

République algérienne démocratique et populaire
Ministère de l'enseignement supérieur et de la recherche scientifique
Université des Frères Mentouri - Constantine 1

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي و البحث العلمي
جامعة الاخوة منتوري - قسنطينة 1



Faculté des Sciences de la Nature et de la Vie
Département de Biologie Animale

كلية عاوم الطبيعة و الحياة
قسم بيولوجيا الحيوان

Support de cours

Épigénétique

Destiné aux étudiants de Master 1

Spécialité Génétique

Docteur *REZGOUNE Mohamed Larbi*
Maître de conférences (A)

Année universitaire
2021 - 2022

Table des matières

Liste des abréviations

Liste des figures

Liste des tableaux

Préambule

Chapitre I : Définition de l'épigénétique	Page 01
1- Naissance du concept d'épigénétique	01
2- Définition de l'épigénétique	01
3- Information épigénétique	02
4- Définition du concept de l'épigénome	03
Chapitre II : Organisation structurale et fonctionnelle de la chromatine	05
1- Découverte de la chromatine	06
2- Composition de la chromatine	06
2-1- Protéines histones	06
2-2- Variants histones	08
2-3- Protéines non-histones	09
3- Structure de la chromatine	10
3-1- Nucléosome et la fibre F10	10
3-2- Fibre F30	12
3-3- Boucles chromatiniennes	13
3-4- Quatrième niveau de condensation	14
4- Différentes étapes de condensation de la chromatine	15
4-1- Assemblage du nucléosome et formation de la F10	15
4-2- Formation des niveaux d'organisation supérieurs de la chromatine	16
5- Territoires chromatiniens	19
Chapitre III : Les modifications post-traductionnelles des histones	21
1- Régulation de l'état de condensation de la chromatine	21
2- Impact des modifications post-traductionnelles des histones	22
3- Principales modifications post-traductionnelles des histones	24
3-1- Acétylation	24
3-2- Méthylation	25
3-3- Phosphorylation	25
3-4- Autres modifications post-traductionnelles	26
4- Code épigénétique des histones	28

Chapitre IV : Méthylation de l'ADN	30
1- Découverte du processus de méthylation de l'ADN	30
2- Caractéristiques du processus de méthylation de l'ADN	31
3- Rôles biologiques de la méthylation de l'ADN	33
3-1- Développement embryonnaire	33
3-2- Différenciation cellulaire	34
3-3- Empreinte génétique parentale	34
3-4- Inactivation du chromosome X	35
3-5- Inhibition de séquences parasites	35
3-6- Vieillesse cellulaire	36
4- Acteurs moléculaires de la méthylation de l'ADN	36
4-1- Dnmt1	36
4-2- Dnmt2	37
4-3- Dnmt3	37
4-4- Protéines liant les cytosines méthylés (MBD)	38
4-5- Déméthylase de l'ADN ?	38
5- Processus de verrouillage l'expression par méthylation de l'ADN	39
6- Implication de la méthylation de l'ADN en cancérogenèse	40
Chapitre V : Empreinte parentale	41
1- Définition de l'empreinte parentale	41
2- Découverte de l'empreinte parentale	41
3- Les gènes soumis à l'empreinte	44
4- Mise en place de l'empreinte	46
5- Empreinte parentale et pathologies humaines	47
Chapitre VI : Inactivation du chromosome X	48
1- Découverte de l'inactivation du chromosome X	48
2- Compensation de dose chez les mammifères	49
3- Dynamique d'inactivation du chromosome X	49
3-1- Cycle de l'inactivation	49
3-2- Modalités de l'inactivation	50
3-3- Réactivation du X	50
4- Mécanistique d'inactivation	500
Chapitre VII : Le monde des ARN non codants	53
1- ARN dans la cellule	53
2- Découverte des ARNnc	53
3- Définition et classification actuelle des ARNnc	54
4- Interférence ARN (siRNA et miRNA)	55
4-1- siRNA	55
4-2- miRNA	57
5- Longs ARNnc	60
Références bibliographiques	63

Liste des abréviations

ADN : Acide Désoxyribo-Nucléique
APRT : Adénine Phospho-Ribosyl-Transférase
ARN : Acide Ribo-Nucléique
ARNi : ARN interférence
ARNnc : ARN non-codants
ARNr : ARN ribosomiques
ARNt : ARN de transfert
AS : Syndrome d'Angelman
ASF1 : Anti-Silencing Function 1
ATP : Adénosine Tri-Phosphate
ATPase : Adenosine Tri-Phosphatase
CAF1 : Chromatin Assembly Factor 1
CpG : dinucléotide composé d'une cytosine suivi d'une guanine
DBD : DNA Binding Domain
DMR : Differentially Methylated Region
Dnmt : DNA méthyltransférase
ENCODE : Encyclopedia of DNA elements
GNAT : GCN5-N-Acetyltransferaserelated
HAT : Histone-Acétyle-Transférases
HDAC : Histone-DésACétylases
HDM : Histone-Dé-Méthylases
HMT : Histone Méthyl-Transférases
ICR : Imprinting Control Region
IGF2 : Insulin-like Growth Factor 2
LINES : Long Interspersed Nuclear Element
MBD : Methyl CpG Binding Domain
MeCP : méthyl CpG binding domain
miRNA : microRNA
NAP1 : Nucleosome Assembly Protein 1
PAD : Peptidyl-Arginine Deiminase
PARP : Poly (ADP-Ribose) Polymérase
PCNA : Proliferating Cell Nuclear Antigen
PGC : Primordial Germ Cells
PWS : Prader-Willi Syndrome
SAE : SUMO-activating enzymes
SAM : S-Adénosyl-Méthionine
SAR : Scaffold Attachment Region
SINES : Short Interspersed Nuclear Element
siRNA : small interfering RNA
SMC : Structural Maintenance of Chromosome
snoRNA : Small nucleolar RNA
SRC : Steroid Receptor Coactivator
SUMO : Small Ubiquitin-like Modifier
TRD : Transcriptional Repression Domain
UDP : Uni-Disomie Parentale
Xist : X-inactive specific transcript

Liste des figures

Figure 01 :	L'épigénome, un second niveau d'information	04
02 :	Représentation schématique de la structure des histones majeures	07
03 :	Cliché en microscopie électronique du nucléo-filament	10
04 :	Structure de l'histone cœur du nucléosome	11
05 :	Structure du nucléosome et de la fibre F10	12
06 :	Passage de la F10 à la F30	13
07 :	Modèles de structure de la F30	13
08 :	Représentation schématique de la structure des boucles chromosomiques	14
09 :	Cliché en microscopie électronique de chromosomes métaphasiques	15
10 :	Cliché en microscopie électronique d'un chromosome sans histones	17
11 :	Représentation schématique de structure des condensines	18
12 :	Représentation schématique d'un noyau interphasique coloré au Feulgen	20
13 :	Aperçu des principales modifications post-traductionnelles des histones	23
14 :	Mécanismes des principales modifications post-traductionnelles des histones	24
15 :	Interdépendance des modifications post-traductionnelles des histones	28
16 :	Effet des principales modifications post-traductionnelles des histones	29
17 :	Réactions chimiques de méthylation/déméthylation de la cytosine	33
18 :	Asymétrie dans la fonction du génome paternel et maternel chez les équines	41
19 :	Technique de transfert nucléaire dans des œufs fécondés de souris	43
20 :	Mise en place de l'empreinte parentale chez la souris	46
21 :	Mécanistique de l'inactivation du chromosome X	52
22 :	Mécanisme général de l'ARN interférence	57

23 :	La voie miRNA	59
24 :	Classification des longs ARNnc	61

Liste des tableaux

Tableau I :	Caractéristiques biochimiques des histones	07
--------------------	--	-----------

Préambule

Le présent polycopié est destiné aux étudiants de première année Master de la formation académique Génétique. Il peut constituer également un appui pour tout biologiste désireux de s'initier à l'épigénétique. Seuls sont présentés les principes et les concepts de base de cette nouvelle discipline très large, en s'appuyant le plus possible sur des données actualisées. Appréhender l'épigénétique nécessite des connaissances en deux disciplines fondamentales : la biologie moléculaire et la génomique ; enseignements prodigués en Licence et en Master.

L'importance du principe épigénétique est mise en évidence du fait que toutes les cellules d'un organisme vivant partagent un génome identique. Cependant, elles n'ont pas toutes la même morphologie et ne contribuent pas aux mêmes fonctions biologiques. Cette hétérogénéité est le fruit d'une expression différente du génome en fonction du type cellulaire. La machinerie qui se met en place pour réguler cette expression fait appel à des mécanismes moléculaires complexes appartenant au domaine de l'épigénétique. Le terme épigénétique définit aujourd'hui les modifications transmissibles et réversibles de l'expression des gènes et ne s'accompagnant pas de changements des séquences nucléotidiques. Ce type de régulation peut cibler l'ADN ou l'ARN et agir au niveau du noyau ou du cytoplasme. La transmission d'un état épigénétique peut être réalisée par la méthylation de l'ADN, par des modifications post-traductionnelles des histones, et par la déposition de variants d'histones. De plus, il devient de plus en plus évident que les mécanismes médiés par les petits ARN non codants jouent un rôle central dans la mise en place et le maintien de l'état d'activité des gènes.

Afin de simplifier la lecture de ce polycopié et de susciter chez le lecteur des questions et des réflexions, son attention est attirée par une organisation en plusieurs chapitres disposés selon un enchaînement cohérent. Dans le premier chapitre, nous avons essayé de donner une définition très simplifiée du concept d'épigénétique et d'intégrer la terminologie particulière utilisée dans cette discipline au bagage linguistique de l'étudiant.

Pour qu'un gène puisse conduire à la synthèse d'une molécule (protéine ou ARN fonctionnel), il doit être lisible, c'est-à-dire accessible à différents complexes protéiques qui interviennent dans ce processus. Les marques apposées sur les histones modifient l'état de compactage de la molécule d'ADN, favorisant ou au contraire limitant l'accessibilité aux gènes. Pour comprendre cet aspect, nous avons consacré les chapitres II et III respectivement à la description de la structure de la chromatine ainsi qu'aux modifications post-traductionnelles des histones.

Les marques de méthylation localisées sur l'ADN quant à elles vont le plus souvent obstruer les aires d'arrivée de ces complexes protéiques, conduisant ainsi à l'inactivation des gènes concernés. L'importance de ce mécanisme sera développée dans le chapitre IV. Les chapitres qui suivent sont consacrés à des exemples de processus physiologique où la régulation épigénétique joue, avec ces différentes formes, un rôle important : empreinte parentale (chapitre V) et inactivation du chromosome X (chapitre VI). Dans le chapitre VII nous aborderons la dernière forme de l'information épigénétique, la plus complexe et aussi la plus méconnue, à savoir le monde des ARN non codant. Il est désormais largement admis que des anomalies épigénétiques contribuent au développement et à la progression de maladies humaines, en particulier de cancers. Certains syndromes héréditaires résulteraient eux aussi de mutations dans les gènes codant pour la machinerie épigénétique. Par ailleurs, le rôle de l'épigénétique est soupçonné et très étudié dans le développement et la progression de maladies complexes et multifactorielles, comme les maladies neurodégénératives (Alzheimer, Parkinson, sclérose latérale amyotrophique, Huntington...) ou métaboliques (obésité, diabète de type 2...). La contribution de l'épigénétique en pathologie humaine sera développée sous forme de travaux personnels fournis à l'étudiant et traités de manière interactive pour développer leur esprit de synthèse et leur permettre d'appliquer les enseignements prodigués dans cette matière.

Aujourd'hui, l'épigénétique offre la possibilité de progresser dans la compréhension et la maîtrise de l'expression du patrimoine génétique. C'est un vaste domaine d'investigations puisque de nombreux processus moléculaires sont impliqués, interagissant entre eux et nécessitant d'innombrables acteurs pour mener à bien toutes les étapes d'écrasement, d'apposition des marques, de modifications et de lecture. Malgré, l'avalanche de publications et de données à hauts et moyens débits, les questions résolues sont bien moins nombreuses que celles posées.

Chapitre I

Définition de l'épigénétique

1- Naissance du concept d'épigénétique

Chacune de nos cellules renferme l'ensemble de notre patrimoine génétique. Mais, si toutes nos cellules contiennent la même information génétique, elles n'en font manifestement pas tout le même usage : une cellule de la peau ne ressemble en rien à un neurone, une cellule du foie n'a pas les mêmes fonctions qu'une cellule du cœur. De même, deux jumeaux qui partagent le même génome ne sont jamais parfaitement identiques ! À travers ces exemples et dans bien d'autres, la clef du mystère se nomme « épigénétique ». L'organisme humain est composé de 10^{13} cellules organisées en 106 types cellulaires différenciés ou spécialisés (épithéliales, musculaires, nerveuses ...etc.). Toutes ces cellules sont issues d'une seule cellule ; l'œuf fécondé, qui, par des mitoses successives, va donner tous les types cellulaires. Autrement dit : toutes les cellules qui composent un organisme sont génétiquement identiques, car issues de mitoses ; processus conservateur de l'information génétique qui génère à partir d'une cellule mère deux cellules filles identiques entre elles et identiques par rapport à la cellule mère préexistante. Comment expliquer alors les variations de forme, de physiologie et de fonction qui existent entre les cellules différenciées alors qu'elles possèdent le même génome ?

La réponse réside dans le contrôle de l'expression des gènes, domaine de la biologie qu'on nomme épigénétique. Aujourd'hui il est admis que les cellules diffèrent n'ont pas du fait qu'elles contiennent des gènes différents, mais du fait qu'elles activent ou répriment l'expression de gènes différents. Alors que la génétique correspond à l'étude des gènes, l'épigénétique s'intéresse à une « couche » d'informations complémentaires qui définit comment ces gènes vont être utilisés par une cellule... ou ne pas l'être.

2- Définition de l'épigénétique

Venant du mot épigénèse, théorie selon laquelle l'embryon est indifférencié lors des phases précoces du développement, le terme « épigénétique » qui signifie littéralement « sur », « au-dessus de » ou « ce qui couvre » la génétique, a été introduit pour la première fois par *Conrad Waddington* en 1942, à une époque où on n'avait pas encore établi la nature Acide Désoxyribonucléique (ADN) du support de l'information génétique, pour décrire la discipline de la biologie qui étudie « les interactions des gènes avec leur environnement pour définir le phénotype des êtres ».

De nos jours, le terme épigénétique peut être défini comme l'étude des modifications transmissibles et réversibles du profil d'expression des gènes, autrement dit la manière dont un gène est exprimé, qui ne s'accompagnent pas de changements dans la séquence d'ADN. La génétique, telle qu'elle a été pratiquée pendant de nombreuses années, se focalisait sur les modifications transmissibles et irréversibles de la séquence de l'ADN, comme les mutations, n'a effectivement pas suffi à élucider certains des mystères du vivant. L'épigénétique est donc une nouvelle science capable de combler les insuffisances de la génétique mendélienne classique. Elle représente une vision complémentaire du contrôle de l'information génétique et désigne la grande diversité des mécanismes modulant l'expression des gènes. L'information épigénétique est donc une « extension » de l'information génétique.

3- Information épigénétique

Les modifications épigénétiques sont induites par l'environnement au sens large, à l'échelle cellulaire et celui de l'organisme : la cellule reçoit en permanence toutes sortes de signaux l'informant sur son environnement, de manière à ce qu'elle se spécialise au cours du développement, ou ajuste son activité à une situation particulière. Ces signaux, peuvent conduire à des modifications dans l'expression de nos gènes, sans pour autant affecter leur séquence. Le phénomène peut être transitoire, mais il existe des modifications épigénétiques durables, qui persistent même lorsque le signal qui les a induites disparaît. Concrètement, ces modifications sont matérialisées par des marques biochimiques, apposées par des enzymes spécialisées sur l'ADN ou sur des protéines qui le structurent (histones) et parfois même en faisant appel à des molécules d'Acide Ribonucléique (ARN) particulières qu'on appelle ARN non codant (ARNnc). Alors que l'information génétique s'exprime sous une seule forme ; une séquence d'ADN, l'information épigénétique s'exprime sous quatre formes possibles :

- La modification de la structure de l'ADN par méthylation des cytosines,
- Les modifications post-traductionnelles histones par fixation covalente de différents groupements chimiques (acétylation, méthylation, phosphorylation, ubiquitinylation, ribosylation, etc.),
- L'incorporation de variants d'histones (H2A-X, H2A.Z, H3.3, etc.).
- Les régulations médiées par les petits ARN non-codants.

Un phénomène épigénétique peut être régulé par une ou plusieurs formes de cette information.

4- Définition du concept de l'épigénome

De nos jours, l'épigénétique fait référence à la fois aux changements héréditaires de l'activité et de l'expression des gènes de la descendance d'une cellule ou d'un individu, et aux changements stables sur le long terme du programme transcriptionnel d'une cellule, sans pour autant être nécessairement transmis aux cellules filles. Ainsi, l'épigénétique représente l'ensemble des mécanismes permettant à une cellule de faire face aux différents événements de sa vie (différenciation, stress, etc.) en modulant son programme d'expression génétique. Les modifications de l'information épigénétique sont à la fois stables et modulables par un grand nombre de facteurs cellulaires, reflétant un état physiologique ou pathologique de la cellule ainsi qu'une adaptabilité à son environnement. En comparaison du génome qui représente l'ensemble des gènes d'une cellule, l'épigénome représente l'ensemble des marques et modifications épigénétiques d'une cellule. Cet épigénome est propre à chaque cellule à un instant t , et influe directement sur le génome et donc sur le programme génétique utilisé par la cellule.

En conclusion, un épigénome est défini comme étant : **l'ensemble des modifications épigénétiques propre à chaque type cellulaire différencié à un instant t .**

Il est important de noter qu'il existe une grande connectivité entre ces différentes informations épigénétiques : de nombreux travaux mettent en évidence une synergie et un enchaînement de modifications épigénétiques pour arriver *in fine* à la régulation spécifique de l'expression du génome. Ces différents mécanismes font appel à toute une batterie de facteurs et de complexes :

- **Les « écrivains » et les « effaceurs »** : capables de modifier les marques épigénétiques,
- **Les « lecteurs et recruteurs »** : capables de reconnaître une ou plusieurs marques épigénétiques et de cibler ainsi des régions spécifiques de la chromatine afin d'y recruter différents complexes protéiques,
- **Les « effecteurs »** : capables de remanier la chromatine et/ou d'entraîner différents processus biologiques.

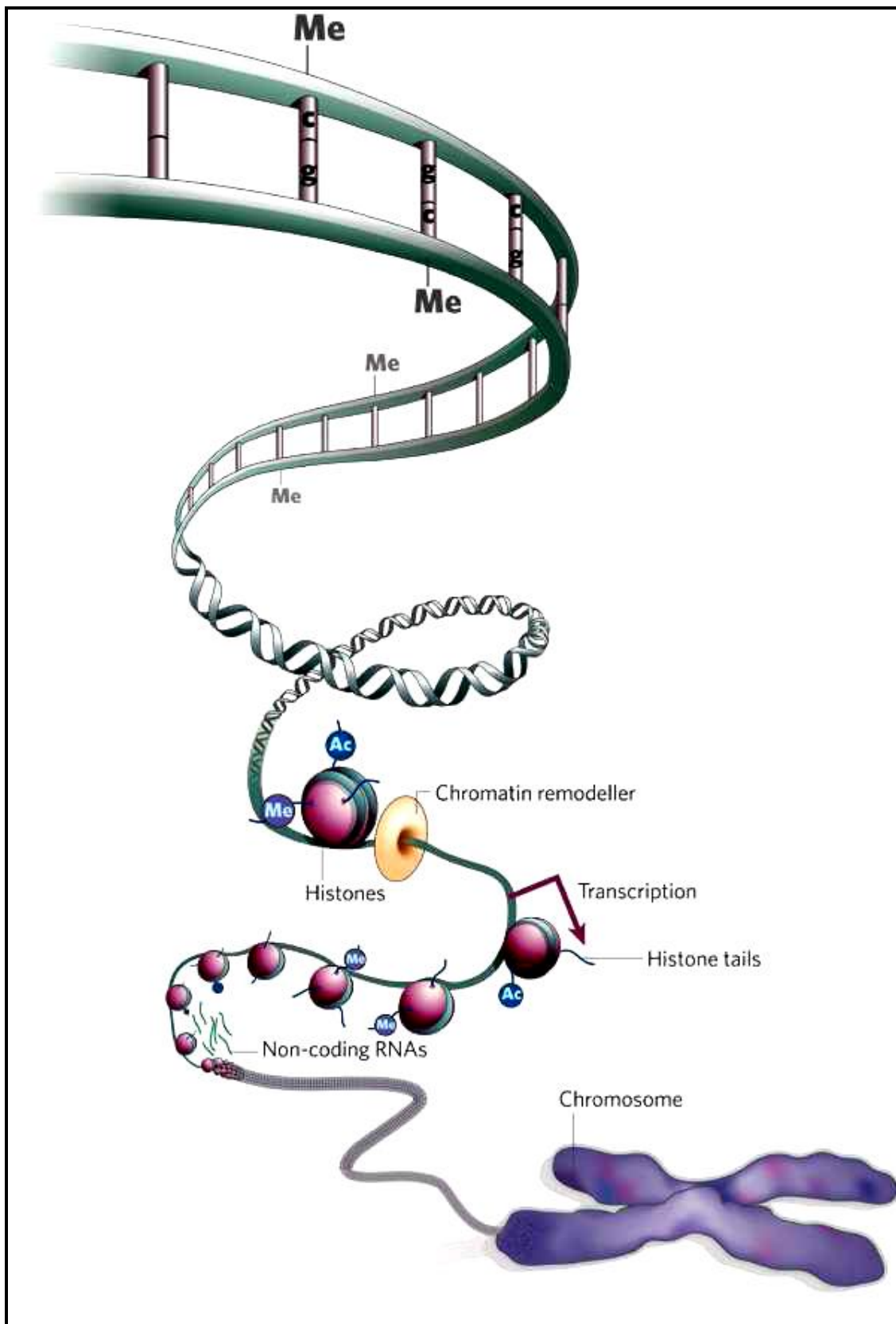


Figure 01 : L'épigénome, un second niveau d'information (Jones *et al.*, 2008).
 (La troisième forme de l'information épigénétique ou variants histones, n'est pas mentionnée dans cette figure)

Chapitre II

Organisation structurelle et
fonctionnelle de la chromatine

La chromatine est la forme sous laquelle se présente l'ADN dans le noyau. C'est la substance de base des chromosomes eucaryotes. Elle correspond à l'association de l'ADN et de protéines structurales et fonctionnelles spécifiques.

À titre d'exemple, l'ADN contenu dans un noyau d'une cellule humaine est composé d'environ 50 000 gènes, repartis au sein de 23 paires de chromosomes pour un total de 6,4 milliards de paires de bases (pb). L'ADN est une macromolécule de taille conséquente, d'environ 2 mètres de long mis bout à bout, pour 2,50 nm de large, et tout le génie d'une cellule réside dans sa capacité à maintenir une si grande quantité d'ADN au sein de l'espace si restreint d'un noyau cellulaire. Une image couramment utilisée pour se représenter une échelle de taille est que si l'on compare le noyau d'une cellule à une orange, l'ADN replié à l'intérieur peut être assimilé à une pelote d'un long fil de pêche d'environ 20 kilomètres. La cellule doit donc structurer son ADN sans s'emmêler, tout en restant capable d'accéder à n'importe quel moment à l'information codée par un gène précis. Ceci réclame une étroite collaboration entre l'ADN et les protéines structurantes de la chromatine, afin de permettre l'accès en temps voulu à la machinerie cellulaire pour diverses activités telles que la réplication, la transcription, la réparation, etc... On peut donc attribuer deux fonctions à la chromatine :

- **une fonction de compaction de l'ADN** : la chromatine permet en effet la compaction d'un polymère d'ADN d'environ deux mètres, dans un espace nucléaire dont les dimensions ne dépassent pas quelques micromètres dans le cas de l'homme. La chromatine est très organisée dans le noyau, et cette organisation résulte de différents niveaux hiérarchiques de compaction. Deux questions auxquelles on va essayer de répondre dans ce chapitre : comment arriver à ce niveau de compaction ? et comment garder quand même une grande activité en ce qui concerne la réplication, la transcription et la réparation ?
- **une fonction dynamique qui permet la régulation de l'expression des gènes** : des marques apposées sur les histones modifient l'état de compactage de la molécule d'ADN, favorisant ou au contraire limitant l'accessibilité aux gènes.

Quelle est la structure de la chromatine et comment est-elle modulable afin d'apporter une information supplémentaire à celle de l'ADN ?

1- Découverte de la chromatine

La substance nucléaire ou chromatine, a été décrite pour la première fois par *Miescher* en 1869 est nommée ainsi par *Fleming* en 1882. L'étymologie du mot chromatine est « corps colorés ». Elle est disposée dans le noyau de la cellule sous forme de réseaux de filaments enchevêtrés sans organisation apparente. Il semble que ce soit *Sutton* en 1903 et *Boveri* en 1904 qui proposèrent pour la première fois d'associer les gènes aux chromosomes qui deviendraient ainsi supports de l'hérédité. Ces deux chercheurs étaient intrigués par le comportement des chromosomes lors de la mitose décrit par *Flemming* en 1879. Au début de 20^{ème} siècle se développèrent les observations cytologiques par l'utilisation de techniques de colorations spécifiques de la chromatine, dont la première était le Feulgen : immersion dans du HCl suivit d'une coloration par la fuschine, la technique de coloration la plus répandue de la chromatine utilise le Giemsa qui est un mélange de deux colorants : le bleu de méthylène à caractère basique et l'éosine à caractère acide. La chromatine présente également une très grande affinité pour l'hématoxyline qui la colore en noir, le bleu de méthylène en bleu, le bleu de toluidine et pyronine qui lui donne une teinte bleu-violacée.

2- Composition de la chromatine

La chromatine est composée d'acides nucléiques (ADN et ARN) et de protéines. Les protéines sont de deux types : essentiellement les histones à caractère basique, et les protéines non-histones dites « acides ». L'ARN est représenté essentiellement par des chaînes d'ARNm naissantes qui sont en transit vers le cytoplasme où ils seront traduits.

2-1- Protéines histones

Petites protéines nucléaires d'environ 120 résidus, à caractère basique riches en acides aminés arginine (R) et en lysine (K) qui représentent plus de 25% de leurs compositions, ils possèdent ainsi de nombreuses chaînes latérales chargées positivement qui se lient aux groupements phosphates de l'ADN chargés négativement ; l'ADN est un polyanion possédant une charge négative globale due à la fonction acide libre de l'acide phosphorique. Leurs poids moléculaires (PM) varient de 11 à 23 kilodaltons (kDa). Cinq classes d'histones ont été caractérisées en fonction de leurs proportions relatives en acides aminés basiques : H1, H2A, H2B, H3 et H4 : H3 et H4 riches en Arginine, H2A et H2B légèrement riches en Lysine et la H1 très riche en Lysine. Récemment, on a pu isoler H5 des érythrocytes aviaires, H6 dans les cellules germinales de poissons, ainsi que de nombreux variantes histones.

Les histones sont les protéines liées à l'ADN les plus abondantes dans les cellules eucaryotes et parmi les protéines connues les plus conservées au cours de l'évolution. Les histones H2A, H2B, H3 et H4 présentes sont appelées histones majeures, histones cœur, ou réplication dépendante. Cependant, l'histone H1, appelée histone inter-nucléosomale ou histone de liaison, présente des caractéristiques fondamentalement différentes des autres classes d'histones (**tableau I**).

Tableau I : Caractéristiques biochimiques des histones (**Ridgway, 2006**).

Histones	Acides aminés basiques		Acides aminés Acides	Rapport basiques / acides	PM (Da)
	Lysine	Arginine			
H1	29 %	1 %	5 %	5,4	23000
H2A	11 %	9 %	15 %	1,4	13960
H2B	16 %	6 %	13 %	1,7	13774
H3	10 %	13 %	13 %	1,8	15342
H4	11 %	14 %	10 %	2,5	11282

Les histones majeures possèdent un domaine central hydrophobe de repliement caractéristique des histones appelé « motif histone-fold » formé de 70 acides aminés environ. Ce domaine est composé de trois hélices α reliées entre elles par deux boucles L1 et L2, ainsi que de deux queues flexibles aux extrémités N-terminales (N-ter) et C-terminales (C-ter) linéaires dépourvues de structure secondaire. Ce domaine commun à toutes les histones assure leur dimérisation selon une configuration dite en poignée de main (**figure 02**).

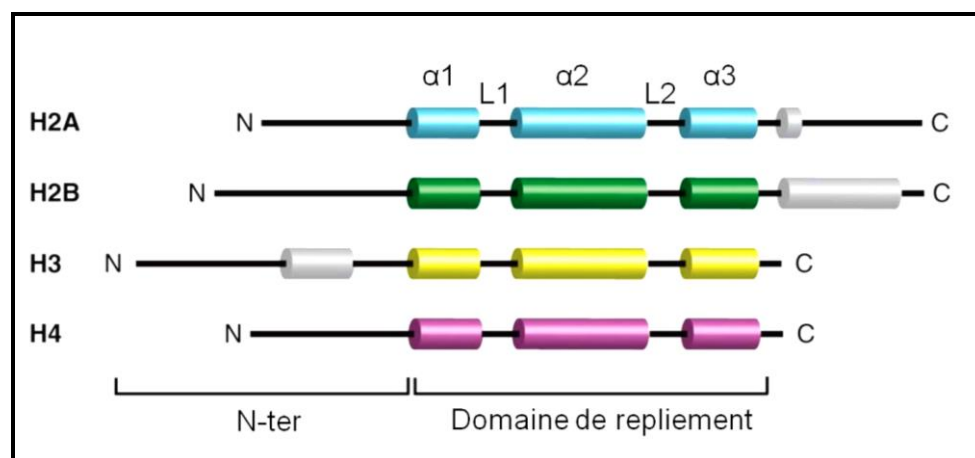


Figure 02 : Représentation schématique de la structure tridimensionnelle des histones majeures (**Young, 2014**).

Les histones H3 et H4 sont les protéines les plus conservées au cours de l'évolution chez toutes les espèces eucaryotes. La région la plus conservée de ces histones est leur domaine central. Par contre, les extrémités N-ter et C-ter, sont plus variables et sont dépourvues de structure secondaire. Ces extrémités sont particulièrement riches en résidus lysine et arginine et sujettes à de nombreuses modifications post-traductionnelles. Les H2A et H2B, de structures similaires à H3 et H3, présentent néanmoins des variations interspécifiques non négligeable.

L'histone de liaison H1 présente quant à elle une structure radicalement différente des autres classes d'histones. Elle possède le PM le plus élevé, très riche en lysine ; c'est la plus basique de toute. Elle possède un domaine globulaire hydrophobe GH1 très conservé, aussi appelé « wing hélice », composé de trois hélices α et d'un feuillet β responsable de la liaison à l'ADN. Les extrémités N-ter et C-ter, très peu conservées, sont impliquées dans de nombreuses interactions protéine-protéine. Cette histone, au niveau de ces extrémités, présente des variations importantes entre tissu et espèces. La H1 ne présente pas d'homologie structurale avec les autres histones et doit son nom du fait qu'elle est associée *in vivo* avec les nucléosomes dans un rapport de 1/1.

Les gènes des histones sont transcrits en phase S du cycle cellulaire de manière synchrone avec la réplication de l'ADN, elle est toujours étroitement couplée avec celle de l'ADN, cette synthèse est rapide puisqu'on ne trouve jamais de segments d'ADN nu au niveau des fourches de réplication : l'inhibition de la synthèse des histones diminue le taux de production de l'ADN et réciproquement.

Les gènes des histones se situent au niveau des chromosomes : 1, 6 et 12, très répétés, de 10 à 1000 copies suivant les espèces, ce qui reflète leurs importances pour la survie de l'organisme. Ces gènes sont regroupés en tandem, mais transcrits séparément. La plupart des gènes des histones sont sans introns et leurs ARNm ne portent pas de queue de poly A.

2-2- Variants histones

Les histones font partie des protéines les mieux conservées au cours de l'évolution. Étant codées par de multiples allèles, les histones possèdent de nombreux variants définis comme étant des isoformes non-alléliques d'une histone canonique (dite aussi typique) qui diffèrent par leurs séquences. Il existe plusieurs variants différents pour chaque sous-type d'histone, excepté pour l'histone H4. Actuellement, ce ne sont pas moins de 55 variants uniques d'histones qui ont été décrits chez l'Homme.

Les histones sont ainsi classées en deux groupes majeurs : les histones canoniques et les variants d'histones, encore appelés histones de remplacement. Au-delà de leur séquence en acides aminés, la distinction entre ces deux groupes se fait selon que leur incorporation au sein de la chromatine est dépendante ou non de la réplication de l'ADN. Les histones canoniques sont exclusivement exprimées au cours de la phase S du cycle cellulaire et incorporées dans la chromatine pour répondre à un besoin important en histones lors de la réplication de l'ADN. Les variants d'histones peuvent, quant à eux, être synthétisés tout au long du cycle cellulaire et sont, la plupart du temps, incorporés dans la chromatine indépendamment de la synthèse d'ADN. Par ailleurs, les pré-ARNm des variants d'histones sont polyadénylés et contiennent des introns et des exons. De plus, il est important de noter qu'il existe des isoformes canoniques d'histones H2A, H2B, H3 et H1 qui ne diffèrent parfois que d'un seul acide aminé.

L'incorporation des différents variants d'histones a une répercussion directe sur les interactions entre histones au sein des nucléosomes. Ce phénomène peut influencer la stabilité des nucléosomes et avoir des conséquences sur la conformation de la chromatine, et sur sa perméabilité à la transcription. Certains variants jouent même directement un rôle dans divers processus cellulaires (mitose, réparation de l'ADN, etc.). Par conséquent, le remplacement des histones canoniques par des variants d'histones représente à lui seul un niveau d'information épigénétique. Des variants d'histones ont été découverts chez plusieurs espèces eucaryotes, différant de façon plus ou moins importante des protéines histones conventionnelles au niveau de leurs séquences, mais également dans leur assemblage au sein des nucléosomes. Ces différences ont une répercussion au niveau de la fonction de ces variants, et présentent des rôles spécifiques dans de nombreux processus biologiques, tels que l'organisation centromérique (H3 CenpA), la réparation de l'ADN, l'inactivation du chromosome X (macro-H2A) ou encore la transcription.

2-3- Protéines non-histones

Sous l'appellation générique de protéines non-histones, dite aussi protéines « acides » on regroupe toutes les protéines nucléaires à l'exception des histones. Il en existe plusieurs et jouent de nombreux rôles :

- protéines de structure contribuant à la formation du squelette chromosomique.
- protéines intervenant dans la réplication de l'ADN, sa transcription et sa réparation.
- protéines jouant le rôle de facteurs d'assemblage et de remodelage de la chromatine comme les protéines chaperonnes.

3- Structure de la chromatine

L'ADN contenu dans le noyau est fortement compacté. Cette compaction de l'ADN résulte de plusieurs niveaux d'organisation. Le premier niveau de repliement est fourni par la double hélice d'ADN elle-même. Puis, l'enroulement de l'ADN autour d'un cœur protéique va constituer ce que l'on appelle un nucléosome. La répétition de cette structure tout au long de l'ADN donnera à la chromatine l'aspect d'un « collier de perles » d'une épaisseur de 11 nm. Dans un second temps, cette fibre de 11 nm se replie pour adopter une structure en hélice de 30 nm de diamètre. Enfin, elle est à son tour repliée en une fibre de 300 nm pour constituer des domaines en boucles de 15 000 à 100 000 pb. Au cours de la mitose et de la méiose, il existe transitoirement un niveau extrême de compaction, les chromosomes. Ils sont tellement condensés qu'on peut étudier leur morphologie au microscope optique.

3-1- Nucléosome et la fibre F10

La chromatine est constituée de particules régulièrement espacées appelées nucléosomes. L'existence de cette unité fondamentale a été décelée dans les années 1970, suite à la digestion de l'ADN par des nucléases et l'obtention de fragments résiduels d'environ 200 pb et leurs multiples et ne sont pas obtenus lorsque la même expérience est effectuée sur de l'ADN purifié. Ceci a permis de proposer qu'il existe, au sein du noyau, une unité structurale de base. Cette interprétation a été largement confortée par des observations de microscopie électronique. Une digestion plus fine permet d'isoler un complexe de 200 kDa, composé d'ADN et de protéines. Peu de temps après, l'enchaînement des nucléosomes en dont l'aspect rappelle celui d'un chapelet a pu être visualisé grâce au progrès de la microscopie électronique en 1974. La même année, *Kornberg* proposera l'existence d'une unité de base au sein du noyau, de deux cents paires de bases d'ADN complexées avec 4 paires d'histones, ce que *Chambon* appellera en 1975 le nucléosome (**figure 03**).

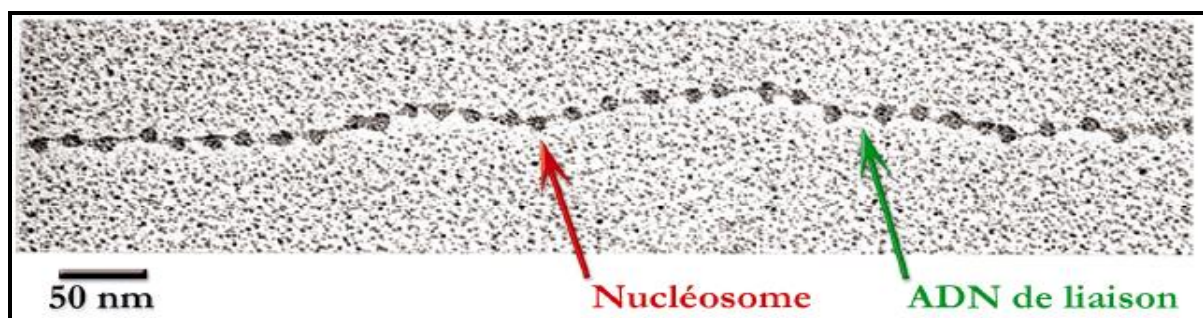


Figure 03 : Cliché en microscopie électronique du nucléofilament (**Adkins, 2004**).

Un nucléosome est composé d'une particule cœur et une région de liaison appelée région inter-nucléosomale ou ADN linker qui relie les particules cœurs adjacentes. La particule cœur dont la structure est très conservée parmi les espèces, est composée de 146 pb d'ADN enroulés selon environ 2 tours (1,7 plus exactement) autour d'un octamère protéique formant deux disques arrangés en parallèle comprenant chacun des histones H3, H4, H2A et H2B ; la molécule d'histone H1 est placée à l'extérieur et joue le rôle de socle. La longueur de région inter-nucléosomale varie de 8 à 116 pb selon les espèces ; elle est de 60 pour l'humain.

Le nucléosome ainsi formé est maintenu par un jeu d'interactions électrostatiques et de liaisons hydrogène entre les groupements phosphates de l'ADN et les histones, et par des contacts non polaires avec le sucre désoxyribose de l'ADN et les histones. Comme aucun contact direct n'existe entre les bases azotées et les histones, le nucléosome ne présente aucune spécificité à la séquence d'ADN.

Cette structure a un diamètre de l'ordre de 11 nm, c'est le premier niveau d'organisation de la chromatine à l'intérieur du noyau et qu'on appelle : Fibre de 10 nm ou F10. Cette organisation qui provoque une compaction de l'ADN selon un facteur 7 : une structure condensée ainsi va se retrouver à 7 fois moins sa longueur initiale (**figures 04 et 05**).

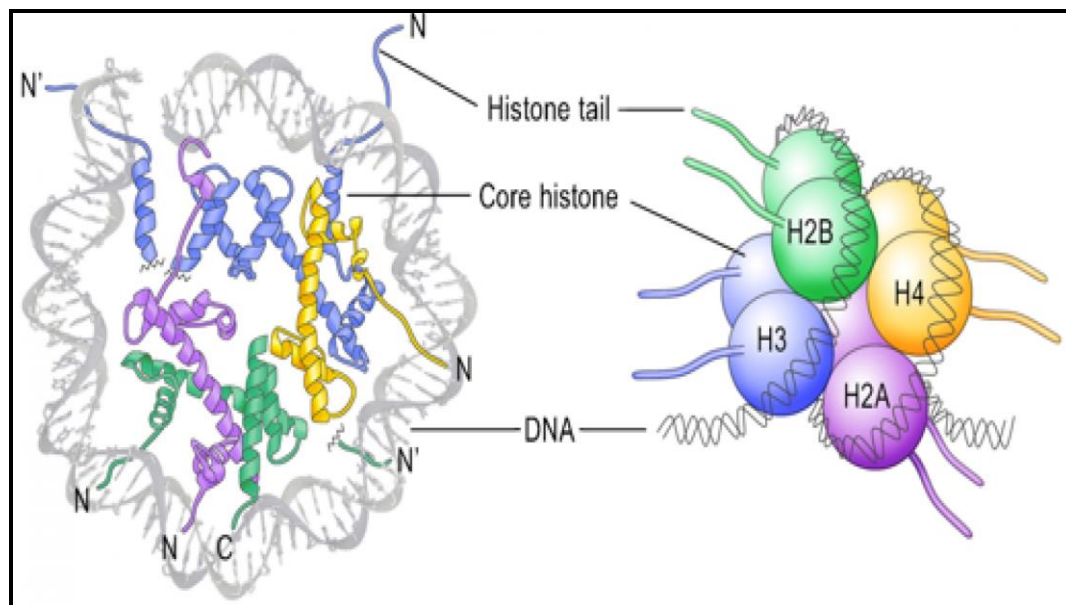


Figure 04 : Structure de l'histone cœur du nucléosome (Ridgway, 2006).

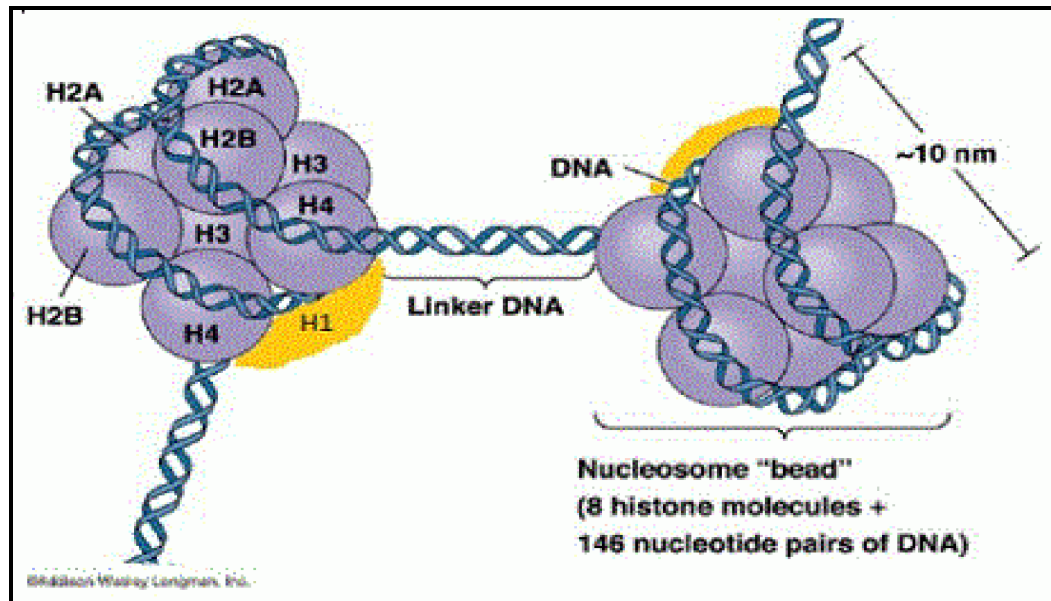


Figure 05 : Structure du nucléosome et de la fibre F10 (Young, 2014).

Il est important de signaler que l'histone de liaison H1 n'intervient pas dans la formation de la sous-unité fondamentale de la chromatine. Cependant, avec des protéines non-histones, elles peuvent interagir avec les nucléosomes pour former des structures de plus haut degré d'enroulement.

3-2- Fibre F30

Les niveaux supérieurs d'organisation des fibres de chromatine ont été visualisés dans différentes conditions salines. À une force saline basse, la chromatine se structure en collier de perles. À une force saline plus élevée (100 mM NaCl), elle forme un cylindre irrégulier d'environ 30 nm de large, communément appelé fibre de 30 nm ou F30.

La F10 présente une structure secondaire enroulée. Chaque tour comprend environ 6 nucléosomes ce qui correspond à un facteur de condensation d'environ 40 : chaque μm le long de l'axe de la fibre contient 40 μm d'ADN. Dans cette fibre, les nucléosomes s'enroulent de façon hélicoïdale en une structure appelée solénoïde ; c'est le deuxième niveau de compaction de la chromatine. Pour cette fibre, deux modèles sont actuellement discutés : le premier dit en solénoïde, décrit précédemment, et qui le plus largement accepté. Le deuxième modèle dit en Zig-Zag (**figure 06 et 07**).

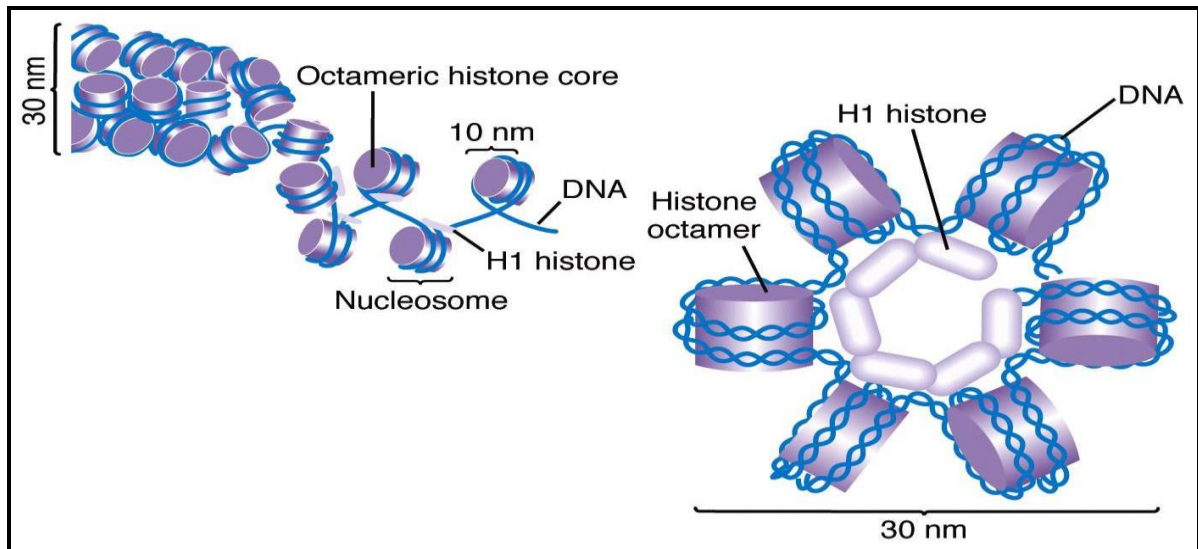


Figure 06 : Passage de la F10 à la F30 (Ridgway, 2006).

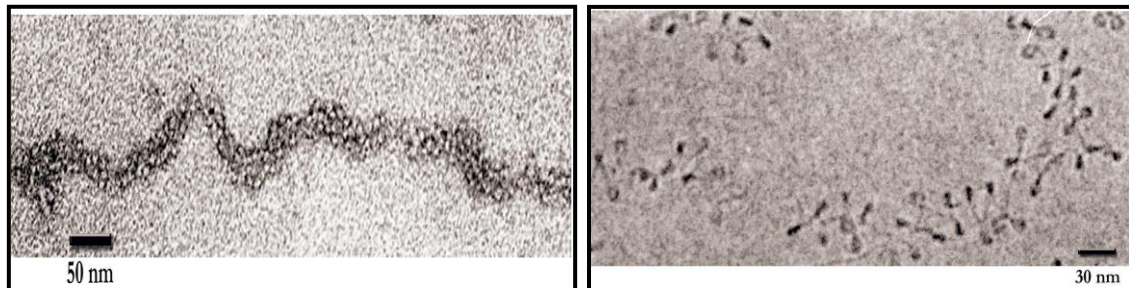


Figure 07 : Modèles de structure de la F30 (Adkins, 2004).

À gauche, un cliché en microscopie électronique de la fibre de 30 nm à salinité physiologique (solénoïde). À droite, un cliché en cryo-microscopie électronique à 5 mM d'ion monovalent.

Dans leur modèle dit de solénoïde, les nucléosomes consécutifs se localisent les uns à proximité des autres dans la fibre, formant une hélice simple. Depuis, cette fibre de 30 nm a été observée par microscopie électronique par de nombreux groupes, mais sa structure intrinsèque est encore débattue. Dans ce second modèle, les nucléosomes sont arrangés en Zig-Zag entre eux, de sorte à former une hélice double brin dont chaque nucléosome se retrouve à proximité du nucléosome.

3-3- Boucles chromatiniennes

La F30 peut également se replier et former des structures dont la topologie est encore moins bien décrite. Ces repliements d'ordre supérieur forment ce que l'on appelle l'organisation à grande échelle de la chromatine, qui repose sur la formation de boucles géantes de chromatine.

Le troisième niveau de compaction est assuré par le repliement de la fibre de 30 nm en boucles, ou domaines topologiques fermés, d'un diamètre de 250 à 300 nm environ, et contenant de l'ordre de 40 à 100 kb d'ADN, ces supers-tours apparaissent sous microscope électronique ancrés autour d'une partie centrale appelée armature ou matrice nucléaire, formée de protéines non-histones. De cette armature protéique partent des boucles d'ADN, les fibres formants ces boucles sont les solénoïdes. La compaction atteint un facteur 1 000. Chaque boucle d'ADN commence et finit au niveau de l'armature, cette dernière est constituée en grande partie de protéines non-histones. Les boucles sont fixées à l'armature au niveau de régions spéciales le long de l'ADN appelées : région de liaison à l'armature ; SAR (Scaffold Attachment Région). Ces régions sont riches en AT et correspondent fréquemment à des fragments d'ADN courbés sujets à des ouvertures locales de la double hélice (comme les origines de réplication) (**figure 08**).

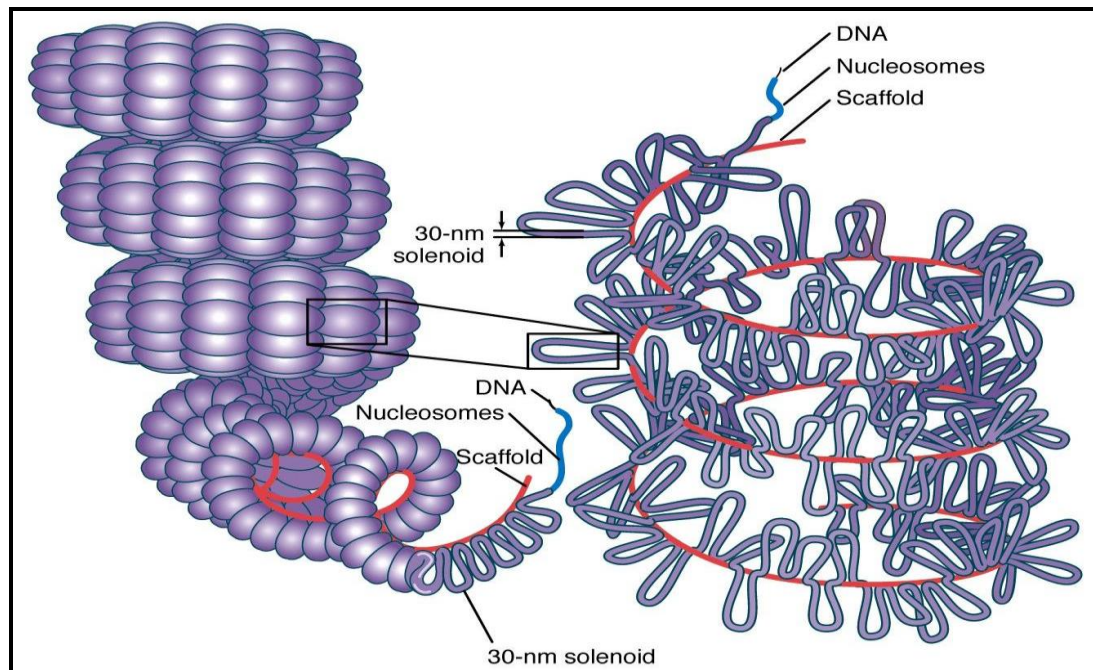


Figure 08 : Représentation schématique de la structure en boucles (**Ridgway, 2006**).

3-4- Quatrième niveau de condensation

Enfin, un quatrième niveau de compaction correspond à l'enroulement des boucles en une hélice de l'ordre de 700 à 850 nm de diamètre qui constitue un chromatide. Le degré de compaction total de l'ADN atteint ainsi plus de 10 000 dans la chromatine condensée. Le chromosome métaphasique observé en mitose représente le niveau de compaction maximale et est capable d'être à la fois résistant, maniable et compact pour assurer une bonne ségrégation du matériel génétique (**figure 09**).

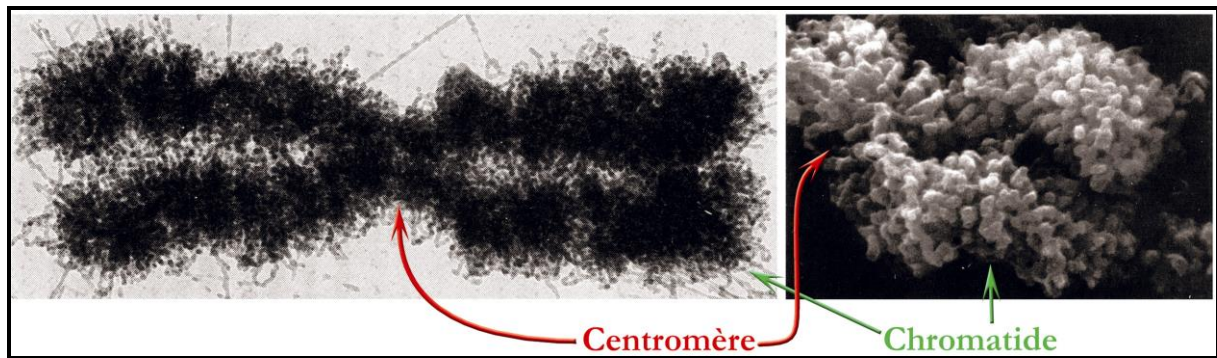


Figure 09 : Cliché microscopie électronique d'un chromosome métaphasique (Adkins, 2004).

4- Différentes étapes de condensation de la chromatine

Le facteur de condensation général du matériel génétique suggère d'emblée que l'ADN ne peut être condensé directement dans la structure finale de la chromatine. Une hiérarchie au sein de l'organisation doit exister. L'assemblage de l'ADN en chromatine comprend plusieurs étapes successives initiées dès la réplication et qui commencent par la formation de la sous-unité fondamentale, le nucléosome, pour aboutir à des niveaux de compaction supérieurs organisés en domaines spécifiques dans le noyau.

La formation de la chromatine peut se découper en deux étapes principales : l'assemblage du nucléosome et la formation de la F10 et la F30, processus qui ne nécessitent pas l'hydrolyse de l'Adénosine Tri-Phosphate (ATP) (ATP-indépendant), puis l'organisation en structures de plus en plus compactes (supertours et boucles) par des processus qui mettent en jeu l'hydrolyse de l'ATP (ATP-dépendant), il s'agit du remodelage de la chromatine.

4-1- Assemblage du nucléosome et formation de la F10

L'assemblage du nucléosome s'effectue grâce à des facteurs d'assemblage de la chromatine qui fonctionnent en coordination avec des protéines chaperonnes qui accompagnent les histones jusqu'à leur dépôt sur l'ADN. Selon le modèle généralement admis, l'assemblage du nucléosome se déroule en deux étapes : un tétramère d'histones H3 et H4 est d'abord déposé sur l'ADN nouvellement répliqué, puis c'est le tour des deux dimères de type H2A-H2B. Ces deux étapes sont sous le contrôle du même facteur d'assemblage de la chromatine CAF1 (Chromatin Assembly Factor 1). Le couplage avec la réplication se ferait par le recrutement de CAF1 au niveau de la fourche de réplication par le facteur PCNA (Proliferating Cell Nuclear Antigen) : complexe protéique joue le rôle d'une pince glissant le long de l'ADN et de facteur de progressivité de l'ADN polymérase.

La première étape est assistée par une chaperonne d'histone ASF1 (Anti-Silencing Function 1) quant au dépôt des histones H2A et H2B est assisté par le facteur NAP1 (Nucleosome Assembly Protein 1). Enfin, l'assemblage du nucléosome se termine par l'addition d'une histone H1 aux brins entrants et sortants de l'ADN entourant l'octamère. Le nucléosome est alors formé. La F30 est formée par l'enroulement spontané de la F10 sous l'effet du fort caractère basique de la H1. Cette F30 peut subir des compactions supplémentaires pour atteindre le repliement des chromosomes mitotiques. La structure précise de ces formes compactes de la chromatine n'est pas connue. Néanmoins, des facteurs de remodelage ATP-dépendant sont impliqués.

4-2- Formation des niveaux d'organisation supérieurs de la chromatine

L'utilisation de la microscopie électronique a permis d'obtenir les premières informations sur l'architecture interne du chromosome métaphasique. Après extraction des histones par une solution de forte force ionique très riche en polyanions, on observe un halo de boucles d'ADN organisées en F30, super-enroulées en périphérie d'un résidu protéique appelé « matrice nucléaire » ou « squelette chromosomique » central ayant grossièrement la forme du chromosome métaphasique. Ces boucles représentent environ 100 kb d'ADN et correspondent aux rosettes de la fibre de 30 nm, insérées par leur base dans la structure protéique centrale au niveau des SAR. En faisant varier les conditions ioniques du milieu, on observe des variations morphologiques du squelette protéique, qui s'étire ou se contracte de façon réversible (**figure 10**).

Deux principales protéines constituent le squelette interne du chromosome : la topoisomérase II et les condensines I et II.

La Topoisomérase est une protéine qui présente une dualité fonctionnelle : de structure, elle fait partie des composants du squelette chromosomique, enzymatique car elle permet de supprimer les supertours positifs créés par l'ouverture localisée de la double hélice. Cette fonctionnalité est essentielle pour permettre une séparation correcte des deux chromatides sœurs pendant l'étape de condensation des chromatides en prophase. L'inactivation du gène de la Topoisomérase II se traduit alors par un défaut de condensation de la chromatine, qui survient de façon imparfaite et incomplète. Cette suppression a pour conséquence une séparation incomplète des deux lots chromosomiques en anaphase et la persistance de masse de chromatine au niveau du plan de division cellulaire. Cette protéine n'est cependant pas un composant fixe de l'axe protéique, elle en constitue au contraire un élément dynamique en renouvellement permanent.

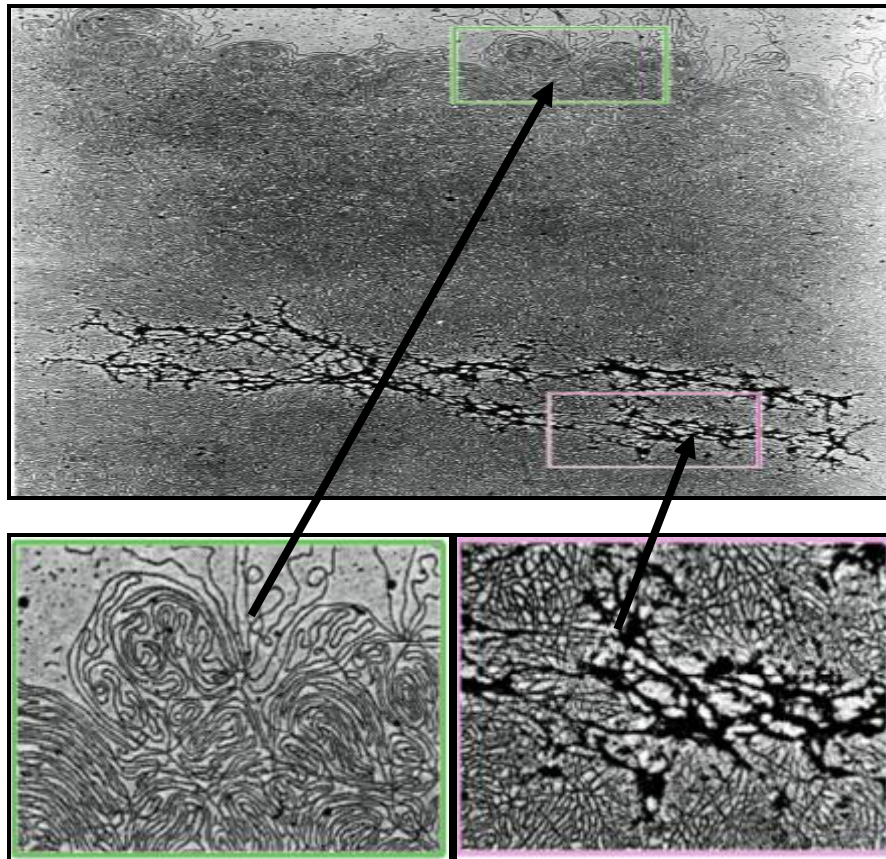


Figure 10 : Modèle d'organisation du chromosome (Adkins, 2004).

Cliché en microscopie électronique d'un chromosome mitotique dont les histones ont été enlevées. Les zones, encadrées en bas de la figure, correspondent à des agrandissements des zones de la même figure. Des boucles d'ADN d'environ 50 à 100 kb (à gauche) sont reliées à une structure d'apparence squelettique (à droite) formée de protéines non-histones.

Le deuxième constituant essentiel de l'axe chromosomique est la condensine, dont il existe deux sous-types : I et II. Ces protéines sont en fait des complexes multiprotéiques associant 2 sous unités SMC (SMC2 et SMC 4 pour Structural Maintenance of Chromosome) et 3 sous unités non SMC (CAP H/H2, CAP D2/D3, CAP G/G2). Les deux sous unités SMC s'associent pour former un V porteur d'un site de fixation pour l'ATP à l'extrémité de chacune des deux branches. Les sous-unités non SMC permettent de contrôler l'ouverture de la pince constituée par les deux protéines SMC. Grâce à l'énergie libérée par l'hydrolyse de l'ATP, les condensines ont la capacité de se lier à l'ADN et de générer des supertours et des boucles, ce qui entraîne une condensation de la chromatine (**figure 11**).

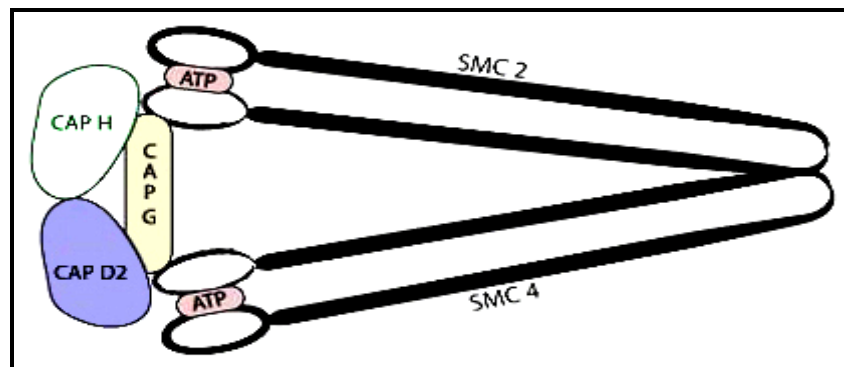


Figure 11 : Représentation schématique de structure des condensines (Becker, 2002).

Le rôle exact des deux sous-types condensine I et II n'est pas clairement identifié, mais il semble que la condensine II intervienne en premier dans les étapes précoces de la condensation, alors que la condensine I interviendrait plus tard pour stabiliser la chromatine dans cet état compacté. De manière surprenante au regard du rôle présumé des condensines, leur inactivation n'entraîne pas une abolition de la compaction, qui survient malgré tout, mais de façon retardée et incomplète. Cette compaction imparfaite entraîne cependant une morphologie anormale des chromosomes en raison d'une désorganisation de l'axe protéique des chromatides, ainsi qu'une ségrégation anormale en raison de masses chromatiniennes persistantes au niveau de la plaque équatoriale en fin d'anaphase.

Un dernier type de protéine ne faisant pas partie du squelette protéique, joue néanmoins un rôle fondamental dans la structure et la physiologie du chromosome métaphasique : il s'agit des cohésines. Ces protéines appartiennent à la même famille que les condensines et ont une structure semblable, constituée de 4 sous-unités spécifiques (2 sous unités SMC1 et 3, et deux sous unités non SMC), RAD21 (REC8 pour les Cohésines méiotiques) et SA1. Les cohésines sont associées à la chromatine pendant la phase S, au moment de la réplication de l'ADN et servent à maintenir ensemble les deux chromatides sœurs jusqu'à leur séparation.

La topoisomérase II et condensine II sont présentes pendant l'interphase dans le noyau. En revanche, la condensine I a une localisation cytoplasmique et ne peut s'associer aux chromosomes qu'après la disparition de l'enveloppe nucléaire. En début de prophase, on observe une phosphorylation de la topoisomérase II et de la condensine II entraînant leur association avec la périphérie de la chromatine en cours de compaction. Cette association permet un premier niveau de compaction de la fibre de 30 nm pour aboutir à une fibre de 200 à 250 nm de diamètre. Puis, un deuxième niveau de compaction apparaît en fin de prophase suite à la réorganisation des protéines de structures qui deviennent axiales et assurent la stabilisation de l'enroulement de la fibre de 250 nm en une fibre de 700 nm.

La condensine I semble jouer un rôle prépondérant dans la stabilisation finale de la structure du chromosome. Un point essentiel de cette architecture chromosomique est son caractère dynamique, caractérisée par un flux permanent de protéines qui s'associent et se dissocient.

5- Territoires chromatiniens

Exception faite de la mitose où elle existe sous forme de chromosomes individuels compactés, la chromatine est diffuse dans le noyau. Elle apparaît sous forme de régions plus ou moins condensées et cet arrangement en compartiments distincts influence les activités fonctionnelles du noyau. Ainsi, la chromatine n'est pas organisée de manière chaotique au sein du noyau, bien au contraire, cette structure est hautement régulée.

En 1928, basé sur des observations exclusivement histologiques, *Heitz* remarqua « l'existence de deux régions contiguës dans le noyau interphasique, mais non délimitées, présentant des affinités tinctoriales distinctes ». En effet, cette différence de coloration est due au fait que certaines zones du noyau cellulaire qui sont plus condensées que d'autres. Il a introduit le terme d'euchromatine (dite active), représentant la chromatine lâche, et d'hétérochromatine (dite inerte), représentant les zones plus denses et compactes : ces deux types de chromatine dans les cellules eucaryotes présentant des critères structuraux et fonctionnels différents. Par la suite, l'euchromatine a été décrite comme associée à une transcription active des gènes, alors que la transcription est majoritairement réprimée au sein de l'hétérochromatine. Il a été communément admis depuis un certain nombre d'années que l'euchromatine, contiendrait des gènes particulièrement actifs, contrairement à l'hétérochromatine qui serait de la chromatine transcriptionnellement inactive. En fait, cette idée est une simplification de la réalité.

Le degré de condensation de la chromatine varie au cours du cycle cellulaire. Pendant l'interphase, l'euchromatine est assez bien décondensée et dispersée dans tout le volume du noyau. Durant cette phase du cycle cellulaire, les gènes sont transcrits et l'ADN se réplique. Une partie de la chromatine interphasique conserve un état très condensé, c'est l'hétérochromatine. Elle échappe à la transcription et contient des séquences d'ADN non codantes, hautement répétitives comme celles que l'on trouve au niveau des centromères et des télomères.

Dans les cellules somatiques, on peut distinguer également, au cours de l'interphase, les zones de réplication précoce correspondant à l'euchromatine, et celles de la réplication tardive qui concernent l'hétérochromatine : c'est ce qu'on appelle le profil répliatif.

On distingue deux sous-catégories d'hétérochromatine, en fonction de leur stabilité : l'hétérochromatine constitutive et l'hétérochromatine facultative.

- Toutes les cellules d'une même espèce vont regrouper des régions similaires de l'ADN sous la forme d'une hétérochromatine constitutive. Par exemple, pour l'Homme, les chromosomes 1, 9, 16 et Y contiennent de larges régions d'hétérochromatine constitutive, majoritairement réprimées. Cette hétérochromatine constitutive est généralement constituée de séquences répétées, et semble jouer un rôle préférentiel de structuration de régions telles que les centromères et télomères. Elle peut également avoir un impact direct sur l'expression des gènes à proximité, en les réprimant plus ou moins en fonction de la propagation de l'hétérochromatine.
- L'hétérochromatine facultative quant à elle n'est pas répétée, et peut structurer des régions d'ADN différentes au sein de cellules d'une même espèce. Bien qu'elle partage la même structure compactée que l'hétérochromatine constitutive, elle peut se décondenser sous certaines conditions et permettre ainsi la transcription de gènes normalement réprimés. La formation de l'hétérochromatine est un moyen de réprimer spécifiquement certaines régions du génome, et participe ainsi à l'élaboration d'un programme génétique propre à chaque cellule. Cette organisation de la chromatine est donc extrêmement régulée et fait appel à de nombreux acteurs épigénétiques. Le meilleur exemple est le chromosome X inactif chez la femelle des mammifères, on peut l'observer sous forme d'un point dense dans le noyau interphasique appelé : corps de Barr (**figure 12**).

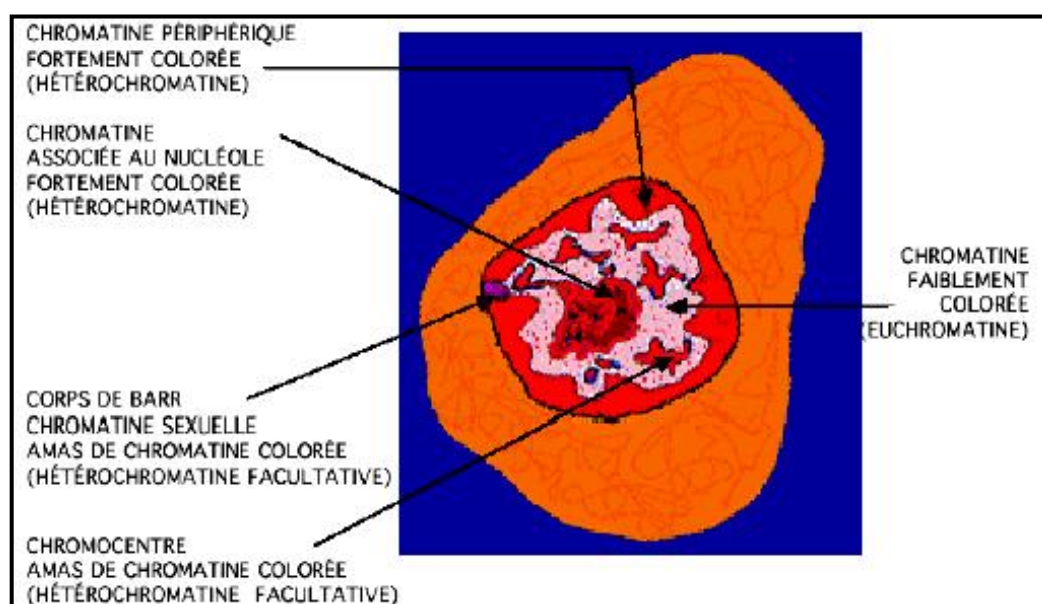


Figure 12 : Représentation schématique du noyau coloré au Feulgen (Baxter, 2002).

Chapitre III

Modifications post-traductionnelles des histones

1- Régulation de l'état de condensation de la chromatine

L'état de condensation de la chromatine n'est pas figé au cours de la vie d'une cellule. Il existe deux types de chromatine qui se distinguent sur la base de critères structuraux et fonctionnels : l'euchromatine et l'hétérochromatine. L'euchromatine correspond à la forme relâchée, permissive à la transcription, alors que l'hétérochromatine correspond à une forme condensée, répressive à la transcription et majoritairement présente dans des régions non-codantes du génome. L'état de condensation de la chromatine évolue de façon dynamique par l'intermédiaire de modifications tant de l'ADN que de ses protéines de liaison, notamment les histones. C'est en jouant sur ces modifications de la chromatine que la variabilité cellulaire et tissulaire de l'expression des gènes est assurée. Les histones forment le corps du nucléosome et donc de la chromatine. Les changements dans la structure ou dans la composition du corps du nucléosome se traduisent par un changement de la conformation de la chromatine. Les modifications post-traductionnelles des histones jouent sur leur structure et donc sur celle des nucléosomes et par conséquent sur l'accessibilité de l'ADN au sein de la chromatine.

L'enroulement de l'ADN autour des histones interfère avec ces interactions. Le masquage d'un promoteur d'un gène par un nucléosome est réfractaire à la transcription. L'ADN étant globalement chargé négativement (polyanion), il y a donc une interaction forte entre cette molécule et les histones, faisant que ces protéines peuvent ainsi être considérées comme des répresseurs généraux de la transcription. De ce fait, pour qu'un gène puisse être exprimé, il faut qu'il y ait un remodelage de la chromatine au niveau de sa séquence. Il y a nécessité d'un signal, d'un code, afin d'indiquer à la cellule de décompacter la fibre de chromatine à un locus donné. Ce code se trouve dans les protéines histones et plus particulièrement par des modifications covalentes mais réversibles portées par ces protéines.

2- Impact des modifications post-traductionnelles des histones

Une modification post-traductionnelle est la modification covalente d'une protéine introduite par l'intermédiaire d'une enzyme et qui intervient après l'étape de la traduction. Chez les eucaryotes, il existe une grande diversité chimique de modifications post-traductionnelles qui peuvent influencer la localisation, la demi-vie ou encore l'activité de la protéine cible. Dans le cas des histones, toutes les modifications sont réversibles : l'état de modification d'un résidu d'histone est contrôlé par un équilibre entre les enzymes qui catalysent l'ajout et celles qui catalysent le retrait d'une modification donnée.

Les histones sont sujettes à de nombreuses modifications post-traductionnelles, incluant, chez l'Homme, neuf types décrits dans la littérature et de nombreux sites de modification ont été identifiés pour chacune d'entre elles. Toutes les classes d'histones sont concernées par ces modifications. Cependant elles concernent préférentiellement H3 et H4, à un moindre degré H2A et H2B, très rarement l'histone H1. Ces modifications se produisent essentiellement sur le domaine NH₂-terminal, rarement sur le domaine COOH, exceptionnellement sur le domaine central. Les cibles de ces modifications sont les résidus lysine, arginine, sérine, thréonine ou tyrosine (ordre décroissant de probabilité d'occurrence). Classiquement, ces modifications sont réparties en deux groupes selon leur taille :

- **le premier groupe inclut les modifications chimiques de petite taille :** à savoir l'acétylation, la méthylation (mono-, di- et tri-méthylation), la phosphorylation et la citrullination.
- **le deuxième groupe inclut les modifications de taille plus importante :** telles que l'ubiquitinylation, la sumoylation, la ribosylation ou la biotinylation, mais également la crotonylation.

Les conséquences fonctionnelles de ces modifications dépendent du type de groupement chimique apposé (méthyle, acétyle, phosphate, ubiquitine, glycosyle, sumo, etc.), ainsi que du type de résidu ciblé (lysine, arginine, sérine, etc.) et de sa position le long de l'histone. La diversité des groupements et des résidus qui peuvent être ciblés donne lieu à une combinatoire potentiellement très riche. Ces modifications influent, d'une part, sur l'interaction entre l'ADN et les histones et par conséquent sur la compaction de la fibre de chromatine ; et d'autre part, sur l'interaction de l'ADN avec les autres protéines associées à la chromatine. Ainsi, elles agissent sur les fonctions liées au génome, telles que la transcription ou la réparation. Chacune de ces modifications a un rôle biologique précis.

Les modifications post-traductionnelles des histones sont souvent classées en deux catégories selon qu'elles sont corrélées avec une activation de la transcription ou une répression :

- **les modifications associées à une transcription active :** telles que l'acétylation des histones H3 et H4, ou la méthylation de H3K4, H3K36 et H3K79 sont considérées comme des marques de l'euchromatine.
- **les modifications fréquemment observées sur une chromatine compactée et transcriptionnellement réprimée :** comme la tri-méthylation de H3K9, H3K27 et H4K20, sont référencées comme des modifications de l'hétérochromatine.

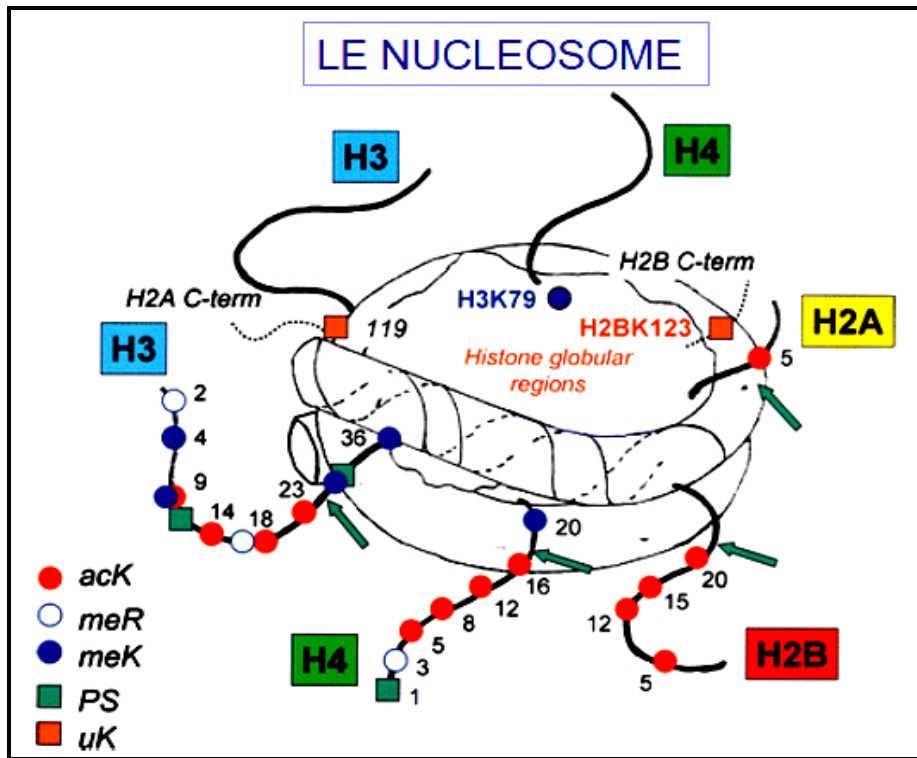


Figure 13 : Aperçu des principales modifications post-traductionnelles des histones (Graff et Mansuy, 2008).

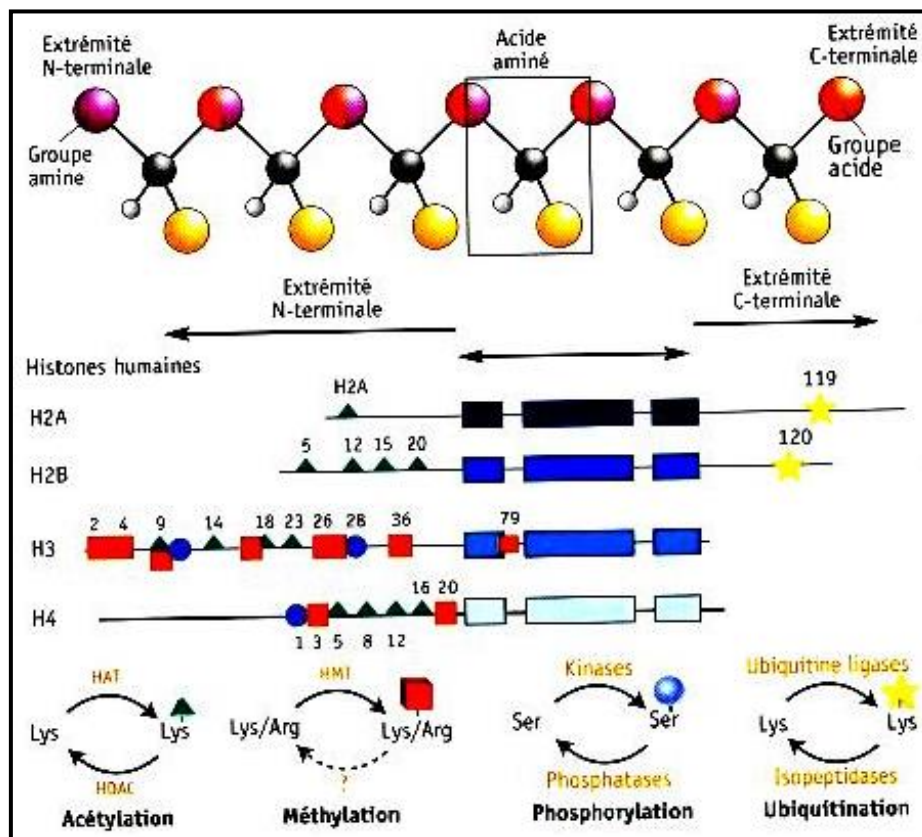


Figure 14 : Mécanismes des principales modifications post-traductionnelles des histones. (Graff et Mansuy, 2008).

Les fonctions des modifications post-traductionnelles des histones peuvent également être subdivisées en deux catégories en fonction de leurs répartitions :

- **d'une part, les fonctions que l'on peut qualifier de globales** pour la structuration des grands domaines chromatinien : euchromatine et hétérochromatine.
- **d'autre part, les fonctions plus localisées** comme la régulation spécifique d'un gène.

La combinaison entre les variants d'histones incorporés dans la chromatine et les modifications post-traductionnelles qu'ils portent constituent une information combinatoire complexe résumée sous le terme de « code histone » déchiffrée *in vivo* par des mécanismes moléculaires qui permettent d'associer à chaque combinaison un état fonctionnel de la chromatine. Une fois établi, ce code marquerait de façon épigénétique un locus soit d'une manière transitoire ponctuelle, soit de façon stable sur de grandes régions et serait finalement traduit en un état chromatinien particulier : **actif, permissif, restrictif ou inactif**.

3- Principales modifications post-traductionnelles des histones

3-1- Acétylation

L'acétylation des histones est le résultat d'une réaction enzymatique réversible qui correspond au transfert d'un groupement acétyle $-COCH_3$ sur la chaîne latérale d'un résidu lysine, exclusivement. Toutes les histones sont susceptibles d'être acétylées sur des lysines spécifiques dans leur domaine amino-terminal. Cette réaction de transfert se fait à partir d'une molécule d'acétyl-coenzyme A et est catalysée par des enzymes appartenant à la famille des Histone-Acétyl-Transférases (HAT). Ces enzymes se trouvent pour la plupart au sein de complexes multiprotéiques qui peuvent assurer d'autres fonctions en parallèle de l'acétylation (co-activateur transcriptionnel). Il existe quatre familles de HAT : la superfamille des GNAT (*GCN5-N-Acetyltransferaserelated*), la superfamille CBP/p300, la superfamille MYST et la superfamille SRC (*Steroid Receptor Coactivator*). La réaction inverse est catalysée par des enzymes appartenant à la famille des Histone-DésAcétylases (HDAC) dont il existe six classes majeures. Il existe donc une balance entre l'activité des HAT et celle des HDAC dont résultera le niveau global d'acétylation des histones. L'acétylation est considérée comme une marque d'activation de la transcription puisque l'addition d'un groupement acétyle neutralise la charge positive de la chaîne latérale des lysines et diminue ainsi les interactions électrostatiques entre les histones et l'ADN chargé négativement. Lors d'une acétylation, la chromatine est donc décondensée et l'ADN est davantage accessible à la machinerie transcriptionnelle.

Certaines protéines de remodelage de la chromatine peuvent également se servir des résidus acétylés comme points d'ancrage. À l'échelle cellulaire, son niveau délimite topographiquement l'euchromatine et l'hétérochromatine au moyen d'un gradient d'acétylation/ désacétylation. Elle joue aussi un rôle dans la progression du cycle cellulaire, la réparation de l'ADN, l'apoptose et peut être sujet à un dysfonctionnement en oncogène.

3-2- Méthylation

La méthylation des histones se fait par transfert d'un groupement méthyle -CH₃ à partir d'une molécule de S-Adénosine-Méthionine (SAM) vers un résidu lysine ou arginine. Cette réaction est catalysée par des enzymes de la famille des Histone Méthyl-Transférases (HMT) dont certaines sont spécifiques des résidus K, les HKMT, et les autres des résidus R, les HRMT. Elles peuvent ajouter entre 1 et 3 groupements méthyle sur le même résidu, aboutissant donc à une mono- ou une di-méthylation symétrique ou asymétrique des arginines et à une mono-, une di- ou une tri-méthylation des lysines. Cependant, seules les réactions de mono- et di-méthylation des lysines sont réversibles, la déméthylation étant catalysée par des enzymes de la famille des Histone-Dé-Méthylases (HDM) spécifiques des lysines. Contrairement à l'acétylation, les méthylations n'affectent pas la charge globale du résidu modifié mais peuvent par contre ajouter un encombrement stérique et une hydrophobicité réduisant la stabilité du nucléosome. Les résidus méthylés peuvent également servir de point d'ancrage à certaines protéines impliquées dans le remodelage de la chromatine.

D'un point de vue biologique, le rôle de la méthylation est beaucoup plus complexe que celui de l'acétylation. Tout d'abord, la méthylation d'une même protéine peut avoir deux effets opposés selon le résidu concerné. Par exemple, la mono-méthylation de la lysine 4 de l'histone H3 (H3K4me) entraînera une activation de la transcription des gènes tandis que la mono-méthylation de la lysine 9 (H3K9me) entraînera une répression de la transcription. Ensuite, la méthylation potentielle d'un même résidu à trois degrés différents ajoute un niveau de complexité supplémentaire puisque l'impact sur la chromatine ne sera pas le même.

3-3- Phosphorylation

Considérant l'ensemble des protéines d'un organisme, la phosphorylation est la plus répandue des modifications post-traductionnelles. Elle correspond au transfert d'un groupement phosphate -PO₄ à partir d'une molécule d'ATP vers un résidu sérine, thréonine ou tyrosine. La réaction de phosphorylation est catalysée par des enzymes de la famille des sérine/thréonine kinases qui ne sont pas spécifiques des histones.

L'ajout du groupement phosphate qui possède une charge négative intrinsèque confèrera une charge négative aux résidus cibles, diminuant l'affinité des histones pour l'ADN et induisant ainsi une décondensation localisée de la chromatine. Comme c'est le cas avec les autres protéines, la phosphorylation des histones est une réaction réversible. Cette réaction inverse de déphosphorylation est assurée par des phosphatases non spécifiques des histones. La plupart des phosphorylations sont dépendantes du cycle cellulaire et servent à guider sa progression à travers les différentes phases qui le composent. Un même résidu modifié peut ainsi être impliqué dans différents phénomènes cellulaires en fonction du temps. Les phosphorylations sont également très impliquées dans le « dialogue » entre les différentes modifications post-traductionnelles, notamment au niveau de l'histone H3. La phosphorylation de la sérine 139 du variant H2A.X est quant à elle classiquement associée aux cassures double-brin de l'ADN et est indispensable au recrutement de la machinerie de réparation.

3-4- Autres modifications post-traductionnelles

Les autres modifications post-traductionnelles que peuvent subir les histones sont beaucoup moins documentées dans la littérature, mais font l'objet de plus en plus d'études. Il s'agit de l'ubiquitinylation, de la sumoylation, de la ribosylation, de la biotinylation, de la crotonylation, de la citrullination ou encore de la succinylation, de la malonylation et de la carbonylation. Bien que beaucoup moins rencontrées que les modifications précédentes, elles n'en jouent pas moins un rôle biologique significatif. Hormis la citrullination, toutes ces modifications surviennent sur des résidus lysine. L'ubiquitinylation, la biotinylation, la ribosylation et la crotonylation sont réversibles tandis que la sumoylation ne l'est pas.

- **La citrullination** : est un cas particulier puisqu'il s'agit en réalité d'une réaction de déimination de l'arginine catalysée par des déiminases PAD (Peptidyl-Arginine Deiminase). Cette réaction de citrullination empêche la méthylation des résidus arginines concernés et peut donc affecter la transcription des gènes.
- **L'ubiquitinylation et la sumoylation** : correspondent respectivement au transfert d'une unité ubiquitine (Ub) de 8,5 kDa et d'une unité SUMO (Small Ubiquitin-like Modifier) de 12 kDa. Ces deux entités protéiques possèdent une homologie de séquence d'environ 20% et présentent une structure tridimensionnelle très similaire. L'ubiquitinylation est catalysée par des enzymes de la famille des lysine-ubiquitinases et la sumoylation par des enzymes de la famille des SAE (*SUMO-activating enzymes*).

À l'image de la méthylation, ces deux modifications peuvent aléatoirement être des marques d'activation ou de répression de la transcription. Leur effet activateur de la transcription est dû à l'encombrement stérique que ces deux gros polypeptides génèrent ce qui stabilise la chromatine dans une conformation ouverte, tandis que leur effet répresseur s'explique en partie par le masquage de certains résidus cibles d'autres modifications post-traductionnelles qui peuvent être activatrices de la transcription. Cependant, la sumoylation semble être davantage associée à une répression de la transcription en favorisant le recrutement d'histone-désacétylases. L'ubiquitinylation est pour sa part considérée comme une marque répressive de la transcription lorsqu'elle affecte H2A et comme une marque alternativement activatrice ou répressive lorsqu'elle affecte H2B. Par ailleurs, les histones ne peuvent être que mono-ubiquitinylées et non poly-ubiquitinylées comme c'est le cas lors de l'adressage des protéines au protéasome.

- **La ribosylation** : est une modification post-traductionnelle assez courante qui consiste au transfert d'un groupement poly (ADP-ribose) sur un résidu lysine depuis une molécule de NAD⁺ par l'intermédiaire d'une enzyme de la famille des PARP (Poly (ADP-Ribose) Polymérase). Dans le cas des histones, la ribosylation affecte la charge portée par les lysines et entraîne une perturbation des interactions électrostatiques entre l'ADN et les histones, aboutissant à un relâchement de la chromatine. Elle joue également un rôle important dans les mécanismes de réparation des cassures de l'ADN.
- **La biotinylation** : est aussi une modification post-traductionnelle qui affecte exclusivement les résidus lysines. Elle correspond au transfert d'un groupement biotine d'environ 244 Da catalysé par des enzymes de la famille des biotinidases. Du point de vue biologique, la biotinylation des histones semble être associée à l'hétérochromatine et donc impliquée dans la répression de la transcription des gènes. Elle joue également un rôle dans la signalisation des dommages à l'ADN.
- **La crotonylation** : découverte en 2011, est la modification la plus récemment mise en évidence sur les résidus lysines des histones. Le groupement crotonyle est transféré depuis une molécule de crotonyl-CoA, mais les enzymes qui catalysent cette réaction restent inconnues à ce jour. De par sa structure proche d'une acétylation, la crotonylation aura la même répercussion sur la conformation de la chromatine. Elle est donc associée à une augmentation de l'activité transcriptionnelle des gènes (**figure 15**).

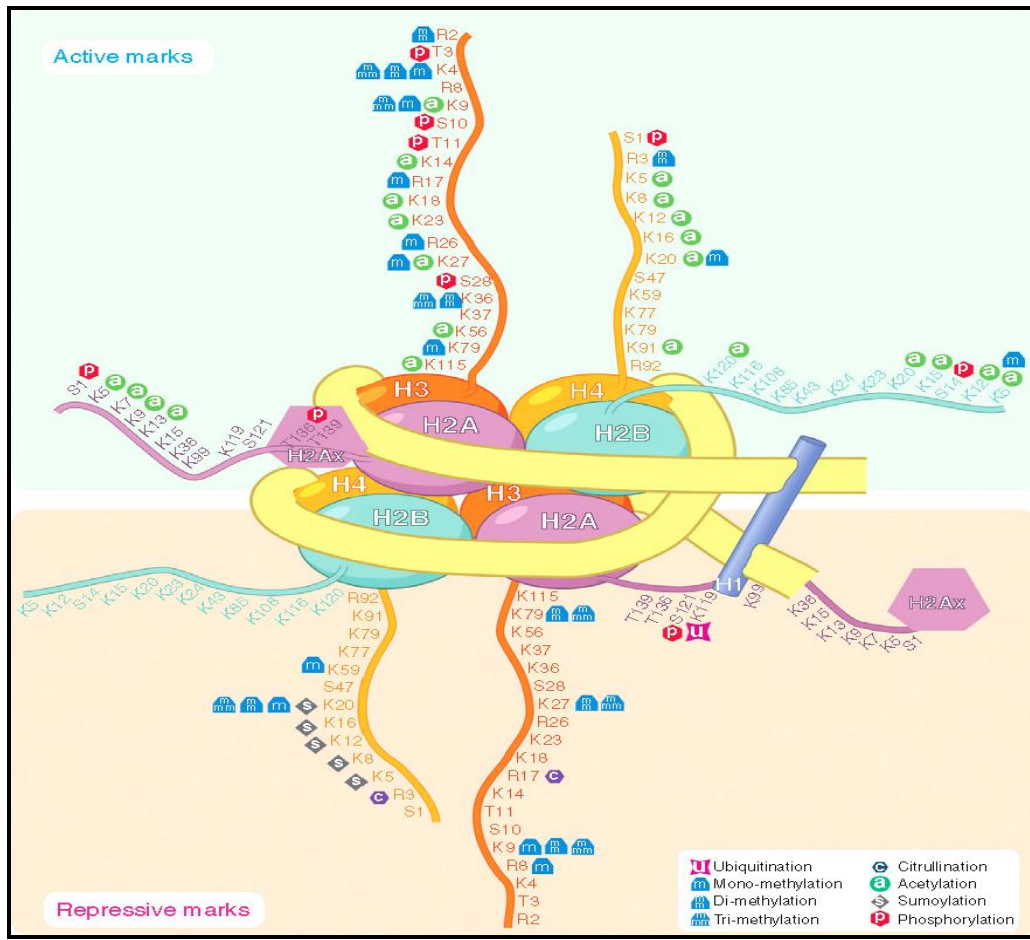


Figure 15 : Effet des principales modifications post-traductionnelles des histones (Young, 2014).

4- Code épigénétique des histones

D'un point de vue biologique, les modifications les plus étudiées sont l'acétylation, la méthylation, la phosphorylation et l'ubiquitinylation. Chacune de ces modifications peut être considérée individuellement ou collectivement. De manière générale, les modifications post-traductionnelles des histones auront un impact sur la conformation de la chromatine et sur sa perméabilité à la transcription ce qui permet de relier la présence d'une modification précise à l'activité de certains gènes. En parallèle, la modification d'un résidu peut influencer directement la survenue d'une modification d'un autre résidu voisin au sein de la même protéine. Certaines modifications seront donc mutuellement exclusives, tandis que d'autres seront à l'inverse positivement corrélées. Cet aspect combinatoire doit obligatoirement être pris en compte lorsque l'on s'intéresse à l'impact de ces modifications sur la transcription des gènes. En effet, elles agissent de concert et non de manière isolée, et la présence simultanée de plusieurs modifications différentes aura un effet sur la transcription qu'il sera difficile de prédire (**figure 16**)

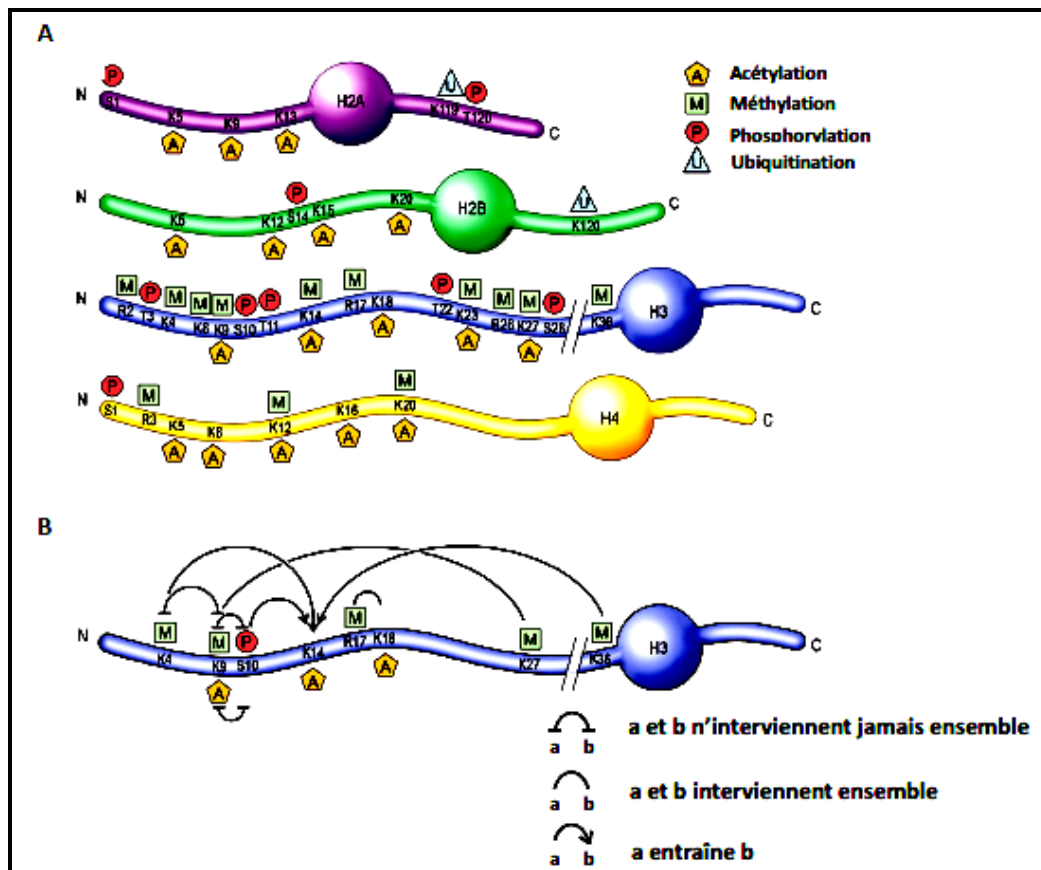


Figure 16 : Interdépendance des modifications post-traductionnelles des histones (Bannister et Kouzarides, 2011).

- (A) des modifications intervenant sur chaque résidu des histones composant le nucléosome.
- (B) interdépendance des modifications sur la terminaison N-terminale de l’histone H3.

In fine, la diversité des formes d’histones, leur richesse en sites potentiellement modifiables ainsi que la diversité chimique des modifications post-traductionnelles décrites laisse entrevoir le nombre de combinaisons possibles (plus d’un million) et donc la complexité de l’information portée par le code histone.

Par exemple, la méthylation d’une arginine peut être simple ou double de manière symétrique ou asymétrique. Chaque lysine peut subir différentes modifications, et la lysine 12 de l’histone H4 peut notamment être détectée sous six états différents : non modifiée, acétylée, mono- di ou tri- méthylée, ou biotinylée. Le nombre de combinaisons possibles au sein d’un seul nucléosome est évalué à plus d’un milliard de milliards de combinaisons. Cependant, toutes les modifications ne se retrouvent pas en même temps sur une histone et au sein d’un nucléosome. La spécificité d’une marque, et son environnement chromatinien avec d’autres, diffère en fonction de la région génomique ciblée et des conditions dans laquelle se trouve la cellule.

Chapitre IV

Méthylation de l'ADN

La méthylation de l'ADN est une modification épigénétique, qui consiste en l'ajout d'un groupement méthyl (-CH₃) à la place d'un atome d'hydrogène. Bien que les quatre bases azotées de l'ADN puissent être méthylées, l'adénine est la plus fréquemment modifiée chez les procaryotes et la cytosine chez les eucaryotes. Chez les procaryotes, la méthylation est portée préférentiellement au niveau du sixième azote de l'adénine, alors que chez les eucaryotes cette modification covalente est ajoutée au niveau du cinquième carbone des cytosines. La méthylation des cytosines est le premier mécanisme épigénétique connu permettant à une cellule de rendre certains gènes silencieux (inactif). C'est un phénomène réversible qui est impliqué dans de nombreux processus cellulaires clefs, et ce dès les phases précoces du développement embryonnaire. C'est un mécanisme fondamental pour le contrôle épigénétique de l'expression des gènes et du maintien de l'intégrité du génome. Par conséquent, l'évaluation du statut de la méthylation de l'ADN est importante dans toutes les situations faisant intervenir la variabilité de l'expression de gènes et la maintenance de l'intégrité du génome : la régulation de la croissance cellulaire, la différenciation tissulaire spécifique et la carcinogénèse.

La méthylation des résidus cytosines en position 5' du noyau pyrimidique est l'une des modifications post-répliquatives, covalentes les plus importantes de l'ADN des cellules eucaryotes. Par ailleurs, cette méthylation de l'ADN d'une cellule est reproduite après synthèse d'ADN et est transmise à ses cellules filles. La méthylation de l'ADN est donc préservée après la division cellulaire.

1- Découverte du processus de méthylation de l'ADN

En 1948, *Hotchkiss* a découvert la 5-méthylcytosine dans l'ADN d'un thymus de veau en utilisant la chromatographie sur papier, à une époque où la structure et la fonction de l'ADN comme matériel héréditaire n'avait pas encore été parfaitement comprises. D'ailleurs, à l'époque, on parlait d'elle comme étant « la cinquième base azotée ». Le déchiffrement du code génétique par *Nirenberg* et *Matthaei* a constitué le début de la biologie moléculaire moderne et le commencement de la recherche sur la méthylation de l'ADN. Après élucidation du code génétique, nous avons constaté que cette modification de la structure de ce nucléotide de l'ADN ne change pas la séquence : la 5-méthylcytosine est reconnue par l'ADN polymérase lors de la réplication et l'ARN polymérase lors de la transcription comme étant une cytosine normale.

La méthylation de l'ADN a un rôle différent chez les organismes procaryotes et eucaryotes. Chez les bactéries, la méthylation de l'ADN est un véritable mécanisme de défense contre les infections de phages, permettant la discrimination entre l'ADN endogène et l'ADN de l'agent infectieux, grâce à un système de méthylases et d'enzymes de restriction méthyle-sensitives capable de dégrader tout ADN étranger non correctement méthylé. Chez les procaryotes, la méthylation de l'ADN est également utilisée comme un moyen de contrôler la fidélité de la réplication. En effet, au cours de la réplication de l'ADN, le brin néo-synthétisé n'est pas immédiatement méthylé et permet ainsi de discriminer le brin parental afin qu'il serve de référence à de possibles corrections d'erreurs de réplication par le système de réparation des mésappariements.

Vers l'année 1975, pour la première fois, *Roy* et *Weissbach* purifient et caractérisent une méthylase de l'ADN à partir de noyaux cellulaires humains. Au début des années 1980, l'observation que le gène de hamster APRT (Adénine Phospho-Ribosyl-Transférase) méthylé *in vitro* et transfecté dans des cellules de souris, resta transcriptionnellement inactif dans les cellules, suggère pour la première fois, un rôle de la méthylation de l'ADN dans la répression génique.

La méthylation de l'ADN fut longtemps considérée comme un épiphénomène biologique, sans grand intérêt, mais depuis la caractérisation des enzymes responsables de la méthylation de l'ADN, d'énormes progrès ont été réalisés et ont permis de démontrer, au cours des dernières années, le rôle essentiel joué par la méthylation de l'ADN dans divers processus biologiques. D'autre part, il semble de plus en plus clair qu'une méthylation aberrante tient un rôle primordial dans diverses pathologies (particulièrement dans la cancérogenèse).

2- Caractéristiques du processus de méthylation de l'ADN

La méthylation de l'ADN est une modification épigénétique réversible conservée au cours de l'évolution, mais n'est toutefois pas universelle. On la retrouve chez les bactéries, les champignons, les plantes et les mammifères. Malgré le caractère essentiel de la méthylation de l'ADN chez les organismes qui ont développé cette marque épigénétique, celle-ci n'est pas ou peu trouvée chez certains organismes comme la levure *Saccharomyces cerevisiae*, la drosophile *Drosophila melanogaster*, ainsi que le nématode *Caenorhabditis elegans*. Ces organismes subissent pourtant des régulations épigénétiques. Au sein des génomes animaux, le taux de méthylation varie de façon substantielle ; de l'indétectable chez *C. elegans* à une méthylation pratiquant globale de l'ADN chez les vertébrés.

Dans les cellules humaines 2 à 5 % des résidus cytosine de l'ADN génomique sont méthylés sous forme de 5-méthyl cytosine. Chez l'Homme, et les autres mammifères supérieurs, la méthylation est essentiellement imposée sur des résidus cytosines qui précèdent une guanosine dans la séquence de l'ADN, appelés encore les di-nucléotides CpG (cette désignation est pour faire la distinction avec les paires de bases CG situées sur les brins complémentaires). Ces sites méthylables suivent une distribution non uniforme. Dans une cellule somatique humaine, environ 70 à 80 % des résidus 5-méthyl cytosines se trouvent au niveau de séquences CpG ; lorsque ces séquences CpG sont fréquemment répétées dans le génome, les régions correspondantes sont appelées îlots CpG. Les îlots CpG sont fréquemment retrouvés dans le promoteur et le premier exon de gènes. Les îlots CpG des gènes de maintenance (gènes de ménage ou « housekeeping ») exprimés dans toutes les cellules ne sont pas méthylés, alors que les îlots CpG des nombreux gènes tissus-spécifiques sont méthylés à l'exception des tissus dans lesquels ils sont exprimés.

La fréquence générale des di-nucléotides CpG dans le génome est considérablement inférieure à sa fréquence théorique. La méthylation de l'ADN a probablement conduit au cours de l'évolution à l'extinction de beaucoup des di-nucléotides CpG du génome. Cette élimination est reliée à la tendance des cytosines méthylées à perdre spontanément un groupement amine et ainsi à se transformer en thymidine (T). Si cette mutation n'est pas réparée, le changement cytosine-thymidine persiste. Ce changement de base constitue également le plus commun des polymorphismes génétiques dans la population humaine. Certes la méthylation de l'ADN est réversible, mais peu est connu à l'heure actuelle sur la déméthylation de l'ADN, qui conduit probablement à un état déréprimé et activable de l'ADN. Les mécanismes par lesquels la déméthylation a lieu, ne sont pas encore compris et aucune protéine impliquée dans un tel processus n'a pu être mise en évidence. Reste donc ouverte la question de savoir si la déméthylation de l'ADN résulte d'une perte de la méthylation au cours des réplifications cellulaires ou d'une substitution de la base méthylée ou, le plus probablement, d'une déméthylation active par une de(s) enzyme(s) encore non identifiée(s) (**figure 17**).

Lorsqu'on se réfère à l'établissement d'un nouveau profil de méthylation dans le génome, on parle d'une méthylation *de novo* de l'ADN. Ensuite ces profils de méthylation sont conservés tout au long des cycles cellulaires et le plus souvent tout au long de la vie de l'organisme par une méthylation de maintien de l'ADN, sauf si des facteurs extérieurs et/ou génétiques perturbent les profils de méthylation, tel que par exemple pendant la cancérogenèse de la cellule.

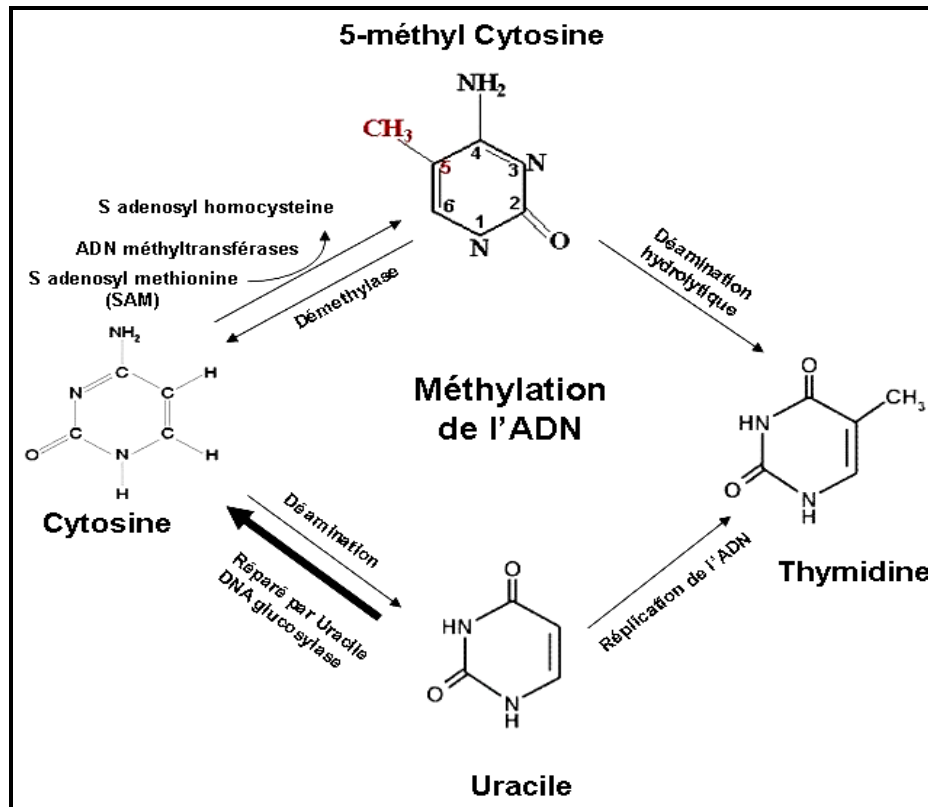


Figure 17 : Réactions de méthylation/déméthylation de la cytosine (Song et He, 2011).

3- Rôles biologiques de la méthylation de l'ADN

La méthylation de l'ADN a un rôle différent chez les organismes procaryotes et eucaryotes. Chez les bactéries, la méthylation de l'ADN est un véritable mécanisme de défense face à l'introduction d'un ADN étranger et un moyen de contrôler la fidélité de la réplication. Chez les organismes eucaryotes, la méthylation joue essentiellement un rôle régulateur de l'expression des gènes. Chez les mammifères, la méthylation de l'ADN est impliquée dans de multiples fonctions biologiques

3-1- Développement embryonnaire

Au cours du développement précoce, des changements du profil de méthylation s'opèrent de façon orchestrée et précise. Un aspect important du rôle de la régulation est le processus de déméthylation lors de l'embryogenèse. Il a été montré que durant le développement initial de l'embryon, avant implantation, immédiatement après la fécondation, une réduction importante de la méthylation globale était observée : une vague de déméthylation globale qui efface des chromosomes maternels et paternels du zygote pratiquement toute trace de méthylation.

Il n'y a que le profil de méthylation des gènes à empreinte qui sera conservée. Cet état était ensuite suivi par une vague de méthylation *de novo* concernant la plupart des résidus CpG, mais laissant les îlots CpG non méthylés au moment de l'implantation. Après l'implantation, la plupart de l'ADN génomique était méthyle. Les profils de méthylation de l'ADN varient selon les tissus et évoluent lors du développement. La régulation épigénétique opère très activement dans le fonctionnement cellulaire normal chez les mammifères, à partir de la conception jusqu'au vieillissement et la mort. La reprogrammation de la méthylation de l'ADN du zygote totipotent à lieu dans les cellules germinales et pendant la pré-implantation. Le génome des spermatozoïdes et des ovules sont hautement méthylés comparé à celui des cellules somatiques. Cependant, quelques heures après la fécondation, une déméthylation rapide du génome paternel a lieu par des mécanismes dynamiques, mais pourtant indéfinis jusqu'à aujourd'hui, en plus de l'acquisition de modifications des histones. Le génome maternel suit un procédé plus lent et plus passif. La fonction biologique exacte de cette reprogrammation dynamique de l'ADN essentielle au développement harmonieux, mais elle reste encore inconnue jusqu'à maintenant.

3-2- Différenciation cellulaire

Selon la définition, la différenciation est une restriction du potentiel cellulaire menant à la spécialisation de fonction. Lors de la différenciation, les cellules évoluent du stade de totipotence en passant par des étapes pluripotence de pour aboutir à un état bien différencié. Au cours de cette progression, elles acquièrent des propriétés spécifiques reflétant un profil d'expression précis. Ceci nécessite des modifications de la répartition de la méthylation de l'ADN qui par la suite seront transmises aux cellules filles, garantissant la pérennité de l'état différencié. Pendant la différenciation cellulaire, quelques gènes dans certains tissus subissent une déméthylation permettant à la cellule d'exprimer ces gènes et d'acquérir ainsi des fonctions spécialisées qui deviennent ensuite transmissibles aux cellules filles. En effet, ce qu'on appelle gènes de ménage ou « housekeeping gene » sont déméthylés dans tous les tissus alors que les gènes tissu-spécifiques sont déméthylés dans les tissus où ils sont exprimés, mais bien méthylés ailleurs.

3-3- Empreinte génétique parentale

La méthylation de l'ADN est également impliquée dans les phénomènes dits d'empreinte parentale. C'est un processus par lequel l'expression de certains gènes est restreinte à un allèle, paternel ou maternel.

L'allèle non exprimé est rendu et maintenu silencieux par la méthylation de sa séquence. Elle est responsable d'une régulation différentielle de l'expression des deux allèles d'un petit nombre de gènes regroupés sur le génome. Ainsi, certaines zones sont dites soumises à empreinte maternelle, car seuls certains allèles maternels sont exprimés, et les allèles paternels correspondants sont réprimés. De la même manière, on parle d'empreinte paternelle lorsque les allèles d'origine paternelle sont exprimés alors que les allèles maternels sont réprimés. L'étude détaillée de ces régions fait ressortir que l'état de méthylation d'une région régulatrice, appelée ICR (Imprinting Control Region), est responsable de cette régulation. La méthylation de l'ICR d'un allèle inhiberait sa transcription. De plus, le niveau de méthylation de chaque ICR est corrélé aux modifications post-traductionnelles des histones associées.

3-4- Inactivation du chromosome X

Chez les mammifères, l'égalité du dosage des gènes liés à X est réalisée grâce à l'inactivation du chromosome X chez les femelles. L'inactivation du chromosome X s'effectue très tôt au cours de l'embryogenèse et est maintenue de manière stable dans les cellules somatiques au cours de la vie de l'individu. Le chromosome inactif est visible en microscopie. Il se présente sous forme de chromatine très condensée à la périphérie du noyau et est nommé Corps de Barr. L'initiation de l'inactivation du chromosome X est contrôlée par la production d'un ARN non codant *Xist* qui recouvre le chromosome en *cis* et déclenche la répression. Cette accumulation engendre une cascade complexe de modifications chromatiniennes comme la désacétylation et la méthylation des histones et l'incorporation de variants d'histones. Le maintien de l'état inactif est dépendant de la méthylation de l'ADN au contraire de l'initiation. La méthylation de l'ADN semble donc surtout jouer un rôle dans la maintenance et l'irréversibilité de l'inactivation du chromosome X. L'inactivation du deuxième chromosome X chez les mammifères femelles est établie pendant l'embryogenèse précoce et permet de rendre transcriptionnellement silencieux de façon permanente un des deux chromosomes X. L'euchromatine est alors transformée en hétérochromatine facultative.

3-5- Inhibition de séquences parasites

Une grande partie des CpG méthylés réside dans les séquences répétées dispersées du génome c'est à dire, les rétrovirus endogènes, les éléments LINES et SINES (Long/Short Interspersed Nuclear Element).

La méthylation de l'ADN permet de stabiliser le génome en masquant ou inhibant les recombinaisons entre les séquences répétées. La méthylation des promoteurs des rétrotransposons entraîne une répression de leur expression. La méthylation de l'ADN stabilise le génome en limitant l'activation des transposons. La majeure partie de l'ADN méthylé eucaryote est formée de séquences répétées. Les transposons constituent 45 % du génome humain. Leur méthylation inactive l'expression des gènes nécessaires à leur propagation, limitant ainsi les risques de mutagenèse par insertion dans une zone codante.

3-6- Vieillesse cellulaire

Le rôle de la méthylation de l'ADN semble aussi important lors du vieillissement cellulaire. Le niveau génomique total de méthylation décroît progressivement avec l'âge chez la plupart des vertébrés. L'hypo-méthylation progressive au fur et à mesure de l'âge a été prouvée dans de nombreuses espèces et dans des tissus aussi variables que le cerveau, le foie, la muqueuse de l'intestin grêle, le cœur, la rate et les lymphocytes T. Un groupe de gènes semblent changer leur statut de méthylation avec le vieillissement et de façon spécifique à certains tissus. D'une manière plus générale, il est considéré que certains loci dans des tissus spécifiques subissent une hypo-méthylation avec l'âge alors que d'autres subissent une hyper-méthylation. Ceci pourrait correspondre à des modifications de fonction cellulaire et de l'expression de gènes associées au vieillissement. Les mécanismes contribuant aux modifications cellulaires dépendant de l'âge ne sont cependant pas encore clairement établis.

4- Acteurs moléculaires de la méthylation de l'ADN

La caractérisation des enzymes responsables de la méthylation de l'ADN ont permis de commencer à comprendre comment la méthylation de l'ADN fonctionne. Les méthyl-transférases de l'ADN utilisent le S-Adénosyl-Méthionine (SAM) comme substrat donneur de groupement méthyle. Chez les mammifères, les Dnmts sont divisées en trois classes : Dnmt1, Dnmt2 et Dnmt3.

4-1- Dnmt1

Dnmt1, la première Dnmt caractérisée, est celle qui est la plus abondante dans la cellule. Dnmt1 méthyle préférentiellement l'ADN hémi-méthylé (par rapport à l'ADN non-méthylé) *in vitro* et ainsi on considère souvent Dnmt1 comme la méthyl-transférase de maintien qui permet de garder le même profil de méthylation au fil des divisions cellulaires.

Effectivement, lors de la réplication cellulaire, Dnmt1, qui est généralement répartie de façon diffuse dans le noyau, se lie à PCNA et se concentre dans les origines de réplication dès le début de la phase S. Comme la présence de Dnmt1 n'est pas restreinte aux cellules en prolifération, mais qu'elle est également localisée dans des tissus ayant un faible taux de prolifération (comme le cœur et le cerveau), il est fort probable que Dnmt1 puisse également intervenir dans la méthylation *de novo*.

L'inactivation du gène Dnmt1 chez la souris entraîne une létalité embryonnaire précoce associée à une hypo-méthylation génomique sévère et une dérégulation des gènes, soulignant son importance pendant le développement embryonnaire précoce. Dnmt1 interagit avec d'autres protéines intervenant dans le remodelage de la chromatine (HDAC et HKMT).

4-2- Dnmt2

La protéine Dnmt2 constitue la classe la moins bien connue des Dnmt. Lorsque Dnmt2 est invalidée chez la souris, les profils de méthylation de l'ADN ne semblent pas être modifiés et la cellule est toujours capable de méthyler son génome. Néanmoins, des études récentes ont pu mettre en évidence que Dnmt2 a effectivement une activité de méthyl-transférase de l'ADN qui semble être reliée à une certaine spécificité de séquence. Jusqu'à présent aucun rôle biologique n'a pu être attribué à Dnmt2, mais il semble que cette protéine ait une fonction plus spécifique que les autres Dnmt.

4-3- Dnmt3

En 1998, Okano caractérise une troisième famille de Dnmt capable de catalyser le transfert de résidus méthyles sur des substrats nucléotidiques hémi-méthylés ou non méthylés avec la même efficacité. Cette famille regroupe aujourd'hui trois membres : Dnmt3a, Dnmt3b et Dnmt3L. Cette classe d'enzymes est très conservée et est représentée aussi bien chez plusieurs espèces. On appelle Dnmt3a et Dnmt3b, les méthyl-transférases *de novo* car elles sont responsables de la vague de méthylation post-implantatoire chez l'embryon. Elles sont très exprimées lors de l'embryogenèse et la gamétogenèse. Au contraire, dans les tissus somatiques, elles présentent un taux plus bas.

Une inactivation de *Dnmt3a* ou *Dnmt3b* dans des cellules embryonnaires précoces n'engendre pas de phénotype spécifique. En revanche, quand les deux sont inactivés simultanément, il y a absence de la méthylation *de novo* et une perte progressive de la méthylation globale de l'ADN, ce qui permet de suggérer que Dnmt3a et Dnmt3b ont non seulement un rôle dans la méthylation *de novo* mais également dans le maintien des profils de méthylation.

À l'échelle de l'organisme, lorsque Dnmt3a ou Dnmt3b sont invalidés chez la souris, celles-ci ne sont pas viables (les souris Dnmt3a^{-/-} meurent 4 semaines après la naissance tandis que les souris Dnmt3b^{-/-} meurent *in utero* au jour embryonnaire 9,5 au moment où la gastrulation se met en place).

En ce qui concerne Dnmt3L, cette protéine est dépourvue d'une grande partie de son domaine catalytique et n'est ainsi plus capable de méthyler l'ADN. Dnmt3L, par contre, semble influencer de façon indirecte la méthylation de l'ADN et joue probablement un rôle dans l'empreinte parentale.

4-4- Protéines liant les cytosines méthylés (MBD)

La méthylation de l'ADN implique non seulement les enzymes qui catalysent l'ajout du groupement méthyl à l'ADN, les Dnmt classiques, mais aussi les protéines qui lient spécifiquement les di-nucléotides CpG méthylés, les MBD (Methyl CpG Binding Domain). Les protéines MBD sont regroupées au sein d'une famille par homologie de séquences. Elles ont la particularité d'être capables de se lier aux résidus méthyl-CpG (sauf MBD3) et entraînent une diminution de la transcription par recrutement d'autres machineries répressives. MeCP2 fut le premier membre identifié. Il présente un domaine MBD capable de lier un di-nucléotide CpG symétriquement méthylé. MeCP2 réprime la transcription via son domaine TRD (Transcriptional Repression Domain) par interaction avec la machinerie de la transcription. Par l'étude de MeCP2, le premier lien moléculaire entre la méthylation de l'ADN et la structure de la chromatine fut établi.

4-5- Déméthylase de l'ADN ?

La méthylation de l'ADN est un phénomène réversible, mais à ce jour, très peu d'éléments sont connus en ce qui concerne la déméthylation. Plusieurs hypothèses ont été avancées sur ce sujet : une déméthylation passive résultant de la réplication de l'ADN en absence de méthyl-transférases ou par la dissimulation des sites de méthylation par des facteurs de transcription ou encore par une diminution générale de l'activité méthyl-transférase. Mais la déméthylation ne semble pas nécessiter la division cellulaire. En 1999, *Ramchandani* a extrait une protéine à activité déméthylase. La même année, *Bhattacharya* a cloné une déméthylase homologue à MBD2. Une déméthylation active a été rapportée dans un seul cas en 2003, celui du gène de l'interleukine 2 (IL-2) durant l'activation des lymphocytes T. Ces résultats attribuant une activité de déméthylation de l'ADN à MBD2, sont encore très contestés dans le monde de l'épigénétique et nécessitent d'être confirmés.

5- Processus de verrouillage l'expression par méthylation de l'ADN

L'importance de la méthylation de l'ADN dans de nombreux processus biologiques est probablement à rapprocher de son rôle important dans la répression et le verrouillage de la transcription de certaines régions spécifiques du génome. En revanche, les mécanismes par lesquels cette méthylation peut inhiber l'expression génique sont peu compris. Trois principaux mécanismes ont été proposés :

Le premier mécanisme : constitue une interférence directe des CpG méthylés empêchant la liaison de facteurs de transcription spécifiques sur leurs sites actifs *cis* de reconnaissance des promoteurs respectifs. Il a été proposé que le rajout d'un groupement méthyl sur l'ADN puisse engendrer un encombrement stérique, ne permettant plus la fixation sur l'ADN de protéines essentielles à la transcription génique, comme les facteurs de transcription. Plusieurs facteurs de transcription reconnaissent des séquences contenant des résidus CpG. Il a été ainsi montré que la méthylation de l'ADN inhibait la liaison de chacun de ces facteurs de transcription sur le génome. Cependant, cette notion ne semble pas toujours être valable, car il a été montré que d'autres facteurs de transcription ne sont pas sensibles à la méthylation de leurs sites de liaison. Bien que ce modèle soit attractif, il semble qu'il ne soit pas toujours valable. Ce modèle explique en effet uniquement une minorité de situations où la méthylation induit une diminution stable de la transcription. Ce mécanisme est peu fréquent *in vivo*.

Un second mécanisme correspond à la liaison directe de répresseurs spécifiques transcriptionnels sur l'ADN méthylé. Ce mécanisme, beaucoup plus général que le premier, se base sur les protéines reconnaissant et liant spécifiquement les di-nucléotides CpG méthylés, les MBD 1 et 2 qui ont comme capacité de se lier aux résidus CpG méthylés dans n'importe quel contexte de séquence. La protéine MeCP-1 se lie sur de multiples sites CpG méthylés symétriquement. Il a été proposé que les gènes faiblement méthylés forment un complexe instable avec MeCP-1 qui préviendrait la transcription lorsque le promoteur est faible. Ce complexe pourrait être modifié par un promoteur fort, autorisant alors le gène méthylé à être transcrit. Il a aussi été montré que MeCP-2 était capable de diminuer la transcription d'ADN nu *in vitro* suggérant que la structure de la chromatine n'était pas toujours nécessaire pour son action répressive.

Le troisième mécanisme par lequel la méthylation pourrait diminuer la transcription est l'altération de la structure de la chromatine. Ce modèle alternatif met en jeu l'apparition de modifications de l'architecture du nucléosome comme conséquence de la méthylation.

Ce modèle mettrait en jeu les histones déacétylases associées à MeCP-2, mais aussi d'autres facteurs qui pour l'instant sont inconnus. Les histones déacétylases catalysent le retrait des groupements acétyles du noyau des histones, conduisant ainsi à transformer une structure de chromatine ouverte et compétente du point de vue transcriptionnel en une structure fermée ne pouvant pas être accessible à la transcription. Les liens entre les protéines méthyl-CpG-binding, les histones déacétylases et les modifications de structures de la chromatine sont donc des éléments importants expliquant comment la méthylation de l'ADN peut modifier la transcription. Par ailleurs, il a aussi été montré que les enzymes ADN méthyltransférases pouvaient aussi recruter les histones déacétylases sur les sites méthylés. Le rôle respectif des DNA méthyl-transférases et des protéines méthyl-CpG binding dans la diminution de la transcription reste à évaluer.

6- Implication de la méthylation de l'ADN en cancérogenèse

Le développement d'un cancer résulte de mutations génétiques, phénomène bien connu, mais aussi d'altérations épigénétiques qui désorganisent les principales voies de régulation cellulaire. On s'est rendu compte ces dernières années que la méthylation de l'ADN peut être une voie alternative aux mutations dans le développement de cancers. En effet, les cellules cancéreuses peuvent présenter une altération du profil de méthylation (méthylation aberrante). L'impact de la méthylation de l'ADN dans la cancérogenèse est encore mal connu. S'il semble avéré que les tissus cancéreux sont globalement hypo-méthylé, il n'en demeure pas moins que l'hyper-méthylation semble être un processus très répandu d'inactivation de gènes spécifiques au cours de l'oncogenèse. Au cours de la tumorigénèse, il existe deux aberrations de la méthylation de l'ADN : l'hypo-méthylation globale et l'hyper-méthylation ciblée. L'hyper-méthylation de gènes suppresseurs de tumeurs, l'hypo-méthylation du génome conduisant à une instabilité chromosomique ou à l'expression d'oncogènes. Divers gènes suppresseurs de tumeurs sont des cibles de cette inactivation par méthylation dans les cancers humains.

Ces changements peuvent être exploités en pratique médicale pour le diagnostic (ou éventuellement le pronostic) et le traitement de ces pathologies. Une meilleure détection des tumeurs à un stade précoce est un objectif majeur pour améliorer le pronostic des patients atteints de cancer. En rétablissant l'expression de gènes critiques, inactivés du fait des signaux épigénétiques erronés, cette approche ouvre une voie prometteuse dans la thérapie anticancéreuse par la modulation du profil d'expression génique et donc du phénotype des cellules tumorales.

Chapitre V

Empreinte parentale

Tout organisme diploïde contient deux lots de chromosomes, l'un hérité du père et l'autre de la mère. Un gène donné est donc composé de deux allèles : l'allèle paternel et l'allèle maternel. En accord avec les règles classiques de l'hérédité définies par *Mendel*, le comportement des deux génomes est similaire, c'est-à-dire que pour un gène donné, quand le gène s'exprime, les deux allèles sont actifs, et quand il ne s'exprime pas, les deux allèles sont silencieux (inactifs). Jusqu'au début des années 1980, d'après le schéma de la génétique mendélienne, on admettait que les deux copies s'exprimant de manière équivalente. Cependant, plusieurs découvertes en biologie du développement dans les années 1980 ont montré que tel n'était pas le cas pour tous les gènes.

En effet, l'hypothèse d'une contribution non équivalente, chez les mammifères, des génomes parentaux à la formation d'un embryon viable a été avancée. Un exemple connu de cette asymétrie dans la fonction du génome paternel et maternel est le cas du mulet et du bardot, le premier étant issu du croisement d'une jument avec un âne et le second du croisement d'une ânesse avec un étalon. Bien que possédant tous deux une part de génome issu de l'âne, et une autre issue du cheval, leur phénotype est visiblement différent selon la provenance maternelle ou paternelle de ces parts (**figure 18**).

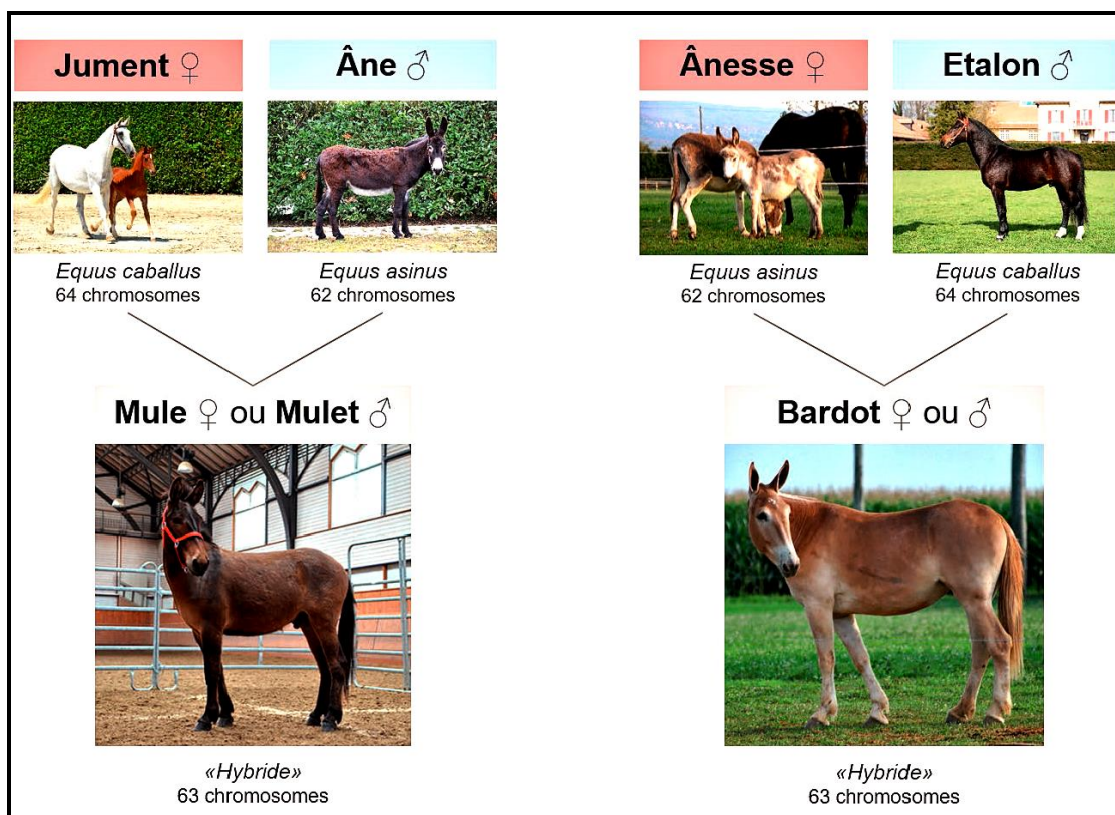


Figure 18 : Asymétrie dans la fonction du génome paternel et maternel chez les équines (Guénet, 2009).

1- Définition de l’empreinte parentale

Chez les mammifères placentaires, aussi bien chez quelques marsupiaux et angiospermes (plantes à fleurs), certains gènes font l’objet d’une empreinte parentale qui se caractérise par une différence d’expression entre les deux allèles selon qu’ils proviennent du père ou de la mère. Cette asymétrie fonctionnelle, qui se traduit par l’expression d’un seul allèle sur les deux, n’est pas la conséquence d’une modification de la séquence d’ADN, mais résulte au contraire de modifications épigénétiques réversibles de la structure de la chromatine et du profil de méthylation de l’ADN. L’empreinte génomique parentale ne concerne qu’une faible proportion des gènes, mais elle a une grande importance pendant les étapes précoces du développement embryonnaire et son dérèglement peut conduire à des maladies graves. Sa finalité n’est pas totalement élucidée et fait encore l’objet de spéculations. Différentes théories ont été avancées pour expliquer l’avantage qui pourrait être conféré par l’expression mono-allélique de certains gènes, alors qu’à *priori* un tel mécanisme serait défavorable en cas de mutation.

2- Découverte de l’empreinte parentale

En 1984, *Davor Solter* a mis au point une technique de transfert nucléaire permettant d’échanger, dans des œufs fécondés de souris, les noyaux issus du spermatozoïde et ceux issus de l’ovocyte. Des gynogénètes ou embryon avec deux noyaux femelles et des androgénètes ou embryon deux noyaux mâles, ont ainsi été produits. Le résultat obtenu est que ni les androgénètes, ni les gynogénètes ne se développaient jusqu’à terme, et ce malgré leur composition diploïde.

Au stade le plus avancé de leur développement, les deux types de zygotes ont des phénotypes opposés : les gynogénètes présentaient un développement à peu près normal de l’embryon lui-même, tandis que les annexes embryonnaires (amnios, allantoïde, vésicule ombilicale et placenta) étaient anormales; la situation inverse a été observée pour les androgénètes avec un embryon anormal, mais des annexes embryonnaires normalement développées. Cette expérience permet d’évaluer le rôle respectif des deux génomes : les génomes mâles et femelles semblaient donc jouer un rôle symétrique et complémentaire. Ces études ont mené à la notion que le génome dérivé de la mère et celui dérivé du père n’étaient pas fonctionnellement équivalents et que les deux étaient nécessaires au développement. Le terme « empreinte génomique parentale » a été introduit la même année par *Surani* pour décrire ces différences fonctionnelles entre les deux haplo-génomes.

Ainsi, les génomes maternel et paternel, malgré leur constitution génique identique, ne fonctionnent pas de façon équivalente, car ils sont marqués par cette empreinte, apposée durant la gamétogenèse. Il en résulte une différence fonctionnelle des deux génomes parentaux, qui rend la présence de chacun d'eux indispensable au développement à terme de l'embryon (**figure 19**).

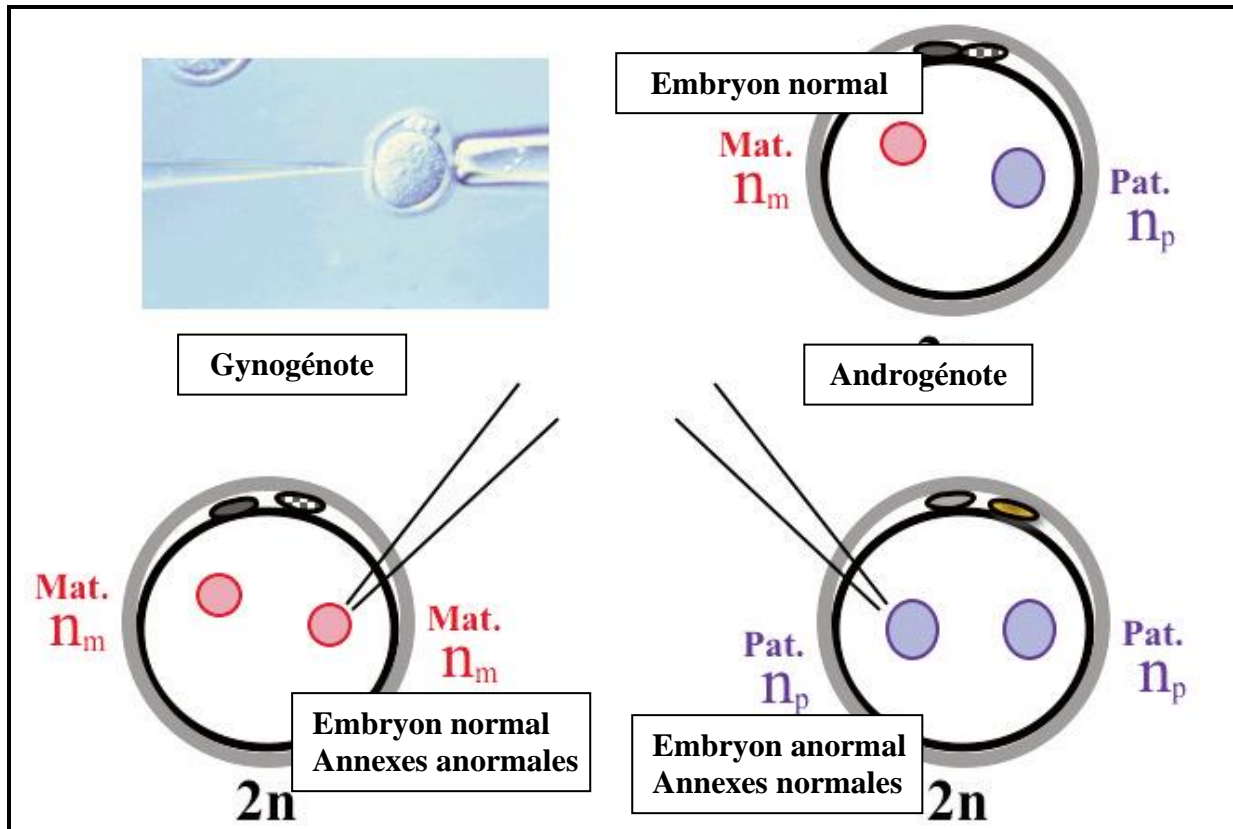


Figure 19 : Technique de transfert nucléaire dans des œufs fécondés de souris (**Reik, 2001**).

Ces observations ont amené un ensemble de questions, et, en premier lieu, à se demander comment fonctionne l'empreinte parentale.

Par la suite, des études génétiques furent réalisées par *Cattanach* et son équipe à partir de croisements de souris portant des remaniements chromosomiques: ces croisements permettaient d'obtenir des individus chez lesquels tout ou partie d'une paire de chromosomes provenait d'un seul des deux parents ; c'est ce qu'on appelle en cytogénétique l'Uni-Disomie Parentale (UDP) où les deux chromosomes sont issus du même parent : délétion du chromosome maternel et duplication du chromosome paternel ou inversement. Au cours des vingt dernières années, a été établie une carte complète des onze régions chromosomiques non équivalentes selon qu'elles sont héritées du père ou de la mère, et qui affectent la croissance, le comportement ou la viabilité des souris.

Ces résultats ont permis de préciser les hypothèses dérivées des premières expériences de transfert nucléaire : certains gènes, essentiels pour le développement, ne sont exprimés qu'à partir d'un seul chromosome, maternel ou paternel, et sont soumis à une empreinte parentale.

Ces expériences pionnières ont été suivies par l'identification des premiers gènes soumis à empreinte vers 1990. Les expériences d'inactivation de gènes, utilisant les cellules ES (cellules embryonnaires du stade de blastocystes) prenaient leur essor à ce moment-là. Un gène candidat a été ainsi invalidé, celui de l'*IGF2* (Insulin-like Growth Factor 2).

De façon surprenante, les croisements réciproques des souris dont ces gènes ont été invalidés ne donnent pas le même résultat : par exemple, si la mutation nulle du gène *IGF2* est transmise par la femelle, les descendants sont de taille normale ; en revanche, si la mutation nulle est transmise par le mâle, les souriceaux obtenus sont 40 % plus petits que la normale. L'explication vient du fait que le gène *IGF2* ne s'exprime qu'à partir de l'allèle d'origine paternelle, car il est soumis à une empreinte parentale. Chez les hétérozygotes portant la mutation nulle d'*IGF2* sur le chromosome d'origine paternelle, aucune expression d'*IGF2* à partir du chromosome d'origine maternelle ne compense la délétion, et ces souris hétérozygotes ont un phénotype identique à celui de souris homozygotes pour la mutation nulle.

3- Les gènes soumis à l'empreinte

Si la plupart des gènes des mammifères sont exprimés de façon bi-allélique, certaines régions chromosomiques portent des gènes dont l'expression est mono-allélique et dépend de l'origine parentale. Ces gènes marqués par une empreinte parentale jouent, pour la plupart, un rôle important dans le développement et la croissance fœtale et placentaire.

La plupart des gènes soumis à empreinte sont regroupés en grandes régions chromosomiques (clusters). Ces domaines, qui peuvent couvrir jusqu'à 4 Mb, comportent également des gènes exprimés aussi bien à partir de l'allèle maternel qu'à partir de l'allèle paternel. D'abord identifiées chez la souris, ces régions sont parfaitement conservées chez l'homme, certaines d'entre elles étant associées à des maladies liées à l'origine parentale. L'empreinte parentale est caractérisée par une méthylation différentielle des di-nucléotides CpG de l'ADN. Des éléments de contrôle, appelés DMR (Differentially Methylated Region), ont été identifiés au niveau de la plupart des gènes soumis à empreinte ; la méthylation de ces séquences contrôle la transcription des gènes concernés.

La première mise en évidence de l'importance de la méthylation dans le mécanisme d'empreinte parentale est venue de l'analyse de souris portant une délétion de *Dnmt1*. Cette délétion entraîne une létalité des embryons au jour 9 (E9), preuve de l'importance de la méthylation dans le développement embryonnaire des mammifères. Plus spécifiquement, les embryons *Dnmt1*^{-/-} pris avant E9 ont été analysés pour la méthylation des quelques gènes soumis à empreinte connus à cette époque : tous ces gènes avaient une expression totalement perturbée.

L'établissement de la méthylation sur une séquence d'ADN s'accompagne également d'un changement de conformation de la chromatine, en raison de modifications post-traductionnelles des histones. Des enzymes capables de modifier ces histones, comme les HDAC ou les HMT, participent au processus complexe de remodelage de la chromatine. Ces modifications des histones sont également impliquées dans la régulation des gènes soumis à empreinte, et sont étroitement associées à la méthylation de l'ADN. Les DMR, cibles de la méthylation différentielle, sont considérés comme des éléments de régulation en *cis* capables de contrôler l'expression mono-allélique de l'ensemble des gènes. Des facteurs agissant en *trans* ont également été identifiés. Associées aux *Dnmt* et aux protéines de modification des histones, on trouve les MBD, ainsi que des protéines isolatrices (insulator), capables de constituer des barrières au sein d'une unité transcriptionnelle, comme le facteur CTCF (CCCTC-binding factor). De plus, les produits de certains gènes soumis à empreinte sont des ARN non codants dont le rôle n'est pas encore bien défini. Ces ARN pourraient intervenir dans cette régulation.

Les gènes soumis à l'empreinte génomique parentale ont une expression mono-allélique et spécifique de l'origine parentale : quand un gène soumis à l'empreinte s'exprime, un seul des deux allèles est actif, en fonction de l'origine parentale du chromosome qui le porte. Cette définition est peut-être à nuancer par rapport à la réalité de certaines situations. En effet, pour certains gènes considérés comme soumis à l'empreinte, l'expression n'est pas strictement restreinte à un allèle, mais seulement biaisée : le gène est plus fortement exprimé à partir d'un des deux allèles parentaux spécifiquement. D'autre part, il n'est pas rare que des gènes aient une expression soumise à l'empreinte dans certains tissus et au cours de certaines phases du développement seulement, et avoir une expression bi-allélique par ailleurs. Actuellement, environ une centaine de gènes soumis à empreinte sont décrits chez la souris, dont la plupart sont conservés chez l'humain. Cependant, une approche bio-informatique du génome de souris a permis de prédire la découverte possible de 600 gènes (sur 23788 gènes analysés) dont 64 % seraient à expression maternelle.

4- Mise en place de l’empreinte

L’empreinte parentale est un phénomène qui permet de différencier le génome paternel du génome maternel dans l’œuf fécondé. Il doit donc être mis en place à un moment où ces deux génomes sont localisés dans des compartiments différents ; durant la spermatogénèse chez le mâle, et l’ovogenèse chez la femelle. Son déroulement comprend dans un premier temps un effacement des marques épigénétiques dans les cellules germinales, suivi de l’établissement d’un nouveau marquage par méthylation de l’ADN. Ainsi, les DMR maternelles seront méthylées durant l’établissement de l’ovogenèse alors que les DMR paternelles le seront durant la spermatogénèse. Cette première reprogrammation épigénétique dans les cellules germinales est suivie par une seconde qui apparaît après la fécondation et qui concernera l’ensemble du génome de l’embryon nouvellement formé à l’exception des gènes soumis à empreinte. Avant l’établissement des différences épigénétiques, entre les deux chromosomes parentaux, les empreintes des parents sont effacées et les nouvelles empreintes sont établies selon le sexe du nouvel embryon. Cet effacement, qui a lieu dans les cellules germinales primordiales PGC (Primordial Germ Cells), s’effectue par la perte de méthylation de l’ADN. La perte de méthylation spécifique d’allèle liée à l’empreinte à partir du jour 13,5 dans l’embryon de souris. On observe donc un processus d’effacement de l’empreinte dans ces PGC, suivi d’une méthylation *de novo* dans les cellules germinales mâles et femelles. Le moment précis de cette mise en place de la méthylation a été déterminé dans chaque lignée germinale. Elle intervient chez le mâle dans les gonocytes quiescents, pour être complétée au stade pachytène des spermatogonies après la naissance. Chez la femelle, la méthylation n’est complète qu’après la naissance, dans les ovocytes matures (**figure 20**).

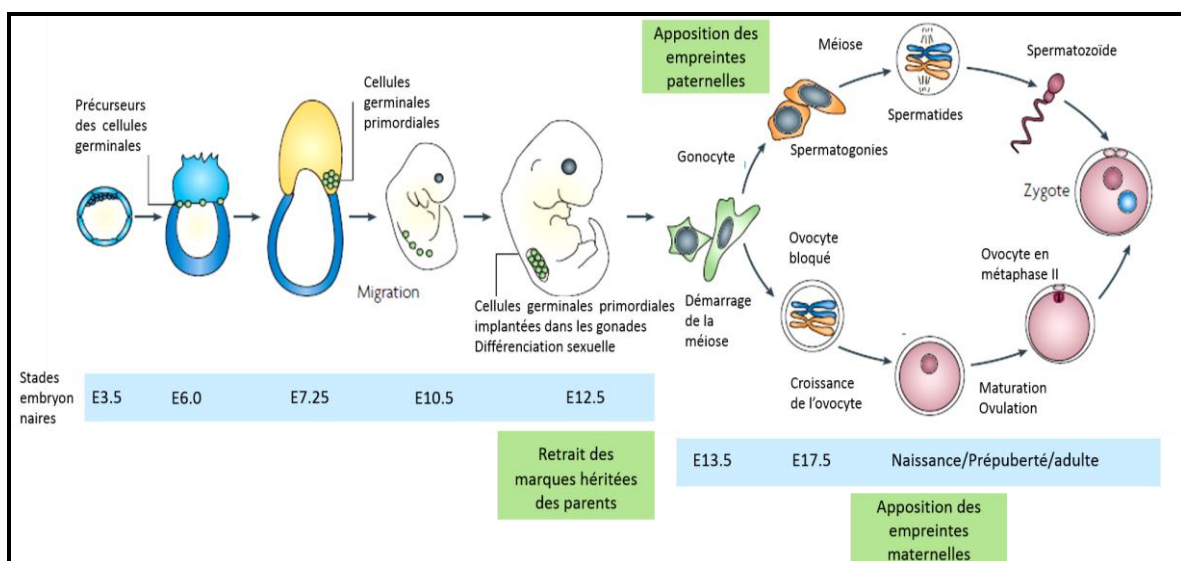


Figure 20 : Mise en place de l’empreinte parentale chez la souris (Sasaki et Matsui, 2008).

5- Empreinte parentale et pathologies humaines

L'empreinte parentale est contrôlée par différents mécanismes étroitement corrélés. Leur dérèglement peut perturber l'équilibre d'expression un ou plusieurs gènes d'une région, conduisant à l'apparition de plusieurs maladies humaines associées à ces domaines d'empreinte. Lorsqu'un gène est soumis à empreinte, un seul des deux allèles est normalement actif, l'autre restant silencieux. Les maladies impliquant des gènes soumis à empreinte sont donc la conséquence :

- La perte d'expression de l'unique allèle actif par délétion ou UDP

- **Syndrome de Prader-Willi (PWS)** : ce syndrome présente une incidence de 1 pour 15 000 naissances. Il est caractérisé par une hyperphagie entraînant une obésité et un retard mental modéré. Ce syndrome résulte de l'inactivation d'un ensemble de gènes de la région PWS (15q11-q12) du chromosome 15 paternel qui est normalement activé. L'anomalie génétique est soit une micro-délétion dans la région PWS du chromosome 15 paternel, soit une UDP maternelle du chromosome 15.

- **Syndrome d'Angelman (AS)** : ce syndrome, présentant une incidence de 1 pour 20 000 naissances, se caractérise par un retard mental sévère, une absence de langage, une microcéphalie, une ataxie et des convulsions. L'anomalie génétique est une absence de UBE3A (ubiquitine-protéine ligase 3A) entraînant un défaut dans la dégradation de certaines protéines par le protéasome au cours du développement cérébral. Cette absence est due soit à une micro-délétion dans la région AS du chromosome 15 d'origine maternelle qui, normalement, est seule active, soit à une UDP paternelle du chromosome 15.

- L'expression de l'allèle normalement silencieux

- **Syndrome de Beckwith-Wiedeman** : syndrome poly-malformatif se caractérise par une croissance excessive du fœtus, une grosse langue (macroglossie), des oreilles malformées, une hypertrophie des organes (viscéromégalie), une prédisposition à développer certaines tumeurs (surtout tumeurs de Wilms). Dans la majorité des cas, on retrouve une expression bi-allélique du gène *IGF2*. Pour que le développement embryonnaire se déroule normalement, un seul allèle d'*IGF2* doit être actif. L'anomalie génétique est le plus souvent une UDP paternelle du chromosome 11.

Chapitre VI

Inactivation du chromosome X

L'inactivation du chromosome X chez les femelles est le processus de normalisation de l'expression des gènes liés à l'X entre mâles et femelles chez les mammifères. C'est aussi un exemple élégant de mécanisme épigénétique régi par plusieurs formes de cette information, mais essentiellement par un ARN non-codant très particulier : Xist (X-inactive specific transcript).

1- Découverte de l'inactivation du chromosome X

La première découverte historique à l'origine de l'inactivation du X remonte à 1949, quand *Barr* et *Bertram* décrivent une structure morphologique condensée associée à l'enveloppe nucléaire dans des neurones de chat en interphase, présente dans les cellules d'individus femelles, mais pas mâles. Dix ans après, on apprend que cette structure, baptisée « corpuscule de Barr » en l'honneur de la personne qui l'a découverte, correspond à l'un des deux chromosomes X qui se trouve à l'état d'hétérochromatine. En parallèle, le comportement particulier d'un gène lié à l'X, dépendant du genre mâle ou femelle de l'individu qui l'exprime, est décrit chez le chat. En effet, les femelles hétérozygotes Mottled (ou pelage tacheté) /Tabby (ou pelage rayé), ont un pelage bigarré hétéroclite avec des régions exhibant le phénotype [Mottled] et d'autres exhibant le phénotype [Tabby]. Chez les mâles, qui sont homozygotes pour l'un ou l'autre de ces allèles, le pelage est systématiquement uniforme. Ceci suggère que les femelles sont mosaïques, avec des clones de cellules qui expriment l'allèle Mottled et d'autres clones qui expriment l'allèle Tabby ; le mode d'hérédité de ce gène n'est pas de type Mendélien. Enfin, de manière inattendue, chez la souris, les femelles (45, XO), sont viables, ce qui montre que l'expression génique à partir d'un des deux chromosomes X est indispensable à ces femelles. Chez l'Humain, on a caractérisé le syndrome de Turner chez l'humain qui est défini comme étant la seule monosomie viable. Mettant en relation ces observations, la généticienne anglaise *Mary Lyon* émet sa célèbre hypothèse en 1961 selon laquelle chez les individus femelles, l'un des deux chromosomes X est inactivé. En l'honneur de cette chercheur, l'inactivation du chromosome X est dite aussi « lyonisation ». Ces découvertes marquent le début d'intenses recherches sur ce phénomène. Il est important de préciser que dans l'espèce humaine, la cinétique d'inactivation du chromosome X au cours du développement embryonnaire est peu connue du fait des interdictions ou des limitations de la recherche sur l'embryon humain qui existent dans la majorité des pays et du manque d'embryons humains disponibles pour la recherche.

2- Compensation de dose chez les mammifères

Chez les mammifères, la paire d'hétérosomes (chromosomes sexuels) diffère selon le sexe de l'individu : les femelles sont homogamétiques et portent deux chromosomes X, alors que les mâles sont hétérogamétiques et portent un chromosome X et un chromosome Y. Cette situation pose un problème de dosage : pour les gènes portés par le X, les femelles auraient deux doses du produit du gène, alors que les mâles n'en auraient qu'une. Afin de normaliser l'expression des gènes liés au X, une « compensation de dose » est assurée. Dans ce but, l'un des chromosomes X des cellules de mammifères femelles est mis en silence. Cependant, des découvertes récentes démontrent que la notion de compensation de dose par inactivation relève de la simplification plutôt que d'une réalité biologique.

- **Premièrement**, bien que l'inactivation du chromosome X soit un processus global qui affecte l'ensemble du chromosome, certains gènes y échappent. C'est ce qu'on appelle les « gènes fugitifs » ou « escapes genes », qui gardent une expression bi-allélique dans les cellules de femelles. Chez les rongeurs, le nombre de gènes fugitifs est limité. Chez l'homme, à l'inverse, ce sont environ 15 % des gènes du chromosome X qui échappent à l'inactivation, et 10 % qui montrent une inactivation variable selon les individus (polymorphisme).
- **Deuxièmement**, les gènes du chromosome X sont monosomiques (une seule copie) alors que les gènes des autosomes sont disomiques (deux copies). Afin de compenser ce déséquilibre, il avait été prédit que la transcription à partir du chromosome X devait être doublée chez les mâles et les femelles. Et en effet, plusieurs études récentes ont démontré que le ratio entre l'expression des gènes de l'X et l'expression des gènes autosomiques était proche de 1 dans des tissus somatiques. On est donc en présence d'un deuxième niveau de compensation de dose qui affecte le chromosome X des deux genres (mâle et femelle).

3- Dynamique d'inactivation du chromosome X

3-1- Cycle de l'inactivation

Les chromosomes X des cellules des mammifères femelles traversent plusieurs cycles d'inactivation et de réactivation au cours du développement embryonnaire, qui varient selon les tissus. Chez la souris, dans les tissus extra-embryonnaires, le chromosome X paternel est spécifiquement inactivé.

Dans l'embryon, au cours des premières divisions cellulaires, c'est aussi le X paternel qui est inactivé ; au stade préimplantatoire, il est réactivé dans la masse cellulaire interne du blastocyste ; s'ensuit alors l'inactivation au hasard d'un des deux chromosomes X dans ces mêmes cellules qui donneront l'embryon lui-même. Enfin, dans la lignée germinale de l'embryon, le chromosome X inactif est réactivé avant de rentrer dans un nouveau cycle après la fertilisation.

3-2- Modalités de l'inactivation

Deux formes d'inactivation du chromosome X sont connues :

- **inactivation aléatoire** : cette inactivation affecte de manière aléatoire le chromosome X maternel ou paternel. Une telle inactivation prend place dans les cellules de la masse cellulaire interne du blastocyste et est maintenue tout au long de la vie de l'individu. Chez l'humain, elle a été particulièrement étudiée notamment dans des cellules embryonnaires souches (ES) en différenciation. Ces cellules ES humaines proviennent de l'embryon aux tout premiers stades de son développement, quelques jours seulement après la fécondation. Les cellules ES femelles indifférenciées sont dérivées de la masse interne du blastocyste et ont deux chromosomes X actifs. Quand leurs conditions de culture sont changées de façon à induire leur différenciation, l'un des deux chromosomes X est inactivé de manière aléatoire.
- **inactivation soumise à l'empreinte** : le chromosome X hérité du père est inactivé spécifiquement. Ce phénomène a lieu chez les marsupiaux, et dans les tissus extra-embryonnaires de la souris, plus précisément dans le tropho-ectoderme et l'endoderme primitif au stade implantatoire du développement de l'embryon.

3-3- Réactivation du X

Chez la souris, le chromosome X est réactivé à deux occasions. La première a lieu au stade post-implantatoire et concerne le chromosome X paternel qui a subi l'inactivation soumise à l'empreinte des stades précoces du développement. Pendant la croissance du blastocyste, les cellules de l'épiblaste commencent à ne plus exprimer Xist et perdent progressivement les marques associées à l'inactivation. Cette réactivation a lieu dans les cellules de l'épiblaste spécifiquement (qui donneront l'embryon à proprement parler), et pas dans celles du tropho-ectoderme ou de l'endoderme. Au moment de l'implantation, le X paternel a été totalement réactivé dans l'épiblaste.

La deuxième vague de réactivation du X a lieu dans les cellules de la lignée germinale femelle. Des études anciennes avaient défini le moment de la réactivation dans les PGC ayant atteint les crêtes génitales, peu avant la méiose (entre E11,5 et E13,5 selon les études). Récemment, on a redéfini le moment de la réactivation à un stade beaucoup plus précoce. Il a été prouvé que des marqueurs du chromosome X inactif, à savoir l'expression de *Xist* et des marques épigénétiques associées, commençaient à être perdues avant le stade E9,5. La réactivation du chromosome X inactif dans la lignée germinale est donc amorcée dans les PGC naissants avant leur migration. Pour rappel, la période de gestation chez la souris est de 20 jours.

4- Mécanistique d'inactivation

L'inactivation du X dans la vie de l'embryon est un processus multi-étapes initié après la fécondation. L'ARN non-codant *Xist*, qui s'accumule sur le chromosome X qui va être inactivé dès le stade 4 cellules parallèlement avec l'exclusion de l'ARN polymérase II, puis s'ensuit l'installation des marques de la chromatine spécifique du X inactif. La répression des gènes localisés sur le X inactivé est observée à partir du stade 8 cellules. Ce processus d'inactivation se déroule comme suit :

- « **coating** » du chromosome X par l'ARN *Xist* : cet ARN a la faculté remarquable de s'accumuler de manière stable le long de son chromosome. Il s'étend en *cis* pour recouvrir tout le chromosome X qui sera inactivé.
- **répression transcriptionnelle** : par exclusion de la machinerie de l'ARN polymérase II, des facteurs d'épissage et des transcrits naissants du territoire chromosomique X.
- **perte de modifications d'histones typiques de la chromatine transcriptionnellement active** : comme l'acétylation H3K9 et la méthylation H3K4.
- **acquisition de marques des histones associées avec la chromatine inactive** : H4 hypo-acétylées, H3K27 méthylée, H3K9 méthylée, H4K20 méthylée, H2AK119 mono-ubiquitinylée.
- **recrutement de protéines qui participent aux modifications de la chromatine** : en parallèle aux changements des marques des histones.
- **asynchronie de réplication en phase S** : durant les divisions cellulaires qui surviennent après cette inactivation, la réplication du chromosome X inactivé sera tardive comme pour l'hétérochromatine.
- **incorporation d'un variant d'histone spécifique** : la macro-H2A.
- **méthylation de l'ADN des séquences promotrices (figure 21)**.

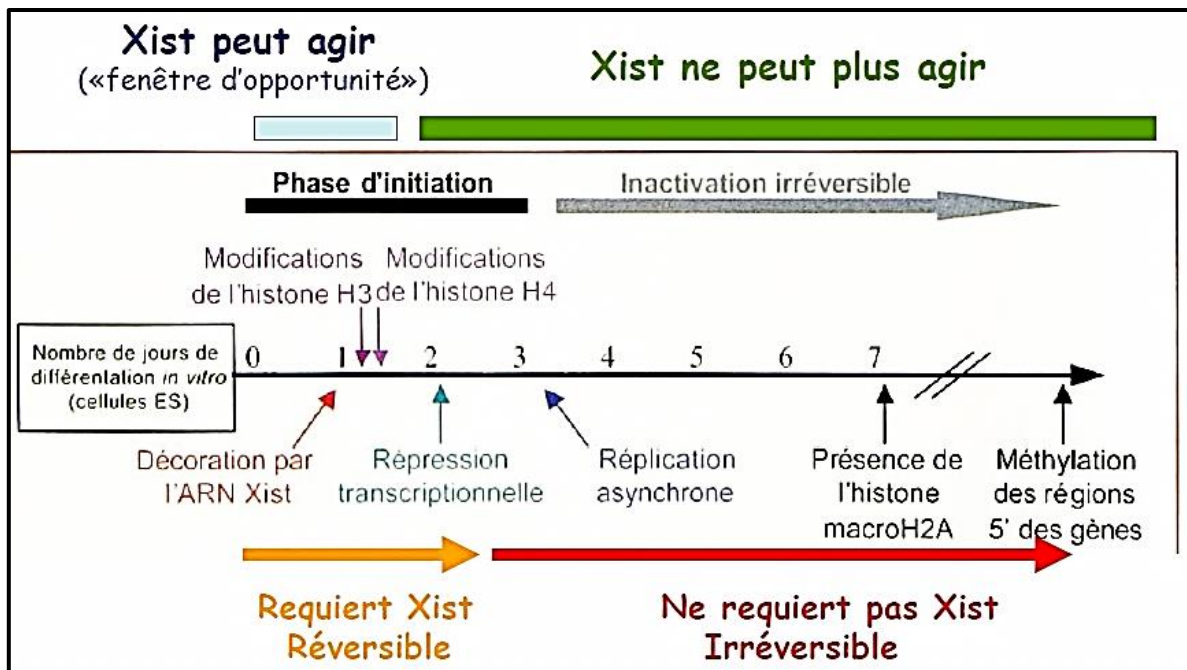


Figure 21 : Mécanistique de l'inactivation du chromosome X (Heard, 2004).

L'étude du profil d'expression des gènes se trouvant au sein du chromosome X chez l'humain a permis d'isoler et de caractériser le gène *Xist*, révélé comme étant le seul gène connu étant uniquement exprimé à partir du chromosome X inactif. Cependant, il est important de signaler que l'ARNnc *Xist* est requis pour établir l'inactivation, mais pas pour son maintien. Le gène *Xist* produit un ARNnc de 17 kb. Le chromosome X inactif produit l'ARN *Xist* alors que l'expression du gène est réprimée sur le X actif. Le chromosome X sans *Xist* ne peut pas être inactivé. De même si on greffe un gène *Xist* sur un autre chromosome il sera inactivé. Avant l'inactivation, les deux chromosomes X produisent l'ARN *Xist* en petite quantité. Durant le processus d'inactivation, le futur X actif stoppe la production de cet ARN alors que l'X inactif l'augmente considérablement. Sur le futur X inactif, l'ARN *Xist* recouvre progressivement le chromosome et induit son inactivation. Tout comme *Xist*, le gène *Tsix* produit un ARN non traduit à partir du brin complémentaire au gène *Xist* : c'est un gène anti-sens à *Xist*. *Tsix* est un inhibiteur de *Xist* : un chromosome X portant une mutation qui abolit l'expression de *Tsix* est systématiquement inactivée.

Chapitre VII

Le monde des
ARN non codants

Des études récentes comme le Projet ENCODE (Encyclopedia of DNA elements), démontrent qu'entre 33 et 75 % du génome humain serait transcrit. Parmi les ARN transcrits, on retrouve une importante fraction qui ne code pas pour des protéines. Actuellement, plus de 600 familles d'ARN non-codants (ARNnc) sont connues et répertoriées dans des banques de données disponibles sur le web. Les ARNnc interviennent dans de nombreux processus essentiels de la cellule. Plus récemment, plusieurs études ont révélé l'existence de nombreux petits ARN.

1- ARN dans la cellule

Dans chacune des cellules d'un organisme vivant, il existe trois grands types de molécules informationnelles : l'ADN, l'ARN et les protéines. L'ADN stocke l'information génétique transmise au cours des générations cellulaires. Les protéines sont le résultat de l'expression de cette information ; elles assurent la plupart des tâches nécessaires au bon fonctionnement de l'organisme. L'ARN a la particularité d'être une molécule très polyvalente. Le dogme central proposé par *Francis Crick* à la fin des années 50, fait apparaître une des fonctions des ARN : servir d'intermédiaire dans la synthèse protéique avec les ARN messagers. Il existe néanmoins d'autres types d'ARN, issus de la transcription, mais qui ne subissent pas de traduction. Ces ARN sont fonctionnels par eux-mêmes, sans coder pour une protéine. Pour les distinguer des ARN messagers, on les appelle des ARN non-codants. Suite à la découverte des mécanismes du dogme central, *Watson* et *Crick* ont mis en évidence en 1961 les deux premiers types d'ARN non-codants : les ARN ribosomiques (ARNr) et les ARN de transfert (ARNt). Ces ARN sont les plus célèbres et sont présents dans tous les organismes vivants. La fraction des ARN qui n'étaient ni des ARNm, ni des ARNr et ni des ARNt a été peu étudiée pendant longtemps, car celle-ci était peu abondante et la plupart du temps instable. Il y avait peu de motivation et peu de capacité pour déterminer si cette fraction d'ARN contenait un autre type d'information.

2- Découverte des ARNnc

Dans les années 1970, plusieurs autres petits ARN ont été détectés et isolés biochimiquement. Ceux-ci incluent la RNaseP qui est nécessaire pour la maturation des ARNt, les ARN « U » qui participent entre autres à l'épissage des ARNm et les petits ARN nucléolaires (snoRNA) qui guident des modifications chez d'autres ARN. Au milieu des années 80, *Altman* et *Cech* ont démontré que les ARNnc pouvaient catalyser des réactions chimiques par eux-mêmes, un rôle exclusivement réservé aux protéines (enzyme).

Altman a démontré que la composante ARN de la RNaseP pouvait cliver l'extrémité 5' des précurseurs d'ARNt. Ils ont ainsi démontré que l'ARN pouvait avoir un rôle catalytique de clivage et ligation sans la présence de cofacteurs de protéine. Cette découverte mène à un prix Nobel de chimie en 1989, car elle a ouvert tout le domaine d'étude des ribozymes et la perspective de thérapies basées sur l'ARN.

3- Définition et classification actuelle des ARNnc

Les ARNnc sont des molécules ARN qui sont transcrites à partir de l'ADN génomique, mais non traduites en une protéine et qui ont une fonction dans la cellule. Ces ARNnc assurent diverses fonctions pour la plupart déterminées par la structure spatiale qu'adopte la molécule d'ARN en se repliant sur elle-même. Les ARNnc qui ne se replient pas de manière spécifique s'associent le plus souvent à d'autres molécules pour former des complexes. Les fonctions de ces derniers sont alors en lien étroit avec leur séquence. Les ARNnc peuvent être divisés en fonction de la longueur du transcrit. En effet, il y a les petits ARNnc, d'une longueur maximale de 200 pb et les longs ARNnc qui sont d'une longueur de plus de 200 pb. De par la fonction, actuellement, nous pouvons classer ARNnc en trois classes :

- **les ARNnc structuraux ou adaptateurs** : ARNt et ARNr,
- **les ARNnc avec un rôle enzymatique ou ribozymes** :
 - **ribonucléase P** : complexe composé d'un ARN catalytique et d'une ou plusieurs protéines accessoires. Sa fonction est de catalyser la maturation des ARNt.
 - **ribosome** : le ribosome peut aussi être considéré comme « un ribozyme » car le site actif du ribosome, qui catalyse la réaction de synthèse des liaisons peptidiques, n'est en effet composé que par l'ARNr.
 - **riboswitch** : ARN que l'on trouve sur certains ARNm ont une activité auto-lytique : ils se clivent eux-mêmes lorsqu'ils sont liés à un ligand activateur. C'est le cas par exemple du riboswitch spécifique de la glucosamine-6-phosphate.
 - **introns auto-épissables** : ces introns sont capables de se cliver spécifiquement en l'absence de protéines. Même s'ils sont capables de promouvoir leur épissage, ce ne sont pas des catalyseurs au sens strict, puisque c'est une réaction unique, sans recyclage du catalyseur.
- **les ARNnc régulateurs** : siRNA (small interfering RNA) et miRNA (microRNA) ; médiateurs de l'interférence ARN (ARNi). Un autre groupe celui des longRNA.

4- Interférence ARN (siRNA et miRNA)

L'ARNi est un mécanisme hautement conservé entre espèces et permet d'éteindre l'expression d'un gène de manière spécifique en utilisant un ARN double brin (ARNdb). Le rôle naturel du « silencing » par l'ARN semble être la régulation de l'expression des gènes, mais aussi dans la défense contre les éléments transposables et les virus. Deux classes ont été trouvées comme étant des régulateurs séquence spécifique post-transcriptionnelle : les siRNA et les miRNA. Les siRNA sont produits à partir d'un d'ARNdb qui est synthétisé *in vitro* ou *in vivo* à partir des virus ou des séquences répétitives introduites par ingénierie génétique. Plusieurs miRNA endogènes participent à la régulation de l'expression des gènes et sont des importantes molécules régulatrices dans les plantes et les animaux.

4-1- siRNA

a- Découverte de l'interférence ARN

En 1990, le premier phénomène de type interférence ARN est découvert par hasard chez le pétunia par *Jorgensen et al.* En introduisant des copies supplémentaires du gène de la *CHS* (CHalcone Synthase), celui-ci pensait renforcer la couleur violette des pétunias. Or, les fleurs obtenues se sont avérées au contraire blanches ou partiellement blanches et le niveau d'ARNm de la *CHS* cinquante fois inférieur aux fleurs sauvages. Ce mécanisme, encore incompris, est alors appelé « co-suppression » puisque les deux copies du gène, endogène et transgène, semblent interagir de manière négative en se « co-supprimant » mutuellement.

En 1995, *Guo et Kempfues* injectent un ARN anti-sens à l'ARNm de *par-1* dans les gonades du ver nématode *Caenorhabditis elegans*, l'ARN anti-sens étant supposé s'hybrider à l'ARNm et inhiber ainsi l'expression du gène dans la descendance. De manière surprenante, le contrôle effectué avec l'ARN sens de *par-1* donne des résultats similaires, mimant le phénotype d'un mutant *par-1*.

En 1998, une explication à ces résultats étranges a été apportée par la publication des travaux de *Fire et Mello* qui décrivent pour la première fois le mécanisme d'interférence ARN (ARNi) chez le nématode. Leurs résultats montrent que l'expression de gènes endogènes est inhibée de manière spécifique et très efficace par l'injection d'ARN double brin (ARNdb) homologue au gène visé, tandis que l'ARN sens ou l'ARN anti-sens seuls sont relativement inefficaces. Peu de temps après, il a été démontré que le mécanisme de RNAi est également fonctionnel chez la drosophile où de l'ARNdb est utilisé de manière efficace sur des embryons pour inactiver des gènes et mimer des phénotypes mutants.

b- Identification des siRNA

L'utilisation de cellules de drosophile a permis de mettre en évidence des ARN de petite taille (21-25 nt) responsables du mécanisme de RNAi. *Hammond et al* mettent ainsi en évidence, dans les extraits cellulaires préalablement transfectés avec un ARNdb, une activité nucléase, responsable de la dégradation de l'ARNm homologue cible. Cette activité nucléase est nommée RISC (RNA-Induced Silencing Complex), et le fractionnement des extraits cellulaires permet de montrer que cette activité RISC co-purifie avec des petits ARNs de 25 nt correspondant à des fragments sens et anti-sens de l'ARNm cible.

À l'aide de ce système *in vitro*, *Zamore et al* montrent quant à eux que l'ARNdb est clivé en petits fragments de 21-23 nt même en absence de l'ARNm cible. Ils montrent aussi que l'ARNm cible est clivé dans la région correspondant à l'ARNdb, et tous les 21-23 nt environ, suggérant que les ARNs de 21-23 nt sont les intermédiaires guidant ce clivage.

Par la suite, *Elbashir et al* montrent que des fragments synthétiques d'ARNdb de 21-23 nt, avec des extrémités 3' sortantes de 2 nt, sont suffisants pour éteindre de manière spécifique de la séquence l'expression d'un gène. Ceci provoque alors l'émission d'un photon dont la lumière résultante est jaune-verte. Ils confirment ainsi que ces petits ARNdb sont les intermédiaires de la RNAi et les nomment siRNAs. Leurs résultats montrent aussi que l'ARNm cible est clivé à peu près au centre de la région complémentaire au siRNA.

À cette période, seules de rares études ont montré la possibilité d'éteindre des gènes par RNAi chez les mammifères, et uniquement dans l'ovocyte et l'embryon précoce de souris, les cellules ES indifférenciées. Les difficultés rencontrées sont liées au fait qu'à des stades plus avancés du développement, les cellules de mammifères déclenchent normalement une réponse de défense antivirale non spécifique appelée « réponse interféron » lorsque des ARNdb de plus de 30 nt entrent dans la cellule. L'ARNdb active la Protéine Kinase dépendante de l'ARNdb (PKR) et la 2'-5'-oligoadenylate synthétase. La PKR phosphoryle le facteur d'initiation de la traduction eIF2 α , déclenchant ainsi une inhibition globale de la traduction, tandis que la 2'-5'-oligoadenylate synthétase active indirectement la RNaseL responsable de la dégradation non spécifique des ARNm. Lorsque cette réponse a lieu, si elle ne conduit pas à l'apoptose, elle masque le phénomène spécifique de RNAi par une inhibition globale et non spécifique de l'expression des protéines. La découverte des siRNA permet alors de montrer que, de par leur petite taille, les ARNdb de 21-22 nt, avec des extrémités 3' sortantes de 2 nt sont capables d'éteindre spécifiquement l'expression de gènes exogènes et endogènes dans des cellules de mammifères sans déclencher de réponse interféron.

c- Voie siRNA

L'enzyme responsable du clivage de l'ARN double brin en petits duplex de 21-23 nt avec des extrémités 3' sortantes de 2 nt est une RNase de type III appelée Dicer. Ces siRNA sont reconnus par le complexe multiprotéique RISC. En présence d'ATP, le complexe activé utilise le brin anti-sens du siRNA comme guide pour reconnaître et cliver la séquence homologue sur l'ARNm cellulaire. La dégradation de l'ARNm provoque l'inactivation de l'expression de la protéine pour laquelle il code normalement (**figure 22**).

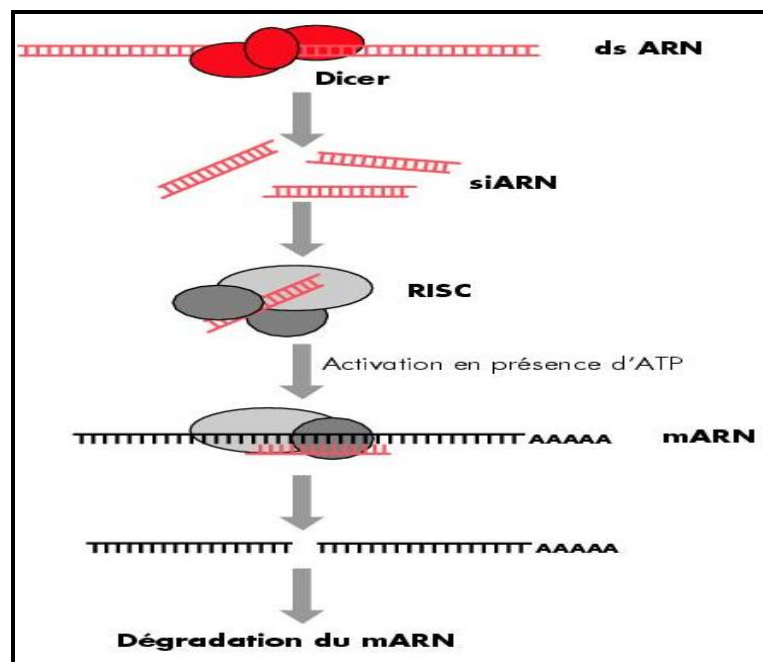


Figure 22 : Mécanisme général de l'ARN interférence (Davidson et McCray, 2011).

4-2- miRNA

a- Découverte des micro-ARNs

Parallèlement à la découverte du mécanisme de RNAi, des recherches menées sur des mutants du ver *C. elegans* ont permis de mettre en évidence l'existence de petits ARNs interférents endogènes, transcrits à partir du génome cellulaire : les micro-ARNs (miRNAs). Le premier miRNA, *lin-4*, a été décrit de manière concomitante par les équipes d'Ambros et Ruvkun en 1993. Chez ce Nématode, la séquence des étapes du développement larvaire peut être affectée chez certains mutants : les gènes pour lesquels ils sont mutés sont appelés « hétéro-chroniques ». Ainsi, des mutants peuvent exécuter un programme de divisions et de différenciations spécifiques d'un stade larvaire donné à un stade antérieur ou postérieur.

L'étude des mutants de *C. elegans* avait déjà permis d'identifier les gènes *lin-4* et *lin-14* comme essentiels pour le contrôle temporel du développement post-embryonnaire de ce nématode. Il existe en effet chez ce ver quatre stades larvaires (L1 à L4) qui peuvent être caractérisés par l'état de division et de différenciation des cellules. Il a ainsi été montré que des mutations de *lin-4* entraînaient des divisions cellulaires caractéristiques du stade L1 à des stades plus avancés de développement. Au contraire, des mutations de *lin-14* entraînaient de manière prématurée des caractéristiques de stades larvaires plus avancés. *Ruvkun* et *al* ont montré que *lin-14* codait pour une protéine nucléaire dont la diminution graduelle d'expression à la fin du stade larvaire L1 initiait le passage au stade L2. Les phénotypes opposés des mutants de *lin-4* et *lin-14* avaient déjà permis à *Ambros* de proposer que *lin-4* soit responsable de l'inhibition de *lin-14*. En 1993, son équipe montre que le gène *lin-4* ne code pas pour une protéine, mais pour deux transcrits de petites tailles (environ 61 et 22 nt), dont une partie de la séquence est partiellement complémentaire à des séquences contenues dans le 3'UTR (Untranslated region) de l'ARNm de la protéine *lin-14*. Au même moment, *Whightman* et *al* montrent que la région 3'UTR de *lin-14* est suffisante à sa régulation par *lin-4* au niveau post-transcriptionnel. Il est ensuite montré que c'est une interaction directe par un appariement partiel anti-sens du petit ARN *lin-4* avec les séquences contenues dans le 3'UTR de l'ARNm de *lin-14* qui est responsable de l'inhibition de *lin-14* au moment de la traduction, initiant ainsi le passage des stades larvaires L1 à L2. Entre temps, il est montré que *lin-4* inhibe de la même manière, via un élément de réponse contenu dans le 3'UTR, la protéine *lin-28*, responsable du passage des stades larvaires L2 à L3.

En 2000, *Ruvkun* découvre un nouveau miRNA chez *C. elegans*. Il s'agit de *let-7*, responsable de l'inhibition de l'expression de la protéine *lin-41*, permettant cette fois la transition du stade larvaire L4 à l'âge adulte. Contrairement à *lin-4*, des homologues de *let-7* sont rapidement découverts dans d'autres organismes, y compris les mammifères. En 2001, *Ambros*, *Bertel* et *Kusch* publient simultanément des études montrant l'existence de nombreux autres gènes, conservés du ver à l'humain, codant pour des petits ARN comme *lin-4* et *let-7*, qui sont alors appelés micro-ARNs ou miRNA. On compte aujourd'hui plus de 1000 miRNAs humains identifiés, séquences et annotés dans le génome (www.mirbase.org).

b- Biogenèse et voie des miRNA

Les gènes codant pour des miRNAs sont transcrits par l'ARN polymérase II (ou III) en un transcrit primaire appelé pri-miRNA.

La structure tige-boucle du pri-miRNA est reconnue et excisée par le microprocesseur : un complexe nucléaire composé de la RNase III Drosha et de la protéine de liaison à l'ARNdb DGCR8, générant ainsi le pre-miRNA (voie canonique). Alternativement, quelques pre-miRNAs constituant de courts introns sont générés directement par épissage (mirtrons). Le pre-miRNA est exporté dans le cytoplasme par le complexe Exportine 5-RanGTP, où il est reconnu par un complexe composé de la RNase III Dicer, et la ou les protéine(s) de liaison à l'ARNdb TRBP et/ou PACT. Dicer clive le pre-miRNA de manière à éliminer la boucle et générer un duplex miRNA : miRNA. La présence préalable ou le recrutement d'une protéine AGO permet de former le complexe pre-RISC, activé par l'élimination du brin « passager ». Le complexe miRISC activé s'associe à l'ARNm cible grâce à l'appariement partiel du miRNA à une séquence cible située généralement dans le 3'UTR. La fixation de miRISC sur l'ARNm cible a pour conséquence(s) une inhibition de la traduction et/ou une déadénylation suivie d'une dégradation de l'ARNm (**figure 23**).

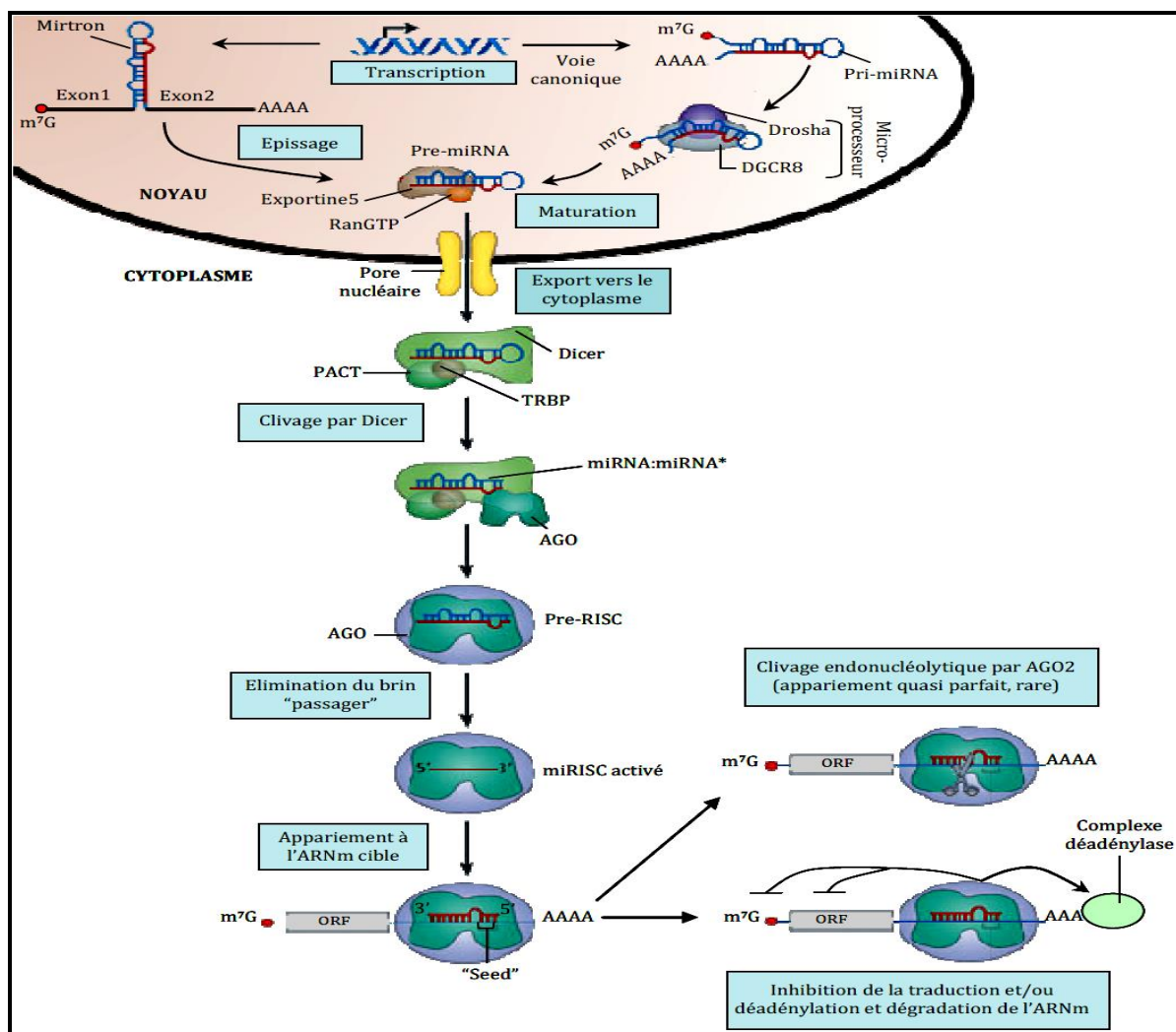


Figure 23 : La voie miRNA (Davidson et McCray, 2011).

5- Longs ARNnc

5-1- Caractéristiques générales

Les longs ARNnc sont des transcrits étant exprimés chez plusieurs sinon tous les organismes et ces derniers se comptent par dizaine de milliers. Malgré le fait qu'ils ne codent pour aucune protéine, ces derniers ont une importance biologique soulignée par le fait de leur régulation. La transcription de plusieurs longs ARNnc se fait par l'action de l'ARN polymérase II. Ils peuvent subir également quelques modifications dont un ajout d'une coiffe à l'extrémité 5', une poly-adénylation à l'extrémité 3' ainsi qu'un épissage. Toutes ces modifications sont subies également par les ARNm expliquant ainsi la similarité structurale entre certains longs ARNnc et ARNm. De plus, certains longs ARNnc possèdent un court cadre de lecture ouvert (ORF) ne servant pas à la production de protéines, ou du moins, d'une quantité très faible et non fonctionnelle. Les longs ARNnc ont été longtemps délaissés par la communauté scientifique du fait qu'ils ne codaient pour aucune protéine et ensuite du fait que la longueur excessive du transcrit produit, allant généralement à plus de 1000 pb pour certaines classes. Le premier long ARNnc, dénommé *Imprinted Maternally Expressed Transcript (H19)*, a été découvert en 1990, et ce, avant même l'utilisation du terme. À l'époque, le gène *H19* était décrit comme étant inhabituel et son produit, une petite molécule d'ARN, était caractérisé comme une molécule différente d'un ARNm classique. Quelques longs ARNnc ont été découverts par la suite, comme *Xist* et son long ARNnc anti-sens associé *Tsix*. Toutefois, l'émergence de la plupart des longs ARNnc s'est effectuée suite à certains avancements technologiques et connaissances génomiques. La découverte de plusieurs milliers de longs ARNnc a amené les scientifiques à pousser encore plus loin les recherches portant sur ces longs transcrits afin d'établir les bases de leur classification et ainsi élucider certains de leurs mécanismes. Malgré l'avancement des connaissances dans ce sens, plusieurs facettes des longs ARNnc restent encore à être plus étudiées. En effet, peu de choses sont connues quant aux différentes fonctions et mécanismes de ces derniers.

5-2- Classification

La classification des longs ARNnc se fait en fonction de leur localisation dans le génome et selon leur emplacement par rapport à un gène codant. Il existe actuellement 5 classes de longs ARNnc : sens, anti-sens, intronique, divergent et inter-génique (**figure 24**).

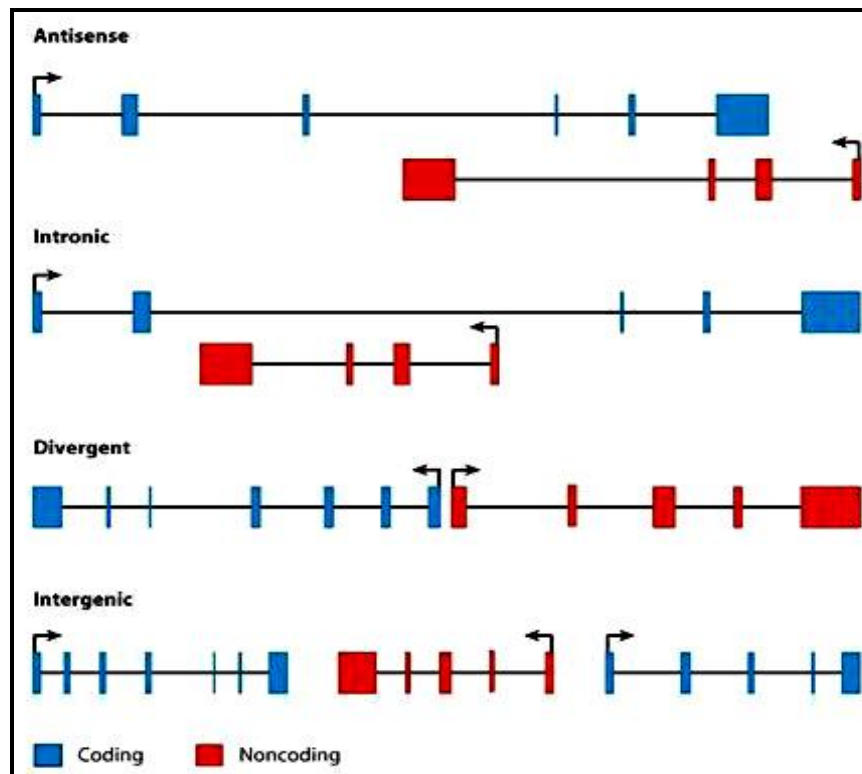


Figure 24 : Classification des longs ARNnc (Davidson et McCray, 2011).

Les portions bleues représentent les parties codantes d'un gène et les portions rouges les parties non codantes. Seule la classe sens n'est pas représentée dans cette figure.

- **Les longs ARNnc sens et anti-sens** : sont initiés dans une portion d'un gène codant pour une protéine ou dans sa portion 3'. Ces derniers sont transcrits dans la même direction que le gène (sens) ou dans la direction opposée (anti-sens) et chevauchent généralement au moins un exon codant.
- **Les longs ARNnc intronique** : ne chevauchent aucun exon codant. Ils sont initiés dans un intron d'un gène et leur transcription peut être effectuée dans les deux directions.
- **Les longs ARNnc divergents** : la transcription peut être également effectuée dans les deux directions. Ce qui caractérise cette classe de longs transcrits est que le site d'initiation se fait à proximité d'un promoteur d'un gène codant pour une protéine. Ce site se situe généralement à une centaine de pb du promoteur, mais cette distance n'est toutefois pas encore définitivement établie.
- **Les longs ARNnc inter-génique** : contiennent des longs ARNnc dont le site d'initiation est généralement éloigné de 5000 pb d'un gène codant pour une protéine. Cette distance empêche tout chevauchement. L'initiation de leur transcription peut se faire dans les deux directions et les longs ARNnc faisant partie de cette classe sont parmi les plus conservés à travers les organismes.

5-3- Fonctions

Plusieurs rôles essentiels à la biologie de la cellule sont exercés par les longs ARNnc. Le fait que plusieurs d'entre eux possèdent une faible expression comparativement à celle des ARNm, permet de suggérer un rôle de régulation. De plus, de nombreux longs ARNnc ont une expression différentielle et/ou restreinte à certains compartiments cellulaires et tissus spécifiques. Un nombre considérable d'études convergent vers un rôle de régulation épigénétique ainsi qu'un rôle de régulation des gènes autant au niveau transcriptionnel que post-transcriptionnel. Les longs ARNnc utilisent différents mécanismes :

- **un mécanisme de guide ou encore de molécule d'échafaudage** : ce mécanisme permet la réunion de complexes protéiques n'ayant pourtant aucun domaine d'interaction commun. La plupart de ces longs ARNnc impliqués dans ce type de régulation va entraîner une répression de la transcription de leurs gènes cibles. Les longs ARNnc principalement connus pour ce type de régulation sont entre autres les ARNnc suivant : *HOTAIR* (Hox transcript anti-sens RNA), *ANRIL* (Anti-sens Non-coding RNA in the *INK4* Locus) et *Xist*.
- **un mécanisme d'interaction directe entre le transcrit d'ARNnc et certains facteurs de transcription** : un long ARNnc connu pour utiliser ce genre de mécanisme est *PANDA* (Promoter of *CDKN1A* anti-sense DNA damage activated RNA). Ce dernier permettrait l'inhibition de certains gènes apoptotiques en se liant directement au facteur de transcription nucléaire Y (*NF-YA*), qui active la mort cellulaire par apoptose.
- **mécanismes de régulation des gènes au niveau post-transcriptionnelle** : certains longs ARNnc vont s'associer directement avec un ARNm spécifique, comme c'est le cas pour *BACE1-AS* (Beta-site APP-cleaving enzyme 1 anti-sense RNA) ce qui favorise une diminution de l'accessibilité de l'ARNm cible.
- **mécanismes de régulation des gènes au niveau post-transcriptionnelle par association avec un miRNA** : empêchant la liaison de ce dernier à l'ARNm ayant ainsi un impact sur la synthèse protéique. Le long ARNnc est *Linc-MD1* (Long non-coding RNA muscle differentiation 1) se lie à deux miRNA (mir-133 et mir-135).

Malgré le fait que plusieurs mécanismes d'action de certains longs ARNnc restent encore à être déterminés ou approfondis, l'impact de ces derniers sur l'expression des gènes est non-négligeable. La dérégulation d'expression des longs ARNnc amène une perturbation au niveau cellulaire et peut entraîner plusieurs conséquences extrêmement néfastes comme entre autres l'apparition de plusieurs maladies.

Références bibliographiques

1. **ADKINS NL, WATTS M, GEORGEL PT et al.** 2004. To the 30-nm chromatin fiber and beyond. *Biochimica Biophysica Acta*. 1677:12-23.
2. **ALBERTS B, JOHNSON A, LEWIS J et al.** 2002. *Molecular Biology of the Cell*. 4th édition. *Garland Science, New York*. ISBN:0-8153-3218-1.
3. **BANNISTER et KOUZARIDES.** 2011. Regulation of chromatin by histone modifications. *Cell Research*. 21:381-395.
4. **BARLOW DP et BARTOLOMEI MS.** 2014. Genomic imprinting in mammals. *Cold Spring Harbor perspectives in biology*. 1:6(2).
5. **BARTOLOMEI MS et FERGUSON-SMITH AC.** 2011. Mammalian genomic imprinting. *Cold Spring Harbor perspectives in biology*. 30:561-580.
6. **BAXTER J, MERKENSCHLAGER M et FISHER AG.** 2002. Nuclear organization and gene expression. *Current Opinion in Cell Biology*. 14:372-376.
7. **BECKER PB et HORZ, W.** 2002. ATP-dependent nucleosome remodeling. *Annual Review of Biochemistry*. 71:247-273.
8. **BEDNAR J, HOROWITZ RA, GRIGORYEV SA et al.** 1998. Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proceedings of the National Academy of Sciences-USA*. 95:14173-14178.
9. **CALLEN JC et PERASSO P.** 2005. *Biologie cellulaire : des molécules aux organismes*. 2^{ème} édition. *Dunod, Paris*. ISBN:2-1004-9236-5.
10. **CECH TR et STEITZ JA.** 2014. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell*. 157:77-94.
11. **CHURCH D.** 2013. *Le génie dans vos gènes : médecine épigénétique et nouvelle biologie de l'intention*. *Éditions Dangles*. Pagination multiple. ISBN:2-7033-0972-4.
12. **CREMER T et CREMER C.** 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Reviews Genetics*. 2:292-301.
13. **DAVIDSON BL et MCCRAY PB.** 2011. Current prospects for RNA interference-based therapies. *Nature Reviews Genetics*. 12:329-340.
14. **EBERHARTER A et BECKER PB.** 2004. ATP-dependent nucleosome remodeling: factors and functions. *Journal of Cell Science*. 117:3707-3711.
15. **GEISLER S et COLLER J.** 2013. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nature reviews Molecular cell biology*. 14:699-712.

16. **GRAFF J et MANSUY I.** 2008. Epigenetic codes in cognition and behavior. *Behavioural Brain Research*. 192:70-87.
17. **GRIFFITHS A, CARROLL S, WESSLER S et al.** 2013. Introduction à l'analyse génétique. 6^{ième} édition. *Éditons De Boeck Supérieur*. Pagination multiple. ISBN:2-8041-6013-0.
18. **GRIFFITHS A.** 2001. Analyse génétique moderne. *Editions De Boeck Université, Bruxelles*. Pagination multiple. ISBN:2-7445-0111-5.
19. **GUÉNET JL.** 2009. Empreinte génomique parentale. *Bulletin de l'Académie Vétérinaire de France*. 162:4/5.
20. **HEARD E.** 2013. Épигénétique et mémoire cellulaire. *Éditions Fayard*. Pagination multiple. ISBN : 2-2136-7948-7.
21. **HIROYUKI SASAKI et YASUHISA MATSUI.** 2008. Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nature Reviews Genetics*. 9(1):129-140.
22. **HUDSON QJ, KULINSKI TM, HUETTER SP et al.** 2010. Genomic imprinting mechanisms in embryonic and extraembryonic mouse tissues. *Heredity* 105:45-56.
23. **JONES PA, ARCHER TK, STEPHEN B et al.** 2008. Moving AHEAD with an international human epigenome project. *Nature*. 454:711-715.
24. **LEE JT et BARTOLOMEI MS.** 2013. X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell*. 152:1308-1323.
25. **MORRIS KV et MATTICK JS.** 2014. The rise of regulatory RNA. *Nature reviews Genetics* 15:423-437.
26. **PÀLDI A.** 2018. L'épigénétique ou la nouvelle ère de l'hérédité. *Éditions le Pommier*. Pagination multiple. ISBN:2-7465-1684-5.
27. **PETERS J.** 2014. The role of genomic imprinting in biology and disease: an expanding view. *Nature reviews Genetics*. 15:517-53.
28. **REIK W, DEAN W, WALTER J et al.** 2001. Epigenetic reprogramming in mammalian development. *Science*. 293:1089-93.
29. **SONG CX et HE C.** 2011. The hunt for 5-hydroxymethylcytosine: the sixth base. *Epigenomics*. 3:521-523.
30. **YOUNG ZK.** 2014. Altered Histone Modifications in Gliomas. *Brain Tumor Res Treat*. 2(1):7-21.