

À ce stade, la question du rôle de ces remaniements épigénétiques, à la fois du côté mâle et femelle, se pose. Une des hypothèses actuellement privilégiées est que ces remaniements épigénétiques au cours de la reproduction joueraient un rôle de protection contre d'éventuelles insertions d'éléments transposables pouvant être néfastes à la génération suivante (Tirot et Jullien, 2022). En effet, si un élément transposable peut s'insérer dans l'ADN des gamètes, il sera transmis à la génération suivante. Cela peut être bénéfique pour l'élément transposable, mais pas forcément pour la génération suivante : les éléments transposables s'insèrent n'importe où, y compris dans les gènes, et certaines familles d'éléments transposables s'insèrent même préférentiellement dans les gènes. Il y a donc un risque important d'effets délétères pour la génération d'après si cela se produit. Les remaniements décrits ci-dessus aboutissent à l'accumulation de petits ARN dans les gamètes et au maintien de la méthylation des éléments transposables dans ces cellules. Par ailleurs, dans l'embryon durant son développement, on observe une forte vague de méthylation, notamment des éléments transposables, et certains chercheurs pensent que c'est une conséquence directe de l'accumulation de petits ARN dans les gamètes, même si cela reste à démontrer.

Dans l'albumen, les mécanismes épigénétiques sont uniques comparé aux autres organes. Comme le montre la figure 6.1, du fait des événements épigénétiques différents se passant dans la cellule spermatique et la cellule centrale, la première cellule de l'albumen comporte une asymétrie épigénétique entre allèle maternel et allèle paternel, notamment au niveau des éléments transposables. Cette asymétrie sera maintenue au cours des divisions cellulaires lors du développement de l'albumen.

Si un gène se trouve à proximité d'un élément transposable, il va présenter une asymétrie d'expression entre allèle maternel et paternel, car la régulation de ce gène va être influencée par les marques épigénétiques alentour, spécialement si ces marques se situent dans le promoteur de ce gène. Cependant, même si la relation « méthylation de l'élément transposable sur l'allèle paternel = répression du gène associé = allèle maternel exprimé uniquement » est parfois vérifiée, quelque fois cette méthylation de l'ADN va « protéger » l'allèle paternel du complexe enzymatique Polycomb, qui dépose une autre marque répressive, la méthylation des histones (H3K27). Pour des raisons mal comprises, les Polycomb sont exclus de zones génomiques méthylées. Dans ce cas, l'allèle maternel, qui ne possède pas de méthylation de l'ADN, sera réprimé par la marque histone, alors que l'allèle paternel, sur lequel la méthylation de l'ADN exclut les Polycomb, pourra s'exprimer, aboutissant à un gène sous empreinte paternelle.

On ne comprend pas encore tous les mécanismes aboutissant à l'expression d'un seul allèle parental à partir de l'asymétrie épigénétique de départ. Mais ce qui semble sûr, c'est que cette asymétrie est la base de la mise en place de l'empreinte parentale. Par ailleurs, cette asymétrie épigénétique n'existe presque pas dans l'embryon, ce qui explique que l'empreinte parentale a lieu essentiellement dans l'albumen. Pourtant, pour une dizaine de gènes et à des stades très restreints du développement embryonnaire, l'empreinte parentale a été montrée dans l'embryon (Raissig *et al.*, 2013). Cela reste cependant très marginal et est encore en débat.

► Mort de la graine hybride : l'empreinte parentale, *usual suspect*

L'être humain est diploïde (deux copies de chaque chromosome) comme la plupart des animaux. Mais chez les plantes, on trouve non seulement des espèces diploïdes (comme le maïs) mais aussi d'autres qui ont plus de deux copies de leurs chromosomes : certaines sont tétraploïdes (4 copies comme l'armoise annuelle), hexaploïdes (6 copies comme le chou commun ou le blé tendre), dodécaploïdes (12 copies comme la célosie argentée), pour n'en citer que quelques exemples. On regroupe ces espèces sous le terme « polyploïdes » (ayant plus de deux copies de chromosomes), et on retrouve bon nombre d'entre elles dans les espèces cultivées (banane, canne à sucre, fraise, blé, etc.). Très tôt en agronomie, des croisements entre espèces diploïdes et polyploïdes ont été tentés. Or, depuis les premiers œillets hybrides de Thomas Fairchild et jusqu'à aujourd'hui, la production d'hybrides de plantes se heurte à un problème important : les graines issues de croisements entre niveaux différents de ploïdie sont souvent « mort-nées » (non viables). De même, les graines hybrides issues de croisements entre espèces différentes ne sont souvent pas non plus viables.

Ceci a deux conséquences majeures. La première concerne l'évolution des espèces, qui, si elles ne peuvent pas produire d'hybrides viables, resteront séparées génétiquement, aboutissant à la cohabitation et à la perpétuation d'espèces différentes (c'est ce que l'on appelle une barrière reproductive). La seconde concerne l'amélioration des plantes : celle-ci est largement basée sur la production de lignées hybrides, qui nécessite l'obtention de graines hybrides viables.

Les communautés scientifique et agronomique ont étudié la mort de la graine hybride en détail depuis plus d'un siècle, permettant d'avoir un tableau assez détaillé de ce

phénomène. De façon intéressante, quel que soit le cas, hybride entre espèces ou hybride entre diploïdes et polyploïdes, les mêmes phénomènes sont observés, à savoir :

- le problème de la graine hybride vient toujours de l'albumen, qui a un développement anormal et donc ne remplit pas son rôle de nourrissage de l'embryon. On le sait car, notamment depuis le début du xx^e siècle, la technique de « sauvetage d'embryon » est utilisée en agronomie. Elle consiste à extraire l'embryon d'une graine après fécondation et à placer cet embryon dans un milieu nutritif *in vitro*, appelé par certains « albumen artificiel » : en prenant l'embryon d'une graine hybride qui serait normalement non viable, on arrive à cultiver l'embryon en lui donnant les substances nutritives dont il a besoin et ainsi obtenir une plante hybride en parfaite santé. Le problème de la graine hybride ne vient donc pas de l'embryon, mais bien de l'albumen qui ne remplit pas son rôle;
- en fonction du sens du croisement, les graines obtenues ne sont pas exactement les mêmes : en effet, croiser une espèce/ploidie A en tant que plante mère et une espèce/ploidie B en tant que donneur de pollen n'aboutit pas au même résultat si on inverse, c'est-à-dire utiliser A comme donneur de pollen et B comme plante mère (figure 6.2). Parfois même, les graines hybrides sont viables dans un sens du croisement, mais pas dans l'autre. Cela suggère que les génomes paternels et maternels ne sont pas équivalents, ils n'ont pas le même effet sur la graine. Si ce n'était pas le cas, utiliser une plante comme père ou mère dans un croisement hybride ne changerait rien.

Ces deux caractéristiques de la graine hybride correspondent à l'empreinte parentale : mort de la graine hybride et empreinte parentale prennent place dans l'albumen, et elles ont un effet dépendant du parent d'origine. Le fait que les génomes paternels et maternels n'ont pas le même effet sur la graine peut s'expliquer par le fait que ces deux génomes expriment des gènes différents : pour certains gènes, l'allèle paternel sera exprimé, alors que pour d'autres gènes, ce sera l'allèle maternel.

L'empreinte parentale semble donc une cause moléculaire plausible pour expliquer la mort de la graine hybride (Städler *et al.*, 2021). Il faut cependant davantage d'arguments dans ce sens, mais la recherche fondamentale de ces dernières années y aide. Il est aujourd'hui facile de muter des gènes chez un certain nombre d'espèces de plantes, notamment *Arabidopsis*. Chez cette espèce, plusieurs travaux ont montré que si on mute des gènes sous empreinte parentale, les taux de viabilité des graines hybrides sont proches de la normale (Adams *et al.*, 2000). L'empreinte parentale est donc directement impliquée dans la mort de la graine hybride. De la même façon, si l'on mute le donneur de pollen pour une des méthyltransférases, la viabilité des graines hybrides issues du croisement augmente, indiquant que la méthylation de l'ADN, probablement *via* la régulation de l'empreinte parentale, joue un rôle important dans la mort de la graine hybride.

► Causes évolutives de l'apparition de l'empreinte parentale chez les plantes

Cette question reste relativement ouverte. Le fait qu'on retrouve l'empreinte parentale chez toutes les espèces de plantes à fleurs étudiées jusqu'à aujourd'hui pourrait suggérer que l'empreinte parentale est très importante. Et pourtant, ce n'est pas si évident. En effet, plusieurs chercheurs ont proposé que l'empreinte ne soit qu'un sous-produit de mécanismes plus importants, et qu'elle n'ait pas vraiment de rôle biologique. Leurs arguments sont tout d'abord fondés sur le fait que l'empreinte parentale ne concerne que plusieurs dizaines, voire quelques centaines de gènes maximum ;

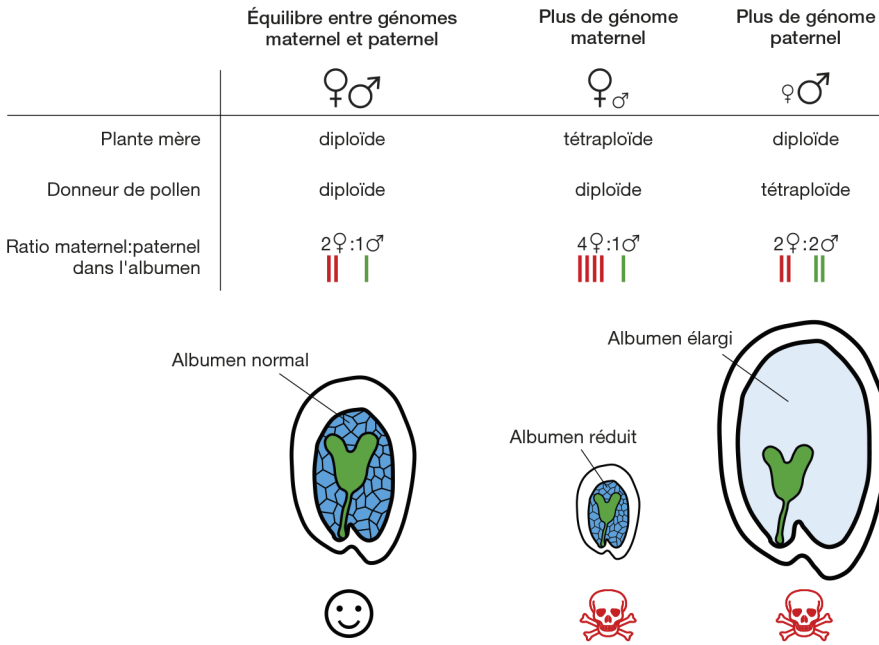


Figure 6.2. Mort de la graine hybride après croisement entre plantes diploïdes et tétraploïdes. Chaque barre rouge et verte représente une copie de chromosomes maternels et paternels, respectivement. Les graines avec des parents de même ploïdie sont viables (représentées par un smiley), alors que les graines avec des parents de ploïdies différentes ne sont généralement pas viables (représentées par une tête de mort). À noter que même si dans les deux sens de croisement entre diploïdes et tétraploïdes les graines ont des problèmes de survie, les problèmes que chaque croisement a ne sont pas les mêmes, montrant l'effet dépendent du parent propre à l'empreinte parentale.

comparé à 20 000-40 000 gènes dans l'ensemble du génome selon l'espèce, cela représente moins de 1% des gènes. C'est peu. L'autre argument est que les mécanismes épigénétiques qui aboutissent à l'empreinte parentale visent avant tout les éléments transposables (voir figure 6.1). Ces mécanismes ont probablement pour fonction la défense contre les éléments transposables et le maintien de l'intégrité du génome à travers les générations. Les gènes à empreinte parentale le seraient uniquement par chance : dans ces quelques cas, un élément transposable se serait inséré près d'un gène, et par conséquent, les mécanismes visant cet élément transposable auraient un impact sur l'expression du gène adjacent (Gehring *et al.*, 2009). Comme ces marques épigénétiques sont asymétriques en ce qui concerne les génomes paternel et maternel dans l'albume, cela aboutirait à l'empreinte parentale sur ce gène.

Entre temps, des études ont montré que l'empreinte parentale joue un rôle dans la mort de la graine hybride. Cependant, il apparaît ici aussi probable que l'empreinte parentale soit d'abord apparue comme un sous-produit de la défense contre les éléments transposables, et qu'elle ait ensuite acquis un rôle plus fonctionnel. Cette hypothèse est soutenue par une étude chez *Arabidopsis* dans laquelle 15 gènes sous empreinte ont été mutés : les graines des plantes mutées ne montraient aucune différence avec les plantes sans la mutation (Wolff *et al.*, 2015). Cela suggère donc que, hormis le cas des graines hybrides, les gènes à empreinte parentale ne jouent pas un rôle important pour le développement

de la graine. Cependant, il est possible que muter un seul gène à empreinte parentale à la fois ne suffise pas à voir un changement sur la graine : l'ensemble des gènes à empreinte pourraient jouer un rôle dans la graine, de manière additive.

Ce rôle pourrait entre autres être lié à une particularité des plantes à fleurs (aussi appelées angiospermes) comparé aux gymnospermes tels que les conifères : chez ces derniers, le tissu nourricier de l'embryon, appelé « endosperme », provient uniquement de la plante mère, et se développe pour l'essentiel avant même que l'oosphère ne soit fécondée par le pollen. Le problème avec cette stratégie reproductive est que si le pollen ne vient jamais (ce qui arrive souvent), la plante mère aura produit en vain un tissu riche en ressources, ce qui a un coût pour la plante : celle-ci n'a qu'une certaine quantité de substances nutritives à sa disposition, et ce qu'elle donne pour les graines, elle ne peut l'utiliser pour sa survie. Ce problème est résolu avec la double fécondation chez les angiospermes : le tissu nourricier ne sera produit que s'il y a fécondation. Ici entre en jeu l'empreinte parentale : pour se développer, l'albumen a besoin de facteurs de croissance, notamment la phytohormone auxine. Chez plusieurs plantes comme le riz et *Arabidopsis*, les gènes à empreinte paternelle incluent des gènes responsables de la synthèse d'auxine tels que les gènes *YUCCA* et *TAR* (Chen *et al.*, 2018). Autrement dit, le développement de l'albumen est ici soutenu par un apport paternel d'auxine *via* des gènes exprimés par l'allèle paternel, alors que l'allèle maternel est réprimé, empêchant la formation d'albumen sans génome paternel. Par manipulation génétique ou par application externe d'auxine, des chercheurs ont pu déclencher le développement de l'albumen sans fécondation, que l'on appelle albumen « autonome » (Figueiredo *et al.*, 2015). Cet albumen autonome finit cependant par dégénérer, probablement en raison de l'absence de gènes exprimés par le génome paternel.

Par ailleurs, comme illustré en figure 6.2, lorsqu'une plante diploïde est croisée avec une plante polyploïde, l'albumen produit induit des problèmes de développement aboutissant à la mort de la graine. Comme mentionné précédemment, ce phénotype dépend du sens du croisement et illustre que les contributions des chromosomes paternels et maternels doivent respecter le ratio de 2 maternel pour 1 paternel. Il s'agit d'un système clé-serrure apporté par l'empreinte parentale, avec les génomes maternel et paternel apportant leur contribution propre, en quantité propre, et de manière complémentaire et interdépendante. Si l'un est absent ou si l'apport de l'un est disproportionné par rapport à l'autre, l'albumen ne se développera pas proprement, ou pas du tout.

Cette hypothèse clé-serrure n'est cependant pas totalement satisfaisante pour expliquer un fait surprenant : dans deux systèmes *a priori* complètement différents, mammifères placentaires et plantes à fleurs, l'empreinte parentale est apparue de manière indépendante. Il pourrait donc s'agir d'un cas d'évolution convergente en réponse à la même pression sélective. Mais laquelle ? Pour répondre à cette question, il faut rappeler les quelques similarités des mammifères placentaires et des plantes à fleur :

- les deux systèmes sont vivipares, avec l'embryon se développant au sein de la mère jusqu'à maturité, avec une connexion mère-embryon pour le transfert de ressources vers l'embryon ;
- le génome paternel contribue au tissu nourricier, placenta/albumen ;
- une mère peut donner naissance à des descendants provenant de pères différents.

Ces trois éléments donnent la possibilité de l'émergence du conflit parental (voir chapitre 5). En bref, d'après cette théorie, la valeur sélective du génome paternel

sera maximale si ce génome peut influencer la quantité de ressources données à son embryon au détriment des embryons fécondés par les autres contributeurs paternels, amenant sa descendance à être de meilleure qualité comparée aux autres. En revanche, ce scénario avec des descendants moins favorisés que d'autres n'est pas bénéfique pour la femelle, car tous les descendants portent ses allèles. La *fitness* du génome maternel sera donc meilleure si un plus grand nombre de descendants ont une *fitness* suffisante pour survivre. On entre donc dans un conflit d'intérêts sélectifs entre mère et père, et le fait d'avoir des gènes exprimés uniquement du côté paternel ou uniquement du côté maternel (empreinte parentale) permet une influence spécifique de chaque parent sur les ressources données à l'embryon. Et le tissu nourricier, l'albumen, est le tissu le plus direct pour influencer la quantité de ressources attribuées à l'embryon (Haig et Westoby, 1989).

Cependant, contrairement aux mammifères, ce scénario n'est pas satisfaisant chez les plantes au regard de la fonction des gènes sous empreinte parentale. En effet, les fonctions de ces gènes n'ont pas de lien clair avec le métabolisme, et ils sont plutôt impliqués dans la régulation de la chromatine et dans la régulation des gènes en général. Bien entendu, il se pourrait que les gènes sous empreinte parentale régulent des gènes en aval qui, eux, sont impliqués dans l'accumulation de substances nutritives, et, bien que quelques études aillent dans ce sens, ce lien reste encore à démontrer clairement. En revanche, les croisements entre plantes de ploïdies différentes apportent un argument en faveur du conflit parental : si le donneur paternel a plus de chromosomes que la plante mère, l'albumen hybride aura donc plus de chromosomes venant du père, et le résultat est que l'albumen va croître de manière importante (avant de mourir). À l'inverse, quand la plante mère a une ploïdie plus importante que le donneur paternel, l'albumen hybride, qui a davantage de chromosomes venant de la plante mère, sera plus petit que la normale. Ceci suggère que les gènes de l'albumen exprimés par les chromosomes paternels favorisent la croissance de l'albumen et donc l'alimentation de l'embryon, alors que les gènes exprimés par les chromosomes maternels ont l'effet opposé. Ceci correspond à ce que l'on attend dans le cas d'un conflit sélectif où donneurs paternels sont sélectionnés pour accroître les ressources données à leurs propres embryons, alors que l'avantage sélectif de la plante mère est de limiter l'inégalité de distribution des ressources entre graines. Davantage de démonstrations formelles du conflit parental en tant que cause évolutive de l'apparition de l'empreinte parentale restent néanmoins nécessaires.

► Application pour l'amélioration des plantes

Il semble clair aujourd'hui que l'empreinte parentale est impliquée dans le développement de l'albumen et dans la mort de la graine. Peut-on utiliser ce savoir à des fins agronomiques ?

Concernant la production de graines hybrides viables, on sait que l'épigénétique, et notamment la méthylation de l'ADN, est impliquée dans la mort de la graine hybride. En manipulant la méthylation de l'ADN, il est possible en théorie d'améliorer la viabilité de ces graines. Ceci permettrait notamment de réduire le travail de sauvetage d'embryons, qui marche certes bien, mais qui requiert de longues heures de travail minutieux pour disséquer les embryons et les placer en milieu *in vitro*. Cependant, en Europe, à cause des régulations actuelles, il est impossible d'utiliser des mutants de gènes impliqués dans l'empreinte parentale. Mais il existe d'autres moyens que la

mutagenèse pour manipuler la méthylation de l'ADN, et l'un d'eux est l'utilisation de produits hypométhylants tels que la 5-azacytidine. Ce composé est un analogue de la cytidine, un des nucléosides de l'ADN. Si on l'applique sur une plante, la 5-azacytidine sera incorporée dans l'ADN de la plante. À la différence de la cytidine, elle causera un blocage et une dégradation des enzymes méthyltransférases. Le résultat sera une hypométhylation du génome de la plante. Mais peut-on obtenir des graines hybrides viables en appliquant de la 5-azacytidine sur des plantes avant de les croiser avec d'autres espèces? La réponse est oui, en tout cas pour *Arabidopsis* et la capselle (une espèce proche du chou), pour lesquelles une étude a montré que le taux de viabilité de graines hybrides entre espèces et entre ploïdies a augmenté après l'application de la 5-azacytidine (Huc *et al.*, 2022). Ces espèces de plantes sont assez petites, et donc l'absorption de 5-azacytidine se fait assez facilement. Il reste maintenant à savoir si d'autres plantes, plus grandes et notamment d'intérêt agronomique, montreraient la même réponse au traitement par la 5-azacytidine...

Dans un autre registre, pour l'amélioration des plantes et des céréales en particulier (blé, maïs, orge), un phénomène bien connu depuis le XIX^e siècle est l'effet *xenia* (Denney, 1992). La racine grecque de ce mot veut dire «étranger», car l'effet *xenia* décrit l'influence du donneur de pollen sur le développement de la graine fécondée. Cela a longtemps intrigué les chercheurs, car il était compliqué d'expliquer comment le pollen peut avoir un effet direct sur des tissus et des organes portés par la plante mère. Cela se traduit par un changement de couleur, de taille ou de qualité de la graine dépendant du donneur de pollen, et pas de la plante mère. Plusieurs dizaines de cas *xenia* différents ont été identifiés à ce jour. Bien sûr, certains de ces cas s'expliquent simplement par de la dominance. Mais parfois, si une plante est utilisée en donneur de pollen ou en plante mère, on n'obtient un effet sur la graine uniquement dans le cas où la plante est le donneur de pollen.

Cet effet *xenia* correspond une fois encore aux propriétés de l'empreinte parentale, et c'est pour cette raison qu'il a été proposé que certains des cas *xenia* puissent être expliqués par des gènes à empreinte paternelle. Pourtant, les gènes en cause restent mystérieux. Les approches QTL (*quantitative trait locus*) réalisées par les sélectionneurs pour identifier les régions du génome responsables de l'effet *xenia* ont défini des régions génomiques très vastes, qui contiennent plusieurs dizaines ou même centaines de gènes. Identifier ensuite le gène responsable est un travail de plusieurs années. En revanche, parce qu'on connaît désormais les gènes à empreinte paternelle dans plusieurs espèces cultivées (maïs, riz, orge ou blé) et que leur nombre est relativement restreint, la liste des gènes candidats contenus dans les QTL *xenia* peut être grandement réduite. Des chercheurs travaillent donc actuellement sur l'identification et la caractérisation de ces gènes. Certains ont un effet sur la qualité de la graine. Il est donc envisageable à l'avenir d'utiliser la sélection assistée par marqueurs basée sur le séquençage ou la mesure d'expression de ces gènes à empreinte paternelle pour jouer sur la qualité des graines.

► Références bibliographiques

- Adams S., Vinkenoog R., Spielman M., Dickinson H.G., Scott R.J., 2000. Parent-of-origin effects on seed development in *Arabidopsis thaliana* require DNA methylation. *Development*, 127, 2493-2502.
- Chen C. *et al.*, 2018. Characterization of imprinted genes in rice reveals conservation of regulation and imprinting with other plant species. *Plant Physiology*, 177, 1754.

- Denney J.O., 1992. Xenia Includes Metaxenia. *HortScience*, 27, 722-728.
- Figueiredo D.D., Batista R.A., Roszak P.J., Köhler C., 2015. Auxin production couples endosperm development to fertilization. *Nature Plants*, 1, 1-6.
- Gehring M., Bubb K.L., Henikoff S., 2009. Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science*, 324, 1447-1451.
- Haig D., Westoby M., 1989. Parent-specific gene-expression and the triploid endosperm. *American Naturalist*. <https://doi.org/10.1086/284971>
- Huc J. *et al.*, 2022. Bypassing reproductive barriers in hybrid seeds using chemically induced epimutagenesis. *The Plant Cell*, 34, 989-1001.
- Ibarra C.A. *et al.*, 2012. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science*, 337, 1360-1364.
- Martínez G., Panda K., Köhler C., Slotkin R.K., 2016. Silencing in sperm cells is directed by RNA movement from the surrounding nurse cell. *Nat Plants*, 2, 16030.
- Raissig M.T., Bemer M., Baroux C., Grossniklaus U., 2013. Genomic imprinting in the *Arabidopsis* embryo is partly regulated by PRC2. *PLoS Genet*, 9, e1003862.
- Städler T., Florez-Rueda A.M., Roth M., 2021. A revival of effective ploidy: The asymmetry of parental roles in endosperm-based hybridization barriers. *Current Opinion in Plant Biology*, 61, 102015.
- Tirot L., Jullien P.E., 2022. Epigenetic dynamics during sexual reproduction: At the nexus of developmental control and genomic integrity. *Current Opinion in Plant Biology*, 69, 102278.
- Wolff P., Jiang H., Wang G., Santos-González J., Köhler C., 2015. Paternally expressed imprinted genes establish postzygotic hybridization barriers in *Arabidopsis thaliana*. *eLife*, 4, e10074.

Chapitre 7

La vie *in utero* : programmation épigénétique des caractères

ANNE GABORY

Les maladies chroniques, ou maladies non transmissibles, sont des affections de longue durée qui résultent de la combinaison de facteurs génétiques, physiologiques, environnementaux et de modes de vie. D'après l'Organisation mondiale de la santé (OMS, 2022), elles sont à l'origine de 41 millions de morts par an, soit 74% de la mortalité globale de la population humaine. Les principales maladies chroniques sont, dans l'ordre d'importance, les pathologies cardiovasculaires (responsables de crises cardiaques et d'accidents vasculaires cérébraux), les cancers, les maladies respiratoires chroniques (comme la maladie pulmonaire obstructive chronique ou l'asthme chez l'humain) et les diabètes. Les facteurs génétiques entrent en jeu dans le développement des maladies chroniques, mais s'ils peuvent expliquer la maladie pour un individu donné, ils ne sont pas pertinents pour justifier l'augmentation rapide de leur fréquence dans l'ensemble de la population. Celle-ci est plutôt déterminée par l'environnement ou le style de vie des individus : tabac, sédentarité, nutrition trop riche, alcool, constituent les principaux facteurs de risque. Cependant, l'ensemble de ces facteurs ne suffit pas à justifier la croissance très rapide de l'incidence mondiale des pathologies chroniques.

► Les origines développementales des maladies

Dès les années 1960, des expériences chez le rat ont montré que l'environnement nutritionnel de la mère pouvait influencer l'apparition de pathologies des descendants devenus adultes. Dans les années 1980, les travaux de David J. Barker et C. Nicholas Hales ont permis de poser l'hypothèse des origines fœtales des maladies, qui deviendra par la suite le concept plus élargi des « origines développementales de la santé et des maladies » (*developmental origins of health and disease*, DOHaD) (Hales *et al.*, 1991). Cette hypothèse, aujourd'hui bien étayée expérimentalement, postule que l'environnement au cours du développement (la vie *in utero*) peut avoir des conséquences importantes pour la santé de l'individu au cours de sa vie adulte, conduisant aux maladies chroniques. La formulation du concept des « 1 000 premiers jours » a été récemment utilisée pour parler des effets d'une période qui court de la grossesse aux 2 ans révolus de l'enfant. Le concept de DOHaD va en fait au-delà

de cette période, puisqu'il englobe les périodes de la gamétogenèse (donc avant la conception) et du développement embryonnaire et fœtal, et la période postnatale de l'enfance, qui s'étend généralement chez l'humain au-delà des deux premières années de vie (Charles *et al.*, 2016).

L'environnement précoce peut être altéré selon différents paramètres : exposition à des toxiques (alcool, tabac, pollution, produits chimiques, etc.), alimentation (sous-nutrition/famine, carences en micronutriments, régime pléthorique de type occidental, etc.), stress (stress psychosocial, maltraitance, adversité, etc.), activité physique (en particulier sédentarité), etc. L'ensemble des expositions environnementales, précoces et tout au long de la vie de l'individu, est appelé « exposome ». Les recherches initiales sur les origines fœtales étaient concentrées sur l'environnement de la mère, du fait de la relation *in utero* et de l'importance des soins maternels. Ce chapitre sur la vie *in utero* sera consacré à l'environnement maternel mais il est aujourd'hui clair que l'environnement du père a également une influence importante (encadré 7.1).

Encadré 7.1. Transmission paternelle de l'information environnementale

Les premières données, suggérant la transmission intergénérationnelle d'effets environnementaux, basée sur des informations épigénétiques, concerne l'environnement maternel. En effet, les facteurs maternels présents dans l'ovocyte (protéines, petits et longs ARN dans le cytoplasme) au moment de la fécondation, la vie *in utero* et la relation transplacentaire puis l'importance des soins maternels aux petits, sont documentés. Dans les années 1990, côté paternel, le spermatozoïde n'était censé transmettre que son noyau et de l'information épigénétique restreinte à l'empreinte parentale. Aujourd'hui, il est clairement établi que le père transmet bien plus que le seul patrimoine génétique du spermatozoïde.

Au début des années 2010, des études chez la souris, révèlent que le statut métabolique des pères, soit au cours de leur vie *in utero* (Radford *et al.*, 2014), soit de leur vie adulte (Carone *et al.*, 2010), a des conséquences sur le phénotype de leur descendance, suggérant un passage d'information non génétique via le sperme. Cette transmission de phénotypes métaboliques du père aux descendants est aussi observée chez l'homme (Chen *et al.*, 2012). Ces études ont suspecté un mécanisme épigénétique sous-jacent sans pouvoir le démontrer formellement. En effet, le statut métabolique défavorable au cours de la vie *in utero* induit une perte de méthylation sur une centaine de régions de l'ADN des spermatozoïdes. Devenus adultes, la fertilité de ces animaux est affectée et se traduit par une réduction du développement embryonnaire. Mais ces modifications de méthylation ne sont pas transmises à la descendance, qui voyait pourtant l'expression des gènes de ces régions modifiée (Radford *et al.*, 2014). De nombreuses études ont montré depuis que l'information est en partie portée par les petits ARN non codants. Ainsi, sous régime alimentaire standard, les descendants de pères nourris avec un régime occidental (gras et sucré) présentent une augmentation de poids corporel, de glycémie à jeun et une moins bonne tolérance au glucose, caractérisant une prédisposition à l'obésité et au diabète. Les animaux issus d'un embryon tout juste fécondé dans lequel sont micro-injectés des petits ARN non codants prélevés dans le sperme des pères montrent la même prédisposition (Grandjean *et al.*, 2015). De la même manière, une restriction protéique paternelle augmente la quantité d'un type de petits ARN non codants (appelés tRF pour fragment d'ARN de transfert) dans les spermatozoïdes, ce qui module l'activité transcriptionnelle et

le développement précoce de l'embryon. Ces petits ARN non codants sont produits par les cellules épithéliales épидидymaires, ils sont présents dans les vésicules extracellulaires et transférés aux spermatozoïdes en maturation en transit dans l'épididyme (Sharma *et al.*, 2016). Une autre partie de l'information est portée par la chromatine. Bien que la plupart des histones soient remplacées par des protamines au cours de la maturation du spermatozoïde, certaines sont retenues et portent des modifications post-traductionnelles qui peuvent être transmises à l'embryon. Une étude récente a en effet montré que la restriction paternelle en folate induit une modification de la distribution de la marque H3K4me3 dans le sperme. Une partie de ces marques, transmise à la descendance, change l'expression dès l'activation du génome embryonnaire avant l'implantation, entraîne des modifications du développement et a des conséquences jusqu'à l'âge adulte (Lisner *et al.*, 2021). L'ensemble des résultats décrits ci-dessus ont été obtenus dans des modèles murins mais des modifications similaires du sperme sont également retrouvées chez l'humain et pourraient de la même manière, transmettre l'information (Lisner et Kimmins, 2023).

En élevage, et plus particulièrement pour les bovins, la sélection génétique s'appuie sur l'utilisation de la semence des mâles sélectionnés. Dans cette espèce comme pour les espèces modèles ou les humains, l'environnement (*vie in utero*, post-natale, péri-pubertaire, adulte) a des conséquences sur le phénotype des reproducteurs, les capacités reproductrices pouvant être modifiées et avoir des effets sur la génération suivante. Il est possible de prédire la fertilité des taureaux par une analyse de positions génomiques différenciellement méthylées (Costes *et al.*, 2022). Ainsi, en élevage, la combinaison des informations génétique et épigénétique pourrait orienter la sélection vers des animaux plus adaptés ou adaptables aux variations environnementales tout en préservant la production de produits de qualité (viande, lait, etc.), nécessaire à la sécurité alimentaire.

Pour conclure, en santé humaine comme en santé animale, un nouvel aspect à prendre en considération est l'environnement du père. Le déchiffrement des mécanismes en jeu, comme pour la transmission maternelle, est encore à l'étude. L'épigénétique semble être un lien clé entre les générations.

Les effets d'environnement maternels sur la santé et la prédisposition aux pathologies chroniques de la descendance dépendent de :

- la courbe de croissance fœtale, en fonction du développement du placenta et de ses capacités de transfert de nutriments vers le fœtus ;
- l'adaptation des tissus fœtaux à cet environnement maternel.

Afin d'étudier les mécanismes associés à la DOHaD, de nombreux modèles animaux, principalement des rongeurs, mais aussi des lapins, des moutons ou encore des primates non humains (macaques, singes rhésus, etc.), ont été développés. Les mères sont placées soit dans un environnement de restriction (réduction de l'apport calorique et de la proportion de protéines dans le régime alimentaire), soit dans un environnement nutritionnel riche (régime hyperlipidique, riche en lipides et en sucre de type occidental, obésité ou diabète induits chez la mère, etc.). La figure 7.1 présente les observations faites sur les fœtus dans ces modèles animaux, en cas de modification de la balance nutritionnelle de la mère. Le développement et la fonction placentaire sont modulés par l'environnement maternel. Il a été proposé que celui-ci peut jouer en quelque sorte un effet tampon, c'est-à-dire absorber les effets

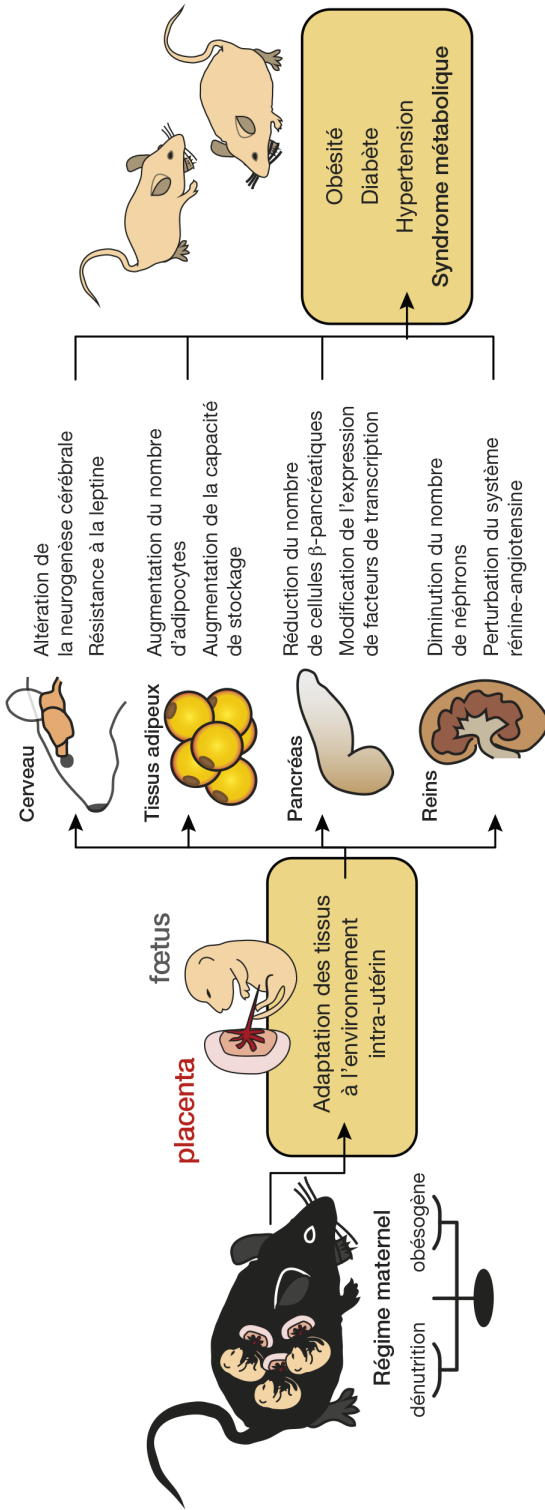


Figure 7.1. Mécanismes de DOHaD : adaptations des tissus en développement à l'environnement intra-utérin.

Lorsque l'alimentation maternelle présente un déséquilibre, le développement du placenta peut être altéré et ainsi sa capacité de transférer des nutriments au fœtus. Cela modifiera la courbe de croissance fœtale, marqueur de la prédisposition d'un individu aux pathologies chroniques. La structure des tissus en développement du fœtus peut également être affectée, par exemple la répartition des types de neurones cérébraux, l'augmentation du nombre d'adipocytes, la réduction du nombre de cellules β-pancréatiques ou de néphrons. Le fonctionnement fin des tissus peut aussi être modulé, entraînant une résistance à la leptine, hormone de la satiété, une capacité accrue de stockage des lipides par le tissu adipeux, une perturbation fine du fonctionnement pancréatique ou de la fonction rénale (activité rénine-angiotensine). L'ensemble de ces petites perturbations prédispose l'individu devenu adulte à l'hyperphagie, au stockage des lipides, à la résistance au glucose et à l'insuline, à l'altération de la fonction cardiovasculaire et donc aux trois composantes du syndrome métabolique que sont l'obésité, le diabète et l'hypertension (Adapté de Panchenko *et al.*, Biologie aujourd'hui, 2015).

de l'environnement. Les nutriments transférés du sang maternel vers le sang fœtal dépendent de cette adaptation placentaire. Le fœtus, à son tour, va s'adapter à cet environnement métabolique, à la présence ou à l'absence de nutriments particuliers. Ainsi, le développement de différents organes, tels que le cerveau, le tissu adipeux, le pancréas et les reins, est modifié. Cela conduit à long terme à une altération fine de la fonction de ces organes. Dans l'exemple de la figure 7.1, les descendants sont prédisposés à l'obésité, au diabète et à l'hypertension, trois pathologies associées dans le syndrome métabolique.

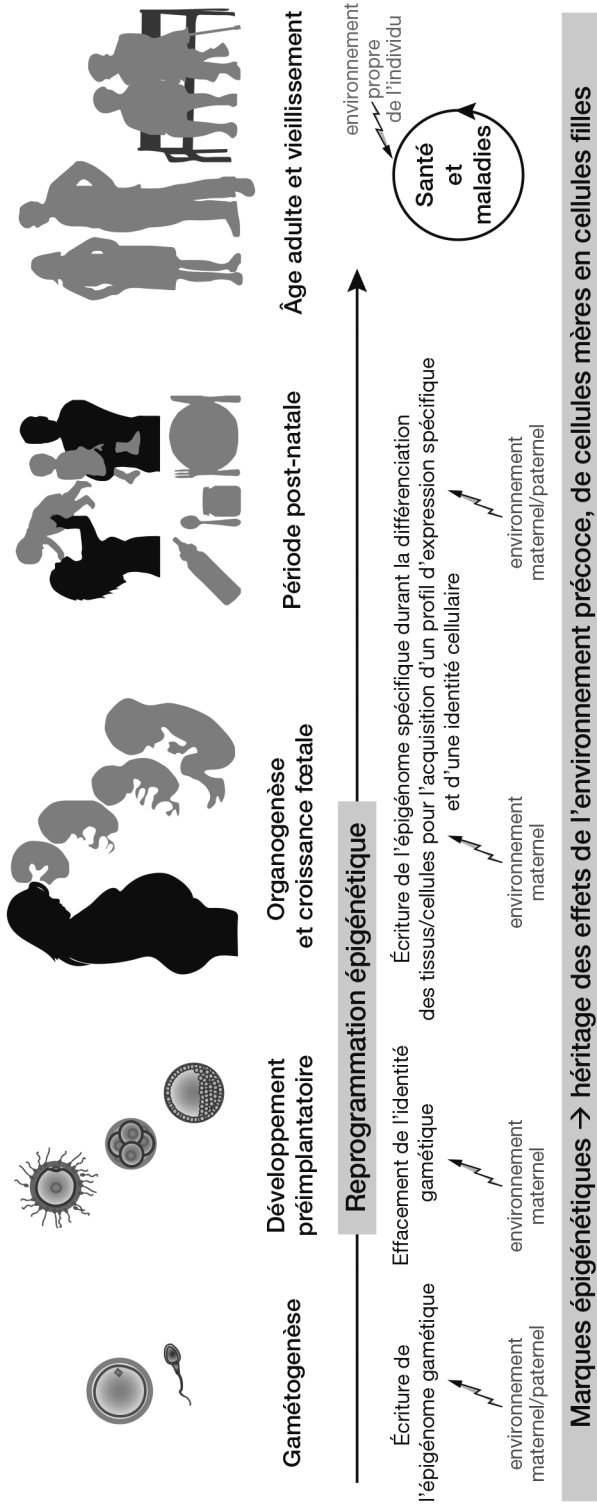
► Épigénétique et DOHaD

Des mécanismes épigénétiques considérés comme fondamentaux dans la DOHaD

L'implication des mécanismes épigénétiques dans la DOHaD s'est rapidement imposée comme étant un élément crucial. Les marques épigénétiques sont en effet recopiées de cellules mères en cellules filles au cours des divisions cellulaires, expliquant leur capacité à « mémoriser » à long terme des environnements changeants (Junien *et al.*, 2016). De fait, l'épigénome subit deux phases majeures de reprogrammation : au cours de la gamétogenèse et pendant le développement (voir chapitres 2, 3 et 4). Ces phases d'effacement et de réécriture permettent d'écrire un épigénome spécifique à chaque type cellulaire sélectionnant les gènes qui doivent fonctionner (profil d'expression génique), et donc de donner l'identité et la fonction propre de chaque cellule, dans un tissu au sein d'un organe (figure 7.2).

Ainsi, au cours de la gamétogenèse, on constate une apposition, *via* l'épigénome, d'une identité gamétique d'ovocyte ou de spermatozoïde. L'effacement de l'épigénome gamétique juste après la fécondation laisse la place à l'apposition de nouvelles marques, concomitantes à la différenciation des premiers lignages cellulaires (trophectoderme et cellules de la masse interne au stade blastocyste), puis des lignages embryonnaires (ectoderme, endoderme et mésoderme), puis des 200 différents types cellulaires composant un individu. La croissance fœtale, la relation materno-fœtale et les échanges transplacentaires sont ensuite des éléments clés du développement et de l'organogenèse. Au cours de la période postnatale, trois systèmes majeurs de l'organisme subissent une maturation importante : le système digestif, le système nerveux central et le système immunitaire. L'environnement nutritionnel change de façon brusque, avec notamment des transitions alimentaires : de l'alimentation *via* le cordon ombilical à l'alimentation orale (allaitement ou biberon), puis de l'alimentation lactée vers une alimentation solide prise de façon autonome. Les soins parentaux sont alors extrêmement importants. Cette période est également marquée par un développement cognitif considérable. Pendant ces périodes fœtales et postnatales, au sein des organes, les cellules acquièrent un destin cellulaire, puis se différencient ; la mémoire de la décision de cette identité cellulaire est ensuite gardée sur le long terme : les mécanismes épigénétiques sont au centre de ce processus.

Si l'environnement paternel ou maternel peut modifier l'effacement ou l'apposition des marques épigénétiques à ces stades de plasticité, alors une mémoire de cet environnement précoce pourrait être conservée, influencer le fonctionnement des organes à long terme et donc conditionner la balance santé-pathologie.



Marques épigénétiques → héritage des effets de l'environnement précoce, de cellules mères en cellules filles

Figure 7.2. Le développement embryonnaire et fœtal : une période de plasticité épigénétique pouvant expliquer la mémoire à long-terme en DOHaD. Le développement commence par la gamétogenèse, qui permet la formation des gamètes, spermatozoïdes ou ovocytes, qui donneront le futur individu. Au cours de cette phase, les marques épigénétiques sont effacées puis les épigénomes spécifiques du spermatozoïde et de l'ovocyte sont apposés. Après fécondation, le développement préimplantatoire est marqué par la mise en place des premiers lignages embryonnaire et extra-embryonnaire. L'identité gamétique est alors effacée afin de permettre à l'embryon de produire l'ensemble des différents types cellulaires. La phase d'organogenèse et de croissance fœtale est marquée par la différenciation de l'ensemble des tissus, et des différents types cellulaires qui les composent. La maturation de plusieurs systèmes essentiels de notre organisme a lieu au cours de la vie post-natale, où l'environnement change brutalement et les soins parentaux sont cruciaux. L'écriture de l'épigénome permet de sélectionner l'information à exprimer dans chaque cellule afin d'en assurer le fonctionnement spécifique. Les environnements du père et de la mère peuvent intervenir dans l'effacement ou l'écriture de l'épigénome à ces stades, modifier la différenciation des tissus, le profil d'expression à long terme et le fonctionnement des organes et le développement de pathologies. (Adapté de Safi-Stibler et al., Seminars in Cell and Developmental Biology, 2020).

Les liens entre environnements et régulation épigénétique du génome

Les liens entre épigénétique et environnement sont maintenant clairement démontrés *via* plusieurs voies distinctes.

- La signalisation cellulaire : l'exposition à des toxiques, la nutrition ou l'activité physique peuvent influencer notre épigénome, suite à l'activation de voies de signalisation cellulaire, qui correspond au système de communication entre cellules (direct ou à courte/longue distance) afin de coordonner leur activité. Les récepteurs membranaires ou nucléaires sont des protéines qui peuvent lier une molécule, appelée « ligand », et induire des cascades de signalisation au sein desquelles les enzymes de la machinerie épigénétique vont être impliquées. Par exemple, la fixation d'un ligand sur un récepteur membranaire peut, entre autres, modifier l'activité de protéines (enzymes) intervenant dans l'écriture, la lecture ou l'effacement des marques épigénétiques (voir chapitre 2), et donc augmenter ou réduire les marques épigénétiques de leurs gènes cibles. Dans le cas des récepteurs nucléaires, leur fixation sur l'ADN au niveau de gènes cibles entraînera non seulement l'expression ou la répression du gène, mais également le recrutement de la machinerie de remodelage de la chromatine, qui placera ce gène dans un espace ouvert ou fermé du noyau, avec la possibilité de mémoriser cet état dans les générations successives de divisions cellulaires (Gadaleta et Magnani, 2014).
- La transduction mécanique du signal entre les différents organites de la cellule : sur la membrane externe de la cellule, la matrice extracellulaire est en contact avec la membrane plasmique, elle-même en contact avec le cytosquelette, lui-même en contact avec la membrane du noyau cellulaire qui est en interaction avec la chromatine. Une variation de l'une de ces structures (par exemple, un régime enrichi en oméga-3 renforcera les membranes plasmiques dans ce type de lipides et modifiera leur fluidité) peut affecter le signal mécanique envoyé vers la chromatine et de ce fait moduler l'expression des gènes (Safi-Stibler et Gabory, 2020).
- Les métabolites de l'alimentation : ils peuvent avoir une action directe sur le fonctionnement des enzymes de la machinerie épigénétique (Lempradl *et al.*, 2015). Cela a, par exemple, été décrit pour le sulforaphane présent dans le brocoli, ou pour l'allyl-mercaptane contenu dans l'ail, deux métabolites inhibiteurs des histones désacétylases, ou encore pour la génistéine que l'on trouve dans le soja, inhibiteur des ADN méthyltransférases (DNMT). Enfin, comme décrit dans le chapitre 2, le métabolisme énergétique et l'épigénétique sont tout à fait interconnectés. Plusieurs de ces réactions comportent des molécules donneuses de groupements chimiques pour l'apposition des marques épigénétiques sur notre génome. Ainsi, un des substrats majeurs du cycle de Krebs, l'acétyl-coenzyme A (acétyl-CoA), est le donneur universel de groupement acétyle pour l'acétylation des histones (Figure 2.3). Par ailleurs, la S-adénosylméthionine (SAM), membre du cycle du métabolisme mono-carboné, est le donneur universel de groupement méthyl pour la méthylation de l'ADN et des histones (Figure 2.5). Ce cycle est lui-même régulé par la vitamine B9 (folates).

Les preuves effectives de l'implication des mécanismes épigénétiques dans la DOHaD

Si les relations environnement-épigénome existent pour un individu, il en est de même pour l'environnement parental, qui agit sur l'épigénome de l'individu en devenir *in utero*.

Les évidences scientifiques selon lesquelles les mécanismes épigénétiques sont impliqués dans ce conditionnement ont été apportées par l'étude de modèles animaux. Ainsi, un article a été pionnier dans la discipline de la DOHaD et de l'épigénétique (Weaver *et al.*, 2004). Il a montré que les soins maternels entraînent l'activation, dans le cerveau des rats nouveau-nés, du récepteur à la sérotonine. Il en résulte une cascade de signalisation dont un des aboutissements est la fixation de l'histone acétyltransférase CBP (*CREB binding protein*) au niveau du gène du récepteur aux glucocorticoïdes. Localement, sont alors induites l'acétylation des histones et la déméthylation de l'ADN, augmentant ainsi l'expression du gène. Selon la qualité et la quantité des soins prodigués par la mère pendant la période de lactation, le gène du récepteur aux glucocorticoïdes est donc plus ou moins exprimé dans le cerveau des rats devenus adultes, la présence de récepteur modulant leur réponse au stress de façon pérenne (Weaver *et al.*, 2004). Une autre étude a montré qu'enrichir le régime de mères souris en donneurs de méthyl (choline, bêtaïne et folates) augmente le nombre de descendants présentant une forte méthylation de l'ADN au niveau d'un gène contrôlant la couleur du poil, modifiant ainsi l'aspect du pelage. Cet article était un des premiers à démontrer formellement que l'alimentation de la mère, pendant la gestation et la lactation, pouvait avoir un effet ferme et définitif sur un phénotype de la descendance, *via* des mécanismes épigénétiques (Waterland et Jirtle, 2003).

Chez l'humain, ce type d'études mécanistiques qui démontrent des liens de cause à effet n'est pas possible : il n'est pas envisageable de réaliser des biopsies répétées de tissu hépatique tout au long de la vie ou des prélèvements de cerveau afin de faire des études de biologie moléculaire. En revanche, les études épidémiologiques permettent de montrer des corrélations entre exposome précoce, épigénome et pathologie de l'adulte : il s'agit principalement d'études d'association entre la méthylation de l'ADN dans le sang circulant et l'environnement prénatal. Des stratégies biostatistiques, dites « de médiation », permettent par ailleurs d'estimer la part de l'effet portée par la modification de la méthylation au niveau d'un gène dans le phénotype. Des analyses bio-informatiques apportent également des informations sur les fonctions biologiques associées aux gènes différentiellement méthylés. Ainsi, il a été montré que l'exposition *in utero* à la famine est associée, à l'âge adulte, à une augmentation du risque de pathologies mentales et métaboliques, ainsi qu'à une modification de la méthylation de l'ADN pour un ensemble de régions génomiques. Dans le cas de la famine de 1945 en Hollande, il a été montré que, près de soixante ans plus tard, la méthylation de 180 régions est modifiée suite à l'exposition prénatale et que les gènes de ces régions sont impliqués dans la croissance et le métabolisme (Tobi *et al.*, 2014). De même, les liens entre d'autres types d'environnements (pollution de l'air ou stress aigu, notamment), épigénome et pathologies chroniques de l'adulte apparaissent dans un nombre croissant de publications scientifiques.

Il est à noter que tous les exemples décrits ici sont une transmission intergénérationnelle : de l'environnement des parents sur le phénotype des enfants. S'il existe des évidences d'une transmission transgénérationnelle (c'est-à-dire sur les générations suivantes, en l'absence d'une exposition directe de l'individu) chez les plantes, les levures, les insectes (*Drosophila*) ou les nématodes, avec des mécanismes épigénétiques sous-jacents, elle est longtemps restée controversée chez les mammifères (Fitz-James et Cavalli, 2022; Heard et Martienssen, 2014). Plus récemment chez les mammifères modèles, des travaux ont recherché une transmission transgénérationnelle des

effets de l'environnement sur souris, rat et homme¹³. Il pourrait s'agir d'un effet de l'environnement sur la séquence génique comme une mutation de l'ADN, et donc une héritabilité génétique difficile à mettre en évidence.

Des mécanismes épigénétiques seuls en jeu ?

La relation [environnement prénatal → modifications épigénétiques → fonction fine des tissus/organes → balance santé/maladie de l'adulte] est donc aujourd'hui admise.

Est-ce à dire que l'intégralité des mécanismes liant l'environnement précoce aux pathologies de l'adulte est de nature épigénétique? Non, cela reviendrait à faire la même erreur qu'à l'ère du « tout génétique ». Ainsi l'environnement métabolique et nutritionnel maternel, par exemple, peut influencer sur la composition des membranes cellulaires et la multiplication des mitochondries au sein des cellules. La composition du microbiote transmis par la mère à son enfant à la naissance, pendant l'allaitement et dans l'environnement de la petite enfance, est aussi un élément primordial. Nous pouvons penser de même à l'immunité acquise par le bébé, *via* les anticorps et autres facteurs qui passent dans le lait maternel, et qui peut dépendre de l'état inflammatoire de la mère. Tous ces éléments ont un rôle en soi dans les origines développementales de la santé.

Cependant, de la même façon que nous avons précédemment vu que le métabolisme énergétique et l'épigénétique sont intimement interreliés, il en est de même pour ces éléments. Un enrichissement en oméga-3 du régime maternel aura pour effet d'enrichir les membranes plasmiques du cerveau, du cœur ou encore du foie des descendants en oméga-3, de façon pérenne jusqu'à l'âge adulte. Cet enrichissement a des effets physiologiques en lui-même, mais également des effets sur la structure de la chromatine, *via* une modification de la transduction mécanique du signal, modifiant l'expression génique.

Les mitochondries, quant à elles, sont des organites qui jouent le rôle de « centrales énergétiques » des cellules, en permettant la production d'ATP, et ont également un rôle important dans la division ou la mort cellulaire, ou encore dans le métabolisme des lipides et la signalisation cellulaire. Différentes études ont montré que, selon l'environnement maternel, la quantité d'ADN mitochondrial (ADNmt) dans les tissus des descendants était altérée. Par exemple, chez l'humain, le stress psychosocial est associé à une réduction de l'ADNmt dans le sang du cordon des bébés dans une première étude, ou encore la quantité d'ADNmt est inversement corrélée au poids de naissance dans une autre étude. Chez le rat, une restriction calorique de la mère pendant la période fœtale est associée à une restriction de la croissance fœtale et à une augmentation de l'ADNmt, mais aussi à une réduction de la production d'ATP dans le placenta, signant une dysfonction de cet organite. Par ailleurs, le lien entre le microbiote de la mère et celui de son bébé est désormais établi. Chez les modèles de rongeur ainsi que chez l'humain, la surnutrition maternelle est associée à un microbiote intestinal favorisant le développement de l'obésité chez l'enfant. Ce microbiote peut avoir des effets variés, et les mécanismes par lesquels il dialogue et influence le fonctionnement de l'organisme hôte sont encore à l'étude. Une partie de

13. On peut noter les travaux du groupe d'Isabelle Mansuy, de M.K. Skinner, d'A.C. Ferguson-Smith, etc. Voir colloque *Epigenetic Inheritance*, août 2023.

ces effets n'est ni de nature génétique, ni de nature épigénétique. Cependant, aussi bien les mitochondries que les bactéries qui peuplent notre organisme sécrètent un certain nombre de métabolites, parmi lesquels le folate, le butyrate, la biotine, l'acétate ou l'acétyl-CoA, qui peuvent avoir un effet sur la méthylation de l'ADN ou sur les modifications post-traductionnelle des histones.

Ces exemples indiquent bien que les mécanismes de la DOHaD ne sont pas tous de nature épigénétique, mais que les mécanismes épigénétiques restent parmi les meilleurs candidats pour expliquer ce phénomène.

► Épigénétique et variation de réponse selon le sexe

Les pathologies chroniques dont il s'agit dans la DOHaD présentent presque toujours des différences entre sexes. Pourtant, les programmes de recherche incluant des modèles animaux dans le contexte de la DOHaD n'étudient que depuis récemment non seulement les mâles mais aussi les femelles. Quand les analyses effectuées sur la descendance sont faites dans les deux sexes, des différences significatives entre mâles et femelles sont presque toujours observées au niveau de la variété des signes cliniques, de leur sévérité ou de leur cinétique d'apparition (Dearden *et al.*, 2018). Cependant, très peu d'études abordent la compréhension des mécanismes sous-jacents.

Les différences phénotypiques entre sexes sont souvent considérées comme une conséquence de l'action des hormones stéroïdes gonadiques. Cependant, des différences entre sexes sont observées avant même la différenciation sexuelle. Par exemple, les blastocystes possédant les chromosomes XY se développent plus rapidement que les blastocystes possédant les chromosomes XX. De plus, au début du développement, avant même l'implantation dans l'utérus, l'expression différentielle des gènes entre embryons XX et XY est déjà présente. En utilisant un modèle murin génétiquement modifié assez complexe, le *Four Core Genotype* (Quatre combinaisons génotypiques, figure 7.3), il est possible de déterminer la part des hormones et du génotype XX/XY dans les différences phénotypiques entre sexes (Arnold, 2020). Cet élégant modèle a été utilisé pour suivre les réponses métaboliques à un régime riche en graisses. Il a été montré qu'elles sont en partie liées aux chromosomes sexuels : la présence de 2 chromosomes X entraîne un gain de poids plus important et une adiposité accrue, indépendamment du sexe gonadique (figure 7.3).

Il existe un petit nombre de gènes, localisés soit sur le chromosome X soit sur l'Y, qui présentent une forte homologie de séquence, sans toutefois coder la même protéine. Ces paires de gènes sont généralement impliquées dans les processus de régulation de l'expression génique. Parmi eux se trouve le gène *KDM5C* sur le chromosome X, qui déméthyle les histones et est donc un régulateur épigénétique. De façon intéressante, les souris femelles portant une délétion du gène *KDM5C* sur l'un de leurs 2 X ont un gain de poids sous régime obésogène plus faible que les souris possédant les 2 copies de *KDM5C*. Ce phénotype est similaire aux souris femelles XY par rapport aux souris femelles XX dans le modèle des Quatre combinaisons (figure 7.3). Ainsi, ce gène porté par le chromosome X participe aux différences de métabolisme liées au sexe, probablement *via* son activité sur la méthylation des histones.

Nous pouvons donc émettre l'hypothèse que si ces enzymes épigénétiques, dont les gènes sont portés par les chromosomes sexuels, participent aux différences de

Modèle des Quatre combinaisons génotypiques

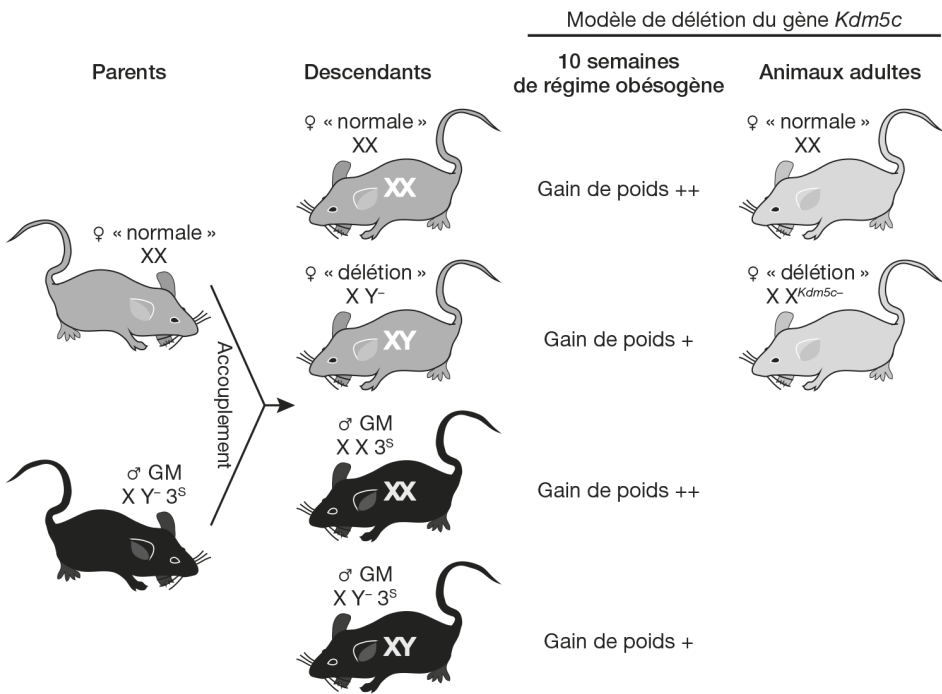


Figure 7.3. Les modèles murins des Quatre combinaisons génotypiques et de délétion du gène *Kdm5c*.

Le modèle implique 2 mutations chez la souris : la mutation Y⁻, délétion (perte du gène) naturelle du gène *Sry* (qui permet à l'embryon de devenir un mâle chez la souris), et la modification génétique 3^S (le gène *Sry* a été inséré sur le chromosome 3). Un croisement entre une femelle normale et un mâle GM (génétiquement modifié), dont le chromosome Y est (Y⁻) et portant la modification 3^S, permet d'obtenir 4 combinaisons génotype-sexe gonadique différentes : les femelles XX, les femelles XY, les mâles XX et les mâles XY. Il est ainsi possible de distinguer les effets dus au sexe (femelles ♀ vs mâles ♂) et ceux dus aux chromosomes sexuels (XX vs XY). Dix semaines de régime obésogène entraînent un gain de poids (pourcentage par rapport au poids initial) similaire entre femelles XX et mâles XX. Ce gain de poids est plus important chez ces individus XX que chez les XY, qu'ils soient femelles ou mâles. Deux copies de l'X sont associées à un gain de poids supérieur à une seule copie. Dans un second modèle, le gène *Kdm5c*, porté par le chromosome X et qui code une déméthylase des histones, a été délété sur un des 2 chromosomes X (indiqué comme X^{Kdm5c-}). L'effet du régime obésogène est similaire au premier modèle : deux copies du gène sont associées à un gain de poids supérieur à une seule copie.

réponses à l'environnement de l'individu entre sexes, il peut en être de même *in utero*. Ainsi, une partie des différences entre sexes présente dans le concept de DOHaD serait portée par les différences de croissance foetale et les hormones, mais une partie serait également portée par des mécanismes épigénétiques.

►► **Conclusion : les origines développementales de la santé**

Le fait que les mécanismes épigénétiques soient impliqués dans la DOHaD n'a jamais vraiment été une question, car la notion même de « mémoire » est intrinsèque à la définition de l'épigénétique. La remarquable plasticité des marques épigénétiques au cours

du développement, qui subissent deux étapes d'effacement/apposition, de la gaméto-genèse à la maturation finale des cellules et des tissus, en fait une interface idéale d'enregistrement de l'environnement précoce. De nombreuses publications rapportent des associations entre les changements épigénétiques, spécifiques d'une ou plusieurs séquences ou pangénomiques, des tissus fœtaux ou adultes et l'environnement maternel ou paternel (Safi-Stibler et Gabory, 2020). Trop peu d'études cependant se positionnent sur une meilleure compréhension de la chaîne causale précise des événements.

Les efforts portés aujourd'hui sur la recherche des mécanismes épigénétiques et la DOHaD évoluent vers la possibilité de réversibilité, avec une intervention proposée à l'individu lui-même. L'espoir est de remodeler positivement les marques épigénétiques et d'inverser le phénotype. Cela peut prendre la forme de conseils sur les modes de vie ou d'«épi-médicaments», mais cela ne permettra pas de guérir toutes les maladies chroniques. Une solution plus rentable pourrait être de prendre soin des parents à risque dans la fenêtre de temps adéquate, afin d'éviter l'établissement d'un épigénotype délétère chez la descendance, empêchant ainsi le développement d'une pathologie. Le concept de DOHaD s'intéresse en effet bien aux maladies, mais aussi à la santé de l'adulte (Hanson et Gluckman, 2014; Rosenfeld, 2017), et il est à espérer que certains aliments ou comportements précoces «sains» que les parents fourniront à leurs enfants leur permettront d'entamer leur vie sur un épigénotype «sain», contribuant ainsi à briser le cercle vicieux de la pandémie des maladies chroniques.

► Références bibliographiques

- Arnold A.P., 2020. Four core genotypes and XY* mouse models: Update on impact on SABV research. *Neuroscience and Biobehavioral Reviews*, 1191-8.
- Carone B.R., Fauquier L., Habib N., Shea J.M., Hart C.E., Li R., *et al.*, 2010. Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals, *Cell*, 143(7), 1084-1096.
- Charles M.-A., Delpierre C., Bréant B., 2016. Le concept des origines développementales de la santé. Évolution sur trois décennies. *Médecine/Science*, 32 (1), 15-20.
- Chen L., Magliano D.J., Zimmet P.Z., 2012. The worldwide epidemiology of type 2 diabetes mellitus—present and future perspectives, *Nature Reviews Endocrinology*, 8(4), 228-236.
- Costes V., Chaulot-Talmon A., Sellem E., Perrier J.-P., Aubert-Frambourg A., Jouneau L., *et al.*, 2022. Predicting male fertility from the sperm methylome: application to 120 bulls with hundreds of artificial insemination records, *Clinical Epigenetics*, 14(1), 54.
- Dearden L., Bouret S.G., Ozanne S.E., 2018. Sex and gender differences in developmental programming of metabolism. *Molecular Metabolism*, 158-19.
- Fitz-James M.H., Cavalli G., 2022. Molecular mechanisms of transgenerational epigenetic inheritance. *Nature Reviews Genetics*, 23 (6), 325-341.
- Gadaleta R.M., Magnani L., 2014. Nuclear receptors and chromatin: an inducible couple. *Journal of Molecular Endocrinology*, 52 (2), R137-R149.
- Grandjean V., Fourné S., De Abreu D.A.F., Derieppe M.-A., Remy J.-J., Rassoulzadegan M., 2015. RNA-mediated paternal heredity of diet-induced obesity and metabolic disorders, *Scientific Reports*, 5(1), 18193.
- Hales C.N., Barker D.J., Clark P.M., Cox L.J., Fall C., Osmond C., Winter P.D., 1991. Fetal and infant growth and impaired glucose tolerance at age 64. *BMJ (Clinical Research ed.)*, 303 (6809), 1019-1022.
- Hanson M.A., Gluckman P.D., 2014. Early developmental conditioning of later health and disease: Physiology or pathophysiology? *Physiological Reviews*, 94 (4), 1027-1076.
- Heard E., Martienssen R.A., 2014. Transgenerational epigenetic inheritance: Myths and mechanisms. *Cell*, 157 (1), 95-109.

- Junien C., Panchenko P., Pirola L., Amarger V., Kaeffer B., Parnet P. *et al.*, 2016. Le nouveau paradigme de l'origine développementale de la santé et des maladies (DOHaD). Épigénétique, environnement : preuves et chaînons manquants [The new paradigm of the developmental origin of health and diseases (DOHaD). Epigenetics and environment: evidence and missing links]. *Médecine/Sciences*, 32 (1), 27-34.
- Lempradl A., Pospisilik J.A., Penninger J.M., 2015. Exploring the emerging complexity in transcriptional regulation of energy homeostasis. *Nature Reviews Genetics*, 16 (11), 665-681.
- Lismer A., Dumeaux V., Lafleur C., Lambrot R., Brind'Amour J., Lorincz M.C., *et al.*, 2021. Histone H3 lysine 4 trimethylation in sperm is transmitted to the embryo and associated with diet-induced phenotypes in the offspring, *Developmental Cell*, 56(5), 671-686.e6.
- Lismer A., Kimmins S., 2023. Emerging evidence that the mammalian sperm epigenome serves as a template for embryo development, *Nature Communications*, 14(1), 2142.
- 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.
- Radford E.J., Ito M., Shi H., Corish J.A., Yamazawa K., Isganaitis E., *et al.*, 2014. In utero effects. In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism, *Science*, 345(6198), 1255903.
- Rosenfeld C.S., 2017. Homage to the «H» in developmental origins of health and disease. *Journal of Developmental Origins of Health and Disease*, 8 (1), 8-29.
- Safi-Stibler S., Gabory A., 2020. Epigenetics and the developmental origins of health and disease: Parental environment signalling to the epigenome, critical time windows and sculpting the adult phenotype. *Seminars in Cell and Developmental Biology*, 97172-180.
- Sharma U., Conine C.C., Shea J.M., Boskovic A., Derr A.G., Bing X.Y., *et al.*, 2016. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals, *Science*, 351(6271), 391-396.
- Tobi E.W., Goeman J.J., Monajemi R., Gu H., Putter H., Zhang Y. *et al.*, 2014. DNA methylation signatures link prenatal famine exposure to growth and metabolism. *Nature Communications*, 5 (1), 5592.
- Waterland R.A., Jirtle R.L., 2003. Transposable elements: Targets for early nutritional effects on epigenetic gene regulation. *Molecular and Cellular Biology*, 23 (15), 5293-5300.
- Weaver I.C.G., Cervoni N., Champagne F.A., D'Alessio A.C., Sharma S., Seckl J.R. *et al.*, 2004. Epigenetic programming by maternal behavior. *Nature Neuroscience*, 7 (8), 847-854.

Chapitre 8

Horloge épigénétique

SARAH VOISIN

Pendant le vieillissement, le profil épigénétique des cellules change de manière remarquable, et ce, chez tous les organismes pourvus d'un épigénome (mammifères, oiseaux, poissons et plantes). Le remodelage de l'épigénome est en effet l'une des signatures moléculaires typiques du vieillissement, au même titre que le taux de mutations somatiques et le rétrécissement des télomères (Lopez-Otín *et al.*, 2023). La méthylation de l'ADN est, de loin, la marque épigénétique la plus étudiée dans le contexte du vieillissement cellulaire.

► Montre-moi ton épigénome, et je te dirai ton âge

Depuis les années 1960, de nombreux laboratoires ont constaté que certaines régions du génome augmentent en méthylation, alors que d'autres voient leur niveau de méthylation diminuer. Ces changements sont extrêmement lents (de l'ordre de ~0,1% de méthylation par an) mais remarquablement universels (Horvath, 2013; Tarkhov *et al.*, 2022). Ils sont, de fait, si prévisibles et stéréotypés qu'en 2011, des chercheurs de l'université de Californie, à Los Angeles, sont parvenus à développer un algorithme capable de prédire l'âge d'un individu à environ 5 ans près, en ne se basant que sur le profil de méthylation des cellules contenues dans un échantillon de salive. La première horloge épigénétique était née.

De nombreux laboratoires ont alors été submergés par une vague d'enthousiasme à l'idée de pouvoir prédire avec une telle précision le degré de vieillissement des organes du corps humain, et ce à l'aide d'un seul biomarqueur, le méthylome. Depuis 2011, pas moins de dix horloges épigénétiques applicables à l'être humain ont vu le jour, la plus connue étant l'horloge de Horvath, également connue sous le nom d'horloge épigénétique « pan-tissus », ou « multi-tissus » (Horvath, 2013). Son créateur, Steve Horvath, un professeur de biostatistiques à l'université de Californie à Los Angeles, fournit un effort colossal pour la construire. Il passa en effet quatre ans, seul, à creuser dans les bases de données en accès libre telles que le *Gene Expression Omnibus* et *The Cancer Genome Atlas* (TCGA) afin de compiler plus de 8000 méthylomes provenant de 51 organes et types cellulaires différents. Il dut par ailleurs essayer de nombreux rejets de la part d'éditeurs et d'experts qui restaient sceptiques face à la précision remarquable de son horloge. Elle est en effet capable de prédire l'âge d'un individu à seulement ~3,6 ans près, quel que soit le tissu d'origine du méthylome en question (peau, sang, cervelet, cortex

occipital, salive, colon, tissu adipeux, reins, foie, poumons ou encore col de l'utérus). Une horloge pan-tissus plus récente, AltumAge, est parvenue, grâce à un modèle mathématique plus adapté, à réduire l'erreur de prédiction à seulement ~ 2,1 ans.

Les horloges épigénétiques peuvent-elles nous éclairer sur les mécanismes fondamentaux du vieillissement cellulaire? Peuvent-elles nous permettre de pronostiquer les maladies du vieillissement, ou de prédire l'agressivité de certains cancers? Peut-on utiliser l'âge épigénétique comme biomarqueur pour déterminer si certaines interventions thérapeutiques ont la capacité de «rajeunir» un organisme? Les horloges épigénétiques peuvent-elles nous aider à élucider les mécanismes du développement et de la différenciation cellulaire? Pourrait-on même utiliser les horloges en science forensique, en déterminant l'âge d'un suspect à partir d'un échantillon collecté sur une scène de crime? Pour pleinement saisir le potentiel (et les limitations) de ces horloges épigénétiques, il est d'abord nécessaire de se plonger dans leur processus de création. Le type de données qui les alimentent, le type d'information qu'on leur demande de prédire, ainsi que le modèle mathématique choisi, sont autant de paramètres dont l'importance est capitale dans l'interprétation des résultats. Êtes-vous prêts à entrer dans les détails mathématiques croustillants qui forment les rouages de ces horloges?

► Une horloge épigénétique, ce n'est pas sorcier !

Les horloges épigénétiques sont basées sur des algorithmes issus d'une approche mathématique appelée «apprentissage automatique» (*machine learning*), une forme d'intelligence artificielle.

L'apprentissage automatique a pour but de prédire, à partir de certaines données et avec la plus petite marge d'erreur possible, une information inconnue mais désirée. Dans le cas des horloges épigénétiques, le but est de prédire l'âge (d'une cellule, d'un tissu ou d'un organisme tout entier) en se basant sur le taux de méthylation à un ou plusieurs CpG (cytosine liée à une guanine *via* un groupement phosphate). L'apprentissage automatique passe par deux phases : la première (apprentissage) consiste à nourrir l'algorithme de profils de méthylation dont l'âge est connu ; l'algorithme trouve alors par lui-même (d'où le terme «automatique») comment combiner les niveaux de méthylation de différents CpG pour prédire l'âge de ces échantillons avec le plus de précision possible (figure 8.1).

De cette phase d'apprentissage résulte une équation, ou modèle mathématique, qui sera alors testée sur un deuxième ensemble de données que l'algorithme n'a jamais rencontré. Cette phase de test permet de voir si le modèle généré fonctionne réellement sur de nouvelles données. La plupart des horloges épigénétiques ont été construites grâce à un modèle mathématique appelé «*elastic-net*» qui pénalise certaines variables pour ne garder que celles qui ajoutent une quantité substantielle d'information au modèle (et lui confèrent donc une précision accrue). Ce modèle est particulièrement adapté dans le cas des horloges épigénétiques, car le nombre de variables potentielles pour la prédiction (ici, les dizaines de milliers de CpG) surpasse de loin le nombre d'échantillons. L'algorithme sélectionne, parmi tous les CpG possibles, ceux dont la combinaison linéaire permet de prédire l'âge de l'échantillon avec la plus grande précision, grâce à une équation du type (figure 8.1) :

$$\hat{\text{Age}} = \alpha + \beta_1 \times \text{CpG}_1 + \dots + \beta_n \times \text{CpG}_n$$

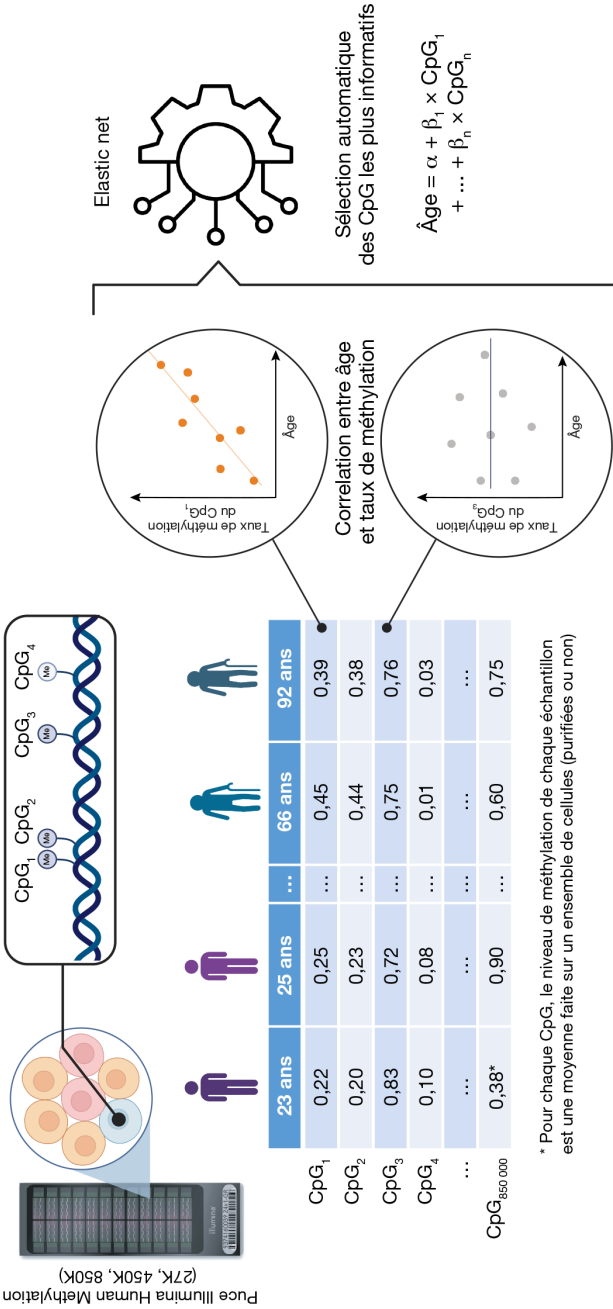


Figure 8.1. Comment construire une horloge epigenetique.

Pour créer une horloge épigénétique, il faut tout d'abord obtenir des profils de méthylation provenant de plusieurs individus d'âge différents (petits bonshommes colorés). Ces profils sont typiquement générés par les puces Illumina *Human Methylation* (en haut à gauche) qui mesurent précisément les niveaux de méthylation de plusieurs dizaines, voire de centaines de milliers de CpG sur des régions bien définies du génome. Ces niveaux de méthylation sont estimés sur une population de cellules qui peuvent provenir de n'importe quel tissu, qu'il soit sanguin, rénal, pulmonaire, cérébral... Ces cellules peuvent avoir été préalablement purifiées pour en faire une population homogène de cellules d'un même type (par exemple une population de lymphocytes T), ou peuvent être un mélange de cellules variées (par exemple un mélange de lymphocytes B, lymphocytes T, monocytes, neutrophiles, etc.). Il résulte de cette expérience une matrice contenant les niveaux de méthylation à des centaines de milliers de CpG pour chaque individu, niveau qui peut aller de 0 à 1 (0 : le CpG en question est entièrement dénué de méthylation dans l'échantillon, et 1 : le CpG en question est entièrement méthylé dans l'échantillon). Cette matrice est alors donnée à l'algorithme d'apprentissage automatique (typiquement elastic-net, mais il en existe d'autres), qui va alors automatiquement sélectionner les CpG dont la combinaison linéaire permet d'arriver à une prédiction de l'âge des individus avec une grande précision. Certains des CpG mesurés sont naturellement corrélés à l'âge (comme le CpG₁), alors que d'autres ne le sont pas (comme le CpG₃). L'algorithme sélectionnera des CpG corrélés à l'âge, mais pas forcément ceux qui le sont le plus fortement ou le plus significativement. L'algorithme produit une équation qui permet alors, pour n'importe quel échantillon dont les niveaux de méthylation à certains CpG sont connus, d'estimer son âge. α : ordonnée à l'origine; β_1, \dots, β_n : coefficient (poids) appliqué à chaque CpG. Un CpG dont la méthylation diminue avec l'âge aura un coefficient négatif, alors qu'un CpG dont la méthylation augmente avec l'âge aura un coefficient positif.

Par exemple, l'horloge pan-tissus utilise les niveaux de méthylation d'exactlyment 353 CpG pour prédire l'âge d'un méthylome (Horvath, 2013), l'horloge de Hannum en utilise 71, et l'horloge musculaire MEAT en utilise 156. La vaste majorité des horloges épigénétiques ont été construites à partir de données de méthylation générées par les puces Illumina *Human Methylation*, qui mesurent précisément les niveaux de méthylation de plusieurs dizaines, voire de centaines de milliers de CpG sur des régions bien définies du génome, supposées d'intérêt fonctionnel (promoteurs, introns, exons, *enhancers*, îlots CpG, etc.). La raison pour laquelle les horloges n'ont pas encore été construites à partir de données provenant d'autres technologies (par exemple, séquençage de type RRBS, *Reduced-representation bisulfite sequencing*, ou MeDIP-seq, *Methylated DNA immunoprecipitation sequencing*), ou d'autres marques épigénétiques comme l'ouverture de la chromatine ou les modifications d'histones, est simplement une question d'extensibilité : ces puces Illumina sont relativement bon marché. Elles ont donc été utilisées massivement dans de nombreuses études de physiopathologie humaine, générant au fil des années des dizaines de milliers de profils épigénétiques issus de multiples organes et cellules purifiées humaines. Ce fut une aubaine pour les horloges épigénétiques, qui nécessitent des centaines, voire des milliers d'échantillons pour parvenir à une prédiction suffisamment précise.

Qu'ont donc de spécial les CpG sélectionnés par les horloges ? Pourquoi sélectionner ces CpG en particulier, et pas d'autres ? Certains groupes de recherche essaient actuellement de déconstruire les horloges épigénétiques pour comprendre quels phénomènes biologiques elles représentent. La capacité de ces horloges, et en particulier des horloges pan-tissus, à prédire l'âge de n'importe quel tissu avec une telle précision, suggère que le passage du temps laisse une empreinte profonde et universelle dans notre génome. Mais à quel phénomène biologique sous-jacent cela correspond-il ? Plus de dix ans après la publication de la première horloge épigénétique, cette question n'est toujours pas résolue. Les études les plus récentes suggèrent que ces horloges sont intimement liées aux processus de développement et de différenciations cellulaires. Les cellules embryonnaires et les cellules souches pluripotentes induites, même après de multiples passages en culture, gardent un âge épigénétique de 0, et l'horloge épigénétique ne se met à tourner que lorsque les cellules entrent en différenciation (Kabacik *et al.*, 2022). Cependant, cela ne signifie pas que le profil de méthylation des cellules souches et des cellules germinales ne change pas au cours de la différenciation.

Quand on parle d'âge épigénétique, on parle de l'âge prédit par le modèle mathématique, et qui ne prend en compte les taux de méthylation qu'à un nombre limité de CpG. Une autre observation intéressante est que la lignée germinale a le même âge épigénétique que l'organisme qui la porte (Kerepesi *et al.*, 2021). Comment explique-t-on que de nouveaux organismes repartent à 0 d'âge épigénétique lors de la reproduction ? On a récemment découvert que le zygote passe par une phase de « réjuvenation épigénétique » au stade de gastrulation. Pendant cette phase critique, on observe d'autres modifications profondes qui préparent l'organisme à la vie, telles qu'une extension des télomères, des vagues successives de déméthylation et de méthylation et une inactivation du chromosome X.

Par ailleurs, il semblerait que le processus biologique qui fait tourner l'horloge pan-tissus soit indépendant de certaines marques typiques du vieillissement telles que les mutations génétiques, la sénescence cellulaire et l'attrition des télomères (Kabacik *et al.*, 2022). L'horloge épigénétique se met à tourner plus vite en cas de perte de

fonction mitochondriale ou de capacité à détecter des fluctuations dans les taux de nutriments comme le glucose, les lipides et les acides aminés (Kabacik *et al.*, 2022). Des souris soumises à une restriction calorique ou traitées avec de la rapamycine (qui inhibe la voie mTOR, *mechanistic Target of rapamycin*) voient en effet leur horloge épigénétique tourner plus lentement. Horvath a suggéré que ce sont les programmes développementaux et de maintenance cellulaire qui laissent une empreinte moléculaire sur la méthylation de l'ADN (Horvath et Raj, 2018). Il a également émis l'hypothèse qu'une partie du signal provient de la perte de fonction et du nombre de cellules souches qui sont présentes dans de nombreux tissus (cerveau, muscle, foie). Des expériences récentes chez la souris (Chondronasiou *et al.*, 2022) ont démontré que les facteurs de Yamanaka (quatre facteurs de transcription appelés « OSKM » pour OCT4, SOX2, KLF4 et c-MYC, qui sont capables de dé-différencier des cellules vers un état pluripotent caractéristique du stade embryonnaire), injectés à faible dose et/ou de manière cyclique, ont la capacité de régénérer de nombreux organes tels que le pancréas, le foie, la rate et les cellules sanguines.

Il est toutefois peu probable que les composants spécifiques de ces horloges nous apprennent quoi que ce soit de révolutionnaire sur la biologie du vieillissement. En effet, les CpG qui constituent les horloges n'ont absolument rien de spécial : ce sont simplement les CpG dont la combinaison permet d'arriver à l'erreur de prédiction la plus réduite. Naturellement, ces CpG sont bel et bien sensibles aux effets de l'âge (sans quoi ils ne seraient porteurs d'aucune information qui permettrait à l'algorithme de parvenir à une telle prédiction), mais il existe des dizaines de milliers de CpG dont le niveau de méthylation change au cours du vieillissement, et ces CpG sont probablement tout aussi pertinents pour comprendre les mécanismes du vieillissement cellulaire. Il faut bien garder en tête que les horloges épigénétiques sont des algorithmes dont le but est de prédire une information, sans aucune considération pour la biologie sous-jacente. Elastic-net sélectionne les CpG de manière opaque, et est loin d'inclure les CpG les plus significativement ou fortement associés à l'âge.

Par ailleurs, l'algorithme aurait tout aussi bien pu arriver à une précision similaire en utilisant un autre groupe complètement distinct de 300, 500, ou même 1 000 CpG. Il suffit d'ailleurs de nourrir l'algorithme avec des données d'apprentissage un tant soit peu différentes, et les CpG sélectionnés sont très différents. Elastic-net a également tendance à éliminer les CpG « redondants » dont les niveaux de méthylation sont fortement corrélés entre eux, afin de ne garder que les CpG qui apportent une nouvelle couche d'information au modèle (Tarkhov *et al.*, 2022). En d'autres termes, Elastic-net va avoir tendance à choisir les CpG qui changent au cours du vieillissement de manière isolée (sans coordination avec d'autres CpG) : les CpG qui forment des clusters de co-régulation parce qu'ils font partie de la même voie métabolique, et/ou parce qu'ils sont physiquement proches sur l'ADN, auront tendance à être éliminés de l'équation, mais ils sont potentiellement tout aussi intéressants à étudier. Les horloges épigénétiques ne sont donc en aucune mesure nécessaires à notre compréhension du vieillissement, car il est possible d'obtenir des connaissances approfondies en étudiant l'ensemble des CpG qui changent au cours du vieillissement, et pas seulement les CpG qui constituent les horloges.

Si ces horloges peuvent mesurer le passage du temps dans nos cellules, cela signifie-t-il qu'une cellule, un organe ou un organisme tout entier est d'autant plus sensible aux pathologies et malade que son âge épigénétique est avancé? Nous avons tous

l'intuition que certains individus vieillissent plus rapidement que d'autres en fonction de certaines prédispositions génétiques ou d'expositions à l'environnement. Mais ces différences de trajectoire de vie se reflètent-elles dans le vieillissement épigénétique ? Est-il possible d'être né il y a exactement trente ans, mais d'avoir en fait 40 ans en « âge biologique » ? Peut-on prédire la mort d'un individu en se basant sur son âge épigénétique ? Est-il possible de « remonter le temps épigénétique » afin de diminuer le risque de maladie, et de retarder, voire éliminer la mort certaine qui nous attend tous ? Si oui, comment ? Ces questions, de prime abord scientifiques, ont des implications philosophiques profondes pour la conception de la vie humaine et animale.

►► Âge biologique *versus* âge chronologique

Les premières études sur les horloges épigénétiques ont révélé que, même en prenant en compte de nombreux facteurs de risques tels que l'obésité, le niveau d'éducation, la prise d'alcool et de tabac, le diabète, le cancer, l'hypertension et le niveau d'activité physique, l'âge épigénétique mesuré dans le sang est associé à un plus grand risque de mortalité, qu'elle qu'en soit la cause. Les individus atteints de maladies neuro-dégénératives (Alzheimer, Parkinson, démence) présentent un âge épigénétique avancé dans le sang et le cerveau, les individus atteints de maladies cardiovasculaires (obésité, maladie coronarienne, hypertension), dans le sang et le foie, et ceux atteints de maladies musculo-squelettiques (arthrose, syndrome de fragilité), dans le sang (Horvath et Raj, 2018). C'est aussi le cas de maladies plus rares telles que la maladie de Huntington, la trisomie 21, le sida et le syndrome de Werner, qui sont associées à une accélération du vieillissement épigénétique dans le sang. Le cas du cancer est équivoque, car, selon le type de cancer et les mutations qui lui sont associées, on observe soit une augmentation, soit une diminution de l'âge épigénétique.

Certains facteurs intrinsèques peuvent également réguler la vitesse à laquelle tourne l'horloge épigénétique. Les centenaires et leur descendance présentent un âge épigénétique sanguin plus jeune qu'attendu, ce qui suggère l'existence de variants géniques qui prédisposent à une plus grande longévité. On a d'ailleurs découvert plus d'une centaine de mutations dans des gènes du métabolisme lipidique et de la fonction immunitaire qui sont associées à un âge épigénétique sanguin plus jeune. Par ailleurs, il semblerait que les femmes vieillissent épigénétiquement moins vite que les hommes, et la ménopause, caractérisée par une diminution de la production d'œstrogènes, s'accompagne d'une accélération de l'horloge épigénétique, ce qui suggère une très probable influence des hormones sexuelles sur le vieillissement épigénétique (Tan *et al.*, 2022). On observe également un ralentissement de l'horloge épigénétique mesurée dans le sang, avec une éducation plus poussée, une activité physique accrue et une moindre consommation de produits animaux et d'alcool. Cependant, ces associations avec des facteurs environnementaux restent relativement faibles.

La pertinence de ces horloges pour élucider les différences de mortalité entre individus du même âge a toutefois été rapidement remise en question par des spécialistes de la biologie computationnelle. Ils ont mis le doigt sur le biais de ces horloges optimisées pour prédire l'âge chronologique (c'est-à-dire le nombre de jours écoulés depuis la naissance) : en effet, l'algorithme qui sous-tend l'horloge a pour mission de prédire l'âge réel d'un individu avec la plus grande précision possible. L'algorithme aura donc tendance à éliminer de son équation les sites de méthylation qui sont sensibles à

d'autres facteurs que l'âge, tels que le tabac, l'alcool, la pollution ou la sédentarité. En effet, ces facteurs introduisent de la variabilité dans le niveau de méthylation entre individus du même âge pour certains CpG, et ces CpG sont donc moins informatifs pour l'algorithme. Un algorithme à qui on demande de prédire l'âge chronologique d'un individu aura tendance à sélectionner les CpG qui changent de manière identique chez tous les individus, et qui ne sont donc pas à même d'expliquer les différences de mortalité entre individus du même âge. Cela signifie que théoriquement, ces horloges ne devraient pas être associées à un risque accru de mortalité.

Mais alors, comment expliquer qu'elles aient été associées à toute une batterie de maladies depuis les années 2010? Il est probable que cela s'explique par l'imperfection des données qui ont alimenté les premières horloges lors de leur construction, et par la confusion statistique entre âge et facteurs qui accélèrent le vieillissement. Par exemple, si les individus plus âgés sont plus sédentaires que les jeunes, le signal que l'algorithme détectera comme étant associé à l'âge est en fait associé à la sédentarité, ce qui conduit à une association fallacieuse entre âge épigénétique et mortalité. Idéalement, l'algorithme devrait être nourri de profils de méthylation venant d'individus pour qui la distribution de ces facteurs environnementaux est répartie de manière équitable entre les différentes classes d'âge. En pratique, de nombreux facteurs de confusion sont totalement inconnus, et seule une augmentation de la taille de l'échantillon peut limiter ce biais statistique. Un article paru en 2019 a d'ailleurs montré qu'en augmentant la quantité de données fournies à l'algorithme – et donc en réduisant la probabilité de confusion statistique grâce à un échantillon plus large –, l'association entre mortalité et âge épigénétique s'atténue au point de devenir insignifiante.

Face à un tel dilemme, les développeurs des toutes premières horloges épigénétiques (dites de « première génération ») ont alors eu l'idée de demander à l'algorithme de prédire non pas l'âge chronologique, mais l'âge biologique de leurs échantillons. Deux horloges épigénétiques, PhenoAge et GrimAge (Lu *et al.*, 2022), ont été créées grâce à une méthodologie en deux étapes : la méthylation de l'ADN est d'abord utilisée pour prédire les niveaux de nombreux biomarqueurs de santé tels que l'indice de masse corporelle (IMC), le ratio taille/hanches, la consommation de tabac, le taux sanguin de triglycérides, de lipoprotéines ou d'hémoglobine glyquée. Ensuite, ces différents prédicteurs passent eux-mêmes par une deuxième phase d'apprentissage automatique optimisé pour prédire le temps écoulé entre la prise de l'échantillon et la mort d'un individu.

Cette méthodologie, au premier abord alambiquée, a révolutionné les applications potentielles des horloges épigénétiques. Par rapport aux horloges de première génération, les horloges de deuxième génération sont beaucoup plus fortement et significativement associées à la mortalité et à de nombreuses pathologies (Lu *et al.*, 2022) telles que la maladie coronarienne, la perte de fonction pulmonaire, la stéatose hépatique, le syndrome métabolique, le diabète, etc. Elles sont aussi beaucoup plus sensibles aux facteurs environnementaux qui influencent les trajectoires de vie comme la consommation de fruits et légumes, l'exercice physique et l'éducation (Lu *et al.*, 2022). Une étude récente a par ailleurs montré que, contrairement aux horloges de première génération, les changements de méthylation aux CpG mesurés par GrimAge sont une cause directe du développement de certains cancers. L'élégance de la construction de ces horloges de deuxième génération fait qu'elles s'appliquent même aux individus jeunes chez qui les symptômes du vieillissement ne sont pas encore visibles.

Le vieillissement est un processus de dégradation moléculaire progressive qui démarre dès la naissance, et les horloges de deuxième génération sont un outil très puissant pour identifier les individus, quel que soit leur âge, qui sont sur une trajectoire de vie inquiétante, afin de mettre en place des interventions pour ralentir le vieillissement. Mais comment savoir si une intervention a bel et bien remis un individu sur la bonne trajectoire de vie et a effectivement diminué son risque de maladie et de mortalité? PhenoAge et GrimAge ont été développées dans des cohortes transversales (dans lesquelles les individus sont échantillonnés à un seul moment dans le temps), et il est donc difficile de savoir si une intervention qui diminuerait l'âge épigénétique estimé par GrimAge se traduirait effectivement par une rééquilibration de la trajectoire de vie à l'échelle individuelle.

Pour répondre à cette question, des chercheurs de l'université de Bristol, en Grande-Bretagne, ont développé DunedinPoAm (et sa version plus récente DunedinPACE), une horloge épigénétique qui mesure la cadence du vieillissement à l'échelle individuelle. Pour ce faire, ils ont suivi l'évolution de biomarqueurs sur vingt ans dans une cohorte d'individus bien phénotypés, afin d'obtenir une estimation longitudinale de l'évolution de l'âge biologique au niveau individuel. Une étude récente vient de révéler un lien entre le vieillissement épigénétique et le concept de DOHaD (voir chapitre 7). L'exposition des fœtus à la crise économique des années 1930 est associée à un vieillissement épigénétique accéléré, et cet effet est remarquablement limité à la période *in utero* (vs prénatale ou postnatale), et n'est détecté qu'avec les horloges de deuxième génération.

Alors que les applications de ces horloges n'en sont qu'à leurs balbutiements, l'engouement s'est répandu bien au-delà de la communauté scientifique. Le millionnaire américain Bryan Johnson a mesuré son âge épigénétique avant et après le régime alimentaire, physique et médical draconien qu'il a suivi ces cinq dernières années. Son âge épigénétique a diminué de 47 à 42,5 ans en sept mois, et, selon l'horloge épigénétique DunedinPACE qui mesure la vitesse du vieillissement épigénétique, il vieillit à un taux de 0,76. En d'autres termes, il ne prend que 277 jours d'âge par an (par rapport à un être humain moyen). Bien que le but ultime des horloges épigénétiques soit précisément d'obtenir un tel suivi du vieillissement à l'échelle individuelle, l'interprétation de telles données est encore un sujet de débat dans la communauté scientifique, et il convient d'être prudent.

Mais au fait, c'est quoi exactement, l'âge biologique? Nous avons tous une intuition de ce que sont les symptômes d'un âge biologique avancé, comme le ralentissement des capacités cognitives, une plus grande fragilité physique, des cheveux blancs, des rides, de la douleur au niveau des articulations. Mais d'un point de vue scientifique, il est en pratique extrêmement compliqué de définir l'âge biologique. Tout d'abord, à quel niveau définir cet âge biologique? Au niveau de l'organisme tout entier, auquel cas il faut combiner des symptômes disparates comme la cognition, les capacités physiques ou la souplesse de la peau? Ou doit-on plutôt définir l'âge biologique à l'échelle d'un organe? Et est-il possible d'avoir un organe mosaïque composé de cellules jeunes et de cellules plus âgées? Doit-on alors définir l'âge biologique au niveau d'une cellule unique? Les études les plus récentes ont révélé que certains individus vieillissent plus vite au niveau immunologique, alors que d'autres vieillissent plus vite au niveau rénal, par exemple. Par ailleurs, les facteurs associés à la longévité ne semblent pas exercer les mêmes effets dans tous les organes, par exemple l'obésité accélère l'âge épigénétique du foie mais pas celui du muscle.

►► Peut-on avoir simultanément un foie vieux et un muscle jeune ?

Si les horloges épigénétiques ont initialement surpris la communauté scientifique à leur découverte au début des années 2010, c'est parce que le méthylome varie considérablement entre organes et types cellulaires. Personne ne s'attendait à ce que le vieillissement exerce un effet universel sur tous les tissus du corps humain, transcendait ces différences épigénétiques liées à la différenciation cellulaire. L'erreur de raisonnement a été de confondre différences « basales » entre tissus, et différence « de sensibilité » de ces tissus face au vieillissement. Les différences épigénétiques basales entre types cellulaires, établies au cours du développement, n'empêchent en rien que ces cellules répondent de manière similaire à un *stimulus* donné. C'est d'ailleurs le but ultime de nombreuses applications cliniques de l'épigénétique, qui cherchent à déterminer quelles marques épigénétiques détectables dans un échantillon de sang reflètent ce qui se passe dans d'autres tissus difficiles d'accès (cerveau, foie, os, etc.).

Il faut bien comprendre que l'applicabilité de l'horloge pan-tissus à de nombreux types cellulaires n'implique en rien que le vieillissement exerce exactement le même effet sur tous les tissus. Il existe une signature universelle du vieillissement, et c'est elle que l'horloge pan-tissus détecte, mais il existe également des horloges épigénétiques spécifiques à certains tissus (peau, muscle, cerveau, os) qui s'appliquent assez mal à d'autres types cellulaires. Si le vieillissement épigénétique n'affecte pas tous les tissus de manière égale, que nous apprennent les horloges épigénétiques sur les différences de vieillissement entre tissus ? Existe-t-il des organes qui vieillissent plus rapidement que d'autres ? Ou différemment des autres ? Cette question a été remarquablement peu abordée, alors qu'elle constitue un pas fondamental vers la compréhension de ce qu'est le vieillissement au niveau moléculaire. Le cervelet, et à un moindre degré le cerveau tout entier, semble vieillir plus lentement, alors que le tissu mammaire semble vieillir plus rapidement, mais nous n'en savons pas plus sur les différences entre tissus. Ces résultats sont également à prendre avec des pincettes, car ils sont basés sur l'horloge pan-tissus de première génération, qui n'est que faiblement associée aux pathologies humaines.

Il est également important de mentionner une critique récurrente des horloges épigénétiques, et du champ de la biologie moléculaire en général : la confusion statistique induite par l'hétérogénéité cellulaire. Les premières horloges furent construites à partir de méthylomes provenant de tissus entiers, et non de cellules purifiées ou uniques. Il est donc tout à fait plausible que les changements de méthylation détectés par les horloges épigénétiques ne reflètent pas une modification épigénétique intrinsèque aux cellules qui constituent le tissu, mais reflètent un changement dans les proportions relatives de ces différents types cellulaires dans le tissu (par exemple, une diminution du nombre de cellules souches). Les fonctions bien distinctes exercées par nos organes, telles que le mouvement généré par les muscles, la régulation de la glycémie par le pancréas ou la filtration de l'urine par les reins, résultent de la coordination de nombreux types cellulaires. Dans le cas du muscle, par exemple, le mouvement provient certes des fibres musculaires qui se contractent, mais cette contraction n'est possible que grâce à la présence de neurones moteurs qui envoient des décharges électriques, de vaisseaux sanguins qui irriguent le muscle pour lui apporter l'oxygène nécessaire à son métabolisme, et de cellules adipeuses qui forment un réservoir d'énergie.

Les organes sont donc des mosaïques de cellules aux profils épigénétiques extrêmement dissemblables, mais qui fonctionnent de manière coordonnée pour remplir une certaine tâche. Pour continuer sur l'exemple du muscle, l'essor récent des technologies à cellule unique a permis de révéler qu'une large partie de l'ADN extrait d'une biopsie de muscle squelettique provient de cellules sanguines (monocytes, lymphocytes), de cellules endothéliales qui tapissent la paroi des vaisseaux sanguins, de fibroblastes qui forment la matrice extracellulaire, et d'adipocytes. Les proportions relatives des différents types cellulaires changent au cours du vieillissement dans tous les tissus, de sorte que certains types cellulaires deviennent moins abondants, et d'autres se mettent à fourmiller. Le vieillissement s'accompagne notamment d'une infiltration de cellules immunitaires dans presque tous les organes (López-Otín *et al.*, 2023), ce qui signifie que les proportions de monocytes et de lymphocytes augmentent avec l'âge. Compte tenu du faible degré de changement en méthylation pour de nombreux CpG disparates ($\sim 0,1\%$ par an), les horloges épigénétiques ne reflètent-elles donc qu'un changement de la composition cellulaire des tissus au cours du vieillissement ?

Dès leur création, les horloges épigénétiques de deuxième génération ont pris en compte les différentes proportions de cellules sanguines. De manière générale, un âge épigénétique plus avancé est bel et bien corrélé à des changements dans les proportions relatives de différentes cellules immunitaires qui s'opèrent au cours du vieillissement (par exemple, une augmentation du nombre de granulocytes et une diminution du nombre de lymphocytes T auxiliaires et de lymphocytes T naïfs). Toutefois, même en prenant en compte ce facteur de confusion, on observe des changements de méthylation intrinsèques aux cellules qui sont fortement associés à une mortalité accrue (Lu *et al.*, 2022 ; Belsky *et al.*, 2022). D'ailleurs, l'existence d'horloges épigénétiques pan-tissus suggère fortement qu'au moins une partie du signal provient de changements de méthylation qui s'opèrent au sein des cellules. De nombreux facteurs environnementaux (alimentation, activité physique, tabagisme) semblent à la fois aggraver l'inflammation chronique qui se produit naturellement au cours du vieillissement (ce qui correspond à un changement dans les proportions relatives des différentes cellules immunitaires), et accélérer les changements de méthylation intrinsèques aux cellules. Cependant, l'hétérogénéité cellulaire n'a été vraiment prise en compte que sur les analyses faites dans le sang, car il est souvent long et coûteux d'estimer les différentes proportions cellulaires dans d'autres tissus moins bien caractérisés (foie, muscle, cerveau).

Le développement de méthodes de déconvolution statistique promet de remédier à ce problème dans les années à venir (Zhu *et al.*, 2022). Pour mieux comprendre comment les facteurs environnementaux modulent la vitesse à laquelle vieillissent nos organes, le laboratoire de Vadim Gladyshev a récemment développé une horloge épigénétique pour cellule unique (Trapp *et al.*, 2021). Cette horloge a démontré qu'au sein d'un même tissu, certaines cellules sont plus ou moins âgées que d'autres, comme les cellules souches musculaires qui sont épigénétiquement plus jeunes que le tissu dans lequel elles résident.

►► Aucune espèce animale n'échappe au vieillissement épigénétique

Bien que les premières horloges épigénétiques aient été développées à partir de tissus humains dans le but d'étudier le vieillissement chez l'homme, il est rapidement devenu évident que ces horloges pouvaient aussi bénéficier aux animaux de compagnie,

au bétail et aux espèces sauvages. Dans les cinq dernières années, des horloges épigénétiques ont été développées pour le chat, le chien, le loup, les ruminants, la chauve-souris, le rat-taupe nu, le grand dauphin de l'océan Indien, les pinnipèdes et l'ensemble des mammifères.

Augmenter la longévité des animaux de ferme pourrait être avantageux pour le secteur agricole en réduisant les coûts économiques et environnementaux associés à la production d'animaux de remplacement, en prolongeant la productivité des animaux à des âges plus avancés (industrie laitière et poules pondeuses, par exemple) et en réduisant par ailleurs le nombre d'animaux abattus car jugés non productifs. Une horloge épigénétique spécifique aux ruminants (chèvre, vache, renne et mouton) viennent d'être publiées. Les horloges épigénétiques développées spécifiquement pour le chat et le chien ont pour ambition première de créer un biomarqueur de la santé canine et féline, afin de mieux prendre en charge nos animaux de compagnie tout au long de leur vie. Le *Dog Aging Project*, initié par des scientifiques et des cliniques vétérinaires de plus de vingt instituts, a prévu de mettre à profit l'horloge épigénétique canine pour déterminer les facteurs environnementaux responsables du vieillissement du chien domestique, et pour tester l'effet de certains composés pharmaceutiques, comme la rapamycine, sur sa longévité. L'horloge canine n'a paradoxalement révélé aucune corrélation entre le vieillissement épigénétique et différentes caractéristiques des races de chien, qui sont pourtant connues pour leurs longévités très disparates.

Les horloges épigénétiques pour les espèces sauvages (loup, dauphin, chauve-souris), et en particulier les horloges de première génération qui prédisent l'âge chronologique d'un individu, ont des applications évidentes pour la conservation et le suivi des populations sauvages. Certains groupes de recherche se sont penchés sur des espèces à la longévité exceptionnelle telles que le rat-taupe nu, en espérant glaner des informations sur les mécanismes basiques du vieillissement. Le rat-taupe nu n'est guère plus gros qu'une souris de laboratoire, mais peut vivre jusqu'à plus de 30 ans en captivité. Il est particulièrement résistant à de nombreuses maladies du vieillissement telles que le cancer et les maladies cardiovasculaires; il ne montre aucun changement métabolique et aucune détérioration physiologique (cataracte, arthrose) en vieillissant. Malgré cette sénescence négligeable, le rat-taupe nu vieillit bel et bien au niveau épigénétique, et les femelles reproductives (appelées « reines » comme chez les fourmis, les termites et les abeilles) vieillissent épigénétiquement moins vite que les femelles non reproductives.

Horvath et le consortium de la méthylation mammalienne viennent tout juste de développer trois horloges épigénétiques applicables à tous les euthériens, et une applicable à tous les marsupiaux, grâce à des profils de méthylation provenant de 59 tissus de 128 espèces de mammifères. Ces horloges, qui se basent sur des CpG ultra-conservés entre espèces, confirment le lien universel, ancestral et étroit entre vieillissement épigénétique et processus développementaux. C'est peut-être ceci que les horloges épigénétiques nous ont vraiment appris : il existe des changements de méthylation universels et remarquablement prédictibles dans des voies métaboliques fondamentales qui se produisent à cause des dommages quotidiens faits à nos cellules, tout simplement parce que notre organisme tente de maintenir une homéostasie. Les oscillations autour de cet état d'équilibre provoquent des chocs métaboliques pour l'organisme, ce qui érode progressivement notre génome, et les dommages accumulés (le *deleterio*) finissent par nous tuer.

►► Le vieillissement s'accompagne d'une érosion du méthylome

Des modélisations mathématiques plus poussées ont permis de révéler que ces changements ne sont en fait pas linéaires, mais exponentiels (ils s'accroissent avec l'âge) (Kabacik *et al.*, 2022). Les régions du génome dont la méthylation est initialement basse ont tendance à voir leur niveau de méthylation augmenter, alors que les régions dont la méthylation est naturellement élevée ont tendance à perdre en méthylation, de sorte que de nombreux CpG dérivent vers un niveau de méthylation de ~50% au cours du vieillissement. Il en résulte un «lissage», un «aplanissement» du paysage épigénétique imaginé par Waddington, dont les conséquences pour la cellule ne sont pas encore très claires.

La théorie la plus répandue est que cette érosion conduit les cellules à perdre progressivement leur identité et à mal fonctionner. Pour comprendre ce concept, il faut se référer à la théorie de l'information de Shannon, qui est une théorie probabiliste permettant de quantifier le contenu moyen en information d'un ensemble de messages ou, dans notre cas, le contenu moyen en information de l'épigénome. En effet, l'épigénome se caractérise par des niveaux spécifiques de méthylation et de modifications d'histones à des endroits bien précis de l'ADN, procurant à la cellule une quantité colossale d'informations qui lui permet d'utiliser les bons gènes au bon moment, et donc d'exécuter le programme génétique correspondant à son identité cellulaire. L'érosion de l'épigénome pendant le vieillissement peut être vue comme une perte d'information et une augmentation de l'entropie (désordre, chaos) du système épigénétique dans son ensemble, à tel point que les cellules finissent par perdre leur capacité à exécuter le programme qui correspond à leur identité, conduisant à un dysfonctionnement profond du tissu auquel elles appartiennent.

La majorité des modifications épigénétiques causées par le vieillissement sont de nature stochastique (Kabacik *et al.*, 2022) au sens où les différentes régions du génome ne changent pas de manière coordonnée au sein d'une même cellule. En d'autres termes, les niveaux respectifs de méthylation de nombreux CpG d'une cellule donnée ont tendance à changer indépendamment les uns des autres. Cependant, les changements qui s'opèrent dans l'ensemble des cellules qui forment un tissu donné sont loin d'être aléatoires. Par exemple, les promoteurs et l'intérieur des gènes activement transcrits sont relativement résistants au vieillissement épigénétique, et le peu de promoteurs affectés voient, en grande majorité, leur niveau de méthylation augmenter (ce qui n'est pas surprenant, car ces promoteurs ont un niveau de méthylation habituellement très bas). *A contrario*, les *enhancers* sont plus sensibles au vieillissement épigénétique, et une hyper- ou une hypométhylation sont toutes deux possibles.

Les changements les plus remarquables se produisent au niveau des régions liées aux protéines du groupe Polycomb. Ces protéines forment des complexes avec la chromatine (que nous appellerons «régions Polycomb» par simplicité), maintenant certains gènes développementaux clés dans un état silencieux. Si les protéines Polycomb se lient aux gènes développementaux, c'est parce que ces gènes présentent de grandes régions dépourvues de méthylation (appelées «vallées», ou «canyons» déméthylés) pour lesquelles les protéines Polycomb ont une affinité particulière. Une fois liées à l'ADN, les protéines Polycomb maintiennent ces vallées dans un état déméthylé, ce qui confère au gène la possibilité d'être activé rapidement au bon moment au cours de la différenciation cellulaire. Certaines régions Polycomb sont appelées «bivalentes»,

car on y observe à la fois des marques d'histones répressives et activatrices. Pendant le vieillissement, les régions Polycomb, et en particulier les régions bivalentes, augmentent en méthylation, car les protéines Polycomb perdent progressivement leur capacité à se lier à l'ADN, ce qui libère la place pour les enzymes DNMT qui méthylent alors l'ADN. Cette hyperméthylation diminuerait la plasticité des gènes développementaux et affecterait notamment les cellules souches, ce qui conduirait de nombreux tissus (foie, muscle, cerveau, intestins, os) à perdre leur capacité régénératrice avec l'âge.

Nous n'avons parlé jusqu'ici que du changement du niveau moyen de méthylation de certains CpG, mais il se trouve que le vieillissement s'accompagne également d'une augmentation de la variance du niveau de méthylation de nombreux CpG. Sur ces sites, la méthylation ne diffère que très peu entre jeunes individus, alors qu'elle diffère de manière croissante entre individus de plus en plus âgés (la méthylation prend des valeurs de plus en plus extrêmes et divergentes au cours du vieillissement). On ne sait pour le moment pas quelles sont les conséquences de cette hypervariabilité pour la cellule, mais il est probable qu'elle conduise à un programme d'expression cellulaire plus « aléatoire » et chaotique.

► Conclusion

Les horloges épigénétiques ont véritablement révolutionné la recherche sur le vieillissement non seulement d'un point de vue fondamental en soulignant le rôle capital des voies développementales dans le processus de vieillissement, mais aussi d'un point de vue translationnel en générant des biomarqueurs d'utilité clinique capables de suivre les trajectoires de vie d'individus avec une précision inégalée. Il reste cependant encore beaucoup d'inconnues à éclaircir, notamment sur l'interprétation des données issues de ces horloges. Les études cliniques doivent à présent déterminer si une réduction de l'âge épigénétique induit par une intervention se traduit réellement en une longévité accrue.

Remerciements

Je remercie le National and Medical Research Council of Australia, l'Australian Research Council, ainsi que la fondation Jack Brockhoff pour avoir financé les projets qui m'ont permis de mener à bien ma recherche sur le vieillissement épigénétique.

► Références bibliographiques

- Belsky D.W. *et al.*, 2022. DunedinPACE, a DNA methylation biomarker of the pace of aging. *eLife*, 11, e73420.
- Chondronasiou D. *et al.*, 2022. Multi-omic rejuvenation of naturally aged tissues by a single cycle of transient reprogramming. *Aging Cell*, 21, e13578.
- Horvath S., 2013. DNA methylation age of human tissues and cell types. *Genome Biology*, 14, 3156.
- Horvath S., Raj K., 2018. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics*, 19, 371-384.
- Kabacik S. *et al.*, 2022. The relationship between epigenetic age and the hallmarks of aging in human cells. *Nature Aging*, 2, 484-493.
- Kerepesi C., Zhang B., Lee S.-G., Trapp A., Gladyshev V.N., 2021. Epigenetic clocks reveal a rejuvenation event during embryogenesis followed by aging. *Science Advances*, 7.
- López-Otín C., Blasco M.A., Partridge L., Serrano M., Kroemer G., 2023. Hallmarks of aging: An expanding universe. *Cell*, 186, 243-278.

Lu A.T. *et al.*, 2022. DNA methylation GrimAge version 2. *Aging*, 14, 9484-9549.

Tan Q., Mengel-From J., Christensen K., 2022. Sex difference in epigenomic instability during human aging. *Aging (Albany NY)*, 14 (15), 5960-5961.

Tarkhov A.E. *et al.*, 2022. Nature of epigenetic aging from a single-cell perspective. *bioRxiv*.<https://doi.org/10.1101/2022.09.26.509592>

Trapp A., Kerepesi C., Gladyshev V.N., 2021. Profiling epigenetic age in single cells. *Nature Aging*, 1, 1189-1201.

Zhu T. *et al.*, 2022. A pan-tissue DNA methylation atlas enables in silico decomposition of human tissue methylomes at cell-type resolution. *Nature Methods*, 19, 296-306.

Chapitre 9

Mécanismes épigénétiques des interactions hôte-agent pathogène

CHRISTOPH GRUNAU, ISABELLE FUDAL, NADIA PONTS

Les informations épigénétiques peuvent être stockées dans au moins quatre supports différents. Trois d'entre eux sont chromosomiques, c'est-à-dire indissociables des chromosomes et donc immobilisés dans le noyau :

- la méthylation de la position 5 du cycle pyrimidine de la cytosine résultant en une 5-méthylcytosine (5mC) ;
- les modifications post-traductionnelles des histones et leurs associations avec d'autres protéines chromatiniennes ;
- la position d'un locus dans les noyaux interphasiques.

La quatrième est extra-chromosomique et peut donc se déplacer à l'intérieur du noyau, entre noyau et cytoplasme et entre cellules, voire entre organismes : c'est l'ARN non codant. Il a été démontré que ces porteurs d'informations épigénétiques peuvent être modifiés lors de l'infection d'un organisme par des agents pathogènes (virus, bactéries, champignons ou parasites). Par ailleurs, les mécanismes de défense des organismes infectés sont, au moins en partie, régulés par des mécanismes épigénétiques, ceux-ci pouvant à leur tour être la cible des agents pathogènes. Enfin, lors de l'infection, les agents pathogènes et les parasites sécrètent des molécules facilitant l'infection dont la production est en partie régulée de façon épigénétique. Nous présentons dans ce chapitre quelques informations clés concernant les mécanismes épigénétiques impliqués dans les interactions hôtes-agents pathogènes.

►► Modification de l'épigénome à la suite d'une infection par des agents pathogènes

Des modifications de la méthylation de l'ADN dans les îlots CpG peuvent entraîner des modifications de la transcription, de l'épissage et probablement d'autres éléments régulant l'expression des gènes. Il a été montré que les modifications de la méthylation de l'ADN régulent l'expression de gènes clés du système immunitaire, à la fois dans les réponses immunitaires innées et dans le système immunitaire adaptatif (Obata *et al.*, 2015). Plusieurs études ont montré que la méthylation de l'ADN change lorsque des êtres humains ou des animaux sont infectés par des parasites. Par exemple, DiNardo *et al.* (2018) ont rapporté que l'infection par des parasites humains, le ver

plat *Schistosoma haematobium* et le ver nématode *Ascaris lumbricoides*, induisait des modifications de la méthylation de l'ADN dans 751 gènes qui affectaient la réponse du système immunitaire de l'hôte, notamment les lymphocytes T CD4⁺.

Modification de l'épigénome à la suite d'une infection par des bactéries

De nombreuses infections bactériennes provoquent des modifications de la méthylation de l'ADN, par exemple les infections par *Mycobacterium tuberculosis* et *M. leprae* (résumé par Pereira *et al.*, 2016), *Helicobacter pylori*¹⁴, des *Escherichia coli* uropathogènes ou *Campylobacter rectus* (revu par Bierne *et al.*, 2012). Les cellules hôtes n'ont pas nécessairement besoin d'être infectées pour qu'on observe des modifications de la méthylation de l'ADN; parfois, un changement environnemental montrant la présence de bactéries peut suffire. Cela a notamment été montré chez des invertébrés tels que l'huître *Crassostrea gigas*, chez qui l'exposition à des bactéries non pathogènes a laissé des traces dans le patron de méthylation de l'ADN sur trois générations (Fallet *et al.*, 2022).

La présence de bactéries pathogènes ou non pathogènes affecte non seulement la méthylation de l'ADN de l'hôte, mais également les modifications des histones. Ainsi, des macrophages exposés à des doses sublétales de la toxine de l'anthrax (*Bacillus anthracis*) ont montré une désacétylation de l'histone H3 au niveau de la lysine en position 27 (H3K27ac), une marque canonique associée à l'euchromatine et aux activateurs, et l'exposition aux lipopolysaccharides a stimulé la méthylation de la lysine en position 4 de l'histone H3, associée en général à des gènes actifs (résumé par Pereira *et al.*, 2016).

Modification de l'épigénome à la suite d'une infection virale

Les changements de méthylation de l'ADN probablement les mieux caractérisés sont ceux observés lors des infections virales, par exemple par le virus de l'hépatite, par le papillomavirus humain, par le virus de la chorioméningite, et par le virus de l'herpès simplex. La méthylation de l'ADN a même été utilisée pour prédire la progression de la maladie et la mortalité chez les patients atteints du VIH (Zhang *et al.*, 2018). La pandémie de Covid-19 a suscité un intérêt accru pour ces traces épigénétiques laissées par une infection virale, ici provoquant des syndromes respiratoires. Il a été démontré que l'infection par le virus de la grippe H1N1-09 ou par le SARS-CoV entraînent une augmentation de la triméthylation de l'histone H3 lysine 4 (H3K4me3) et l'épuisement de la triméthylation de l'histone H3 de la lysine en position 27 (H3K27me3) dans un sous-ensemble de gènes. En revanche, l'infection par le H5N1-VN1203 et par le coronavirus du syndrome respiratoire du Moyen-Orient (MERS-CoV) entraînent l'enrichissement en H3K27me3 et épuisent le H3K4me3 pour un autre sous-ensemble de gènes, ce qui rend probable l'existence de signatures épigénétiques spécifiques, même pour des virus qui sont phylogénétiquement proches et/ou provoquent des syndromes très similaires.

Menachery *et al.* (2018) ont montré que l'infection par le virus MERS-CoV entraînait une méthylation complète de l'ADN de trois promoteurs de gènes associés à la présentation de l'antigène (*HLA-A*, *B2M* et *PDIA3*) *in vitro*. Fait intéressant, la protéine virale

14. À noter que cette modification se produit dans des gènes suppresseurs de tumeurs et fournissent ainsi un lien fonctionnel entre l'infection et l'induction de cancers gastriques.

codée par l'*ORF8* de SARS-CoV-2 interagit directement avec la DNMT1 humaine, responsable de la maintenance de la méthylation au cours de la vie d'un organisme (Gordon *et al.*, 2020). Sur la base de ces découvertes, l'hypothèse a été émise que le virus MERS-CoV et SARS-CoV-2 modulent activement le paysage épigénétique du génome humain, mais il se pourrait aussi que l'épigénome humain agisse sur l'infection virale en tant que stress environnemental.

Même si la nature exacte des mécanismes qui relie l'infection virale à l'épigénétique n'est pas déterminée, il est désormais clair que les patrons de la méthylation de l'ADN dans les gènes liés à l'odorat distinguent les patients ayant été infectés par le Covid-19 des personnes non infectées (Huoman *et al.*, 2022). Dans l'ADN circulant librement dans le sang suite à une lyse de ses cellules d'origine, on trouve des patrons de la méthylation de l'ADN spécifiquement liés aux dommages du foie, des poumons et des reins (Cheng *et al.*, 2021) après une infection par le SARS-CoV-2. Il est désormais possible de prédire le pronostic de l'infection au Covid-19 sur la base d'une signature spécifique de méthylation de l'ADN qui se produit principalement dans les gènes codant pour la réponse de l'interféron à l'infection virale (Castro de Moura *et al.*, 2021). Finalement, l'infection par le Covid-19 accélère « l'horloge » épigénétique (Cao *et al.*, 2022; voir chapitre 8), ce qui pourrait contribuer au syndrome de Covid long.

Étant donné les preuves selon lesquelles les virus et autres agents pathogènes laissent une trace épigénétique de leur passage chez l'hôte, on a fait l'hypothèse que la vaccination, qui représente une sorte d'« infection atténuée », pouvait également avoir un effet sur l'épigénome. En effet, des liens ont été établis entre les marques épigénétiques et une réponse immunitaire durable (« immunité entraînée »; Netea *et al.*, 2020). Les premières preuves sont venues de la vaccination avec le vaccin Calmette-Guerin-Bacilli (BCG), destinée à prévenir l'infection tuberculeuse, qui a également montré des effets protecteurs contre les infections bactériennes et virales non apparentées à la tuberculose (voir pour revue Bierne *et al.*, 2012). En effet, dans les monocytes circulants des nourrissons, une signature de méthylation de l'ADN a été identifiée qui persistait plus de douze mois après la vaccination par le BCG (Bannister *et al.*, 2022). Par ailleurs, de nouvelles techniques permettant d'analyser, cellule par cellule, l'accessibilité à la chromatine dans des cellules immunitaires (*single cell ATAC-Seq*) ont montré qu'après une infection par le SARS-CoV-2, les monocytes CD14⁺ et CD16⁺ ont une structure chromatinienne altérée qui pourrait être à la base d'une immunité entraînée contre une infection secondaire (You *et al.*, 2021).

► Mémoire épigénétique et *priming* des défenses des plantes lors de l'infection

Des changements épigénétiques peuvent aussi être induits chez les plantes lors de l'infection par différents agents pathogènes (bactéries, champignons, insectes). Ainsi, une étude ancienne de Guseinov et Vanuyshin (1975) décrivait une forte altération du niveau de méthylation des cytosines chez le coton à la suite de l'infection par le champignon pathogène *Verticillium dahliae*. De même, l'infection d'*Arabidopsis thaliana* par *Pseudomonas syringae* pv. *tomato* entraîne une hypométhylation de l'ADN de plusieurs régions génomiques, notamment des régions péri-centromériques et des régions contenant des rétrotransposons.

Certaines de ces modifications sont impliquées dans le mécanisme de *priming*, qui permet à la plante, à la suite d'une première exposition à un stress biotique ou abiotique, de réagir plus rapidement et efficacement suite à l'exposition à de nouveaux stress. Le *priming* peut non seulement être induit par des stress biotiques, mais également lors du contact avec un micro-organisme non pathogène ou un *stimulus* chimique (par exemple, la chitine, des molécules de défense des plantes). Parmi les modifications épigénétiques impliquées dans le *priming*, on note notamment un rôle de l'accumulation de la modification H3K4me3 et de l'acétylation de plusieurs lysines de l'histone H4, qui permettent l'ouverture de la chromatine au niveau de gènes de défense et ainsi leur activation plus rapide, *via* des facteurs de transcription spécifiques, lors de l'exposition à de nouveaux stress.

Certaines de ces modifications épigénétiques peuvent par la suite être transmises à la descendance et permettre une résistance accrue à l'infection par des virus, des bactéries, des champignons ou des insectes (voir pour revue López Sánchez *et al.*, 2021). Ainsi, une étude à l'échelle du génome de *A. thaliana* de la méthylation de l'ADN sur trois générations après une infection par *P. syringae* pv. *tomato* a révélé une résistance acquise à la bactérie liée à la présence de méthylation de l'ADN dans un contexte CpG dans des régions géniques (*transgenerational-induced resistance*), le niveau de méthylation étant plus important en cas d'infection par la bactérie sur plusieurs générations.

► Rôle de la méthylation de l'ADN dans l'induction de réactions de défense chez les plantes

Lors de l'infection par un agent pathogène, les plantes peuvent reconnaître des éléments conservés de ces agents pathogènes (par exemple, la chitine des champignons ou la flagelline des bactéries), ce qui déclenche un premier niveau de résistance basale à l'infection. Les plantes peuvent également reconnaître spécifiquement des molécules sécrétées par l'agent pathogène, appelées « effecteurs », éléments clés de la pathogenèse qui modulent l'immunité de la plante et facilitent l'infection. Cette reconnaissance est médiée par des protéines de résistance (R) membranaires ou cytoplasmiques capables de reconnaître, directement ou indirectement, certains de ces effecteurs, ce qui déclenche un ensemble de réponses de défense regroupées sous le terme *effector-triggered immunity* (ETI), conduisant fréquemment à une mort cellulaire rapide et localisée appelée « réponse hypersensible » (HR). Plusieurs molécules de signalisation végétale régulent la défense des plantes contre leurs agresseurs, dont l'acide salicylique (SA) et le méthyl jasmonate (meJA). La signalisation dépendante du SA est essentielle à l'établissement d'une résistance locale et systémique à l'infection par des agents pathogènes biotrophes (se développant sur cellules vivantes), tandis que la signalisation dépendante du meJA est induite en réponse aux blessures mécaniques ou provoquées par des herbivores et à l'infection par des agents pathogènes nécrotrophes (se développant sur cellules mortes). Les voies de défense dépendantes du SA et du meJA sont mutuellement antagonistes.

Chez les plantes, deux types de méthylation de l'ADN ont été décrits : la méthylation de la 5-cytosine (5mC) et, dans de plus rares cas, une méthylation de la N₆-adénosine. La 5mC est retrouvée à la fois dans des contextes CpG, CHG et CHH (H correspondant à n'importe quel nucléotide à l'exception d'un G). Classiquement, dans les régions

codantes, on retrouve principalement de la méthylation en contexte CpG, alors qu'on peut trouver de la méthylation dans tous les types de contextes nucléotidiques dans les régions répétées ou riches en éléments transposables.

Le rôle de la méthylation de l'ADN dans les mécanismes de défense des plantes a été largement étudié, principalement *via* des traitements chimiques ou *via* la génération de mutants hypométhylés ou hyperméthylés (voir pour revue Ramos-Cruz *et al.*, 2021). Une corrélation négative entre le niveau de méthylation de l'ADN et la résistance des plantes à l'infection par des agents pathogènes biotrophes ressort de ces études. Ainsi, lorsqu'on diminue le niveau de méthylation de l'ADN, on augmente la résistance des plantes à l'infection par des agents pathogènes biotrophes (par exemple, lors de l'infection de *A. thaliana* par *P. syringae* pv. *tomato* ou par *Hyaloperonospora arabidopsidis*, ou lors de l'infection du riz par *Xanthomonas oryzae* pv. *Oryzae*) et on observe une augmentation d'expression de gènes de défense et une activation de la voie hormonale du SA. À l'inverse, des mutants de *A. thaliana* hyperméthylés sont plus sensibles à l'infection par *H. arabidopsidis* ou *Fusarium oxysporum*. Il a alors été postulé que la déméthylation de l'ADN pouvait constituer un signal permettant d'induire plus rapidement l'expression de gènes de défense lors d'une réinfection (effet de *priming*). Il convient de noter que cet effet de la méthylation de l'ADN sur la résistance à l'infection est inversé pour les agents pathogènes nécrotrophes. On ne sait cependant pas s'il s'agit d'un effet direct de l'hypo-méthylation de l'ADN sur la sensibilité à l'infection ou d'un effet indirect lié à l'antagonisme entre la voie hormonale médiée par le SA et celle médiée par le meJA, induite lors de l'infection par un agent pathogène nécrotrophe.

Il n'est en effet pas facile de distinguer les effets causaux des effets pléiotropes générés lors de la mutation de gènes impliqués dans des mécanismes épigénétiques. Afin de pallier cet effet, un outil remarquable a été développé chez *A. thaliana* : une collection de lignées quasi isogéniques épigénétiquement recombinantes (lignées epiRIL). Ces lignées ont été obtenues en croisant une lignée sauvage (Col-0) avec des mutants de fond génétique Col-0 affectés dans leur niveau de méthylation de l'ADN (mutants de l'ADN méthyltransférase 1, *met1* ou de *decrease in DNA methylation 1*, *ddm1*). Les descendants de ces croisements sont quasi isogéniques, mais diffèrent pour leurs profils de méthylation de l'ADN (figure 9.1). Ces lignées sont des outils précieux pour identifier des régions du génome de *A. thaliana* différentiellement méthylées impliquées dans la résistance à l'infection par différents agents pathogènes. Ainsi, un criblage d'une centaine de lignées de cette collection d'epiRIL pour leur sensibilité à l'infection par *H. arabidopsidis* a permis d'identifier quatre régions du génome hypométhylées (epiQTL, pour *epigenetic quantitative trait loci*) expliquant 60% de la sensibilité à l'infection (Furci *et al.*, 2019). Cependant, peu de gènes sont présents dans les régions génomiques correspondant à ces epiQTL, suggérant une régulation de gènes de défense localisés dans d'autres régions du génome (régulation en *trans*) par un mécanisme non déterminé.

Rôles des ARN non codants dans les interactions plantes-agents pathogènes

Les ARN non codants jouent un rôle important dans la reprogrammation de l'expression des gènes chez les plantes en réponse à l'infection par une large gamme d'agents pathogènes (voir pour revue Huang *et al.*, 2019). Ainsi, différents types de

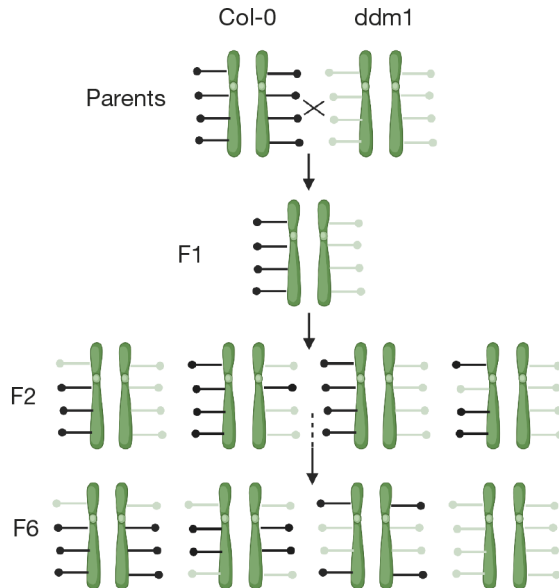


Figure 9.1. Stratégie utilisée pour générer des lignées recombinantes épigénétiques (epiRIL).

Les lignées epiRIL sont issues d'un croisement entre un génotype sauvage d'*Arabidopsis thaliana* (Col-0) présentant de l'ADN méthylé (épingles noires) et un mutant *ddm1* (*decrease in DNA methylation 1*), généré à partir du génotype sauvage Col-0 appauvri en méthylation (épingles vertes). Les plantes issues de ce croisement (F1) ont été autofécondées pour générer plusieurs centaines de lignées F2. Les lignées portant l'allèle homozygote *DDM1* de type sauvage ont été maintenues. Ces lignées ont été multipliées pendant plusieurs générations, ce qui a permis d'obtenir des lignées avec des épigénotypes en mosaïque. Les lignées epiRIL sont donc génétiquement identiques mais épigénétiquement distinctes.

petits ARN régulant spécifiquement l'expression de gènes de résistance cytoplasmiques (NLR, *nucleotide-binding site leucine-rich repeat*) et de gènes de défense ont été identifiés. Parmi eux, les petits ARN interférents hétérochromatiniens (hc-siRNA pour *heterochromatic small interfering RNAs*, généralement d'une longueur de 24 à 30 nucléotides) dérivent de régions hétérochromatiniennes riches en répétitions et en éléments transposables et agissent sur la méthylation *de novo* de l'ADN et/ou sur des modifications d'histones dans ces régions. Les clusters de NLR ou de gènes codant des récepteurs impliqués dans les défenses basales des plantes sont souvent associés à des éléments répétés ou transposables et sont ainsi régulés par des hc-siRNA. Cette régulation fine de l'expression des gènes de résistance et des mécanismes de défense permet de maintenir un équilibre entre croissance végétative et résistance à l'infection, et de limiter ainsi le coût de *fitness* lié à l'immunité végétale. On peut citer comme exemple les travaux de Deng *et al.* (2017) concernant la caractérisation du locus de résistance PIGMR-PIGMS du riz. Alors que *PIGMR* est exprimé constitutivement et confère une résistance large au champignon *M. oryzae*, *PIGMS* n'est exprimé que dans les fleurs et les panicules et empêche l'activation par homodimérisation de *PIGMR* dans ces organes, permettant ainsi d'optimiser la production de grains. L'expression de *PIGMS* est régulée par la présence de deux transposons miniatures (MITE1 et MITE2) dans son promoteur, par un fort niveau de méthylation

de l'ADN de ce promoteur pendant la croissance végétative et par la production de hc-siRNA dérivés de MITE1 et MITE2 également pendant la croissance végétative, ce contrôle étant levé dans les fleurs et panicules.

Par ailleurs, il a été montré que de petits ARN pouvaient être échangés entre la plante et ses agents pathogènes et moduler l'infection. Ainsi, chez *A. thaliana*, des vésicules extracellulaires contenant des petits ARN produits par la plante s'accumulent au niveau des sites d'interaction lors de l'infection par le champignon nécrotrophe *Botrytis cinerea*. Au moment du contact de la plante avec les cellules fongiques, le contenu de ces vésicules extracellulaires est libéré, permettant la modulation de gènes impliqués dans l'agressivité de l'agent pathogène par les petits ARN de *A. thaliana*, limitant ainsi l'infection (Cai *et al.*, 2018). De leur côté, les agents pathogènes ont également la capacité d'utiliser la machinerie de *silencing* de la plante à leur avantage. Ainsi, *H. arabidopsidis* et *B. cinerea* sécrètent de petits ARN dans les cellules de *A. thaliana* et utilisent la machinerie de *silencing* médiée par la protéine Argonaute de la plante pour moduler l'expression de gènes de défense de l'hôte, et ainsi faciliter l'infection (Weiberg *et al.*, 2013; Dunker *et al.*, 2020).

Rôle des modifications d'histones dans les interactions plantes-agents pathogènes

Comme présenté dans le chapitre 2, les modifications post-traductionnelles des queues N-terminales d'histones constituent un véritable code épigénétique. Parmi ces modifications, les acétylations/désacétylations médiées par des histones acétyltransférases (HAT) et désacétylases (HDAC) et les méthylations/déméthylations médiées par des histones méthyltransférases et déméthylases jouent un rôle majeur chez les plantes.

Tout comme pour la méthylation de l'ADN, le rôle de l'acétylation des histones dans les mécanismes de défense des plantes a été largement étudié, principalement *via* la génération de mutants perte-de-fonction *HDAC* ou *HAT* (voir pour revue Ramos-Cruz *et al.*, 2021). Il ressort de ces études une implication de l'acétylation des histones dans les réponses de défense basales des plantes et dans leur résistance à l'infection. Ainsi, les mutants *hdac* montrent une résistance accrue à l'infection par des agents pathogènes biotrophes et une induction de gènes de défense liés au SA (par exemple dans le cas d'interactions entre *A. thaliana* et *P. syringae* pv. *tomato* ou entre le riz et *Magnaporthe oryzae* ou *X. oryzae* pv. *oryzae*). À l'inverse, des mutants *hat* chez *A. thaliana* présentent une sensibilité accrue à l'infection par *P. syringae* pv. *tomato*. Les protéines HDAC ont donc été décrites comme des régulateurs (directs ou indirects) négatifs de la voie du SA et les protéines HAT comme des régulateurs positifs. Certains champignons phytopathogènes produisent des inhibiteurs d'HDAC comme la toxine HC produite par *Cochliobolus carbonum* indispensable à l'infection du maïs ou la dépudécine produite par *Alternaria brassicicola* lors de l'infection d'*A. thaliana*.

Le rôle des méthyltransférases et des déméthylases dans la résistance des plantes à l'infection est moins clair. Cela est sans doute dû à la diversité des protéines qui ont été étudiées et des méthylations d'histones qu'elles déposent, certaines marques ayant un effet répressur et d'autres un effet activateur (Ramos-Cruz *et al.*, 2021). On peut notamment noter l'effet de la présence de la marque répressive H3K9me2 à proximité des sites de polyadénylation des gènes de résistance *RPP7* et *RPP4* (*Recognition of Peronospora parasitica*) d'*A. thaliana*, qui empêche la maturation de transcrits courts

inactifs et favorise l'accumulation de transcrits fonctionnels de pleine longueur. L'équilibre dans la production de ces deux types de transcrits est modulé par la présence d'EDM2 (*enhanced downy mildew 2*), une protéine associée aux histones qui se fixe à la marque H3K9me2 au niveau d'éléments transposables présents à proximité et dans les gènes *RPP7* et *RPP4*. EDM2 régule également l'expression d'autres NLR, permettant ainsi de maintenir l'équilibre entre croissance végétative et résistance à l'infection. De même, la protéine ATRX1 (*alpha thalassemia/mental retardation syndrome X-linked*), qui présente une activité méthyltransférase et est responsable du dépôt du variant d'histone H3.3, régule l'expression d'un grand nombre de gènes de défense, de gènes de résistance et de facteurs de transcription de type WRKY (notamment WRKY70, impliqué dans la régulation de la voie du SA).

Régulation épigénétique de la production de molécules impliquées dans la pathogénèse

L'expression du pouvoir pathogène d'un micro-organisme capable d'infecter un hôte donné est intimement liée à la production de molécules diverses, protéines ou métabolites, appelées « effecteurs ». Ces effecteurs sont produits lorsque l'agent pathogène est au contact de son hôte et permettent l'infection soit en facilitant la pénétration de l'agent pathogène, sa propagation, son approvisionnement en nutriments, soit en agissant directement sur l'hôte en affaiblissant ses défenses, tel que présenté plus haut. Par exemple, les parasites apicomplexes (tous intracellulaires obligatoires) tels que l'agent du paludisme *P. falciparum* sécrètent un important arsenal d'effecteurs protéiques dans la cellule hôte pour les utiliser à leur avantage (voir pour revue Koning-Ward *et al.*, 2016). Notamment, les protéines *P. falciparum erythrocyte membrane protein 1* (PfEMP1) sont des protéines hypervariables codées par une famille d'une soixantaine de gènes *VAR* dont l'expression coordonnée à la surface des érythrocytes infectés permet au parasite d'échapper au système immunitaire de l'hôte. À chaque cycle cellulaire, un seul gène *var* est exprimé (jamais le même), les autres étant réprimés simultanément. Les loci des gènes *VAR* réprimés sont abondamment décorés par les marques hétérochromatiniennes H3K9me3, Heterochromatin Protein 1 (HP1) et H3K36me3, et colocalisent dans un secteur nucléaire transcriptionnellement inactif en périphérie du noyau grâce à la formation de nombreuses boucles, structurant ainsi le génome dans l'espace (voir pour revue Abel et Le Roch, 2019). Le locus de l'unique gène *VAR* actif est décoré par les marques histones H3K4me3, H3K9ac, H3K27ac et H4K8ac, et n'est pas localisé dans le même secteur nucléaire que les *VAR* inactifs.

Les champignons filamenteux phytopathogènes sont également de formidables producteurs d'effecteurs. Le champignon filamenteux *Fusarium graminearum* est un exemple d'agent pathogène du blé très efficace qui, en plus de causer des pertes de rendement très importantes, produit des mycotoxines toxiques par ingestion, dont le déoxynivalénol (DON), qui peuvent contaminer les grains d'apparence saine. Les mycotoxines sont des métabolites secondaires, c'est-à-dire non essentiels à la survie immédiate (par opposition aux métabolites primaires, essentiels au métabolisme cellulaire, comme par exemple les acides aminés) mais jouant néanmoins des rôles écologiques importants, dans la survie dans l'habitat du champignon (par exemple compétitions microbiennes) mais aussi dans l'infection. Le DON est notamment un effecteur métabolique permettant la colonisation rapide de l'hôte blé par *F. graminearum* (Mentges *et al.*, 2020).

Comme tous les métabolites secondaires fongiques, la voie de biosynthèse du DON est constituée de diverses enzymes catalysant les étapes de transformation métabolique. Les gènes codant ces enzymes, et bien souvent ceux codant des facteurs de transcription spécifiques de la voie de biosynthèse en question, sont rassemblés de façon contiguë à un locus commun du génome. On parle de «cluster de gènes de biosynthèse», ou BGC (*biosynthetic gene cluster*). L'expression précise et coordonnée des BGC est alors orchestrée de façon très fine en fonction de l'environnement dans lequel se trouve le champignon (facteurs biotiques et abiotiques, voir plus haut), en faisant notamment intervenir diverses combinaisons de marques histones (voir pour revue Etier *et al.*, 2022). Les BGC exprimés portent la marque H3K4me3, tandis que les BGC réprimés sont décorés par la marque H3K27me3. Les régulations par des marques histones pourraient également être indirectes, comme l'activation de la production de DON chez *F. graminearum*, potentiellement liée à une inhibition dépendante de la marque hétérochromatique H3K9me3 de l'expression d'un facteur répresseur en *trans*.

Le code définissant les effets des marques histones n'est toutefois pas universel. Par exemple, le BGC impliqué dans la biosynthèse de la mycotoxine stérigmatocystine par le champignon phytopathogène *Aspergillus nidulans* est réprimé en présence de H3K9me3. Au-delà de ces quelques exemples, l'étendue du répertoire des marques épigénétiques pouvant réguler la production d'effecteurs, et les combinatoires sous-jacentes, permet des régulations rapides et ajustées de la production d'effecteurs par des micro-organismes pathogènes en réponse à un vaste éventail de *stimuli* biotiques et abiotiques.

Les petits ARN jouent également un rôle important dans la mise en place de l'infection par des agents pathogènes et la colonisation de l'hôte. La localisation génomique dans des régions hétérochromatiniennes (riches en éléments répétées et en transposons) de certains d'entre eux, les hc-siRNA, a été évoquée plus haut. Les différentes familles de petits ARN peuvent en outre être elles-mêmes porteuses de modifications covalentes déposées post-transcriptionnellement, qui viennent affecter leur biogenèse et *in fine* leurs fonctions (voir pour revue Li *et al.*, 2021). Les miARN, par exemple, sont fréquemment méthylés en N₆ des adénosines (m6-A), jouant un rôle catalyseur dans le processus de maturation des miARN.

► Questions futures

Les progrès techniques et les coûts réduits des techniques de profilage épigénomique à haut débit ont permis de réaliser la caractérisation d'épigénomes au niveau individuel. Dans le domaine biomédical, cela a conduit à une démocratisation des outils de médecine individualisée, et le domaine le plus avancé est certainement l'évaluation du risque de cancer et le pronostic de l'issue du traitement. Nous savons maintenant que les maladies infectieuses laissent des traces dans l'épigénome, et nous savons également que certaines conditions épigénétiques sont liées à une plus grande résistance aux infections. C'est le cas chez les animaux et les plantes. En utilisant du «dépistage» à haut débit des profils épigénétiques à l'échelle de la population, il a été montré chez l'espèce modèle *A. thaliana* un effet de modifications épigénétiques (notamment la méthylation de l'ADN) sur la résistance à l'infection. Des recherches sur des espèces d'intérêt agronomique sont désormais nécessaires pour déterminer s'il est possible de sélectionner des plantes cultivées et des animaux d'élevage présentant des niveaux particuliers de

méthylation à des endroits spécifiques de l'ADN, afin d'améliorer la résistance aux maladies, tout en n'altérant pas la quantité et la qualité des aliments produits.

Par ailleurs, il a été montré que l'exposition à un stress biotique ou abiotique entraîne des modifications épigénétiques qui peuvent être transmises à la descendance, la rendant plus résistante à ces stress. Cependant, cette transmission générationnelle de résistance aux stress a principalement été étudiée chez des organismes modèles. Il est donc désormais nécessaire de s'intéresser à la stabilité de ce phénomène au cours des générations et pour une grande diversité de pathosystèmes. L'étude de ces mécanismes pourrait ouvrir la voie à de nouvelles stratégies de protection des écosystèmes exploités. Par exemple, moduler finement les miARN produits par la plante pourrait être une stratégie pour améliorer la tolérance aux stress (abiotiques ou biotiques) dans les cultures et les élevages d'importance économique. Une fois que nous saurons quels types de marques épigénétiques sont liés à une meilleure résistance/tolérance à la maladie, il sera également possible de modifier l'épigénome des animaux et des plantes exploités, et d'augmenter ou de diminuer les signaux épigénétiques à des sites spécifiques, sans modifier le génome de ces organismes. Cette approche est appelée « ingénierie épigénétique », ou édition de l'épigénome (voir chapitre 12), et n'a jusqu'à présent été utilisée que sur une poignée d'organismes montrant sa faisabilité. Il sera nécessaire de développer ces techniques dans des organismes d'élevage et d'évaluer les gains et les risques sanitaires et environnementaux associés à ces techniques, et d'encadrer correctement son utilisation.

Depuis la découverte de la présence de petits ARN dans l'environnement qui peuvent être ingérés par des agents pathogènes (nématodes, champignons), il est envisagé d'utiliser des petits ARN comme pesticides : sRNA (*small RNA*) ou dsRNA (*double-stranded RNA*). Cette stratégie, appelée « *spray-induced gene silencing* » (SIGS), a été testée avec succès pour lutter contre des agents pathogènes eucaryotes pendant la saison culturale (par exemple, *Sclerotinia sclerotiorum* ou *B. cinerea* sur feuilles de colza) ou après récolte (par exemple, *B. cinerea* sur feuilles de laitues, tomates, fraises et roses). Cependant, une des limites de ces traitements est la dégradation rapide des petits ARN dans l'environnement. La découverte récente d'exosomes impliqués dans le transfert de petits ARN entre hôtes et agents pathogènes pourrait aider à développer des vésicules extracellulaires artificielles, inspirées de ces exosomes, permettant des traitements à l'aide d'ARN artificiels ciblant les agents pathogènes dans le but de contrôler les maladies dans diverses cultures avant et après récolte, tout en les protégeant de la dégradation, ou en prenant comme cible spécifiquement certains vecteurs (insectes, mollusques, etc.) des maladies transmissibles.

Enfin, développer des stratégies durables de lutte contre les agents pathogènes implique de ne pas se cantonner aux interactions bipartites hôte-agent pathogène, mais de regarder également l'implication des mécanismes épigénétiques sur les interactions au sein de l'holobionte complet, c'est-à-dire entre l'hôte et tous les micro-organismes présents dans son environnement.

Remerciements

Nous remercions le réseau INRAE REAcTION (Réseau d'échange sur les mécanismes épigénétiques qui façonnent les interactions plante-bioagresseurs et organismes symbiotiques) soutenu par le département Santé des plantes et Environnement et le GDR3E (GDR Épigénétique en écologie et évolution) soutenu par le CNRS INEE/INSB.

► Références bibliographiques

- Abel S., Le Roch K.G., 2019. The role of epigenetics and chromatin structure in transcriptional regulation in malaria parasites. *Briefings in Functional Genomics*, 18 (5), 302-13.
- Bannister S., Kim B., Domínguez-Andrés J. *et al.*, 2022. Neonatal BCG vaccination is associated with a long-term DNA methylation signature in circulating monocytes. *Science Advances*, 8, eabn4002. <https://www.science.org/doi/10.1126/sciadv.abn4002>
- Bierne H., Hamon M., Cossart P., 2012. Epigenetics and bacterial infections. *Cold Spring Harbor Perspectives in Medicine*, 2 (12), a010272. <https://perspectivesinmedicine.cshlp.org/content/2/12/a010272>
- Cai Q., Qiao L., Wang M., He B., Lin F.-M., Palmquist J. *et al.*, 2018. Plants send small RNAs in extracellular vesicles to fungal pathogen to silence virulence genes. *Science (NY)*, 360, 1126-1129. <https://www.science.org/doi/10.1126/science.aar4142>
- Cao X., Li W., Wang T., Ran D., Davalos V., Planas-Serra L. *et al.*, 2022. Accelerated biological aging in COVID-19 patients. *Nature Communications*, 13 (2135). <https://doi.org/10.1038/s41467-022-29801-8>
- Castro de Moura M., Davalos V., Planas-Serra L. *et al.*, 2021. Epigenome-wide association study of COVID-19 severity with respiratory failure. *EBioMedicine*, 66, 103339. <https://doi.org/10.1016/j.ebiom.2021.103339>
- Cheng A.P., Cheng M.P., Gu W. *et al.*, 2021. Cell-free DNA tissues of origin by methylation profiling reveals significant cell, tissue, and organ-specific injury related to COVID-19 severity. *Med (NY)*, 2, 411-422.e5. <https://doi.org/10.1016/j.medj.2021.01.001>
- Deng Y., Zhai K., Xie Z., Yang D., Zhu X., Liu J. *et al.*, 2017. Epigenetic regulation of antagonistic receptors confers rice blast resistance with yield balance. *Science*, 355, 962-965. <https://doi.org/10.1126/science.aai8898>
- DiNardo A.R., Nishiguchi T., Mace E.M., Rajapakshe K., Mtetwa G., Kay A. *et al.*, 2018. Schistosomiasis Induces Persistent DNA Methylation and Tuberculosis-Specific Immune Changes. *The Journal of Immunology*, 201, 124-133. <https://doi.org/10.4049/jimmunol.1800101>
- Dunker F., Trutzenberg A., Rothenpieler J.S., Kuhn S., Pröls R., Schreiber T. *et al.*, 2020. Oomycete small RNAs bind to the plant RNA-induced silencing complex for virulence, *ELife*, 9e56096. <https://doi.org/10.7554/eLife.56096>
- Etier A., Dumetz F., Chéreau S., Ponts N., 2022. Post-translational modifications of histones are versatile regulators of fungal development and secondary metabolism. *Toxins*, 14 (5), 317. <https://doi.org/10.3390/toxins14050317>
- Fallet M., Montagnani C., Petton B., Dantan L., de Lorgeril J., Comarmond S. *et al.*, 2022. Early life microbial exposures shape the *Crassostrea gigas* immune system for lifelong and intergenerational disease protection. *Microbiome*, 10 (85). <https://doi.org/10.1186/s40168-022-01280-5>
- Furci L., Jain R., Stassen J., Berkowitz O., Whelan J., Roquis D. *et al.*, 2019. Identification and characterisation of hypomethylated DNA loci controlling quantitative resistance in *Arabidopsis*. *eLife*, 8e40655. <https://doi.org/10.7554/eLife.40655>
- Gordon D.E., Jang G.M., Bouhaddou M. *et al.*, 2020. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, 583, 459-468. <https://doi.org/10.1038/s41586-020-2286-9>
- Guseinov V.A., Vanyushin B.F., 1975. Content and localisation of 5-methylcytosine in DNA of healthy and wilt-infected cotton plants, *Biochimica et Biophysica Acta*, 395, 229-238.
- Huang C.-Y., Wang H., Hu P., Hamby R., Jin H., 2019. Small RNAs – Big players in plant-microbe interactions, *Cell Host and Microbe*, 26 (173-182). <https://doi.org/10.1016/j.chom.2019.07.021>
- Huoman J., Sayyab S., Apostolou E., Karlsson L., Porcile L., Rizwan M. *et al.*, 2022. Epigenetic re-wiring of pathways related to odour perception in immune cells exposed to SARS-CoV-2 *in vivo* and *in vitro*. *Epigenetics*, 17, 1875-1891. <https://doi.org/10.1080/15592294.2022.2089471>
- Koning-Ward T.F. de, Dixon M.W., Tilley L., Gilson P.R., 2016. Plasmodium species: Master renovators of their host cells. *Nat Rev Microbiol.*, 14 (8), 494-507. <https://doi.org/10.1038/nrmicro.2016.79>
- Li X., Peng J., Yi C., 2021. The epitranscriptome of small non-coding RNAs. *Noncoding RNA Res.* 26, 6(4):167-173. <https://doi.org/10.1016/j.ncrna.2021.10.002>

- López Sánchez A., Pascual-Pardo D., Furci L., Roberts M.R., Ton J., 2021. Costs and benefits of trans-generational induced resistance in *Arabidopsis*. *Frontiers in Plant Science*, 12. <https://doi.org/10.3389/fpls.2021.644999>
- Menachery V.D., Schäfer A., Burnum-Johnson K.E. *et al.*, 2018. MERS-CoV and H5N1 influenza virus antagonize antigen presentation by altering the epigenetic landscape. *Proc Natl Acad Sci USA*, 115, e1012. <https://doi.org/10.1073/pnas.1706928115>
- Mentges M., Glasenapp A., Boenisch M., Malz S., Henrissat B., Frandsen R.J.N. *et al.*, 2020. Infection cushions of *Fusarium graminearum* are fungal arsenals for wheat infection. *Molecular Plant Pathology*, 21 (8), 1070-87. <https://doi.org/10.1111/mpp.12960>
- Netea M.G., Domínguez-Andrés J., Barreiro L.B., Chavakis T., Divangahi M., Fuchs E. *et al.*, 2020. Defining trained immunity and its role in health and disease. *Nat Rev Immunol*, 20, 375-388. <https://doi.org/10.1038/s41577-020-0285-6>
- Obata Y., Furusawa Y., Hase K., 2015. Epigenetic modifications of the immune system in health and disease. *Immunology and Cell Biology*, 93, 226-232. <https://doi.org/10.1038/icb.2014.114>
- Pereira J.M., Hamon M.A., Cossart P., 2016. A lasting impression: Epigenetic memory of bacterial infections. *Cell Host Microbe*, 19, 579-582. <https://doi.org/10.1016/j.chom.2016.04.012>
- Ramos-Cruz D., Troyee A.N., Becker C., 2021. Epigenetics in plant organismic interactions. *Current Opinion in Plant Biology*, 61, 102060. <https://doi.org/10.1016/j.pbi.2021.102060>
- Weiberg A., Wang M., Lin F.-M., Zhao H., Zhang Z., Kaloshian I. *et al.*, 2013. Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. *Science*, 342, 118-123. <https://doi.org/10.1126/science.1239705>
- You M., Chen L., Zhang D., Zhao P., Chen Z., Qin E.Q. *et al.*, 2021. Single-cell epigenomic landscape of peripheral immune cells reveals establishment of trained immunity in individuals convalescing from COVID-19. *Nat Cell Biol*, 23, 620-630. <https://doi.org/10.1038/s41556-021-00690-1>
- Zhang X., Hu Y., Aouizerat B.E. *et al.*, 2018. Machine learning selected smoking-associated DNA methylation signatures that predict HIV prognosis and mortality. *Clinical Epigenetics*, 10 (155). <https://doi.org/10.1186/s13148-018-0591-z>

Chapitre 10

Mémoires épigénétiques et santé des plantes

PHILIPPE GALLUSCI, MARGOT M.J. BERGER

Le changement climatique se traduit par une augmentation de la fréquence et de l'intensité d'épisodes de stress intense, et par une montée progressive de la température globale. Ces phénomènes tels que les vagues de chaleur, la sécheresse, les inondations, ou la pression exercée par les agents pathogènes, influent directement sur la croissance et le rendement des plantes cultivées et mettent en danger la sécurité alimentaire mondiale.

Au-delà de la nécessaire adaptation des pratiques culturales, l'amélioration des plantes cultivées repose sur l'utilisation de la diversité génétique des espèces cultivées et des espèces sauvages apparentées pour répondre à de tels défis (Renzi *et al.*, 2022). En effet, elle vise à identifier les combinaisons d'allèles les plus favorables afin d'optimiser des caractères d'intérêt agronomique, tels que le rendement, la qualité des récoltes (fruits, graines) ou encore la tolérance aux stress abiotiques et/ou la résistance à différents agents pathogènes. Alors que ces traits phénotypiques sont particulièrement importants en agriculture, seule une partie de leur héritabilité semble déterminée par les variations génétiques (Brachi *et al.*, 2011). À cette héritabilité manquante, il faut ajouter une érosion progressive de la diversité génétique, du fait de processus récurrents de sélection variétale, ou encore lorsque les méthodes de création de nouvelles variétés reposent sur la multiplication végétative, comme pour la vigne et de nombreuses autres espèces d'arbres fruitiers ou forestiers.

Dans ce contexte, les mécanismes épigénétiques apparaissent comme une nouvelle source de variations phénotypiques hérissables, générées indépendamment des changements de la séquence d'ADN. En outre, ils jouent un rôle important dans la mémoire des plantes, capacité qu'elles ont acquise leur permettant de stocker les informations générées par un signal endogène ou lié à leur environnement, puis de les réutiliser (Gallusci *et al.*, 2022). Grâce à ces propriétés, les plantes sont souvent considérées comme des organismes ayant acquis une forme d'« intelligence », bien qu'elles ne soient pas pourvues de capacité cognitive (van Loon, 2016).

Dans ce chapitre, nous présentons un résumé de l'état des connaissances concernant la mémoire épigénétique des plantes, et des opportunités qu'elle ouvre dans le but d'améliorer la tolérance des plantes aux contraintes générées par le changement climatique. Nous discutons dans quelle mesure l'utilisation de la mémoire des végétaux et la mise en place d'approches telles que l'*epi-breeding* (amélioration épigénétique des plantes) pourraient être à la base de stratégies innovantes à même d'améliorer la santé des plantes.

► La mémoire des plantes : importance des mécanismes épigénétiques

Il est communément accepté que les plantes sont des organismes capables de stocker, puis de réutiliser de façon appropriée l'information qu'elles perçoivent de l'environnement. Elles possèdent des mécanismes de mémoires particulièrement performants qui reposent sur un ensemble de processus physiologiques, cellulaires, métaboliques et moléculaires. Parmi ceux-ci, des travaux de plus en plus nombreux montrent que les mécanismes épigénétiques sont l'un des supports majeurs de la mémoire des plantes (Gallusci *et al.*, 2022).

Chez les plantes, comme chez les autres eucaryotes, l'information épigénétique est portée par les variants d'histones, les modifications post-traductionnelles des histones (HPTM), et par la méthylation de l'ADN. Cette dernière présente certaines spécificités, puisque dans les génomes végétaux les cytosines peuvent être méthylées dans tous les contextes de séquences, qu'ils soient symétriques car méthylés sur les deux brins d'ADN, comme dans le cas des motifs CG et CHG (où H peut être A, T ou C), ou non symétriques pour les motifs CHH où seul un brin d'ADN porte la méthylation (figure 10.1 ; voir chapitre 2).

Ainsi, chez les végétaux, la mise en place et le maintien de la méthylation de l'ADN génomique sont en partie assurés par des mécanismes spécifiques. La méthylation *de novo* est réalisée dans les trois contextes de séquences par les mécanismes de RdDM (*RNA directed DNA methylation*) qui impliquent les enzymes DRM1 et 2 (*domain rearranged DNA methyltransferase*) et des petits ARN interférents (siRNA, *short interfering RNA*) de 24 nucléotides, ou par la chromométhylase 2, cette dernière agissant surtout dans les régions hétérochromatiques. Ces enzymes assurent aussi le maintien des méthylations CHH après réplication de l'ADN. Le maintien des méthylations dans les contextes CG et CHG est assuré respectivement par l'enzyme ADN méthyltransférase 1 (MET1) et par la chromométhylase 3 (CMT3). Enfin, les méthylations de l'ADN peuvent être perdues par dilution lors de la réplication de l'ADN, en l'absence de mécanismes de maintien ou par élimination active. Dans ce dernier cas, ce sont des ADN glycosylases/lyases qui éliminent les cytosines méthylées, qui seront ensuite remplacées par des cytosines non méthylées (figure 10.1). En ce qui concerne les HPTMs, les mécanismes permettant leur mise en place et leur élimination sont aussi bien décrits chez les végétaux (Lauria et Rossi, 2011). En revanche, peu de choses sont connues sur leur maintien au cours des divisions cellulaires, bien que dans le cas de la vernalisation, par exemple, les marques histones sont clairement conservées lors des mitoses des cellules méristématiques. Ainsi, l'information épigénétique contribue aux différents types de mémoires identifiées chez les végétaux. D'une part, elle est maintenue au cours des mitoses et participe à leur « mémoire somatique » (figure 10.2). Elle est aussi conservée, du moins en partie, dans les méristèmes lors de la multiplication végétative des plantes, et contribue à la mémoire transgénérationnelle des végétaux lors de la reproduction non sexuée (Latutrie *et al.*, 2019). D'autre part, une partie de l'information épigénétique peut être conservée à travers les méioses, puis la fécondation, et est ainsi susceptible de participer à la mémoire trans- et/ou intergénérationnelle pendant la reproduction sexuée. Enfin, l'existence d'une mémoire épigénétique transannuelle portée par les cellules des méristèmes a été suggérée dans les bourgeons dormants chez les plantes pérennes comme le peuplier (Le Gac *et al.*, 2018).

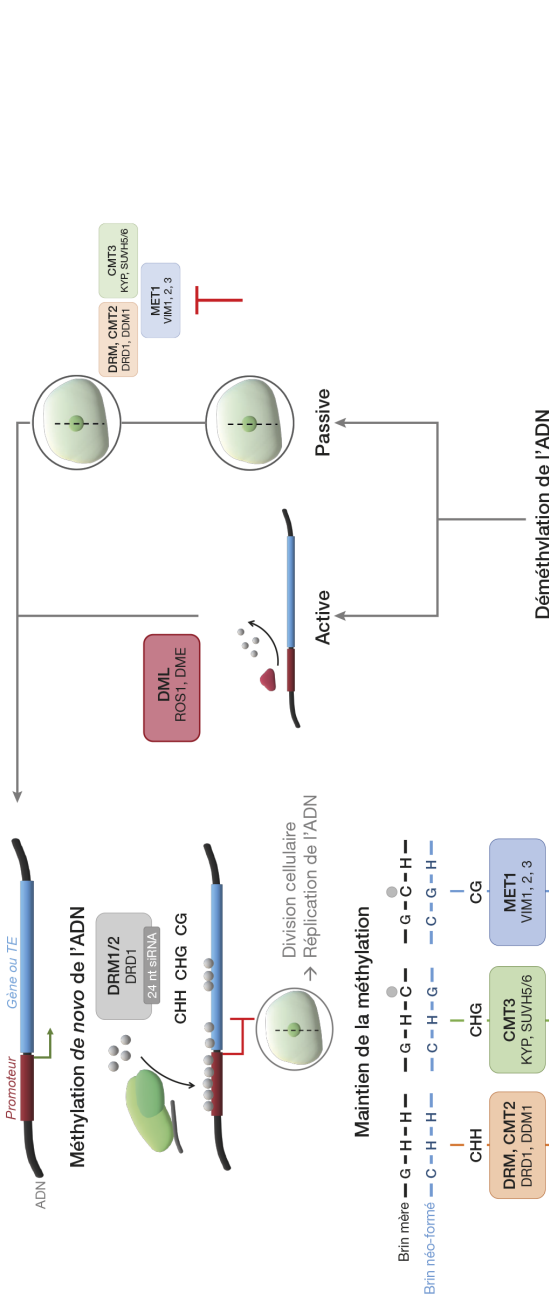


Figure 10.1. Mise en place et maintien de la méthylation chez les végétaux.

La mise en place de la méthylation, aussi appelée « méthylation *de novo* » sur l'ensemble du génome (promoteurs, gènes, élément transposable, régions non-codantes...), repose sur la voie de la méthylation dirigée par ARN (RdDM, *RNA-directed DNA methylation*), impliquant un ensemble d'acteurs incluant les protéines DRM1/2 et de petits ARN interférents (siRNAs). Le maintien de la méthylation à travers les divisions cellulaires dépend du contexte de méthylation des cytosines. La méthylation dans le contexte CG est maintenue par l'activité de la protéine MET1, alors que CMT3 et CMT2 sont responsables du maintien de la méthylation CHG. La méthylation dans le contexte CHH est maintenue *via* la voie des RdDM, impliquant également dans certains cas l'enzyme CMT2. La méthylation de l'ADN peut être perdue de manière passive et non ciblée après réplication de l'ADN si les mécanismes de maintien de la méthylation ne sont pas actifs. En outre, les cytosines méthylées peuvent être activement retirées par l'action d'ADN glycosylases/lyases, encore appelées déméthylases (DML), enzymes ciblant des loci spécifiques. La méthylation de l'ADN peut impacter l'expression des gènes en modulant l'accessibilité de la chromatine à la machinerie transcriptionnelle lorsqu'elle se trouve au niveau des promoteurs des gènes et, dans une moindre mesure, à l'intérieur des gènes. Elle est très forte au niveau des transposons et contribue à inhiber leur activité. NB : Les encadrés indiquent le nom des protéines actrices (en gras) dans les processus décrits, et des exemples d'intervenants qui leur sont associés.

DRM : Domains Rearranged Methylase; DRD : Defective in RNA Directed DNA methylation; CMT : Chromométhylase; MET : Methyltransferase; DDM1 : Decrease in DNA Methylation; KYP : kryptonite; VIM : Variant In Methylation; DML : DEMETER-like; DME : DEMETER; ROS1 : Repressor Of Silencing 1; TE : Élément Transposable, transposon.

En conclusion, les mécanismes épigénétiques sont porteurs d'une information qui est conservée, du moins en partie, suite aux divisions cellulaires (mitoses et éventuellement méiose), contribuant ainsi à la mémoire de la plante. En ce sens, ils peuvent contribuer à son acclimatation et/ou à son adaptation à l'environnement.

► L'acclimatation des plantes : mémoire somatique et santé des plantes

Le stress est souvent défini comme un changement des conditions environnementales, capable de perturber l'équilibre physiologique d'un organisme. Ces perturbations, qui peuvent altérer le développement de la plante, se traduisent souvent par une réduction

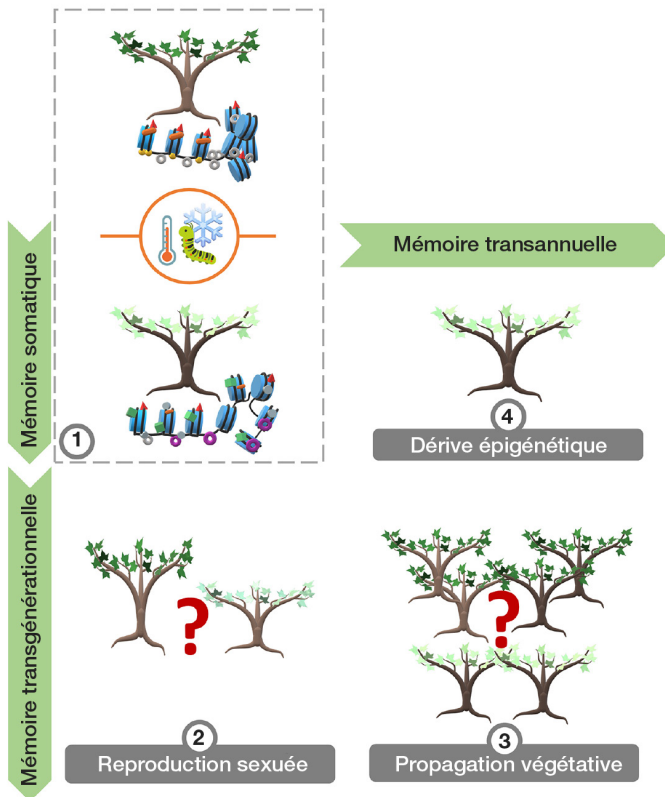


Figure 10.2. Mécanismes épigénétiques et mémoires des plantes (d'après Berger *et al.*, 2023).

L'information épigénétique générée au cours du développement et/ou en réponse aux stress environnementaux est mise en place lors de la différenciation cellulaire et est maintenue lors des divisions cellulaires (mitoses). Les marques épigénétiques (méthylation de l'ADN, modification des histones, etc.) peuvent être maintenues pendant la croissance annuelle de la plante, participant à la mémoire somatique (1). Les marques épigénétiques peuvent contribuer à la mémoire trans-intergénérationnelle, car elles peuvent être transmises à travers la méiose puis la fécondation sur au moins une génération lors de la reproduction sexuée (2), ou à une descendance obtenue par propagation végétative (3). De récents travaux sur les plantes pérennes suggèrent également l'existence d'une mémoire transannuelle (4), permettant de transmettre les marques épigénétiques à travers les cycles de développement et de dormance des plantes au fil des saisons. Cette mémoire transannuelle pourrait être à l'origine d'une dérive épigénétique.

du rendement des récoltes et/ou de leur qualité, et peuvent, dans les cas extrêmes, entraîner la mort. Cependant, les plantes, qui sont des organismes sessiles et de ce fait subissent leur environnement, ont mis en place des mécanismes leur permettant de répondre et de s'adapter aux stress environnementaux. Ces derniers entraînent le changement de nombreux paramètres physiologiques et moléculaires, tant au niveau cellulaire que de l'organisme entier (Zhang *et al.*, 2022). Dans ce contexte, les mécanismes épigénétiques semblent jouer un rôle important aussi bien dans la réponse des plantes aux stress que dans leur mémoire, qu'elle soit somatique ou impliquée dans le transfert d'information aux générations suivantes (figure 10.2).

Vernalisation : un exemple de mémoire somatique

La mémoire somatique des végétaux a été particulièrement bien décrite dans le cas de la vernalisation (figure 10.3). Il s'agit d'une période de froid nécessaire à la floraison de certaines plantes, parmi lesquelles *Arabidopsis*, le chou ou encore la betterave à sucre. Ce phénomène illustre la capacité des plantes à mémoriser les conséquences

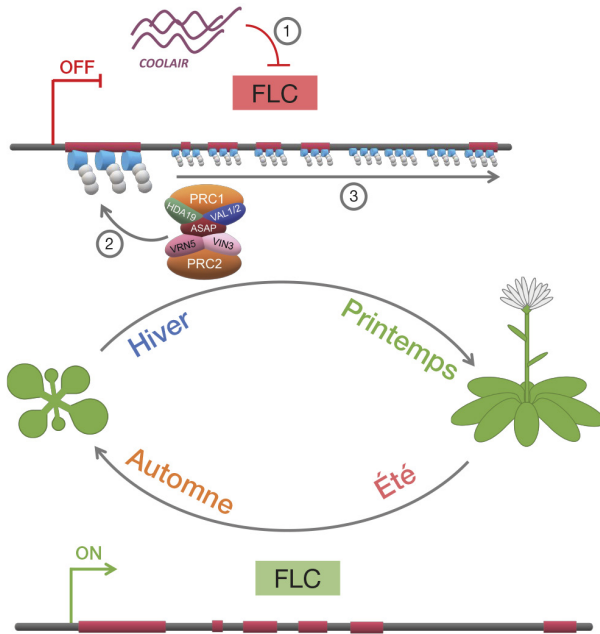


Figure 10.3. Contrôle de la vernalisation chez *Arabidopsis*.

La répression du gène *FLC* est essentielle pour permettre la floraison. La baisse des températures pendant l'hiver induit, dans un premier temps, l'expression de l'ARN antisens *COOLAIR* (1), puis la mise en place de la marque répressive H3K27me3 au niveau de trois nucléosomes situés dans le premier exon du gène *FLC* (2), marque qui s'étendra sur l'ensemble du gène lors de la hausse de température printanière (3). La mise en place de cette marque est médiée par l'action du complexe PRC2, recruté par le biais de l'interaction entre les protéines VRN5 et VIN3, toutes associées au PRC2, et de la protéine ASAP. Cette dernière protéine est associée au complexe PRC1 et aux protéines HDA19 et VAL1/2. Les marques associées à la répression du gène *FLC* sont ensuite perdues au cours de la reproduction sexuée.

FLC : *flowering locus C*; *PRC* : *Polycomb repressive complex*; *VRN5* : *vernalization 5*; *VIN3* : *vernalization insensitive 3*; *ASAP* : *apoptosis and splicing-associated protein*; *HDA19* : histone désacétylase 19; *VAL* : *viviparous1/ABI3-like*.

d'un signal environnemental sur leur processus de développement. Brièvement, dans le cas d'*Arabidopsis*, le traitement froid subi par la plante se traduit par l'induction forte mais transitoire de l'ARN antisens *COOLAIR*, associée à une baisse d'expression du gène *FLC* (*Flowering Locus C*), un répresseur de la floraison, et à la mise en place de la marque répressive H3K27me3 au niveau de 3 nucléosomes situés dans l'exon 1 du gène *FLC*, par le complexe PRC2 (*Polycomb Repressive Complex 2*) (Luo et He, 2020). Ce processus nécessite les protéines VAL1/2, qui interagissent avec l'histone désacétylase HDA19, le complexe PRC1 et la protéine ASAP. Cette dernière assure le recrutement du complexe PRC2, car elle interagit avec VIN3 et VRN5 qui y sont toutes deux associées. Ce processus se fait de façon aléatoire au niveau des cellules du méristème, le nombre de cellules affectées dépendant de la durée du froid. Pendant la période de froid, la marque H3K27me3 reste localisée au niveau de l'exon 1 et est transmise entre cellules par division mitotique, assurant ainsi la mémoire somatique de la répression du gène *FLC*. Elle s'étendra à l'ensemble du gène *FLC* lorsque les températures remonteront (figure 10.3). L'ensemble de ce processus permet la floraison au printemps lorsque les conditions environnementales sont favorables. En revanche, l'information épigénétique mise en place au gène *FLC* n'est pas transmise à la génération suivante et est reprogrammée au cours de la reproduction sexuée.

L'acclimatation des plantes : principes généraux

La mémoire somatique des végétaux peut aussi être mise en œuvre dans le cas de l'acclimatation des plantes (*priming*) (Liu *et al.*, 2022). Ce processus permet aux plantes d'optimiser leur réponse aux stress (*acquired stress tolerance*). Il se caractérise par une mémoire cellulaire des changements d'état physiologique de la plante, générés en réponse à un premier stress (biotique ou abiotique) ou suite à l'exposition à un éliciteur¹⁵. L'acclimatation peut être décomposée en différentes étapes :

- une première exposition au stress ou à une molécule éliciteur induit une réponse de la plante et potentialise sa mémoire ;
- une seconde phase dans des conditions de croissance normales permet à la plante de récupérer ;
- l'exposition à un second stress au cours duquel la mémoire de la plante est mobilisée amène une réponse plus rapide et/ou plus adaptée au stress (Mozgova *et al.*, 2019).

De nombreux événements peuvent induire l'acclimatation des plantes. Si un stress de faible intensité est en général suffisant, différents éliciteurs peuvent être utilisés, incluant des biostimulants naturels ou de synthèse, comme l'acide salicylique (SA) et ses analogues de synthèse (INA, ASM), l'acide β -aminobutyrique (BABA) naturellement produit par les plantes, ou encore la systémine. Certains micro-organismes présents dans la rhizosphère tels que les *plant growth promoting microorganisms* (PGPM) ou des agents de biocontrôle tels que *Trichoderma* jouent aussi un rôle dans l'acclimatation des plantes aux stress. Dans d'autres cas, des composés volatils émis suite à l'attaque d'une première plante peuvent préparer les plantes situées à proximité à répondre au stress. Ces procédés permettent à la fois une meilleure adaptation de la plante à son environnement, mais peuvent aussi avoir un effet positif sur leur croissance et leur rendement (Kerchev *et al.*, 2020).

15. Un éliciteur est une substance capable, dans certaines conditions, de stimuler des mécanismes de défense naturelle.

Un aspect essentiel au phénomène d'acclimatation réside dans la capacité des plantes à mémoriser l'information reçue lors des premiers stress et/ou suite à l'application d'un éliciteur. Si divers mécanismes participent à la mémoire des stress, le rôle des processus épigénétiques apparaît ici déterminant.

Mémoire somatique des stress : importance des mécanismes épigénétiques

Les mécanismes moléculaires impliqués dans l'acclimatation des plantes ne sont pas toujours bien décrits. Cependant, certains gènes semblent porter la mémoire des stress (figure 10.4). Ces « gènes mémoire » ont été classés en deux catégories : les gènes de type I, qui se caractérisent par le maintien d'un état d'expression induit par un premier stress, et les gènes de type II, qui présentent un niveau d'expression modifié lors de stress récurrents par rapport à celui induit par le premier stress (Bäurle, 2018). À l'heure actuelle, des travaux de plus en plus nombreux visent à déterminer la contribution des régulations épigénétiques dans la mémoire des stress, en particulier en lien avec la régulation des gènes mémoire (Crisp *et al.*, 2016).

Différentes études réalisées chez la plante modèle *Arabidopsis*, le riz ou encore *Brachypodium distachyon* montrent une modification du profil épigénétique de gènes dits « mémoire » en réponse aux stress hydrique, froid ou thermique (Liu *et al.*, 2022). Par exemple, chez la plante modèle *Arabidopsis*, l'analyse de la réponse à un stress hydrique répété a permis de mettre en évidence que les gènes « mémoire » sont enrichis dans la marque histone H3K4me3, caractéristique d'états transcriptionnels actifs, et dans la forme phosphorylée de l'ARN polymérase II (Ser5P). Ce phénomène, encore appelé « *stalling* »¹⁶, permet une induction accélérée de ces gènes lors d'un second stress. Ces résultats démontrent l'existence d'un lien entre optimisation de la réponse transcriptomique lors de stress répétés et modifications épigénétiques (Ding *et al.*, 2012).

De même, différentes études montrent une augmentation générale de la méthylation à la suite d'une attaque pathogène, ou après traitement avec un éliciteur. Cette augmentation pourrait résulter d'un ralentissement général de l'expression des gènes et du métabolisme, conséquence de l'infection par l'agent pathogène. L'hyperméthylation de l'ADN génomique est associée à la déméthylation des régions promotrices de gènes de résistance aux agents pathogènes, ou de transposons situés à proximité de ces gènes, et à l'activation des gènes concernés. L'importance de la méthylation de l'ADN est aussi attestée par les différences de sensibilité aux attaques pathogènes de mutants affectés dans l'homéostasie de l'ADN (*met1*; *ddc*, *drm1 drm2 cmt3*; *ros 1*, *repressor of silencing 1*) comparée au contrôle sauvage. Les mutants d'*Arabidopsis*, affectés soit dans la (dé) méthylation, soit dans l'acétylation/désacétylation des histones, présentent une sensibilité accrue à l'infection par différents agents pathogènes tels que *Pseudomonas syringae*. La modification des marques histones au niveau de gènes participant à la défense des plantes a été observée en réponse à des attaques par des agents pathogènes ou après un pré-traitement d'*Arabidopsis* avec un éliciteur de défense (Joseph et Shah, 2022).

16. Alors que le « *priming* » est un phénomène global impliquant une combinaison de processus moléculaires, hormonaux et métaboliques chez les plantes (décrits dans la figure 10.4), le « *stalling* », lui, décrit un mécanisme moléculaire particulier essentiellement associé aux « gènes mémoire » et à la régulation de leur transcription. Le « *stalling* » est un des processus impliqués dans le phénomène du « *priming* » chez les plantes.

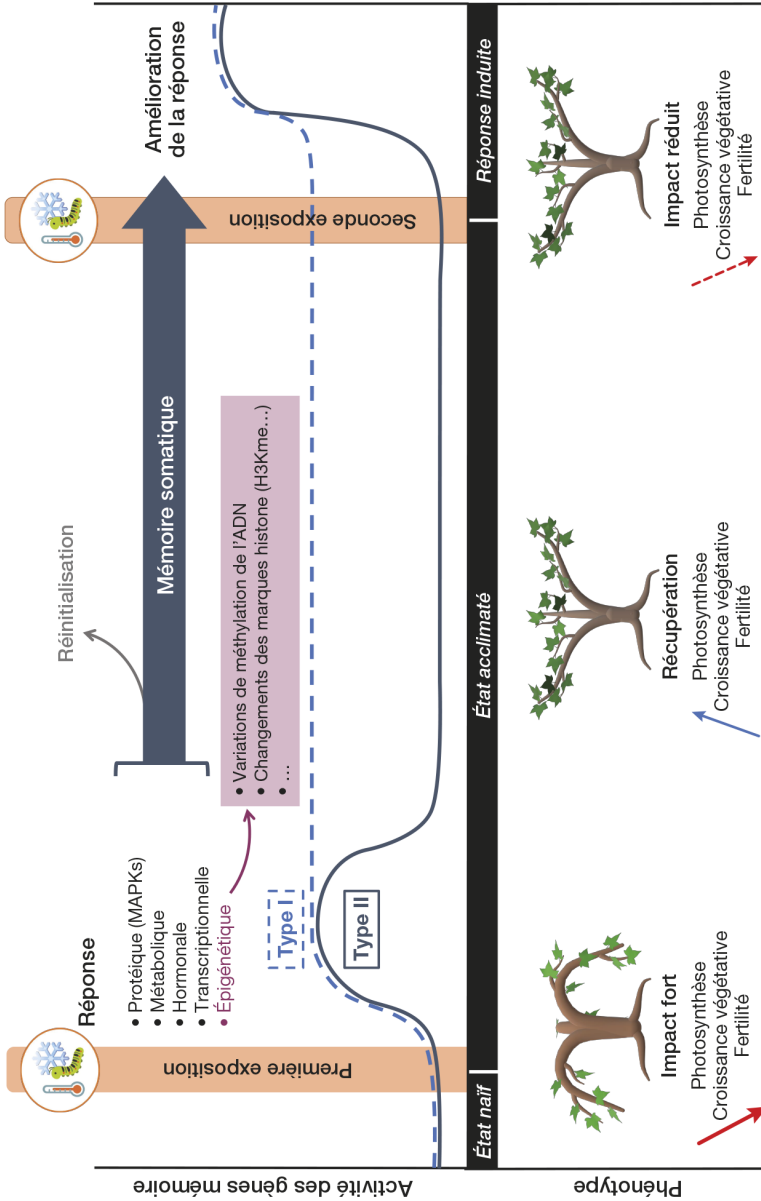


Figure 10.4. Acclimation et mémoire des plantes (d'après Berger *et al.*, 2023).

Exposées au stress pour la première fois, les plantes « naïves » peuvent montrer une réaction phénotypique très forte, affectant leur physiologie et leur croissance (photosynthèse, croissance végétative, fertilité, etc.). Cette réponse phénotypique est associée à des changements de l'état des protéines (modifications post-traductionnelles), métaboliques, hormonaux, transcriptomiques et épigénétiques. Ces modifications peuvent être réversées et revenir à leur état initial après le stress, ou être (partiellement) maintenues pendant la période de récupération, constituant une mémoire somatique des stress subis par la plante, éventuellement amenant à une modification de sa réponse aux stress. Dans ce contexte, les plantes possèdent des gènes dits « de mémoire ». Ces gènes sont répartis dans deux catégories : les gènes mémoire de « type I » (courbe bleue en pointillés) sont induits par un premier stress et leur expression est maintenue lors de la période de récupération. L'expression des gènes de « type II » (courbe bleu foncé) retourne à l'état initial pendant la période de récupération et leur réponse à un second stress est plus intense ou plus rapide. Les plantes acclimatées (« *primed* ») présentent une tolérance accrue par rapport aux plantes naïves, qui se caractérise par un impact réduit du stress sur la vigueur de la plante.

Dans ce contexte, il a été suggéré que les HPTMs soient les principales marques impliquées dans la mémoire à court/moyen terme des stress, tandis que la méthylation de l'ADN, plus stable, participerait plutôt à la mémoire à long terme (Pastor *et al.*, 2013). Cette dernière pourrait éventuellement être transmise aux générations suivantes, contribuant ainsi à l'acclimatation trans- et intergénérationnelle des plantes.

► Mémoire épigénétique inter-transgénérationnelle des stress : conséquences pour la santé des plantes

Mémoire inter et/ou transgénérationnelle des stress et reproduction sexuée

S'il est communément admis que les marques épigénétiques peuvent être transmises de façon stable par reproduction sexuée au fil des générations, l'importance des mécanismes épigénétiques dans la mémoire inter et/ou transgénérationnelle des stress fait encore débat (Quadrana et Colot, 2016).

En effet, pour qu'une information épigénétique générée au cours de la vie d'une plante puisse être transmise par reproduction sexuée à la génération suivante, plusieurs conditions doivent être remplies :

- l'information épigénétique reçue par la plante au cours de son développement doit être mémorisée au niveau des cellules du méristème apical caulinaire. Cette condition est nécessaire tant pour la mémoire somatique (cas de la vernalisation) que pour la transmission entre générations d'une information épigénétique. En effet, chez les plantes, la définition des lignées germinales se fait de façon tardive, lors de la formation de la fleur, les cellules à l'origine des gamétophytes, et donc des gamètes, dérivant directement des cellules du méristème apical caulinaire;
- lors de la formation des gamétophytes mâles et femelles, les cellules subissent la méiose, ce qui provoque une réorganisation chromatinienne potentiellement à l'origine d'une perte d'information épigénétique;
- la formation des gamètes entraîne une modification de l'information épigénétique, car les plantes possèdent des mécanismes qui limitent la transmission d'information épigénétique non désirée. En effet, d'une part, les protéines MOM1 et DDM1 limitent la transmission de l'information épigénétique parentale par le gamète femelle (Iwasaki et Paszkowski, 2014). D'autre part, les mécanismes de (dé)méthylation permettent de reprogrammer la méthylation dans le gamète mâle. Cependant, il n'existe pas comme chez les animaux d'élimination globale de la méthylation de l'ADN lors de la reproduction (voir pour une revue récente Ono et Kinoshita, 2021).

L'existence d'une transmission inter-transgénérationnelle d'information épigénétique a cependant été démontrée lors de la création de populations epiRIL (*epigenetic recombinant inbred lines*; lignées épigénétiques recombinantes; voir chapitre9) chez *Arabidopsis* (voir paragraphe « Diversité épigénétique et amélioration des plantes cultivées ») et est aussi démontrée par l'existence d'épiallèles stables chez les plantes (Catoni et Cortijo, 2018). La transmission d'informations épigénétiques générées par les stress a également été démontrée chez *Arabidopsis* pour différents stress abiotiques incluant les stress UV, osmotique ou hydrique, ou encore après un voyage spatial, mais aussi chez le riz après exposition aux métaux lourds ou à un stress hydrique, mais

sur une seule génération (transmission intergénérationnelle). Le gamète femelle joue un rôle essentiel dans la transmission de cette information (Wibowo *et al.*, 2016). En revanche, il n'existe pas à ce jour de démonstration d'une transmission transgénérationnelle (plusieurs générations) de l'information épigénétique générée par une situation de stress (voir pour revue récente Gallusci *et al.*, 2022).

En conséquence, si les conditions de croissance parentales peuvent influencer le comportement de la population obtenue en première génération, en absence de stress dans la génération suivante cette information ne semble pas conservée. Ceci montre la capacité des plantes à « oublier » une information qui ne semble plus nécessaire. En matière d'application, il paraît nécessaire d' « éliciter » les plantes à chaque génération et de façon régulière pour maintenir une forme de mémoire.

Mémoire épigénétique des stress et multiplication végétative

Chez les végétaux à propagation non sexuée, la descendance est générée à partir des plantes parentales sans méiose ni fécondation. Dans cette situation, l'information épigénétique peut être transmise à la descendance par les cellules situées dans les méristèmes ou bien par apomixie (forme de multiplication asexuée, sans fécondation ni méiose, qui permet la formation de la graine à partir des cellules de l'appareil reproducteur). Il en est de même lorsque ce mode de propagation est utilisé en agriculture ou en agroforesterie, comme chez la pomme de terre (tubercule), les arbres fruitiers, la vigne ou encore le peuplier (boutures).

La capacité à transmettre une information épigénétique par multiplication végétative (reproduction non sexuée des plantes effectuée à partir d'un organe de type feuille, tige ou bourgeon) a initialement été montrée lors de la régénération par organogenèse (obtention d'un organe, tige ou racine, directement à partir d'un fragment d'organe, ou à partir d'un amas de cellules non différenciées appelé « cals ») de plantes d'*Arabidopsis in vitro*. Ce processus implique une dédifférenciation cellulaire suivie d'une reprogrammation des cellules afin de former les axes caulinaires et racinaires. Une partie des empreintes épigénétiques de l'organe d'origine est conservée et influence le phénotype des plantes nouvellement générées (Wibowo *et al.*, 2018). Ceci démontre clairement à la fois la transmission d'informations épigénétiques parentales et leur influence sur le développement de la plante.

De façon similaire, l'existence d'une transmission de marques épigénétiques induites par les stress a été démontrée chez le trèfle blanc, le pissenlit ou encore le peuplier, bien que les modalités de reproduction non sexuée diffèrent entre ces plantes (Yang et Kim, 2016). Par exemple, des boutures, collectées à partir de peupliers de Lombardie ayant poussé dans des environnements différents, ont été mises en culture dans un même lieu afin de comparer leur phénologie et leurs paramètres épigénétiques. Les marques épigénétiques parentales sont retrouvées dans les boutures, confirmant une transmission de l'information épigénétique parentale (Vanden Broeck *et al.*, 2018). De la même manière, l'analyse des profils de méthylation de feuilles de vigne de clones cultivés dans différents vignobles argentins montre une influence plus marquée de l'origine clonale des plantes que de leur environnement (Varela *et al.*, 2021). Ces travaux suggèrent bien l'existence d'une transmission d'information épigénétique au cours de la multiplication des plantes, reflétant le contexte de croissance des plantes parentales, bien que leur influence sur le phénotype de la descendance n'ait pas été déterminée.

À l'inverse, d'autres études suggèrent que cette transmission est soit limitée, soit progressivement remplacée du fait des nouvelles conditions de croissance de la descendance. Par exemple, l'étude des profils de méthylation de vignes prélevées à travers différentes régions australiennes montre que ceux-ci varient en fonction de l'emplacement géographique et des modes de gestion des vignobles, suggérant que l'environnement de la plante en façonne le méthylome (Xie *et al.*, 2017).

En conclusion, le paysage épigénétique de plantes obtenues par reproduction non sexuée semble être le résultat non seulement de l'origine parentale des plantes, mais aussi de l'environnement dans lequel elles se développent.

► Diversité épigénétique et amélioration des plantes cultivées

L'amélioration génétique des plantes prend tout son sens dans le cas des plantes ayant une diversité génétique importante et un temps de génération assez court. Elle est beaucoup plus complexe à mettre en œuvre chez les plantes dont le temps de génération est long et/ou pour lesquelles le greffage et/ou la propagation clonale jouent un rôle important. Dans ce dernier cas, la diversité phénotypique observée n'est pas uniquement d'origine génétique. En effet, les mutations restent souvent à l'état hétérozygote et silencieux, sauf cas particulier (mutation dominante, mutations successives sur les deux allèles). La variabilité épigénétique pourrait alors, chez de telles plantes, jouer un rôle essentiel dans la formation de la diversité phénotypique. Par exemple, la diversité clonale de *Vitis vinifera* (cv Malbec) a récemment été associée à des changements de méthylation, illustrant par ailleurs l'importance des marqueurs épigénétiques dans la diversité intravariétale (Varela *et al.*, 2021).

L'analyse du rôle de la variabilité épigénétique sur la définition des phénotypes et son utilisation éventuelle pour améliorer la santé des plantes sont cependant complexes. En effet, dans de très nombreux cas, les variations épigénétiques sont associées à des variations génétiques et leurs contributions respectives se superposent (voir par exemple Quadrona et Colot, 2016). Cependant, les études réalisées ces dernières années sur *Arabidopsis* ont démontré que la variabilité phénotypique, incluant la tolérance aux stress, est en partie définie par des paramètres épigénétiques. Ces conclusions s'appuient sur l'utilisation de lignées qui diffèrent essentiellement par leur information épigénétique, les epiRIL (figure 10.5A). Les lignées epiRIL d'*Arabidopsis* ont été générées en croisant une plante sauvage avec une plante de même génome, mais portant une mutation sur le gène codant soit pour la protéine MET1, soit pour *decrease in DNA methylation* (DDM1). Les hybrides obtenus en première génération (F1) ont des génomes identiques au génome parental, mais des épigénomes hybrides, puisqu'ils reçoivent un épigénome de type sauvage, et un épigénome hypométhylé provenant du parent muté. Après rétrocroisement et/ou autofécondation sur au moins 7 générations pour fixer les variations épigénétiques, les lignées avec des différences épigénétiques stables sur un même fond génétique sont générées. Ces lignées peuvent alors être utilisées pour analyser les conséquences phénotypiques de la diversité épigénétique générée. Ainsi, les travaux sur ces deux populations epiRIL montrent que les plantes présentent de nombreuses variations phénotypiques pour lesquelles des epiQTL ont été déterminés (forme des feuilles, moment de floraison, taille des plantes; Catoni et Cortijo, 2018). La population epiRIL dérivée du mutant *met1* a aussi été testée pour la tolérance à différents

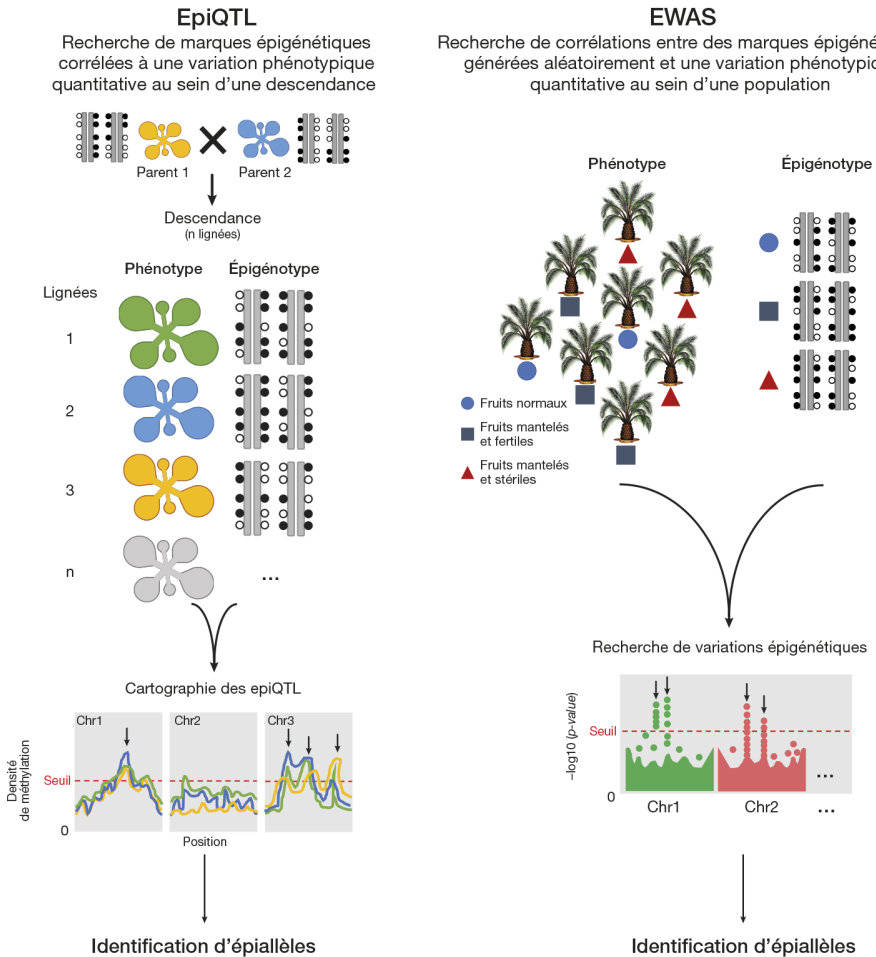


Figure 10.5. Recherche d'association entre variations épigénétiques et différences phénotypiques dans différents types de populations de plantes.

La recherche par association d'un lien entre un caractère phénotypique et des variations épigénétiques peut se faire de différentes manières, en fonction du type de population utilisé. A. La recherche d'epiQTL (*epigenetic quantitative trait loci*) permet d'identifier ce lien au sein de populations issues du croisement de parents ayant des caractéristiques phénotypiques et épigénétiques distinctes. L'analyse conjointe de la ségrégation des lignées produites et de leur épigénome permet d'identifier des epiQTL, et potentiellement des épiallèles. B. L'EWAS (*epigenome-wide association study*) permet de mettre en évidence une corrélation entre variations épigénétiques et caractéristiques phénotypiques quantitatives au sein d'une population non structurée, composée de plantes portant des épiallèles générés de façon aléatoire. Cette méthode a notamment permis d'identifier l'épiallèle responsable du phénotype « mantelé » du fruit du palmier à huile (Ong-Abdullah *et al.*, 2015).

stress, par exemple le stress salin, et montre une plus grande amplitude de réponse que les populations parentales. Plus récemment, l'analyse de la population epiRIL dérivée du mutant *ddm1* montre que des régions hypométhylées localisées dans les régions péri-centromériques des chromosomes confèrent à *Arabidopsis* une résistance quantitative à *Hyaloperonospora arabidopsidis*. Bien que les mécanismes ne

soient pas encore décrits, il semble que certaines epiRIL sont pré-acclimatées au stress, ce qui permet d'activer les gènes de résistance plus rapidement, induisant de fait une réponse précoce à l'infection (Furci *et al.*, 2019).

En revanche, cette approche reste longue à mettre en œuvre, et n'est pas applicable à toutes les plantes d'intérêt agronomique. En effet, une population epiRIL reste complexe à obtenir pour la plupart des plantes cultivées, pour lesquelles la génération de mutants affectés dans l'homéostasie de la méthylation de l'ADN ne semble pas toujours possible. Elle ne peut pas être développée sur les plantes pérennes comme la vigne, ou les arbres fruitiers, pour lesquels les temps de génération sont longs (plusieurs années dans de nombreux cas). La génération de variant(s) épigénétique(s) reste cependant envisageable en utilisant des approches alternatives qui permettent de générer des épiallèles de façon aléatoire. De telles approches incluent le traitement d'explants cultivés *in vitro* en présence d'agents inhibant la méthylation (5 azacytidine, zébularine), ou l'utilisation de contraintes environnementales pour générer de la variabilité épigénétique. Comme les populations générées par de telles approches sont non structurées et la génération d'épiallèles aléatoire, il n'est pas possible de rechercher des epiQTL par approche d'épigénétique quantitative (Catoni et Cortijo, 2018), et l'analyse des effets de ces variations épigénétiques sur le phénotype des plantes nécessite l'utilisation d'approches particulières. Le développement de l'EWAS (*epigenomic wide association study*; figure 10.5B) répond à cet objectif (Rakyan *et al.*, 2011).

► Conclusion

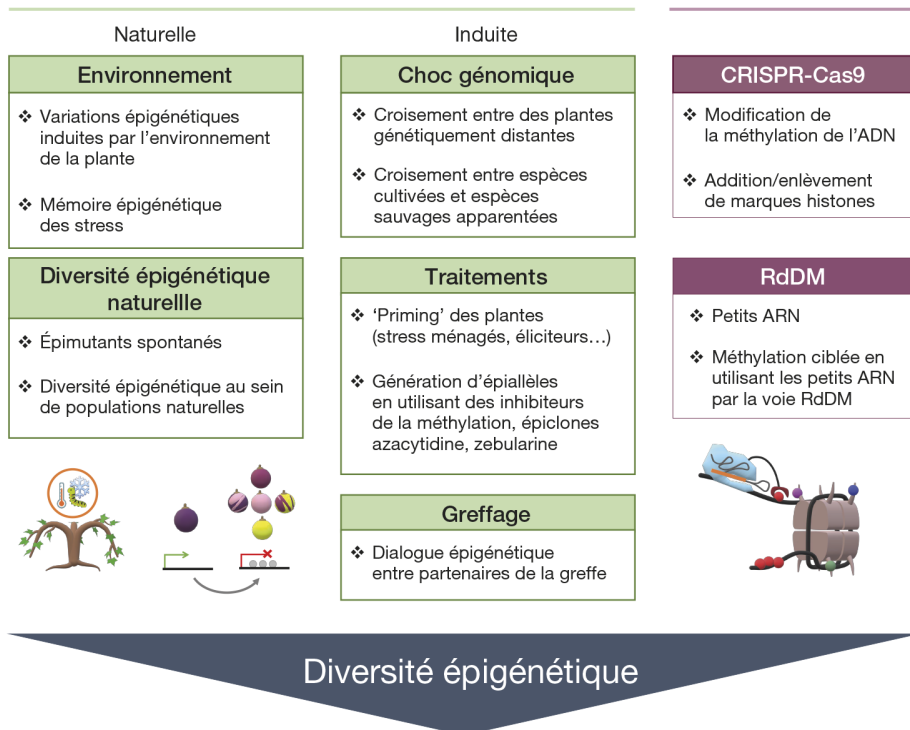
La capacité des plantes à se souvenir des événements passés et à réutiliser cette information ouvre des perspectives particulièrement prometteuses pour l'adaptation des cultures aux enjeux climatiques. Il devient en effet envisageable de «préparer» les plantes à un environnement stressant, sans qu'il soit nécessaire de mettre en place des stratégies d'amélioration génétique, souvent longues et coûteuses, mais en s'appuyant sur la diversité épigénétique et la mémoire qu'elle porte. La figure 10.6 résume différentes approches qui peuvent être utilisées pour atteindre cet objectif.

Une première stratégie consiste à préparer les plantes à mieux répondre aux stress en les acclimatant soit en appliquant des stress ménagés, soit par traitement à l'aide de molécules chimiques utilisées comme éliciteurs. Ces traitements peuvent être effectués au niveau des graines (*seed priming*) ou de la plante, et utilisent la mémoire somatique des plantes afin d'optimiser leur réponse à des stress ultérieurs (voir par exemple Nair *et al.*, 2022). Leur utilisation reste cependant encore limitée, car les mécanismes qui sous-tendent la réponse aux stress induite suite aux mécanismes d'acclimatation des plantes sont souvent encore mal compris, et leur reproductibilité est variable. Pourtant, leur utilisation associée à l'agriculture de précision pourrait permettre d'améliorer les performances au champ. De façon générale, il existe dans les populations naturelles de plantes une variabilité épigénétique souvent mal caractérisée, même si des épiallèles naturels ont été décrits (Catoni et Cortijo, 2018). Cette variabilité, qui traduit les différences d'environnement des plantes, est difficile à exploiter car souvent associée à des différences de la séquence d'ADN. Il est donc difficile de déterminer l'impact réel des différences d'état épigénétique sur le phénotype des plantes. Lorsque des plantes génétiquement distantes sont croisées, comme souvent dans les stratégies d'amélioration génétique des plantes, cela génère un choc génomique potentiellement à l'origine de variations épigénétiques, et donc d'une diversité épigénétique dans la descendance.

Recherche et création de variabilité épigénétique

A. Non ciblée

B. Ciblée



Identification de plantes tolérantes au stress

Figure 10.6. Recherche et création de variabilité génétique : un nouveau moyen pour améliorer la santé des plantes.

Il existe cependant de nombreuses stratégies permettant de générer des épiallèles stables afin de modifier des caractères d'intérêt agronomique, en particulier en lien avec la tolérance des plantes aux stress. Parmi celles-ci, la création de populations de type epiRIL semble un objectif difficile à atteindre pour de nombreuses plantes cultivées, soit parce que les temps de génération sont longs, soit parce que les mutants nécessaires pour générer des épigénomes modifiés sont difficilement viables et/ou stériles. Malgré cela, il s'agit d'une stratégie intéressante qui mériterait d'être testée chez certaines plantes cultivées, au regard des résultats obtenus chez *Arabidopsis*, indiquant que certaines epiRIL obtenues sont pré-acclimatée à certains stress (Furci *et al.*, 2019). Cependant, d'autres stratégies sont susceptibles de donner des résultats comparables. Par exemple, le traitement de plantules ou de jeunes explants avec des inhibiteurs de la méthylation de l'ADN (zébularine ou aza-cytidine) pourrait permettre de générer des épiallèles de façon aléatoire sur le génome. En générant des populations de taille suffisante et en recherchant les associations entre variations épigénétiques et phénotypes par EWAS, il devrait aussi être possible d'identifier des loci dont les différences de statut épigénétique sont associées à une meilleure tolérance ou réponse des plantes au stress. L'existence d'un dialogue épigénétique entre les partenaires de la greffe

(Rubio *et al.*, 2022) permet aussi d'envisager d'utiliser cette méthode pour générer de nouveaux épiallèles dans le greffon et/ou dans le porte-greffe. La greffe pourrait enfin être combinée avec la pré-acclimatation des plantes mères aux stress, dans le but de favoriser la formation de variations épigénétique conférant aux plantes une meilleure tolérance à ces stress.

Il faut ajouter à ces stratégies non ciblées la possibilité de modifier localement le profil épigénétique en développant des approches ciblées. La logique de ce type de stratégie suppose d'avoir identifié un ou plusieurs loci dont le changement d'état épigénétique pourrait être associé à un caractère d'intérêt. Différents travaux montrent la possibilité d'utiliser des outils moléculaires (dCas9 ou d'autres plateformes) pour atteindre cet objectif (Shin *et al.*, 2022; voir chapitre 12). Cependant, de telles approches nécessitent de disposer d'un système de transformation/régénération performant, avec la possibilité d'éliminer le transgène utilisé pour générer les épi-variations. Par ailleurs, la culture *in vitro* peut générer des variations épigénétiques non contrôlées éventuellement avec des effets délétères sur le développement des plantes, comme dans le cas du palmier à huile (Ong-Abdullah *et al.*, 2015). Une autre possibilité serait d'adresser spécifiquement la méthylation à l'aide du système RdDM à un locus spécifique pour générer des épiallèles hyperméthylés.

À l'heure actuelle, l'utilisation de la mémoire des plantes pour améliorer les performances des plantes cultivées au champ, en particulier en réponse aux stress et aux agressions d'agents pathogènes, en est encore à ses débuts. Il est nécessaire d'approfondir la connaissance des mécanismes impliqués pour optimiser les traitements permettant d'acclimater les plantes aux stress qu'elles subissent.

► Références bibliographiques

- Bäurle I., 2018. Can't remember to forget you: Chromatin-based priming of somatic stress responses. *Seminars in Cell and Developmental Biology*, 83, 133-139.
- Berger M.M.J., Stammitt L., Carrillo N., Blancquaert E., Rubio B., Teyssier E., Gallusci P., 2023. Epigenetics: an innovative lever for grapevine breeding in times of climatic changes. *OENO One*, 57(2), 265–282. <https://doi.org/10.20870/oeno-one.2023.57.2.7405>
- Brachi B., Morris G.P., Borevitz J.O., 2011. Genome-wide association studies in plants: The missing heritability is in the field. *Genome Biology*, 12 (10), 232.
- Catoni M., Cortijo S., 2018. EpiRILs: Lessons from *Arabidopsis*. In M. Mirouze, E. Bucher, P. Gallusci (eds), *Advances in Botanical Research*, 88, 87-116, Academic Press.
- Crisp P.A., Ganguly D., Eichten S.R., Borevitz J.O., Pogson B.J., 2016. Reconsidering plant memory: Intersections between stress recovery, RNA turnover, and epigenetics. *Science Advances*, 2 (2), e1501340.
- Ding Y., Fromm M., Avramova Z., 2012. Multiple exposures to drought “train” transcriptional responses in *Arabidopsis*. *Nature Communications*, 3, 740.
- Furci L., Jain R., Stassen J., Berkowitz O., Whelan J., Roquis D. *et al.*, 2019. Identification and characterisation of hypomethylated DNA loci controlling quantitative resistance in *Arabidopsis*. *eLife*, 8. <https://doi.org/10.7554/eLife.40655>
- Gallusci P., Agius D.R., Moschou PN., Dobránszki J., Kaiserli E., Martinelli F., 2022. Deep inside the epigenetic memories of stressed plants. *Trends in Plant Science*. <https://doi.org/10.1016/j.tplants.2022.09.004>
- Hemenway E.A., Gehring M., 2023. Epigenetic regulation during plant development and the capacity for epigenetic memory. *Annual Review of Plant Biology*. <https://doi.org/10.1146/annurev-arplant-070122-025047>

- Iwasaki M., Paszkowski J., 2014. Epigenetic memory in plants. *The EMBO Journal*, 33 (18), 1987-1998. <https://doi.org/10.15252/embj.201488883>
- Joseph J.T., Shah J.M., 2022. Biotic stress-induced epigenetic changes and transgenerational memory in plants. *Biologia*, 77 (8), 2007-2021.
- Kerchev P., van der Meer T., Sujeeth N., Verlee A., Stevens C.V., Van Breusegem F., Gechev T., 2020. Molecular priming as an approach to induce tolerance against abiotic and oxidative stresses in crop plants. *Biotechnology Advances*, 40, 107503.
- Latutrie M., Gourcilleau D., Pujol B., 2019. Epigenetic variation for agronomic improvement: An opportunity for vegetatively propagated crops. *American Journal of Botany*, 106 (10), 1281-1284.
- Lauria M., Rossi V., 2011. Epigenetic control of gene regulation in plants. *Biochimica et Biophysica Acta*, 1809 (8), 369-378.
- Le Gac A.-L., Lafon-Placette C., Chauveau D., Segura V., Delaunay A., Fichot R. *et al.*, 2018. Winter-dormant shoot apical meristem in poplar trees shows environmental epigenetic memory. *Journal of Experimental Botany*, 69 (20), 4821-4837.
- Liu H., Able A.J., Able J.A., 2022. Priming crops for the future: Rewiring stress memory. *Trends in Plant Science*, 27 (7), 699-716.
- Luo X., He Y., 2020. Experiencing winter for spring flowering: A molecular epigenetic perspective on vernalization. *Journal of Integrative Plant Biology*, 62 (1), 104-117.
- Mozgova I., Mikulski P., Pecinka A., Farrona S., 2019. Epigenetic mechanisms of abiotic stress response and memory in plants. In R. Alvarez-Venegas, C. De-la-Peña, J.A. Casas-Mollano (eds), *Epigenetics in Plants of Agronomic Importance: Fundamentals and Applications: Transcriptional Regulation and Chromatin Remodelling in Plants*, Springer International Publishing, 1-64.
- Ong-Abdullah M., Ordway J.M., Jiang N., Ooi S.E., Kok S.Y., Sarpan N. *et al.*, 2015. Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature*, 525 (7570), 533-537.
- Nair A.U., Bhukya D.P.N., Sunkar R., Chavali S., Allu A.D., 2022. Molecular basis of priming-induced acquired tolerance to multiple abiotic stresses in plants, *Journal of Experimental Botany*, 73(11), 3355-3371, <https://doi.org/10.1093/jxb/erac089>
- Ono A., Kinoshita T., 2021. Epigenetics and plant reproduction: Multiple steps for responsibly handling succession. *Current Opinion in Plant Biology*, 61, 102032.
- Pastor V., Luna E., Mauch-Mani B., Ton J., Flors V., 2013. Primed plants do not forget. *Environmental and Experimental Botany*, 94, 46-56.
- Quadrona L., Colot V., 2016. Plant transgenerational epigenetics. *Annual Review of Genetics*. <https://www.annualreviews.org/doi/abs/10.1146/annurev-genet-120215-035254>
- 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.
- Renzi J.P., Coyne C.J., Berger J., von Wettberg E., Nelson M., Ureta S. *et al.*, 2022. How could the use of crop wild relatives in breeding increase the adaptation of crops to marginal environments? *Frontiers in Plant Science*, 13, 886162.
- Rubio B., Stammitti L., Cookson S.J., Teyssier E., Gallusci P., 2022. Small RNA populations reflect the complex dialogue established between heterograft partners in grapevine, *Horticulture Research*, 9, uhab067, <https://doi.org/10.1093/hr/uhab067>
- Shin H., Choi W.L., Lim J.Y., Huh J.H., 2022. Epigenome editing: Targeted manipulation of epigenetic modifications in plants. *Genes and Genomics*, 44 (3), 307-315.
- Vanden Broeck A., Cox K., Brys R., Castiglione S., Ciatelli A., Guarino F. *et al.*, 2018. Variability in DNA methylation and generational plasticity in the Lombardy poplar, a single genotype worldwide distributed since the Eighteenth Century. *Frontiers in Plant Science*, 9, <https://doi.org/10.3389/fpls.2018.01635>
- van Loon L.C., 2016. The intelligent behavior of plants. *Trends in Plant Science*, 21 (4), 286-294.
- Varela A., Ibañez V.N., Alonso R., Zavallo D., Asurmendi S., Gomez Talquenca S. *et al.*, 2021. Vineyard environments influence Malbec grapevine phenotypic traits and DNA methylation patterns in a clone-dependent way. *Plant Cell Reports*, 40 (1), 111-125.

- Wibowo A., Becker C., Durr J., Price J., Spaepen S., Hilton S. *et al.*, 2018. Partial maintenance of organ-specific epigenetic marks during plant asexual reproduction leads to heritable phenotypic variation. *Proceedings of the National Academy of Sciences of the USA*, 115 (39), e9145-e9152.
- Wibowo A., Becker C., Marconi G., Durr J., Price J., Hagemann J. *et al.*, 2016. Hyperosmotic stress memory in *Arabidopsis* is mediated by distinct epigenetically labile sites in the genome and is restricted in the male germline by DNA glycosylase activity. *eLife*, 5. <https://doi.org/10.7554/eLife.13546>
- Xie H., Konate M., Sai N., Tesfamichael K.G., Cavagnaro T., Gilliam M. *et al.*, 2017. Global DNA methylation patterns can play a role in defining terroir in grapevine (*Vitis vinifera* cv. Shiraz). *Frontiers in Plant Science*, 8, 1860.
- Yang Y.Y., Kim J.G., 2016. The optimal balance between sexual and asexual reproduction in variable environments: a systematic review. *Han'gug Hwan'gyeong Saengtae Haghoeji = Korean Journal of Environment and Ecology*, 40(1), 1-18.
- Zhang H., Lang Z., Zhu J.-K., 2018. Dynamics and function of DNA methylation in plants. *Nature Reviews. Molecular Cell Biology*, 19 (8), 489-506.
- Zhang H., Zhu J., Gong Z., Zhu J.-K., 2022. Abiotic stress responses in plants. *Nature Reviews. Genetics*, 23 (2), 104-119.

Chapitre 11

Épigénétique chez les poissons d'élevage

AUDREY LAURENT, DELPHINE LALLIAS, LUCIE MARANDEL,
CATHERINE LABBÉ, PIERRE BOUDRY

Le terme « poisson » désigne les vertébrés aquatiques pourvus de branchies et de nageoires locomotrices. Ce chapitre concerne les téléostéens : les poissons à nageoires rayonnées qui regroupent la grande majorité des espèces (plus de 30 000 espèces recensées appartenant à environ 40 ordres). Parmi ces espèces, certains poissons sont des modèles pour la recherche (poisson zèbre, médaka, etc.), d'autres sont d'importance économique majeure, car leurs populations sont exploitées par la pêche (thon, cabillaud, sardine, etc.) ou domestiquées et élevées dans des piscicultures (carpe, truite, saumon, bar, daurade, etc.). Les recherches visant à optimiser ces élevages ont généré de nombreuses connaissances sur l'importance de l'épigénétique sur leur développement, leur nutrition et leur reproduction, et récemment sur la transmission de certaines informations environnementales (nutrition, domestication) à leur descendance. L'histoire évolutive des poissons explique en outre l'origine de leurs nombreux régulateurs épigénétiques.

► Des génomes dupliqués complexes à l'origine de nouveaux régulateurs épigénétiques

Les téléostéens ont subi, au moment de leur radiation, une duplication complète de leur génome. Certaines sous-familles, telles que celle des Cyprininae, ou même certains ordres, tels que celui des salmonidés, présentent eux-mêmes une duplication complète additionnelle de leur génome. Ces duplications complètes de génomes ont pour conséquence une duplication du matériel génétique avec une possibilité de perte ou de conservation des loci ainsi dupliqués. Un tel événement de duplication complète du génome fournit du matériel génétique supplémentaire à l'espèce et est associé à la diversité phénotypique et aux innovations évolutives. Concernant les acteurs moléculaires épigénétiques, ces duplications complètes du génome ont notamment entraîné une duplication des gènes codants pour les enzymes et les protéines impliquées dans les processus de méthylation de l'ADN, de modification des marques d'histones, mais aussi de biosynthèse et de *turn-over* des ARN non codants (Best *et al.*, 2018), ainsi que des gènes codant directement pour les ARN non codants.

À la suite d'événements de duplication complète du génome, les gènes dupliqués conservés dans le génome au cours de l'évolution peuvent avoir conservé l'intégralité des fonctions

du gène ancestral, n'avoir conservé qu'une partie de celles-ci (sous-fonctionnalisation) ou présenter une ou des nouvelles fonctions par rapport au gène ancestral (néofonctionnalisation). Du fait de la grande diversité des téléostéens, l'évolution de ces gènes est diversifiée et complexe. Même si, globalement, les domaines fonctionnels des protéines impliquées dans les mécanismes épigénétiques sont conservés, le nombre de gènes, les événements de perte/duplication de gènes et la longueur des protéines putatives varient considérablement d'une espèce à l'autre. Ces événements peuvent engendrer des modifications fonctionnelles parfois espèce-spécifiques qui influent sur les différentes voies de régulation épigénétique (Liu *et al.*, 2020). Il a été par exemple récemment découvert une nouvelle classe possible d'histones désacétylases, HDAC12, qui pourrait être spécifique aux téléostéens.

De la même manière, une étude portant sur l'histoire évolutive des gènes codants pour les ADN méthyltransférases chez les chordés a révélé que les téléostéens possèdent entre 5 et 8 gènes codant pour Dnmt3, régulés de manière différentielle au cours de la gamétogenèse et de l'embryogenèse chez les Salmonidés. Ces derniers gènes présentent aussi des profils d'expression différents chez les salmonidés et le poisson zèbre lors de variations de température, suggérant des différences de fonctions. Les gènes codant pour les ARN non codants sont également régulés par des mécanismes épigénétiques. Même si ceux-ci ne sont que peu étudiés chez les téléostéens, il est aisé d'imaginer que la duplication de ces gènes a une influence directe sur les mécanismes de régulation induits par les ARN non codants (notamment sur la quantité d'ARN non codants transcrite en fonction de l'environnement). L'exploration de la fonction de ces gènes dupliqués en est à ses balbutiements, mais permettra, une fois réalisée, de discriminer la contribution relative des outils épigénétiques « généraux » des mécanismes espèce-spécifiques permis par ces duplications complètes du génome chez les téléostéens.

► Originalité du support génétique du méthylome chez les téléostéens

Les téléostéens se différencient des mammifères par les caractéristiques de leur méthylome, et notamment par la fréquence des CpG, le niveau de méthylation globale et la notion d'îlots CpG. Le pourcentage de bases G + C (%GC) est sensiblement différent de celui retrouvé dans le génome des mammifères (entre 39 % et 49 % selon les familles chez les téléostéens, contre 39 % en moyenne chez les mammifères), et le palindrome CpG est plus fréquemment retrouvé dans le génome des téléostéens que dans celui des mammifères. De plus, la méthylation de l'ensemble des cytosines du génome (5mC, incluant les méthylation CpG et non-CpG) a été identifiée comme deux fois plus importante en moyenne chez les téléostéens que chez les mammifères. Le statut poïkilotherme des téléostéens pourrait en être la raison principale, puisqu'une corrélation inverse entre la température interne de l'animal et le niveau de 5mC a été établie, mettant en avant une plus grande propension à la mutation de cette marque à haute température, et ceci même à niveau constant de %GC. Cette corrélation semble cependant indépendante de la distance phylogénétique chez les téléostéens. Enfin, il semble que les téléostéens présentent un niveau de méthylation plus important dans les régions géniques que dans les régions intergéniques, avec, comme chez les mammifères, une réduction de la méthylation autour des sites de démarrage de la transcription. La notion d'îlot CpG n'est en revanche pas conservée chez les téléostéens, car la définition statistique

qui en est faite chez les mammifères ne semble pas adaptée à la grande diversité des génomes des téléostéens, qui diffèrent notamment par leur taille et leur composition en bases ATGC : la définition des îlots CpG dans la littérature repose sur la longueur de la séquence considérée, le pourcentage de C + G et la fréquence des sites CpG. De plus, au sein même des téléostéens, la taille du génome diffère grandement. Cependant, comme chez les mammifères, les régions proximales des gènes présentent un plus faible niveau de méthylation que les autres régions du génome, suggérant que le patron de méthylation est globalement conservé.

► Mécanismes épigénétiques et modulation des réponses physiologiques intragénérationnelles

La plasticité phénotypique, définie comme la capacité d'un organisme vivant porteur d'un génotype donné à faire varier son phénotype selon l'environnement, est une composante clé pour faire face à des changements environnementaux. Les mécanismes épigénétiques sont en particulier impliqués dans la persistance à moyen ou à long terme d'effets physiologiques résultant d'événements survenus antérieurement. Ces mécanismes sont ainsi mis en jeu dans la réponse d'un organisme à des changements de son environnement, et les jeunes stades de vie sont particulièrement sensibles à ces événements. Ces marques épigénétiques peuvent être modifiées par des *stimuli* environnementaux, réguler l'expression des gènes à l'échelle du génome et moduler les phénotypes. L'épigénétique intragénérationnelle correspond ainsi aux modifications épigénétiques ayant lieu pendant le développement ou au cours de la vie d'un individu en réponse à des *stimuli* environnementaux abiotiques (hypoxie, température, salinité, nutrition, contaminants) ou biotiques (interactions sociales, pathogènes) dans la modulation des réponses physiologiques des poissons téléostéens.

Des lignées isogéniques pour étudier les interactions entre génétique, épigénétique et environnement

Comprendre comment des poissons porteurs de différents génotypes peuvent mettre en place différents mécanismes épigénétiques en réponse à un *stimulus* environnemental, et ainsi moduler à plus ou moins grande échelle leur phénotype, nécessite de pouvoir étudier indépendamment les variations génétiques et les variations épigénétiques. Peu d'études chez les poissons ont pu le faire, en se reposant sur des modèles d'études particuliers de poissons asexués, tels que le poisson clonal *Chrosomus eos-neogaeus* (Cyprinidés) ou un poisson hermaphrodite s'autofécondant naturellement, la fondule (*Kryptolebias marmoratus*).

Dans ce contexte d'étude des interactions génétique × épigénétique × environnement, les lignées isogéniques de truite arc-en-ciel (Quillet *et al.*, 2007) constituent un matériel biologique remarquable. Au sein de chaque lignée, les individus sont homozygotes à tous les loci et génétiquement identiques entre eux. Ces lignées permettent donc de comparer à la fois les marques épigénétiques établies en réponse à un *stimulus* environnemental au sein de chaque lignée (c'est-à-dire à génotype constant), mais aussi entre lignées (c'est-à-dire de savoir si les marques établies dépendent du fond génétique). Elles sont produites par gynogenèse en deux générations (figure 11.1). Actuellement, 19 lignées isogéniques de truite arc-en-ciel sont maintenues à INRAE par simples croisements frère-sœur.

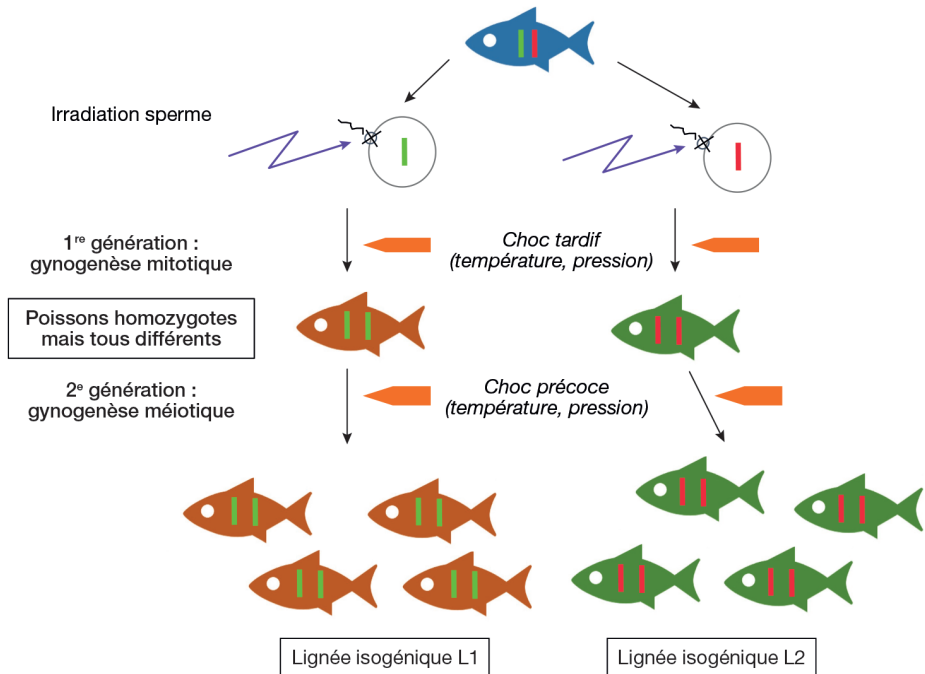


Figure 11.1. Production des lignées isogéniques de truite arc-en-ciel par deux générations de gynogenèse (© Dupont-Nivet).

En 1^{re} génération, des femelles standards sont reproduites par gynogenèse mitotique : un choc de température ou de pression inhibe la première mitose embryonnaire pour restaurer la diploïdie, aboutissant à une population d'haploïdes doublés : des poissons homozygotes mais tous différents. En 2^e génération, les femelles haploïdes doublées produites en 1^{re} génération sont reproduites par gynogenèse méiotique : un choc de température ou de pression inhibe la deuxième disjonction méiotique par rétention du second globule polaire. Chaque femelle double haploïde de 1^{re} génération produit ainsi des centaines d'œufs tous identiques et homozygotes, aboutissant à des centaines de truites génétiquement identiques.

Il existe une forte variabilité entre lignées pour de nombreux caractères : croissance, forme et couleur du corps, résistance aux maladies, reproduction, comportement, réponse au stress, etc. Ces lignées sont un outil puissant en recherche, en permettant la disponibilité du même fond génétique dans le temps et l'espace (standardisation des recherches), et ainsi la possibilité de mener des recherches à différents niveaux (cellule, tissu, animal entier) pour un fond génétique donné. Elles ont servi notamment au développement d'outils génomiques et à l'étude des bases moléculaires de caractères complexes : la résistance aux maladies, la capacité à utiliser l'aliment végétal, la thermotolérance, la réponse au stress, le comportement ou la qualité de la chair.

Impact de la température précoce sur la réponse à un stress aigu de température

Les poissons étant poïkilothermes, le réchauffement global est susceptible d'affecter l'élevage des poissons d'eau froide, ou d'une façon générale les espèces qui vivent à des températures proches de leur limite thermique physiologique. La truite arc-en-ciel, bien que capable de survivre à des températures de 25 °C, peut sérieusement subir les

vagues de chaleur, avec réduction de croissance et augmentation des épisodes infectieux et de mortalité. Il est donc important d'étudier la capacité des truites à s'adapter à des températures plus élevées et/ou fluctuantes, ainsi que les mécanismes moléculaires sous-jacents. De plus, de nombreuses études ont montré l'impact de la température pendant le développement embryonnaire sur la croissance, le développement du muscle et le métabolisme intermédiaire. De manière très intéressante, il a été montré que la capacité d'acclimatation thermique peut être modifiée par la température subie pendant le développement précoce chez le poisson zèbre. Plusieurs études ont par ailleurs montré que la température pendant le développement embryonnaire pouvait modifier la méthylation de l'ADN à l'échelle du génome chez l'épinoche et modifier l'expression des ADN méthyltransférases (*dnmt3*) chez le poisson zèbre.

Certaines lignées isogéniques de truite arc-en-ciel ont été caractérisées pour leur réponse à la température, avec la mise en évidence de lignées sensibles, résistantes ou intermédiaires à la température. Ces lignées ont été utilisées pour répondre aux questions suivantes : le régime thermique appliqué pendant les phases précoces du développement entraîne-t-il des modifications des patrons de méthylation de l'ADN, et ces modifications dépendent-elles du fond génétique ? Le régime thermique précoce influe-t-il sur la réponse à un stress ultérieur ? Pour 8 lignées isogéniques, la moitié des œufs embryonnés a été incubée à température standard (11 °C), l'autre moitié à température chaude (16 °C). Des chocs thermiques aigus ont ensuite été effectués sur les juvéniles de 4 mois pour étudier l'effet de l'exposition précoce à des températures élevées sur la réponse à un stress de température ultérieur. Tout d'abord, il a été montré, par PCR quantitative à la fin de la période d'incubation (stade œuf embryonné avant l'éclosion), que l'expression des *dnmt3* n'était que modérément modulée par la température d'incubation et que cette modulation dépendrait du fond génétique (Lallias *et al.*, 2021).

Ensuite, les patrons de méthylation de l'ADN en réponse à la température d'incubation (11 °C vs 16 °C) ont été analysés à la fin de la période d'incubation (Lallias *et al.*, 2021). Une analyse globale de la méthylation de l'ADN (approche LUMA, *Luminometric Methylation Assay*) a tout d'abord été choisie. Une analyse plus fine des patrons de méthylation de l'ADN à l'échelle du génome a ensuite été réalisée par une nouvelle méthode, EpiRADseq, initialement développée chez un organisme clonal : la daphnie. Cette approche est fondée sur une double digestion enzymatique de l'ADN (l'une des deux enzymes de restriction étant sensible à la méthylation de l'ADN), puis sur un séquençage des fragments générés autour des sites de restriction, aboutissant après des analyses bio-informatiques à un tableau de comptage (nombre de lectures par locus et par échantillon). Un nombre de lectures différent entre échantillons reflète indirectement un pourcentage de méthylation différent au locus donné. L'analyse au LUMA n'a révélé aucun impact de la température d'incubation sur le niveau global de méthylation. Selon les lignées, entre 72 263 et 83 494 loci EpiRADseq, bien répartis sur tous les chromosomes, ont été analysés, représentant environ 1 % du génome de la truite.

Une analyse différentielle a ensuite été effectuée lignée par lignée sur un nombre de loci EpiRADseq variant entre 51 668 et 55 607 après un filtre sur l'abondance des lectures ; 409 loci ont été identifiés comme différentiellement méthylés entre les deux températures d'incubation (11 °C vs 16 °C), dont 385 sur les 29 chromosomes de la truite, démontrant que le régime thermique précoce peut affecter la méthylation de l'ADN. Parmi ces 385 loci, 251 étaient situés dans ou à proximité (1 Kb) d'un ou

plusieurs gènes. Les 268 gènes situés à proximité de ces 251 loci EpiRADseq étaient impliqués dans des fonctions biologiques telles que la transduction de signal, le métabolisme, le système endocrinien, le système immunitaire et la communauté cellulaire, corroborant globalement les résultats d'études transcriptomiques en réponse à la température. De plus, la modulation des patrons de méthylation de l'ADN par la température précoce dépend du fond génétique. En effet, le nombre de loci différenciellement méthylés varie fortement (19 à 155) entre lignées, et seuls 34 des 409 loci étaient différenciellement méthylés dans au moins 2 lignées.

Enfin, les tests aigus de température effectués sur des juvéniles ont permis de confirmer l'existence d'une variabilité génétique de la résistance à la température, avec des lignées résistantes, sensibles ou intermédiaires. En revanche, globalement, aucun effet de la température d'incubation n'a pu être mis en évidence sur la réponse des lignées isogéniques de truite à un stress de température ultérieur (4 mois après l'exposition précoce).

En conclusion, cette étude a mis en évidence que la température pendant le développement embryonnaire pouvait moduler l'expression des *dnmt3* et influencer sur la mise en place des patrons de méthylation de l'ADN. L'utilisation des lignées isogéniques de truite arc-en-ciel a notamment permis de montrer que cette modulation des patrons de méthylation de l'ADN par la température précoce semblait dépendre du fond génétique. En revanche, à ce jour, l'implication des mécanismes épigénétiques dans les différences de réponse à la température des lignées isogéniques de truite n'a pas pu être démontrée.

Impact de la nutrition sur les marques épigénétiques au cours de la vie du poisson

En aquaculture, la quantité et la qualité des aliments utilisés ont des effets directs sur la croissance des poissons et sur la rentabilité des productions. La problématique en aquaculture actuelle est focalisée sur la recherche du maintien des performances des animaux, tout en diminuant le coût de production, mais aussi l'empreinte écologique de celle-ci. La recherche se concentre dans ce sens à diminuer ou remplacer l'apport en matières premières coûteuses (telles les vitamines et les protéines issues de pêches minotières), par des ingrédients alternatifs, moins coûteux mais aussi plus durables (principalement d'origine végétale comme les glucides), et à estimer les conséquences, notamment épigénétiques, d'une telle substitution. L'alimentation, non seulement en matière de quantité que de qualité, si elle est déséquilibrée, peut être à l'origine de changements dans l'homéostasie métabolique et d'une mauvaise croissance, mais aussi provoquer des maladies chroniques et une déficience immunitaire, augmentant la susceptibilité des téléostéens aux maladies infectieuses.

Récemment, une nouvelle voie de recherche a vu le jour, l'épigénétique nutritionnelle, pour chercher à expliquer le lien entre l'alimentation et ses impacts. Cette discipline considère les mécanismes de régulation épigénétique comme connecteurs centraux entre les nutriments ingérés et le contrôle de l'activité des voies du métabolisme intermédiaire de l'organisme. En effet, il est connu de longue date que certains nutriments ou leurs métabolites sont impliqués en tant que cofacteurs des enzymes régissant les mécanismes épigénétiques – comme l' α -cétoglutarate, un intermédiaire du cycle de Krebs et cofacteur des enzymes Tet– (*ten-eleven translocation*) ou en tant que substrat de celles-ci – comme la S-adénosylméthionine (SAM), intermédiaire du cycle

du *One-carbon* et donneur de groupement méthyl dans les processus de méthylation de l'ADN (par les DNA méthyltransférases, Dnmt) ou des histones (par les histones méthyltransférases, Hmt). Les modifications épigénétiques ainsi induites vont conduire à l'expression de phénotypes transmissibles de manière intragénérationnelle (*via* les mitoses pendant la croissance), mais aussi intergénérationnelle (des parents aux descendants *via* les méioses; voir section « Transmission intergénérationnelle d'informations environnementales »), voire transgénérationnelle (des grands-parents à la seconde génération F2) (Skjærven *et al.*, 2022).

Dans l'optique de tendre vers une aquaculture plus durable, la farine et l'huile de poisson sont progressivement remplacées par des ingrédients alternatifs dans les formulations. De forts taux de substitution ont cependant des effets néfastes sur la survie, la croissance et l'efficacité alimentaire des poissons, et peuvent affecter leur métabolisme. Le passage d'un aliment de type marin (à base d'huile et de farine de poisson) à un aliment tout végétal requiert, entre autres, la réévaluation de la composition en acides aminés essentiels. En ce sens, la méthionine, un métabolite central du métabolisme du carbone, et donc un donneur de groupement méthyl, a fait l'objet d'études chez la truite, le saumon et le poisson zèbre. Chez ces espèces, des variations de méthylation de l'ADN sont corrélées au niveau de l'apport en méthionine dans l'aliment (supplémentation ou déficit). De plus, après un *stimulus* précoce déficient en méthionine au 1^{er} repas, une augmentation hépatique des marques épigénétiques H3K36m3 et H3K4me3 est observée chez les juvéniles de truite.

Les glucides, en tant que source énergétique substituant une partie de l'apport en protéines, ont aussi fait l'objet d'études approfondies. Ainsi, une augmentation de la part de glucides dans l'aliment chez la truite arc-en-ciel juvénile induit des modifications hépatiques de la méthylation globale de l'ADN, mais aussi de l'acétylation et de la méthylation globale des histones, effet qui n'a pas été retrouvé chez le tilapia, une espèce bonne utilisatrice des glucides. Des études réalisées sur des alevins de daurade royale, sur le foie de poisson mandarin ou de truites arc-en-ciel ont également mis en évidence des modifications globales des histones et de la méthylation de l'ADN dues à différentes stratégies alimentaires larvaires et juvéniles d'utilisation d'aliments à base d'ingrédients d'origine végétale.

Les lignées isogéniques de truite arc-en-ciel présentées ci-dessus ont également permis de montrer l'existence d'une variabilité génétique pour la capacité à survivre et grandir avec une alimentation exclusivement à base de végétaux. Il a également été montré qu'une exposition précoce à l'aliment végétal améliore, par rapport aux lots témoins, la croissance, l'ingéré et l'efficacité alimentaire des poissons nourris ultérieurement avec ce même aliment végétal. De plus, en étudiant trois lignées isogéniques (deux lignées utilisant mieux l'aliment végétal que l'autre), plusieurs études visaient à comprendre les mécanismes impliqués dans ces performances contrastées entre génotypes.

Dans une étude récente (Marandel *et al.*, 2022), les auteurs ont testé l'implication des modifications de la méthylation de l'ADN dans ces différences de performances entre lignées isogéniques. Trois lignées aux performances contrastées avec l'aliment végétal ont été produites, puis nourries *ad libitum* à partir de la première alimentation, soit avec l'aliment marin soit avec l'aliment végétal. Après cinq semaines, les lots nourris avec l'aliment marin ont été divisés en deux, pour continuer à être nourris *ad libitum* avec l'aliment marin pendant six mois, ou recevoir une ration restreinte de l'aliment

marin (ration reflétant celle consommée avec l'aliment végétal). L'utilisation de ces trois lignées a ainsi permis, d'une part, de comparer l'impact du régime alimentaire sur les profils de méthylation et d'expression des gènes codant pour les acteurs de la méthylation/déméthylation de l'ADN; et, d'autre part, de tester si ces modifications dépendent du fond génétique (c'est-à-dire de la lignée). Les résultats montrent que les profils globaux de méthylation de l'ADN du foie sont modifiés par l'aliment végétal, et que ces modifications dépendent du fond génétique. En effet, selon la lignée, cet impact de l'aliment végétal semble dû à la baisse de l'ingéré alimentaire, seul ou combiné à un effet de la composition de l'aliment végétal. Enfin, l'expression des gènes impliqués dans les processus de méthylation/déméthylation de l'ADN a été modifiée par le régime alimentaire, et pour certains de manière lignée-spécifique.

Sensibilité des gonades et des gamètes mâles aux conditions environnementales

Un exemple phare chez les poissons à grande plasticité phénotypique est la réponse des gonades en développement à des changements de température, avec pour conséquence une inversion sexuelle chez certaines espèces. Il a été montré chez le bar que l'augmentation de la température durant les soixante premiers jours de développement induisait près d'un an plus tard une augmentation de la méthylation du promoteur d'un gène responsable de la différenciation de la gonade femelle. Cette méthylation, en conduisant à la répression locale de l'aromatase, est responsable de la différenciation mâle des individus génétiquement femelles. Chez la sole asiatique, c'est une déméthylation gonadique d'un gène de la différenciation mâle (*doublesex and mab-3 related transcription factor 1, dmrt1*) qui médie l'effet masculinisant de la température. Au-delà des tissus gonadiques, on peut supposer que les gamètes eux-mêmes, à l'instar des cellules de foie, de muscle ou de gonades, seraient eux aussi sensibles à l'environnement auquel est exposé le géniteur.

Cette question est complexe à étudier et n'est abordée que dans un nombre restreint de situations environnementales, notamment lors de l'exposition des poissons à des xénobiotiques (revue par Lombó et Herráez, 2021, et non traité ici) ou sous l'effet de la captivité. La méthylation de l'ADN des spermatozoïdes de truites et de saumons a ainsi été explorée, dans une perspective d'étude d'impact des piscicultures de repeuplement ou de la domestication. Ces études ont démontré que la méthylation de l'ADN des spermatozoïdes dans certaines régions du génome est affectée chez les truites maintenues transitoirement en captivité à des stades jeunes. Cet effet précoce, bien avant la production du sperme, indique une réceptivité des cellules germinales à des stades bien antérieurs au produit final qu'est le spermatozoïde. Plus directement, il a été montré chez la truite qu'une élévation de la température d'élevage pendant la spermiogenèse (phase de la gamétogenèse qui se déroule l'été) modifie également la méthylation de certaines régions du génome du spermatozoïde, les régions les plus significatives ayant trait à la spermatogenèse et au métabolisme des lipides (Brionne *et al.*, 2023).

La communauté scientifique commence aussi à s'intéresser à l'impact des biotechnologies de la reproduction (inductions hormonales, cryoconservation du sperme, etc., revues par Labbé *et al.*, 2017), sans que des effets drastiques aient réellement pu être mis en évidence. La conservation en frais du sperme chez la carpe commune induirait des variations sensibles de la méthylation de l'ADN des spermatozoïdes, mais

les régions affectées dans le génome restent à caractériser. À l'heure actuelle, nous n'avons connaissance d'aucune étude concernant la réponse épigénétique des gamètes femelles à l'environnement.

► Transmission intergénérationnelle d'informations environnementales

Plusieurs études rapportent la transmission d'adaptations environnementales d'individus parentaux (F0) aux générations suivantes chez les poissons. Bien que ces travaux n'abordent pas forcément la nature du vecteur de cette adaptation, ils illustrent parfaitement la transmission d'une réponse environnementale et permettent de suggérer fortement que les gamètes en sont le vecteur, et donc possiblement *via* des marques épigénétiques. La transmission intergénérationnelle concerne des mécanismes assez directs, puisqu'elle traite de la première génération issue de parents soumis aux contraintes environnementales.

Transmission d'informations nutritionnelles

La nutrition des géniteurs a un effet considérable sur la reproduction et la qualité de la descendance chez les téléostéens. Les aliments pour géniteurs n'ont été pourtant que peu modifiés, car l'investissement pour estimer leur efficacité est lourd en matière de temps (dépendant du cycle de reproduction, parfois long comme c'est le cas chez la truite ou le saumon) et de coût. Or un régime déséquilibré par rapport à leurs besoins en micro- et en macronutriments peut être coûteux et délétère pour la croissance et la reproduction. Depuis peu, des recherches s'attachent à réévaluer de nouvelles formulations pour les géniteurs afin de répondre au mieux à leurs besoins et aux nouvelles considérations économiques et environnementales de la production aquacole décrites ci-avant. Ces études incluent logiquement l'évaluation des conséquences de ces changements alimentaires sur la descendance par l'application du concept de programmation nutritionnelle. D'un point de vue mécanistique, la plupart, voire toutes les études portent sur la méthylation de l'ADN.

Les descendants de géniteurs de poisson zèbre nourris avec un régime à base d'ingrédients d'origine végétale pauvre en micronutriments présentent par exemple un phénotype proche de la stéatose hépatique incluant des changements de niveau de méthylation de l'ADN au niveau de gènes clés du métabolisme lipidique. Également, chez la daurade royale, l'ajout d'huiles riches en acide α -linoléique au régime parental modifie la méthylation de l'ADN au niveau de promoteurs hépatiques du métabolisme lipidique. Un tel impact est vraisemblablement épigénétique, dans le sens de l'apparition d'une réponse des descendants F1 à des contraintes environnementales subies par leurs parents.

Chez la truite arc-en-ciel, plusieurs études ont été menées sur la transmission d'informations nutritionnelles. L'une fait état que les descendants des génitrices de truite arc-en-ciel nourries avec un régime déficient en méthionine pendant six mois montrent, dans des alevins entiers, une altération de la méthylation de l'ADN au niveau des promoteurs des gènes *bnip3* (*BCL2 interacting protein 3*) marqueurs de stress cellulaires. De même, des régimes maternels enrichis en sélénium donnent lieu à de nombreuses régions différenciellement méthylées, notamment au niveau de gènes

impliqués dans le métabolisme du *One-carbon*. Enfin, il a été montré récemment que des géniteurs mâles et femelles nourris durant un cycle entier de reproduction avec un régime riche en glucides transmettent des informations à leur descendance qui, au stade alevin avant le 1^{er} repas, présente une hypométhylation globale de l'ADN qui persiste au niveau de leur foie au stade juvénile. Ce phénotype épigénétique semble donc fort et répétable aussi bien en nutrition directe que de manière intergénérationnelle.

Les gamètes, vecteurs d'informations environnementales à leurs descendants : quelle médiation épigénétique ?

Il est connu chez les poissons que l'ovule contient la majorité des composés qui soutiendront le développement de l'embryon après fécondation. Cela inclut notamment les ARNm dits d'origine maternelle, car produits lors de la gamétogenèse femelle, les mitochondries et leur ADN, et les réserves vitellines procurant un substrat nourricier à l'embryon et à l'alevin avant sa première alimentation. Cela souligne le rôle prépondérant du gamète femelle dans la qualité du descendant, et rend difficile l'identification d'une transmission épigénétique spécifique par l'ovocyte. Si le rôle du spermatozoïde a été pendant longtemps cantonné à la transmission de son bagage génétique haploïde, explorer si des informations de nature épigénétique sont également transmises par le spermatozoïde à l'embryon fait désormais l'objet de nombreuses études chez les poissons.

Il a ainsi été proposé que, pour de nombreux gènes, le profil de méthylation établi dans le spermatozoïde sert de guide à la mise en place du profil de méthylation dans l'embryon d'un poisson modèle, le poisson zèbre, et qu'un profil altéré induit des défauts de développement chez un autre animal modèle, le xénope (revue par Labbé *et al.*, 2017). On sait de plus que l'embryon de poisson appose à son tour des marques épigénétiques (histones) particulières sur son génome, marques qui permettent de garder la mémoire du profil (méthylation de l'ADN) qui était présent dans le spermatozoïde (Murphy *et al.*, 2018). Ces données ont permis de voir les marques épigénétiques des gamètes, notamment celles du spermatozoïde, sous un jour nouveau : elles ne sont pas uniquement la signature des remaniements qui ont eu lieu lors de la gamétogenèse et qui visent à établir la mise sous silence des génomes parentaux, elles jouent également un rôle, après fécondation, dans le développement du descendant.

La principale difficulté dans la caractérisation du rôle des gamètes lors d'une transmission d'informations environnementales à la descendance est d'en identifier les vecteurs moléculaires : méthylation de l'ADN, modifications des histones gamétiques, ARN non codants, interactions possibles entre ces vecteurs et avec la machinerie embryonnaire. Une autre difficulté réside dans le tri des informations, pour discriminer si les variations observées dans le gamète sont les résidus de modifications du processus gamétogénétique ou sont au contraire porteuses d'informations majeures pour la descendance. La troisième difficulté réside dans la caractérisation du phénotype de la descendance, sachant qu'on ne peut bien sûr pas prédire à quel stade, dans quel cadre et sur quel caractère la modification des gamètes aura une influence. De plus, on ne peut exclure que des systèmes de réparation ou de stabilisation dans l'embryon restaurent certaines des altérations épigénétiques des gamètes, même si ce processus est encore très peu étudié chez les poissons.

Quelques études démontrent cependant une modification épigénétique des gamètes dans cette chaîne entre environnement et descendance, et ce exclusivement sur le spermatozoïde. Au niveau du gamète femelle, l'instabilité des réserves ovocytaires vis-à-vis de l'environnement complexifie en effet l'identification d'un impact épigénétique strict sur la descendance. Un exemple de transmission intergénérationnelle montre que des saumons sauvages élevés en captivité pendant les deux ans qui précèdent leur maturation sexuelle présentent une méthylation de l'ADN modifiée dans leurs spermatozoïdes par rapport à ceux capturés en rivière juste avant la collecte de leurs gamètes. En parallèle, des modifications de la méthylation de l'ADN sont trouvées dans le foie des descendants F1, dont certains paramètres morphologiques sont par ailleurs modifiés (Wellband *et al.*, 2021). Au niveau intergénérationnel, chez un poisson modèle, le médaka, les spermatozoïdes de mâles de la génération F2 issus de grands-parents soumis à un stress hypoxique présentent une hyperméthylation globale par rapport au contrôle de la même génération, et leur qualité est dégradée.

Ces observations sont complexes à valider, car tout biais génétique doit pouvoir être écarté, d'où l'intérêt des lignées isogéniques. Quand c'est le cas, elles permettent de démontrer, d'une part, que le processus de gamétogenèse n'efface pas toutes les informations épigénétiques somatiques de l'animal, et, d'autre part, qu'au moment de la fécondation, lors des remaniements épigénétiques majeurs subis par les cellules embryonnaires, ces informations perdurent et modifient la performance des individus.

► Conclusion

Les poissons sont par nature profondément exposés à un environnement changeant, ne serait-ce que du fait de variations de qualité de l'eau, sa température et l'oxygène dissous selon les bassins versants et selon la saison. À ces facteurs s'ajoutent les pratiques d'élevage et l'alimentation pour les poissons d'aquaculture. De nombreuses études suggèrent la contribution des régulateurs épigénétiques dans le développement, la reproduction et la croissance des poissons. Pour ce qui concerne la modulation, par des mécanismes épigénétiques, des réponses physiologiques intragénérationnelles à des *stimuli* environnementaux (température par exemple) ou nutritionnels, les lignées isogéniques de truite arc-en-ciel constituent un matériel biologique unique pour étudier ces interactions entre génétique, épigénétique et environnement. Ces études ont ainsi pu montrer que les modifications des patrons de méthylation de l'ADN semblaient dépendre du fond génétique. De nouveaux travaux de recherche devront être menés, en variant par exemple la durée du *stimulus* de température, en étudiant de manière plus fine les patrons de méthylation de l'ADN, ou en étudiant la persistance dans le temps des patrons de méthylation mis en place précocement.

L'étude de l'impact des nutriments sur les marques épigénétiques reste un défi difficile à relever, notamment du fait des variations de composition d'un lot de matière première à l'autre, et plus particulièrement lorsqu'on considère des matières d'origine végétale. Pour ce qui concerne la démonstration d'une transmission épigénétique par le biais des gamètes, une meilleure compréhension des mécanismes moléculaires mis en jeu dans l'établissement des marques épigénétiques à chacune des étapes clés de la gamétogenèse permettra de mieux maîtriser, d'une part, le risque d'apparition et de transmission d'une altération d'origine environnementale et, d'autre part, l'exploitation

d'une plasticité épigénétique favorable à une meilleure adaptation et performance de la descendance. Cela nécessite l'identification de marqueurs épigénétiques spécifiques, indicateurs de l'exposition de larves, juvéniles, géniteurs ou gamètes à différentes conditions environnementales, et leurs éventuelles transmissions intergénérationnelles.

► Références bibliographiques

- Best C., Ikert H., Kostyniuk D.J., Craig P.M., Navarro-Martin L., Marandel L., Mennigen J.A., 2018. Epigenetics in teleost fish: From molecular mechanisms to physiological phenotypes. *Comp. Biochem. Physiol. B, Biochem. Mol. Biol.*, 224, 210-244. <https://doi.org/10.1016/j.cbpb.2018.01.006>
- Brionne A., Goupil A.-S., Kica S., Lareyre J.-J., Labbé C., Laurent A., 2023. Spermatozoa methylome and its sensitivity to water temperature in a teleost fish. *Science of the Total Environment*, 892, 164077. <https://doi.org/10.1016/j.scitotenv.2023.164077>
- Labbé C., Robles V., Herraéz M.P., 2017. Epigenetics in fish gametes and early embryo. *Aquaculture*, 472, 93-106. <https://doi.org/10.1016/j.aquaculture.2016.07.026>
- Lallias D., Bernard M., Ciobotaru C., Dechamp N., Labbé L., Goardon L. *et al.*, 2021. Sources of variation of DNA methylation in rainbow trout: Combined effects of temperature and genetic background. *Epigenetics*, 16, 1031-1052. <https://doi.org/10.1080/15592294.2020.1834924>
- Liu J., Hu H., Panserat S., Marandel L., 2020. Evolutionary history of DNA methylation related genes in chordates: New insights from multiple whole genome duplications. *Sci. Rep.*, 10, 970. <https://doi.org/10.1038/s41598-020-57753-w>
- Lombó M., Herráez P., 2021. The effects of endocrine disruptors on the male germline: An intergenerational health risk. *Biological Reviews*, 96, 1243-1262. <https://doi.org/10.1111/brv.12701>
- Marandel L., Heraud C., Véron V., Laithier J., Marchand M., Quillet E. *et al.*, 2022. A plant-based diet differentially affects the global hepatic methylome in rainbow trout depending on genetic background. *Epigenetics*, 00, 1-12. <https://doi.org/10.1080/15592294.2022.2058226>
- Murphy P.J., Wu S.F., James C.R., Wike C.L., Cairns B.R., 2018. Placeholder nucleosomes underlie germline-to-embryo DNA methylation reprogramming. *Cell*, 172, 993-1006. <https://doi.org/10.1016/j.cell.2018.01.022>
- Quillet E., Dorson M., Le Guillou S., Benmansour A., Boudinot P., 2007. Wide range of susceptibility to rhabdoviruses in homozygous clones of rainbow trout. *Fish Shellfish Immunol.*, 22, 510-519. <https://doi.org/http://dx.doi.org/10.1016/j.fsi.2006.07.002>
- Skjærven K.H., Adam A.-C., Takaya S., Waagbø R., Espe M., 2022. Nutritional epigenetics. *Cellular and Molecular Approaches in Fish Biology*, 161-192. <https://doi.org/10.1016/B978-0-12-822273-7.00006-9>
- Wellband K., Roth D., Linnansaari T., Allen Curry R., Bernatchez L., 2021. Environment-driven reprogramming of gamete DNA methylation occurs during maturation and is transmitted intergenerationally in Atlantic Salmon. *G3 Genes, Genomes, Genet.*, 11. <https://doi.org/10.1093/g3journal/jkab353>

Chapitre 12

Édition de l'épigénome

DAVID L'HÔTE

Il convient de définir tout d'abord ce que l'on entend par « édition de l'épigénome ». Comme vu dans les chapitres précédents, l'épigénome est un ensemble de marques épigénétiques apposées sur le génome, spécifique à chaque type cellulaire. Néanmoins, les cellules cancéreuses présentent souvent des modifications de leur épigénome en comparaison aux cellules saines dont elles sont issues. Ces modifications leur permettent d'acquérir de nouvelles caractéristiques, comme une capacité de prolifération incontrôlée, l'échappement à la mort cellulaire ou au système immunitaire. Ce type de modification de l'épigénome est appelé « épimutation ». Il est possible, en conditions de laboratoire, d'induire des épimutations en modifiant de façon contrôlée l'épigénome d'une cellule. Plusieurs techniques ont été développées : certaines molécules (« drogues ») inhibent ou activent des enzymes ou des acteurs de la chromatine, provoquant des bouleversements importants de l'épigénome. Par exemple, plusieurs médicaments appelés « épidroques » sont actuellement en développement dans le but de corriger des épimutations impliquées dans certains cancers. Ce type d'approche n'est, par nature, pas spécifique : on peut, par exemple, inhiber l'enzyme chromatienne P300 qui régule l'acétylation de l'histone 3 sur la lysine 27 (H3K27ac). Cette inhibition conduira à des modifications de la répartition de la marque H3K27ac de tout l'épigénome de la cellule traitée, et donc modifiera potentiellement l'expression de très nombreux gènes en plus de ceux que l'on souhaiterait cibler. Il n'est donc pas possible, avec une épidroque, de modifier une marque donnée sur un gène ou un ensemble de gènes choisis.

Comment pourrait-on donc modifier une marque particulière sur une région choisie du génome, autrement dit, comment pourrait-on éditer les caractéristiques épigénétiques à un endroit déterminé du génome sans modification épigénétique du reste de la séquence ADN ? De multiples laboratoires de recherche ont inventé différentes technologies regroupées sous le nom d'« édition de l'épigénome ».

Éditer l'épigénome, c'est venir ajouter ou retirer une marque épigénétique sur une ou plusieurs régions régulatrices d'un gène pour en modifier l'expression, afin soit d'étudier l'expression du gène soit de changer la fonction ou l'identité de la cellule.

Il ne faut pas confondre édition de l'épigénome et édition du génome : l'édition du génome consiste à induire des coupures de l'ADN à des endroits précis du génome afin de provoquer l'apparition de mutations. On peut faire une analogie avec le mode de

l'édition littéraire : si le génome était un livre, éditer le génome reviendrait à modifier un mot (enlever une lettre, changer le mot), alors qu'éditer l'épigénome reviendrait à changer la police de caractère (passer le mot en gras, en italique, etc.).

L'édition du génome ne sera pas traitée dans ce chapitre, bien qu'il s'agisse d'un outil puissant de modification de l'information génétique qui peut aussi être utilisé pour modifier indirectement l'épigénome si l'on mute un *enhancer*, ou une séquence cible d'une modification épigénétique. Pour ceux qui seraient intéressés, vous trouverez des informations ou des exemples d'édition du génome à la fin du chapitre (Ballouhey *et al.*, 2020; Beretta *et al.*, 2020).

► Les outils de l'édition de l'épigénome

Si on souhaite mettre en *italique* un mot dans un texte, il faut d'abord retrouver le mot dans le texte, puis avoir l'outil de modification en « Italique » installé dans son traitement de texte. De la même façon, afin de modifier précisément une marque épigénétique sur une région régulatrice donnée, il faut d'abord pouvoir se rendre sur la région du génome à modifier, puis avoir l'outil pour faire la modification épigénétique. Les outils d'édition de l'épigénome se scindent donc en deux modules : un module capable de reconnaître la région du génome à cibler (module de ciblage) et un module capable de modifier l'épigénome (module effecteur) (figure 12.1A).

Concernant les modules de ciblage, il existe « naturellement » plusieurs classes de protéines qui peuvent être utilisées car elles sont capables de s'accrocher à l'ADN sur une séquence particulière. Il s'agit, par exemple, des facteurs de transcription, protéines qui régulent l'expression des gènes. Pour cibler (reconnaître) un endroit précis du génome, il suffit donc de reconnaître la séquence d'ADN d'intérêt (connaître l'enchaînement de A, de T, de C et de G) et de construire le module de ciblage qui reconnaît cette séquence. Dans une première partie, nous traiterons des modules de ciblage couramment utilisés, puis, dans une seconde partie, nous aborderons les différents modules effecteurs classiquement mis en œuvre pour modifier l'épigénome.

Les modules de ciblage

On peut distinguer trois générations successives d'outils d'édition du génome utilisés aussi pour l'édition de l'épigénome : les ZFN (*zinc finger nucleases*), dont les prémisses d'utilisation datent du milieu des années 1990, les TALEN (*transcription activator-like effector nucleases*), disponibles depuis 2010, et enfin le système CRISPR/Cas9 (*clustered regularly interspaced short palindromic repeats/CRISPR-associated 9*), développé par Jennifer Doudna et Emmanuelle Charpentier en 2012, récompensé par un prix Nobel de chimie en 2020. Ces trois outils reposent sur le même concept (figure 12.1) : la création d'un facteur de transcription artificiel (qui n'existe pas naturellement) programmable, couplé à une protéine effectrice. Cette protéine chimère présente donc deux modules : un domaine de liaison à l'ADN que l'on peut modifier à façon pour reconnaître/cibler n'importe quelle séquence (ou presque) du génome, et un module effecteur, par exemple une enzyme de modification de la chromatine dans le cas de l'édition de l'épigénome. Ces trois outils ont été, au départ, développés pour faire de l'édition du génome, puis ont été adaptés pour faire de l'édition de l'épigénome.

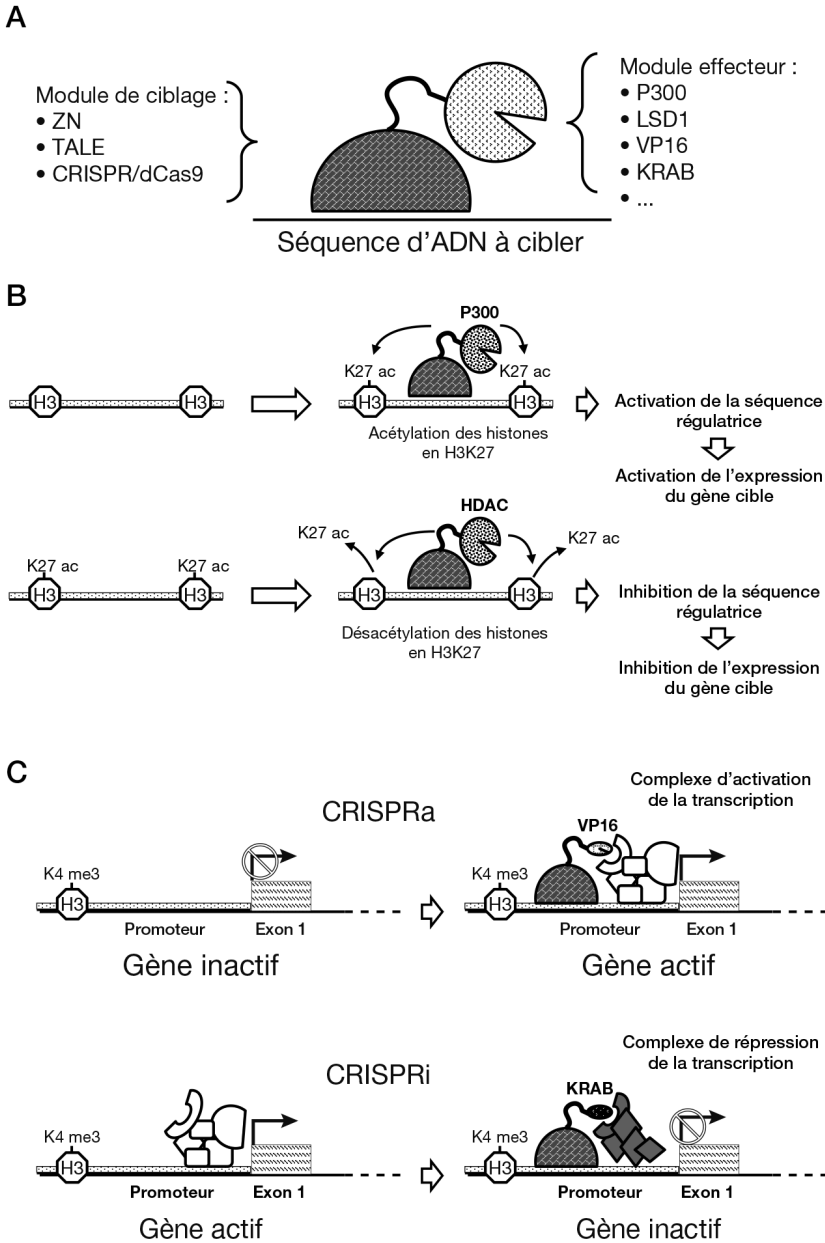


Figure 12.1. Les différentes plateformes d'édition de l'épigénome.

(A) Structure des outils d'édition de l'épigénome. (B) Acétylation et désacétylation des histones. (C) CRISPRa et CRISPRi pour reprogrammer l'expression des gènes.

A) Les outils d'édition de l'épigénome sont formés d'un module de ciblage, se fixant à l'ADN et d'un module effecteur, réalisant la modification épigénétique. B) Exemple d'outils d'édition permettant soit l'ajout (P300) soit le retrait (HDAC) de la marque H3K27ac, associée à l'activation des séquences régulatrice de l'expression des gènes. C) L'outil CRISPRa permet d'activer l'expression des gènes en recrutant la machinerie d'activation de la transcription sur le promoteur du gène cible, alors que l'outil CRISPRi, au contraire, inhibe l'expression du gène cible en recrutant des complexe represser.

Ce qui différencie fondamentalement les trois outils d'édition de l'(épi)génomique est la façon dont le module de reconnaissance de l'ADN est créé :

- les ZF sont des assemblages complexes de protéines fabriquées à partir du domaine de liaison à l'ADN des gènes de la famille des *zinc fingers*, d'où leur nom. Chacune des sous-unités reconnaît un triplet de bases (ATG ou TAA ou CGC, etc.) de l'ADN. Si on souhaite cibler la séquence suivante : AAGTGCAT, il faut donc assembler trois sous-unités reconnaissant dans l'ordre les triplets AAG, TGC et CAT. Cependant, l'assemblage répond à des règles complexes (certaines combinaisons ne fonctionnent pas) et reste fastidieux, ce qui rend cet outil difficile à utiliser ;
- les TALE sont quant à elles des assemblages de sous-unités issues des gènes de la famille des *Transcription activator-like effectors* de l'eubactérie *Xanthomonas*, un agent pathogène des plantes. Chaque sous-unité reconnaît une base ADN spécifiquement (A, T, C ou G), et leur assemblage répond à des règles simples. Du fait de la relative simplicité de leur construction, les TALE sont les premiers outils d'édition du génome qui ont été véritablement démocratisés dans la communauté scientifique. Les TALE, comme les ZF, ont été initialement utilisés pour éditer le génome. Les modules de ciblage ont été associés avec des nucléases, d'où le nom de ZFN ou de TALEN (N = nucléase), comme l'enzyme FOKI, capable de couper l'ADN, et donc d'induire des cassures chromosomiques. La réparation de l'ADN ainsi cassé conduit ensuite à des mutations à l'endroit choisi ;
- la dernière génération d'outils est le système CRISPR/Cas9. Celui-ci, modifié à partir du système de défense des bactéries contre les bactériophages, repose sur un mécanisme différent des deux autres systèmes : en effet, le facteur de transcription programmable est une protéine bactérienne invariable, la protéine CAS9. Cette protéine CAS9 utilise un petit ARN, appelé « ARN-guide » (ou ARNg), de séquence complémentaire à la séquence ADN cible. L'ARNg, associé à la CAS9, s'hybride donc avec l'ADN du génome au bon endroit. Pour cibler une région, il suffit donc de créer un ou plusieurs ARNg. La protéine CAS9 « naturelle » présente une activité nucléasique, c'est-à-dire qu'elle coupe l'ADN cible, ce qui a été mis à profit initialement pour éditer le génome. Afin de réaliser l'édition de l'épigénomique, l'activité nucléasique de la CAS9 n'est pas souhaitable, un mutant de cette protéine, dCas9 (pour *dead Cas9*), a été identifié : dCas9 perd sa capacité à couper l'ADN, mais garde celle de se lier à l'ADN. Le mutant peut ensuite être fusionné à l'effecteur chromatinien désiré pour éditer l'épigénomique.

Si les TALE(N) sont encore souvent utilisées aujourd'hui pour éditer les génomes et les épigénomes, cette technologie a très rapidement été concurrencée par l'émergence des CRISPR/dCas9. L'écrasante hégémonie de ce nouveau système ne repose quasiment que sur la simplicité du système de ciblage (l'ARNg) de la séquence génomique. Un dernier avantage du système CRISPR/Cas9 est, comme nous le verrons plus tard, la possibilité de « multiplexer », c'est-à-dire que l'on peut facilement cibler simultanément un grand nombre de séquences génomiques en utilisant plusieurs ARNg en même temps (on parle de banque d'ARNg). Les banques de TALE(N) sont plus difficiles et plus coûteuses à établir, ce qui limite les possibilités de multiplexage en TALE(N).

Dans la suite de ce chapitre, nous aborderons donc essentiellement l'apport du système CRISPR/dCas9 dans l'édition de l'épigénomique à partir d'exemples principalement décrits sur des modèles humains et murins. Ces approches d'édition ont aussi été

déclinées de nombreuses fois, et avec succès, sur d'autres modèles comme la levure, la drosophile ou les plantes comme *Arabidopsis*, le riz ou le tabac, pour n'en citer que quelques-uns.

Les modules effecteurs

Redessiner l'acétylation des histones

Les séquences régulatrices actives, et notamment les *enhancers*, présentent des enrichissements importants en marque H3K27ac, souvent déposée par l'enzyme histone acétyltransférase P300. C'est donc assez naturellement que des scientifiques ont cherché à recruter artificiellement la protéine P300 sur des séquences régulatrices dans le but de forcer l'acétylation des histones. L'idée est d'activer les séquences régulatrices et de faire en sorte que des gènes jusqu'alors silencieux dans la cellule que l'on veut modifier soient exprimés. En combinant la protéine de ciblage dCas9 avec la partie active de la protéine P300, des chercheurs de l'équipe de Charles Gersbach de l'université Duke, en Caroline du Nord, ont ainsi créé un outil capable d'induire une forte acétylation de l'histone H3, à la fois des promoteurs et des *enhancers*, et d'induire efficacement l'expression des gènes associés à ces séquences régulatrices (figure 12.1B) (Hilton *et al.*, 2015).

Inversement, afin d'éliminer les acétylations des histones, dCas9 a été fusionnée avec différentes protéines HDAC (histones désacétylases), notamment les HDAC3 et 8. Cibler une protéine de fusion dCas9-HDAC3 ou 8 sur une séquence régulatrice d'intérêt permet bien de réduire efficacement l'acétylation des histones sur cette séquence (figure 12.1B). Cependant, la protéine HDAC3 a un rôle complexe et bivalent. En effet, en plus de son rôle d'inhibiteur de la transcription par désacétylation d'histone, elle peut aussi jouer un rôle d'activateur transcriptionnel pour certains gènes dans certains contextes cellulaires, indépendamment de sa fonction de désacétylase. De ce fait, la réduction d'acétylation des histones par la protéine dCas9-HDAC3 ne s'accompagne pas d'une inhibition de l'expression génique, mais au contraire d'une faible augmentation! Lorsque l'on utilise la protéine HDAC8, pour laquelle seul le rôle d'inhibiteur est connu, on observe une réduction de l'acétylation des histones sur la séquence régulatrice ciblée, qui conduit bien à une diminution de l'expression du gène associé.

Cet exemple démontre qu'il faut donc être vigilant lors du choix de l'effecteur pour éditer l'épigénome. Les protéines, et en particulier les enzymes de la chromatine, jouent souvent de nombreux rôles, qui peuvent être parfois contradictoires (activateur et inhibiteur). Ces différents rôles sont souvent assurés par des parties distinctes de la protéine, et une façon d'éviter les problèmes revient souvent à ne prendre que le domaine de la protéine qui assure la fonction d'intérêt.

Pourquoi modifier le paysage épigénétique de l'acétylation des histones dans une cellule? On peut, par exemple, corriger une épimutation associée à une pathologie comme un cancer. On peut aussi utiliser ces outils pour étudier la régulation de l'expression des gènes par l'acétylation des histones. Par exemple, en jouant avec les outils d'acétylation et de désacétylation dCas9-P300 et dCas9-HDAC8, il est possible d'étudier la variation de la régulation de l'expression d'un gène. L'expression des gènes est, en effet, dynamique : elle change au cours du temps dans la même cellule, en fonction de différentes stimulations hormonales ou nerveuses, elles-mêmes pouvant évoluer au cours du temps.

Le gène *FOS*, par exemple, a une expression « pulsatile » dans les neurones, sous la forme de petites périodes d'expression entrecoupées de périodes de silence. La durée de ces périodes de silence est réduite quand le neurone est stimulé. Ceci se traduit par une plus grande expression du gène. Des chercheurs du laboratoire d'Anne E. West de l'université Duke ont utilisé les outils d'édition de l'acétylation des histones pour modifier cette marque sur les *enhancers* du gène *FOS* et, ce faisant, ont réussi à modifier/reprogrammer son profil d'expression (Chen *et al.*, 2019). Cette belle étude démontre ainsi l'importance de la dynamique de l'acétylation des histones dans la régulation dynamique des gènes. Elle suggère, par ailleurs, que les outils d'édition de l'épigénome pourraient être utiles pour moduler finement les profils complexes de l'expression des gènes, notamment dans les neurones, afin d'étudier leur rôle dans des phénotypes cognitifs ou comportementaux, ce qui ouvre des perspectives de recherche passionnantes.

Éditer la méthylation des CpG de l'ADN

L'une des marques épigénétiques les mieux étudiées est la méthylation des cytosines dans les dinucléotides CpG de l'ADN, pour laquelle nous connaissons à la fois la majeure partie des acteurs (méthylases et déméthylases) et la distribution épigénomique dans un nombre croissant de types cellulaires. Comme présenté dans le chapitre 2, la méthylation des CpG est une marque répressive de l'expression génique chez les eucaryotes. Elle est déposée par les enzymes méthyltransférases (les DNMT), impliquées notamment dans le maintien de la méthylation lors des divisions des cellules. À l'inverse, la déméthylation des méthylcytosines fait intervenir plusieurs étapes biochimiques, notamment une première modification de la méthylation par les enzymes TET (*ten-eleven translocation*), qui oxydent le groupement méthyl, puis une seconde étape assurée par l'enzyme thymine DNA glycosylase (TDG), qui convertit la méthylcytosine oxydée en cytosine non méthylée.

Afin d'éditer la méthylation des CpG, des outils TALE et CRISPR ont donc été fusionnés soit avec les domaines catalytiques des méthyltransférases, comme celui des enzymes DNMT3A ou DNMT3B pour déposer de nouvelles méthylations, soit avec le domaine catalytique de l'enzyme TET1 pour retirer localement les méthylations présentes. Si ces outils sont les plus utilisés actuellement, on peut noter que d'autres constructions sont disponibles pour éditer la méthylation de l'ADN, comme la dCas9-MQ1, qui met à profit la méthyltransférase MQ1 d'origine bactérienne, ou encore la dCas9 fusionnée à l'enzyme ROS1, une TDG d'*Arabidopsis thaliana*.

Les approches visant à forcer la méthylation de séquences régulatrices se sont montrées plutôt fructueuses : cibler une séquence régulatrice avec une dCas9-DNMT3A permet de méthyliser efficacement cette séquence (Liu *et al.*, 2016). Ceci a pour conséquence une inhibition importante de l'expression du gène contrôlée par la région régulatrice ciblée. Les approches de déméthylation sont plus difficiles à mettre en œuvre. L'outil dCas9-TET1 doit, en effet, être positionné de façon optimale pour assurer la déméthylation de la méthylcytosine ciblée. De plus, même dans le cas où la déméthylation est efficace, l'impact sur l'expression génique reste souvent très modeste. Ceci est très probablement lié à la présence d'autres marques répressives en plus de la méthylation des cytosines, comme la méthylation des histones H3 sur les lysines K9 et K27. Retirer seulement un verrou, la méthylation des cytosines, ne suffit pas à lever l'inhibition de l'expression du gène.

Une étude de 2019 a montré qu'il est toutefois possible d'améliorer très fortement la déméthylation des CpG notamment, en associant au module TET1 la protéine GADD45A, une protéine connue pour interagir avec la machinerie de déméthylation (les chercheurs ont appelé ce système Casilio-ME) (Taghbalout *et al.*, 2019). De façon intéressante, le système Casilio-ME se révèle certes efficace pour déméthyler les cytosines ciblées, mais aussi pour activer l'expression des gènes cibles. Si ceci suggère qu'un frein à l'activation a été levé, il n'est cependant pas clair si cette activation est uniquement liée à une meilleure déméthylation, ou aussi à d'autres fonctions de GADD45A. Cette protéine est en effet connue pour déstabiliser les interactions histones-ADN, provoquer un relâchement de la chromatine et créer ainsi un environnement conformationnel plus favorable à l'expression génique.

Cet exemple démontre que les techniques d'édition épigénomique ne sont pas encore parfaitement maîtrisées. Comme dit précédemment, les protéines sont multifonctionnelles, et on ne connaît ou on ne contrôle pas toujours toutes leurs fonctions. Les outils d'édition de l'épigénome restent donc à l'heure actuelle des outils en développement en laboratoire. Ils ne sont, pour la plupart, pas encore suffisamment matures pour être utilisés sans danger en thérapeutique humaine.

En laboratoire de recherche cependant, on observe actuellement un déploiement important des outils d'édition du méthylome dans la littérature scientifique. De nombreuses études publiées cherchent à établir de façon non ambiguë la relation de cause à effet entre une épimutation et un phénotype pathologique. Par exemple, une étude conduite *in vitro* a démontré que les défauts d'expression du gène *FRM1* (*Fragile X Messenger Ribonucleoprotein 1*), conduisant au syndrome du X fragile, sont liés à une hyperméthylation pathologique de son promoteur. En utilisant l'outil dCas9-TET1, les chercheurs ont montré qu'il est possible de restaurer l'expression de ce gène dans des cellules en culture, en corrigeant cette épimutation. Ce type d'étude permet donc de mieux comprendre l'étiologie, c'est-à-dire les causes moléculaires responsables de pathologies. Certaines de ces études ouvriront probablement sur le long terme à des applications cliniques innovantes de correction des épimutations comme traitement thérapeutique.

Remodeler la méthylation des histones

La méthylation des histones est un phénomène complexe, car il existe un grand nombre de méthylation possibles avec des combinaisons aux rôles très différents (voir chapitre 2). Certaines méthylation sont activatrices, comme la triméthylation en H3K4me3 retrouvée sur les promoteurs, d'autres servent de système de reconnaissance, comme la monométhylation en H3K4me1 sur les *enhancers* (la cellule « reconnaît » les *enhancers* parce qu'ils portent cette marque), d'autres enfin sont inhibitrices, comme les méthylation en H3K27me3, H3K9me2 et H3K9me3 (ces deux dernières étant retrouvées essentiellement sur l'hétérochromatine). Chacune de ces marques peut être déposée et retirée par une ou plusieurs enzymes spécifiquement. Il est donc possible de créer beaucoup d'outils d'édition épigénétique pour modifier ces marques à façon.

Quelles sont les utilisations actuelles des outils d'édition de la méthylation des histones? Voici quelques exemples déclinés en recherche fondamentale.

- Édition de la marque H3K4me1 au niveau des *enhancers*. L'outil dCas9-LSD1 (*lysine-specific demethylase 1*) est le plus souvent utilisé. Il permet d'inactiver les *enhancers*. En ciblant un *enhancer* avec une dCas9-LSD1, il est possible de réduire

très fortement la présence de la marque H3K4me1, ce qui entraîne un démantèlement de l'*enhancer* (la cellule ne reconnaît plus la région régulatrice comme un *enhancer*). On peut réaliser cette expérience quand on veut démontrer qu'un *enhancer* est bien capable de réguler le gène que l'on étudie. Nous avons conduit une expérience de ce type dans notre laboratoire afin de démontrer qu'un *enhancer* que nous venions de découvrir régulait l'expression du gène *NR5A1*, un gène important pour les cellules gonadotropes de l'hypophyse.

Une autre application intéressante est le criblage fonctionnel à haut débit pour découvrir directement de nouveaux *enhancers* : comme nous l'avons vu, les *enhancers* peuvent être situés n'importe où, ou presque, par rapport aux gènes qu'ils régulent. Savoir qu'une séquence est enrichie en marque H3K4me1 permet certes de l'imaginer comme étant un *enhancer*, mais de quel gène ? Celui directement en amont ou en aval, ou celui situé encore plus loin sur le chromosome ? Tester un à un les *enhancers* plus ou moins proches d'un gène est assez fastidieux, surtout s'il y en a plusieurs centaines ! L'outil dCas9-LSD1 permet de tester les différents *enhancers* : il peut être utilisé pour un criblage fonctionnel à haut débit de plusieurs centaines d'éléments en parallèle. Comme vu précédemment, l'édition épigénétique avec le module de ciblage dCas9 permet de multiplexer, c'est-à-dire de cibler en même temps plusieurs endroits du génome. Il est donc possible d'éteindre au cours d'une même expérience plusieurs centaines d'*enhancers* en parallèle pour en étudier la fonction individuellement. Ce type d'approche permet de disséquer finement la régulation des gènes par leurs *enhancers*. Connaître les *enhancers* d'un gène permet ensuite de mieux comprendre la régulation du gène en situation physiologique ainsi que les perturbations en conditions pathologiques. C'est un prérequis pour proposer ensuite des thérapies ciblées.

À l'inverse, on peut chercher à ajouter la marque H3K4me1 : une étude, menée au laboratoire de Lipeng Ren à l'Institut Karolinska de Stockholm, a mis à profit une fusion dCAS9-MLL3^{SET} (fusion de dCas9 avec uniquement le domaine SET de MLL3) pour induire de façon contrôlée le dépôt de la marque H3K4me1 sur des *enhancers* (Yan *et al.*, 2018). Ceci a ensuite permis d'étudier le rôle de cette marque dans le recrutement des complexes chromatinien et la mise en place d'interactions spatiales entre les *enhancers* et les promoteurs cibles. Ce faisant, les chercheurs ont démontré l'importance de la marque H3K4me1 dans la structuration tridimensionnelle de la chromatine, et le rapprochement des *enhancers* par des boucles de la chromatine vers les promoteurs.

- Édition de la marque H3K4me3 des promoteurs. Une étude effectuée par l'Epigenetic Editing Research Group de l'université de Groningue, aux Pays-Bas, décrit l'utilisation des méthyltransférases PRDM9 (*PR/SET Domain 9*) et DOT1L (*DOT1 like histone lysine methyltransferase*) associées à la protéine de ciblage dCas9, afin d'activer l'expression de gènes silencieux en induisant la triméthylation en H3K4 (par PRDM9) et en H3K79 (par DOT1L) sur les promoteurs (Cano-Rodriguez *et al.*, 2016). L'édition des marques H3K4me3 ou de H3K79me3 seules induit une augmentation modeste de l'expression des gènes cibles, alors que la coédition simultanée des deux marques augmente l'induction de l'expression génique de façon significative. Ces travaux montrent l'importance du couplage de certaines marques épigénétiques dans la régulation de l'expression des gènes et démontrent l'utilité des outils d'édition épigénétique pour étudier l'impact des marques épigénétiques prises isolément ou en association.

Ce type d'étude permettra donc à l'avenir de mieux comprendre les règles régissant le code épigénétique.

- Édition des marques répressives H3K27me3 et H3K9me3. Il est d'usage de présenter les mécanismes moléculaires épigénétiques de la façon suivante : un effecteur de la chromatine est recruté sur une séquence régulatrice, y dépose ensuite la marque épigénétique *ad hoc*, ce qui entraîne l'effet désiré sur l'expression génique. Ce modèle laisse donc penser que l'effecteur n'a pas d'autres fonctions que de déposer la marque, et que la marque assure la fonction. Le groupe de recherche animé par Tamar Segal, à l'université israélienne d'Ariel, en Cisjordanie, qui travaille sur les marques répressives H3K27me3 et H3K9me3, a décidé de tester ce modèle en posant notamment les questions suivantes : l'inhibition de l'expression des gènes est-elle due : à la présence des marques répressives comme H3K27me3, ou à la seule présence des complexes chromatiniens comme le complexe Polycomb, ou enfin à la présence combinée des marques et des complexes (O'Geen *et al.*, 2017) ?

Afin de répondre à ces questions, les chercheurs ont développé deux outils : dCas9-EZH2 (avec le domaine *enhancer of zeste 2* en entier), qui permet d'induire le dépôt de la marque H3K27me3 sur une séquence régulatrice cible, et dCas9-EZH2SET, avec uniquement le domaine SET de EZH2, responsable de la réaction enzymatique aboutissant au dépôt de la marque. Si l'outil dCas9-EZH2 permet bien à la fois le dépôt de H3K27me3 et l'inhibition de l'expression des gènes, l'outil dCas9-EZH2SET, de façon surprenante, ne permet pas le dépôt de la marque H3K27me3, alors qu'il permet bien l'inhibition de l'expression du gène. Cela démontre que EZH2 est capable de réprimer l'expression des gènes en absence de dépôt de la marque H3K27me3. Ces travaux suggèrent donc que les marques de répression ne sont pas obligatoirement requises pour établir la répression génique, et que les effecteurs de la chromatine ont des fonctions plus complexes que nous ne le pensions jusqu'à présent. À quoi servent donc les marques dans ce cas ? Les travaux futurs éclairciront probablement les fonctions réciproques des complexes et des marques dans la régulation génique.

- Édition simultanée de plusieurs marques. Il est possible de cibler plusieurs marques en même temps, par exemple d'éditer à la fois la méthylation de l'ADN avec une méthyltransférase comme DNMT3A et la méthylation des histones avec EZH2. Ce type d'approche permet une meilleure répression de l'expression des gènes et un meilleur maintien de cette répression dans le temps. En effet, lorsque l'on réprime un gène uniquement avec une dCas9-EZH2 par exemple, la répression n'est que transitoire, alors que si on associe EZH2 avec DNMT3A en utilisant une dCas9-DNMT3A-EZH2, la répression transcriptionnelle semble définitive.

Les différents travaux présentés, en étudiant les conséquences fonctionnelles des différents acteurs chromatiniens sur les différentes séquences régulatrices, permettent de soulever un peu le voile de la complexité sous-jacente des mécanismes de la régulation épigénétique, et remettent en question certains des modèles de régulation épigénétique. La bonne connaissance de ces mécanismes, issue des études fondamentales, est l'un des prérequis à une utilisation thérapeutique des technologies d'édition de l'épigénome. Avant de présenter les applications thérapeutiques potentielles, nous allons aborder un dernier outil qui permet de reprogrammer l'expression génique par édition épigénétique : les systèmes CRISPRa et CRISPRi.

Reprogrammer l'expression génique : les systèmes CRISPRa et CRISPRi

Jusqu'à présent, nous avons décrit des outils modifiant une marque de façon précise pour étudier l'effet de cette modification sur l'expression des gènes cibles. Le système CRISPRa, pour activateur, permet l'activation de l'expression des gènes en recrutant de nombreux acteurs de la machinerie de transcription. Alors que le système CRISPRi (pour *interference*, inhibe), permet l'inhibition de l'expression des gènes en recrutant de nombreux inhibiteurs de cette machinerie de transcription. Ces deux systèmes sont donc un peu différents des autres systèmes d'édition de l'épigénome, car les modifications épigénétiques induites sont nombreuses, complexes et souvent mal caractérisées. Cependant, CRISPRa et CRISPRi, en réécrivant profondément les modifications épigénétiques sur les séquences régulatrices ciblées, sont d'une efficacité souvent remarquable, ce qui fait de ces systèmes des outils extraordinaires pour reprogrammer l'expression génique.

- CRISPRa pour induire l'expression d'un gène. L'un des tout premiers modules activateurs à avoir été fusionné avec une protéine de ciblage (que ce soit ZF, TALE ou dCas9) n'était pas un effecteur de la chromatine, mais le trans-activateur fort VP16. Ce que l'on appelle couramment « VP16 » est un peptide d'une dizaine de résidus aminoacides du gène α -TIF du virus de l'herpès. Ce fragment est connu pour recruter la machinerie d'expression des gènes, comprenant le complexe médiateur, de nombreux facteurs de transcription, l'ARN polymérase et même des enzymes de modification de la chromatine (figure 12.2C, complexe activateur de la transcription). VP16 est un peu le facteur de transcription « couteau suisse » du virus de l'herpès qui a été détourné pour les besoins.

- CRISPRi pour réprimer l'expression d'un gène. La dCas9 est souvent fusionnée à un domaine trans-répresseur, majoritairement de type KRAB (*Krüppel-associated box*). Les domaines KRAB sont connus pour recruter des co-répresseurs (complexe de répression de la transcription) ainsi que des complexes de répression de la chromatine impliqués dans la désacétylation des histones et la méthylation de l'ADN notamment (figure 12.2C). Le CRISPRi est donc utilisé pour inhiber de façon réversible l'expression génique, en ciblant le promoteur ou les *enhancers* des gènes. C'est donc une alternative (économique) aux approches ARN interférents (Si ou ShRNA).

De nombreuses études ont montré l'efficacité des CRISPRa et CRISPRi pour reprogrammer l'expression génique en ciblant les séquences régulatrices de gènes d'intérêt, ouvrant la promesse du développement de futurs outils thérapeutiques (Bendixen *et al.*, 2023).

- Les systèmes CRISPRa et CRISPRi pour le criblage fonctionnel à haut débit. L'utilisation principale émergente des systèmes CRISPRa et CRISPRi est le criblage fonctionnel à haut débit. L'idée est de tester la fonction d'un grand nombre de gènes dans un phénotype d'intérêt. Comme nous l'avons vu précédemment, le système CRISPR permet de multiplexer en associant plusieurs ARN guides. Des laboratoires de recherche comme celui de Michael Bassik, du Department of Genetics de l'université Stanford, aux États-Unis, ont poussé cette logique jusqu'au bout, en créant des bibliothèques, ou des banques contenant des ARN guides, pour chacun des plus de 20 000 gènes du génome de la souris ou de l'humain (Han *et al.*, 2020). Ces banques

permettent de tester la fonction potentielle de tous les gènes pour un phénotype d'intérêt. Ce type de criblage fonctionnel peut être mis en œuvre sur n'importe quelle cellule en culture en boîte de Petri et pour n'importe quel phénotype mesurable.

Une étude de 2021 montre comment l'utilisation du système CRISPRi permet de réinterroger la fonctionnalité des pseudogènes (Sun *et al.*, 2021). Les pseudogènes sont les fossiles d'anciens gènes dans notre génome que l'évolution a désactivés par mutation (inhibition permanente de l'expression). Il y aurait 14638 pseudogènes dans le génome humain (à comparer avec les 20805 gènes actifs), mais il est, en fait, difficile d'affirmer qu'un gène est véritablement non actif sur la seule base de sa séquence. Pour définir un pseudogène, il faut être certain que celui-ci n'a plus de fonction active, quel que soit le phénotype considéré. C'est ce qu'a cherché à montrer l'équipe de Yiwen Chen, de l'université du Texas, pour 850 gènes humains décrits comme pseudogènes. En utilisant une banque CRISPRi de 5703 ARNg ciblant les promoteurs putatifs des pseudogènes sélectionnés, les chercheurs ont testé l'éventuelle fonctionnalité de ces pseudogènes dans la croissance cellulaire d'une lignée modèle de cancer du sein. Cette étude révèle que certains pseudogènes sont tout à fait fonctionnels; on peut donc se demander s'ils ne devraient pas être reclassés en gènes actifs!

Les outils de type CRISPRi et CRISPRa nous offrent donc la possibilité d'investiguer la fonctionnalité réelle du génome et permettront probablement d'identifier, parmi les nombreux pseudogènes, un certain nombre de gènes en réalité toujours actifs. Cette question peut sembler purement théorique, mais caractériser de nouveaux gènes actifs peut avoir des retombées importantes en santé humaine notamment : chaque nouveau gène est un candidat potentiel pour mieux comprendre une pathologie, mais aussi, à plus long terme, une cible potentielle pour un médicament.

Le système CRISPRi peut également être utilisé pour étudier la régulation des gènes et en découvrir les *enhancers*, de la même façon que dans le cas des approches d'édition du génome utilisant le système dCas9-LSD1. Une étude récente suggère que nous sommes à l'aube d'une révolution dans le domaine de la génomique fonctionnelle (Gasperini *et al.*, 2019) : l'équipe de Jay Shendure, de l'université de Washington, a réussi le coup de force d'identifier environ 600 relations de régulation directes entre *enhancers* et gènes cibles, en combinant le système CRISPRi pour cibler 5920 *enhancers* potentiels et analyser l'expression des gènes par une approche en cellule unique (qui donne l'expression de tous les gènes dans chaque cellule une à une). L'étude de chacune de ces relations de régulation aurait pris des dizaines d'années de travail avant le développement des outils d'édition de l'épigénome et de séquençage en cellule unique. L'utilisation dans les années à venir de ce type d'étude permettra le décodage à grande échelle des règles fonctionnelles régissant les épigénomes, aussi bien en conditions physiologiques que pathologiques. Les retombées, aussi bien en sciences fondamentales qu'en applications cliniques, seront majeures.

► L'édition de l'épigénome au service de la santé

Avant de décliner les applications possibles de l'édition épigénétique en santé humaine, il convient d'en poser les limites : ce qui est de l'ordre du faisable à moyen terme, ce qui sera peut-être un jour réalisable, et les freins actuels au développement de ces thérapies moléculaires de nouvelle génération.

Dans un premier temps, rappelons qu'actuellement aucune thérapie ne repose sur l'édition épigénomique. Les applications potentielles sont nombreuses, les applications réellement mises en pratique en clinique sont, à ma connaissance, encore inexistantes.

Quels sont les freins à leur développement ? J'en vois de trois natures différentes :

- la compréhension fine des mécanismes épigénétiques et leurs modifications dans un contexte pathologique : on ne connaît pas forcément l'ensemble des épimutations directement responsables d'une maladie ;
- l'efficacité et l'innocuité des outils d'édition épigénomique : comme nous l'avons vu, les protéines ont de nombreuses fonctions pas toujours bien caractérisées. Il faut s'assurer que l'effecteur utilisé fait bien ce que l'on veut et ne présente pas de danger ;
- délivrer l'outil d'édition spécifiquement et efficacement dans les cellules malades du patient. On ne sait pas forcément mettre en contact l'outil d'édition spécifiquement avec les cellules cancéreuses enfouies au cœur d'un tissu d'un malade, sans exposer les cellules saines (dans lesquelles on ne veut pas induire accidentellement des épimutations). *A contrario*, si l'outil d'édition n'est pas capable de cibler toutes les cellules cancéreuses, la récurrence du cancer est assurée...

Une fois ces limites dépassées, il sera possible, en théorie, de soigner un nombre important de pathologies avec une composante épigénétique. Nous allons à présent discuter d'applications potentielles testées en conditions de laboratoire, en utilisant des cellules en culture ou des modèles animaux.

Corriger des épimutations impliquées dans les processus pathologiques

Des anomalies épigénétiques sont impliquées dans des processus pathologiques comme facteurs causaux ou aggravants. De telles anomalies ont été décrites principalement dans un grand nombre de cancers, mais ont aussi été observées dans le cadre de maladies neurodégénératives, métaboliques et immunitaires. Des études récentes montrent que le processus de vieillissement cellulaire s'accompagne lui aussi d'une accumulation d'erreurs épigénétiques caractéristiques.

Concernant le cancer, sur lequel la bibliographie est la plus riche (Muntean *et al.*, 2009, Darwiche, 2020), les travaux de génomique fonctionnelle ont permis d'identifier des aberrations épigénétiques, comme l'acétylation anormalement élevée de *super-enhancers* pro-oncogéniques. Les *super-enhancers* sont des groupes d'*enhancers* souvent associés aux gènes du développement et de la croissance cellulaire. La sur-activation des *super-enhancers* dans les cellules cancéreuses aboutit à l'activation de la prolifération cellulaire, et à l'inhibition de l'apoptose, ce qui permet aux cellules de se diviser de façon non contrôlée.

Comme nous l'avons vu, nous disposons actuellement d'outils qui permettraient de corriger les épimutations rencontrées, par exemple inhiber spécifiquement les *super-enhancers* en désacétylant leurs histones grâce à une dCas9-HDAC. L'intérêt de cette approche réside dans sa double spécificité d'action : cibler une marque particulière sur une région particulière, une combinaison normalement rencontrée uniquement dans la cellule cancéreuse et pas dans les cellules saines autour de la tumeur. On peut donc imaginer que si l'outil dCas9-HDAC entre aussi dans ces cellules saines, il viendra désacétyler un *super-enhancer* non acétylé... ce qui ne devrait pas entraîner trop de problème !

Ce type d'approche a été décliné avec succès *in vitro* sur des modèles de cellules cancéreuses humaines, et la correction de certaines épimutations s'est traduite par une diminution du potentiel cancéreux de ces cellules (diminution des marqueurs de proliférations cellulaires, de migration et d'invasion, promotion de l'apoptose et de la senescence cellulaire, etc.). Ces résultats encourageants ont été reproduits sur des cultures primaires de cellules issues directement de tumeurs humaines. Enfin, la réimplantation des cellules cancéreuses éditées dans des modèles de souris immunodéprimées (xénogreffes) se traduit bien par une diminution *in vivo* du potentiel tumoral de ces cellules.

Il y a cependant peu d'exemple actuel de mise en œuvre d'édition épigénétique ciblée de cellules cancéreuses *in vivo* au sein même d'une tumeur. Le principal obstacle est de délivrer efficacement (et si possible sélectivement) le système d'édition épigénétique, sous forme de précurseur ADN ou ARN, ou de l'effecteur protéique. Une équipe a montré qu'il était possible d'utiliser des nanoparticules pour délivrer de l'ADN codant un système CRISPRa dans un modèle de souris greffées avec des cellules cancéreuses humaines pour induire l'expression d'un gène cible (en l'occurrence un gène anticancéreux). Les nanoparticules sont efficacement endocytées par les cellules cancéreuses qui sont très voraces, mais elles s'accumulent aussi dans les cellules saines d'autres organes comme les reins et le foie, posant le risque potentiel d'effets secondaires indésirables.

Une autre source d'effets secondaires à prendre en compte est le phénomène des hors-cibles (*off-target*), particulièrement pour les systèmes reposant sur l'architecture CRISPR. En effet, il a été observé que la protéine dCas9 peut se fixer sur des séquences génomiques quand celles-ci sont suffisamment similaires à la séquence cible. Si ce phénomène est problématique en recherche fondamentale, il devient une source de préoccupations particulièrement fortes pour les applications *in vivo* des techniques d'édition du génome et de l'épigénome humain à des fins thérapeutiques, car c'est une source très importante de potentiels effets nocifs collatéraux. Des recherches sont en œuvre actuellement afin, d'une part, de réduire le risque de hors-cibles et, d'autre part, de mieux contrôler à la fois l'expression et l'activité de la protéine dCas9 dans le but de limiter les effets indésirables.

L'édition de l'épigénome *in vivo* peut avoir des applications en dehors du cancer, comme rendre la vue à des souris aveugles, en injectant un CRISPRa ciblant le gène du photorécepteur *OPN1MW* dans l'œil (utilisation de souris atteintes de rétinite pigmentaire; Böhm *et al.*, 2020); ou soigner le diabète de souris en reprogrammant des cellules hépatiques en cellules pancréatiques fonctionnelles par CRISPRa (Liao *et al.*, 2017). Ce dernier exemple offre un espoir thérapeutique pour les nombreux diabétiques insulino-dépendants, dont la survie est conditionnée actuellement à l'obtention d'insuline injectable.

Ces exemples vont se démultiplier dans la bibliographie dans les années à venir. Si tout se passe bien, pour certaines pathologies liées à des épimutations ou à des défauts d'expression des gènes, il sera un jour enfin envisageable de passer au test clinique, en espérant que les malades humains réagissent aux nouvelles thérapies épigénétiques aussi favorablement que les modèles animaux.

L'édition épigénétique au service des thérapies cellulaires

Les thérapies cellulaires sont des thérapies émergentes qui consistent à greffer des cellules (ou des tissus) saines afin de restaurer la fonction du tissu endommagé ou malade. Les cellules utilisées sont, dans le meilleur des cas, des cellules différenciées

à partir de cellules souches prélevées chez le patient lui-même. La révolution dans ce domaine vient de la découverte des facteurs de Yamanaka (OCT4, SOX2, KLF4 et c-MYC) (Takahashi *et al.*, 2006). Ces facteurs, quand ils sont exprimés de façon forcée dans des cellules différenciées, permettent de dédifférencier les cellules en cellules pluripotentes induites, ou iPS. Ces iPS sont ensuite redifférenciables dans le type cellulaire désiré (hépatocytes, neurones, astrocytes, myocytes, etc.).

Les outils d'édition de l'épigénome peuvent être mises en œuvre à plusieurs niveaux du processus d'établissement des cellules à des fins thérapeutiques :

- lors de la dédifférenciation en iPS des cellules prélevées. Forcer brutalement l'expression des facteurs de Yamanaka peut être problématique, car certains de ces facteurs ont un rôle pro-oncogénique avéré. On veut réimplanter aux patients des cellules saines sans leur induire un cancer ! Il peut être judicieux d'utiliser l'édition de l'épigénome (CRISPRa ou une dCas9-P300^{core}) pour « réveiller » doucement et de façon contrôlée les gènes codant les facteurs de Yamanaka afin de limiter leur expression, et donc le risque que les cellules deviennent cancéreuses ;

- il est aussi possible d'utiliser l'édition de l'épigénome pour ensuite orienter la redifférenciation des iPS vers le type cellulaire souhaité. En utilisant des systèmes CRISPRa par exemple, on peut forcer la cellule iPS à exprimer un facteur de transcription pro-neural pour la différencier en neurone, ou un autre facteur qui la poussera à devenir une cellule musculaire, ou de la peau... Par exemple, utiliser le système CRISPRa pour l'induction du gène *NEUROG2* dans des iPS de souris par CRISPRa permettrait la différenciation en neurones, alors que celle de *MYOD1* permettrait la reprogrammation des cellules en cellules musculaires squelettiques. L'avantage des CRISPRa sur les techniques traditionnelles (expression forcée directe par un vecteur exogène) reste à étudier ;

- le processus de dédifférenciation des cellules n'est pas parfait : il a été en effet montré que les cellules conservent en partie des éléments de leur épigénome initial, qui peuvent ensuite interférer avec le processus de redifférenciation. La cellule redifférenciée a donc un nombre significatif d'épimutations et souvent une fonctionnalité partielle. L'apport potentiellement majeur de l'édition de l'épigénome réside donc surtout dans la possibilité de corriger ces épimutations afin d'améliorer la différenciation des cellules thérapeutiques *in vitro* avant de les réinjecter dans le patient. Cela devrait permettre *in fine* l'obtention de cellules différenciées les plus proches possibles de cellules « normales ».

On notera que, dans le cas des thérapies cellulaires, qu'il n'y a pas de problème pour délivrer le système d'édition épigénétique, puisque toutes les modifications sont faites en dehors de l'organisme du patient. De plus, il est possible de vérifier si tout s'est bien passé avant de réinjecter les cellules aux patients. L'épimutation est-elle bien corrigée ? Le gène est-il bien ré-exprimé ? La cellule est-elle correctement différenciée et sans danger ? C'est pour ces raisons que les approches de thérapie cellulaire assistées par l'édition épigénomique seront probablement développées dans un futur plus proche que les approches qui proposent démodifier les cellules au sein même du malade. La thérapie cellulaire ne s'applique bien sûr pas à toutes les pathologies, mais, associée à l'édition de l'épigénome, elle présente le potentiel d'être une thérapie révolutionnaire dans tous les cas où elle sera applicable, comme pour certaines maladies neurodégénératives ou métaboliques.

► Conclusion

Les approches d'édition de l'épigénome sont en train de révolutionner le domaine de la génomique fonctionnelle. Pendant longtemps, nous avons été limités à la description des marques épigénétiques et de leurs localisations dans le génome des cellules, nous permettant de poser des hypothèses sur la présence d'*enhancers* et sur la fonctionnalité du génome, mais sans avoir les outils pour tester massivement ces hypothèses. L'édition de l'épigénome nous offre enfin la possibilité de tester la fonctionnalité réelle des *enhancers* et de les associer avec les gènes qu'ils régulent. Cette meilleure compréhension du génome et de ses règles de fonctionnement a pour conséquences directes d'améliorer notre compréhension du lien génotype-phénotype (ou le lien qui existe entre le génome d'un individu et ses caractéristiques physiques ou physiologiques), qui nous a longtemps en grande partie échappé et qui constitue encore aujourd'hui une des grandes questions en biologie.

Les outils d'édition de l'épigénome permettent aussi de décortiquer les mécanismes de la régulation épigénétique, en nous donnant l'opportunité de les modifier avec une certaine subtilité. Notre compréhension de ces mécanismes s'affine, le rôle des complexes chromatinien, des marques, le jeu entre les différentes marques s'éclaircissent peu à peu, et c'est toute la régulation des gènes que nous commençons à mieux comprendre.

Ces révolutions en biologie du génome et de la régulation des gènes sont porteuses d'espoirs thérapeutiques à plus ou moins long terme. L'édition du génome, en permettant de reprogrammer le vivant à des fins thérapeutiques, présente un potentiel infini de soigner le cancer, les maladies métaboliques et dégénératives. Le champ d'application s'accroîtra avec notre compréhension des mécanismes épigénétiques régulant les phénomènes physiologiques et physiopathologiques. Pour ne donner qu'un exemple de perspective à long terme : nous commençons à identifier des épimutations associées au vieillissement. L'édition de ces épimutations porte la promesse de ralentir le vieillissement, ou tout au moins ses symptômes, et d'offrir aux populations âgées l'espoir d'une vie quotidienne en meilleure santé.

Reprogrammer le vivant reste un défi, pour lequel toutes les implications et les menaces sont à prendre en considération : l'amélioration des performances physiques des athlètes est presque à portée de main quand on arrive à stimuler la croissance musculaire chez la souris, par exemple. Que ferons-nous si un jour des variations épigénétiques sont associées à des phénotypes dont le contrôle pose des questions éthiques majeures? La manipulation de l'épigénome n'est pas assimilée à la création d'OGM (organisme génétiquement modifié), qui est strictement encadré par la loi dans la plupart des pays, puisque les marques épigénétiques ne modifient pas la séquence d'ADN. Néanmoins, il a été décrit une hyper-mutabilité des dinucléotides CpG méthylés (en TpG) dans des cas de cancers (Tomso *et al.*, 2003; Misawa *et al.*, 2009), suggérant qu'il est important de suivre l'impact de toute manipulation ponctuelle de l'état de méthylation des CpG. Les progrès techniques sont rapides, et il est nécessaire que le législatif s'en saisisse afin d'élaborer les cadres éthiques de l'édition de l'épigénome, que ce soit à des fins médicales ou vétérinaires, pour l'amélioration des animaux d'élevage et des plantes, pour optimiser l'adaptation à des variations environnementales ou à des fins de profits économiques.

► Références bibliographiques

- Ballouhey O., Bartoli M., Levy N. 2020. CRISP(R)ation musculaire [CRISPR-Cas9 for muscle dystrophies]. *Med Sci (Paris)*, 36 (4), 358-366. <https://doi.org/10.1051/medsci/2020081>
- Bendixen L., Jensen T.I., Bak R.O., 2023. CRISPR/Cas-mediated transcriptional modulation: The therapeutic promises of CRISPRa and CRISPRi. *Mol Ther*, S1525-0016(23)00145-4. <https://doi.org/10.1016/j.ymthe.2023.03.024>
- Beretta M., Mouquet H., 2019. Ingénierie de lymphocytes B humains produisant des anticorps neutralisant le virus VIH-1 par édition génique CRISPR-Cas9 [CRISPR-Cas9 editing of HIV-1 neutralizing human B cells]. *Med Sci (Paris)*, 35 (12), 993-996. <https://doi.org/10.1051/medsci/2019196>
- Böhm S., Splith V., Riedmayr L.M., Rötzer R.D., Gasparoni G., Nordström K.J.V. *et al.*, 2020. A gene therapy for inherited blindness using dCas9-VPR-mediated transcriptional activation. *Sci Adv*, 6 (34), eaba5614. <https://doi.org/10.1126/sciadv.aba5614>
- Cano-Rodriguez D., Gjaltema R.A., Jilderda L.J., Jellema P., Dokter-Fokkens J., Ruiters M.H., Rots M.G., 2016. Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. *Nat Commun*, 7, 12284. <https://doi.org/10.1038/ncomms12284>
- Chen L.F., Lin Y.T., Gallegos D.A., Hazlett M.F., Gómez-Schiavon M., Yang M.G. *et al.*, 2019. Enhancer histone acetylation modulates transcriptional bursting dynamics of neuronal activity-inducible genes. *Cell Rep*, 26 (5), 1174-1188.e5. <https://doi.org/10.1016/j.celrep.2019.01.032>
- Gasperini M., Hill A.J., McFaline-Figueroa J.L., Martin B., Kim S., Zhang M.D. *et al.*, 2019. A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell*, 176 (1-2), 377-390.e19. *Erratum in: Cell*, 176 (6), 1516. <https://doi.org/10.1016/j.cell.2018.11.029>
- Han K., Pierce S.E., Li A., Spees K., Anderson G.R., Seoane J.A. *et al.*, 2020. CRISPR screens in cancer spheroids identify 3D growth-specific vulnerabilities. *Nature*, 580 (7801), 136-141. <https://doi.org/10.1038/s41586-020-2099-x>
- Hilton I.B., D'Ippolito A.M., Vockley C.M., Thakore P.I., Crawford G.E. Reddy T.E., Gersbach C.A., 2015. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol*, 33 (5), 510-7. <https://doi.org/10.1038/nbt.3199>
- Liao H.K., Hatanaka F., Araoka T., Reddy P., Wu M.Z., Sui Y. *et al.*, 2017. In vivo target gene activation via CRISPR/Cas9-mediated trans-epigenetic modulation. *Cell*, 171 (7), 1495-1507.e15. <https://doi.org/10.1016/j.cell.2017.10.025>
- Liu X.S., Wu H., Ji X., Stelzer Y., Wu X., Czauderna S., Shu J. *et al.*, 2016. Editing DNA methylation in the Mammalian genome. *Cell*, 167 (1), 233-247.e17. <https://doi.org/10.1016/j.cell.2016.08.056>
- Misawa K., Kikuno R.F., 2009. Evaluation of the effect of CpG hypermutability on human codon substitution. *Gene*, 431, 18-22.
- Muntean A.G., Hess J.L., 2009. Epigenetic dysregulation in cancer. *Am J Pathol*, 175 (4), 1353-61. <https://doi.org/10.2353/ajpath.2009.081142>
- O'Geen H., Ren C., Nicolet C.M., Perez A.A., Halmaj J., Le V.M. *et al.*, 2017. dCas9-based epigenome editing suggests acquisition of histone methylation is not sufficient for target gene repression. *Nucleic Acids Res*, 45 (17), 9901-9916. <https://doi.org/10.1093/nar/gkx578>
- Sun M., Wang Y., Zheng C., Wei Y., Hou J., Zhang P. *et al.*, 2021. Systematic functional interrogation of human pseudogenes using CRISPRi. *Genome Biol*, 22 (1), 240. <https://doi.org/10.1186/s13059-021-02464-2>
- Taghbalout A., Du M., Jillette N., Rosikiewicz W., Rath A., Heinen C.D. *et al.*, 2019. Enhanced CRISPR-based DNA demethylation by Casilio-ME-mediated RNA-guided coupling of methylcytosine oxidation and DNA repair pathways. *Nat Commun*, 10 (1), 4296. <https://doi.org/10.1038/s41467-019-12339-7>
- Takahashi K., Yamanaka S., 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126 (4), 663-76. <https://doi.org/10.1016/j.cell.2006.07.024>
- Tomso D.J., Bell D.A., 2003. Sequence context at human single nucleotide polymorphisms: Overrepresentation of CpG dinucleotide at polymorphic sites and suppression of variation in CpG islands. *J. Mol. Biol.*, 327, 303-308.
- Yan J., Chen S.A., Local A., Liu T., Qiu Y., Dorighi K.M. *et al.*, 2018. Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions at enhancers. *Cell Res*, 28 (2), 204-220. *Erratum in: Cell Res*, 28 (3), 387. <https://doi.org/10.1038/cr.2018.1>

Liste des auteurs

Margot M.J. Berger, UMR 1287,
Écophysiologie et génomique fonctionnelle
de la vigne, université de Bordeaux, INRAE,
ISVV, Villenave-d'Ornon, France

Maud Borensztein, IGMM, université
de Montpellier, CNRS, Montpellier, France

Pierre Boudry, Département Ressources
biologiques et environnement, Ifremer,
Centre Bretagne, Plouzané, France

Julie Demars, GenPhySE, université
de Toulouse, INRAE, ENVT,
Castanet-Tolosan, France

Véronique Duranthon, université Paris-Saclay,
UVSQ, INRAE, BREED, Jouy-en-Josas, France

Thierry Forné, IGMM, université
de Montpellier, CNRS, Montpellier, France.

Isabelle Fudal, Université Paris-Saclay, INRAE,
UR1290 BIOGER, Palaiseau, France

Anne Gabory, université Paris-Saclay, UVSQ,
INRAE, BREED, Jouy-en-Josas, France

Philippe Gallusci, UMR 1287,
Écophysiologie et génomique fonctionnelle
de la vigne, université de Bordeaux, INRAE,
ISVV, Villenave-d'Ornon, France

Christoph Grunau, Université de Perpignan
Via Domitia, CNRS, Ifremer, université
de Montpellier, UMR5244 Interactions-
Hôtes-Pathogènes-Environnements (IHPE),
Perpignan, France

Hélène Jammes, université Paris-Saclay,
UVSQ, INRAE, BREED, Jouy-en-Josas, France

Catherine Labbé, INRAE, UR1037 LPGP,
Physiologie et génomique des poissons,
Campus de Beaulieu, France

Clément Lafon Placette, université Charles,
Faculté des sciences, Département de
botanique, Benátská 2, Prague, Tchéquie

Delphine Lallias, université Paris-Saclay,
INRAE, AgroParisTech, GABI,
Jouy-en-Josas, France

Audrey Laurent, INRAE, UR1037 LPGP,
Physiologie et génomique des poissons,
Campus de Beaulieu, Rennes, France

David L'Hôte, université Paris Cité,
UMR 8251, Équipe Physiologie de l'axe
gonadotrope-Inserm U 1133,
Paris, France

Stéphane Maury, LBLGC EA1207, USC 1328
INRAE, Université Orléans, France

Lucie Marandel, INRAE, université de Pau
et des Pays de l'Adour, UMR1419 Nutrition
Metabolism and Aquaculture, Aquapôle,
Saint-Pée-sur-Nivelle, France

Frédérique Pitel, GenPhySE,
université de Toulouse, INRAE, ENVT,
Castanet-Tolosan, France

Gaëlle Pontarotti, IHPST - CNRS/ université
Paris 1 Panthéon-Sorbonne, France

Nadia Pons, INRAE, UR1264 MycSA,
Villenave-d'Ornon, France

Clara Roidor, IGMM, université
de Montpellier, CNRS, Montpellier, France

Sarah Voisin, Institute for Health and Sport
(iHeS), Victoria University, Footscray,
Victoria, Australia. Novo Nordisk Foundation
Center for Basic Metabolic Research, Faculty
of Health and Medical Sciences, University
of Copenhagen, Copenhagen, Denmark

Liste des abréviations

3C	chromosome conformation capture
5mC	5-methylcytosine
ADNg	ADN génomique nucléaire
ADNmt	ADN mitochondrial
Airn	Antisense Igf2r RNA
ARNg	ARN-guide
ARNm	ARN messagers
ARNr	ARN ribosomiques
ARNt	ARN de transfert
AS	Angelman syndrome
ATP	adénosine triphosphate
BCG	vaccin Calmette-Guerin-Bacilli
BER	base excision repair
BGC	biosynthetic gene cluster
ChIP	chromatin immunoprecipitation
circARN	ARN circulaires
CMT3	chromométhylase 3
CpG	cytosine liée à une guanine via un groupement phosphate
CRISPR/Cas9	clustered regularly interspaced short palindromic repeats/CRISPR-associated 9
CTCF	CCCTC binding factor
dCAS9	dead Cas9
DDM1	decrease in DNA methylation 1
DLK1	delta like non-canonical Notch ligand 1
DMR	differentially methylated regions
DNase	désoxyribonucléase
DNMT	DNA méthyltransferase
DOHaD	developmental origins of health and disease
DON	deoxynivalenol
dsRNA	double-stranded RNA
epiQTL	epigenetic quantitative trait loci
epiRIL	epigenetic recombinant inbred lines
ERV	endogenous retrovirus
ETI	effector-triggered immunity
EWAS	epigenomic wide association study
FISH	fluorescence in situ hybridization

FLC	flowering locus C
GSE	gènes soumis à l'empreinte
HAT	histones acetyltransferases
hc-siRNA	heterochromatic small interfering RNAs
HDAC	Histone deacetylase
HDAC	histone deacetylase
HDAC	histone deacetylase
HPTM	histone post-translational modifications
HR	hypersensitive response
ICR	imprinting-control region
ICX	inactivation du chromosome X
IGF2	insulin-like growth factor 2
IGF2R	insulin-like growth factor 2 receptor
iPS	induced pluripotent stem cells
kb	kilobase
KDM1A/LSD1	lysine-specific demethylase 1A
KRAB	Krüppel-associated box
KRAB	Krüppel-associated box
LOS	large offspring syndrome (ou AOS, abnormal offspring syndrome)
LTR	long terminal repeats
LUMA	luminometric methylation assay
Mb	mégabase
MBD	methyl-CpG-binding domain
MCI	masse cellulaire interne
MEG3	maternally expressed 3
meJA	methyl jasmonate
MET1	methyltransferase 1
miARN	micro-ARN
NLR	nucleotide-binding site leucine-rich repeat
OSKM	OCT4, SOX2, KLF4 et c-MYC,
PAR	pseudo-autosomal region
PcG	polycomb group
PEG11	retrotransposon-like gene 1
PfEMP1	<i>P. falciparum</i> erythrocyte membrane protein 1 ()
piARN	les ARN Piwi
PRC2	polycomb repressive complex 2
PRMT	protein arginine methyltransferase
PWS	Prader-Willi syndrome
QTL	quantitative trait locus
RNA-FISH	RNA-fluorescence in situ hybridization
SA	salicylic acid
SAM	S-adenosyl methionine

SET	Su var3-9, Enhancer of Zeste and Trithorax
SIGS	spray-induced gene silencing
SNP	single nucleotide polymorphism
sRNA	small RNA
TAD	topologically associating domain
TALEN	transcription activator-like effector nucleases
TCGA	The Cancer Genome Atlas
TDG	thymine DNA glycosylase
TET	ten-eleven translocation
TrxG	trithorax group
Xce	X-choice element
Xic	X inactivation center
XIST	X-inactive specific transcript
ZBED6	Zinc finger BED domain-containing protein 6
ZFN	zinc finger nucleases
ZFP	zinc finger protein

Relecture : xxxxxx

Mise en page : Hélène Bonnet, Studio9

Achévé d'imprimer en xxxxxx 2024
par xxxxxx

Numéro d'impression :

Dépôt légal : xxxxxx 2024