

SV-F-3 poly 02
Texte du cours**Partie F : Génomique structurale et fonctionnelle**
Chapitre 3 : Le contrôle de l'expression du génome

Le phénotype correspond à l'ensemble des caractéristiques observables. Il peut se définir à différentes échelles :

- De l'organisme, dans le cas d'un organisme pluricellulaire ;
- D'une cellule : dans le cas d'un organisme pluricellulaire, les phénotypes cellulaires sont différenciés, permettant ainsi une spécialisation ;
- Des protéines et ARN à rôle fonctionnel (ARNr, ARNt...) présentes dans la cellule ou l'organisme.

Le phénotype est le résultat de l'expression du génome. La diversité des phénotypes dépend de la diversité des génomes, mais aussi des régulations de l'expression du génome.

I. Du génotype au phénotype, des processus régulés

A. Du phénotype moléculaire au phénotype cellulaire

1. Le phénotype moléculaire

L'expression génétique conduit à la présence d'ARN et de protéines. Les protéines peuvent être intracellulaires, au niveau de la membrane plasmique ou sécrétées.

Les protéines présentes sont responsables des diverses propriétés de la cellule :

- Les **enzymes**, en catalysant certaines réactions, sont responsables du métabolisme cellulaire.
- Les protéines **structurales** (ex : protéines du cytosquelette, protéines d'adhérence, etc.) sont responsables de la forme de la cellule, de son ultrastructure, et de ses interactions avec les autres cellules ou la matrice extracellulaire. Ces caractères peuvent inclure, ou pas, une notion dynamique (mobilité cellulaire, etc.).
- Certaines protéines jouent des rôles dans les **échanges** entre les compartiments.
- Les **récepteurs** permettent la réception de messages spécifiques.

Le phénotype moléculaire est aussi constitué d'ARN fonctionnels :

- ARNr et ARNt nécessaires à la traduction.
- Petits ARN, intervenant dans divers mécanismes (excision-épissage, adressage au REG, régulations, etc.).

2. Transcriptome et protéome

L'ensemble des ARN d'une cellule constitue son **transcriptome**. Les ARN fonctionnels et les protéines présentes dans une cellule dépendent directement de son transcriptome : Le phénotype moléculaire découle donc du transcriptome, et donc de l'expression génétique de la cellule.

L'ensemble des protéines de la cellule constitue son **protéome**. Ce protéome dépend donc de l'expression génétique, mais aussi de la maturation et de l'adressage des protéines.

Par exemple, les enzymes digestives nécessitent un clivage protéolytique afin d'être actives : cette étape de leur maturation est donc indispensable à leur fonctionnalité, et donc à la réalisation du phénotype moléculaire (digestion enzymatique) et au niveau de l'organisme (capacité de digérer les aliments).

3. Diversité des phénotypes

Une cellule différenciée possède un protéome différent d'une autre cellule de l'organisme, ce qui conduit à un phénotype différent à l'échelle cellulaire.

Par exemple, chez l'Homme, une cellule hépatique est capable de réaliser le cycle de l'urée, permettant la formation d'urée à partir d'ammoniaque. Les autres cellules humaines n'expriment pas les gènes codant les enzymes de ce cycle, et ne sont donc pas capables de synthétiser de l'urée. De même, un poisson téléostéen, ammoniotélique, ne possède pas dans son génome les gènes correspondants, et n'est donc pas non plus capable d'uréotélie.

De même, les différentes cellules d'un organisme pluricellulaire n'expriment pas toutes les mêmes récepteurs membranaires : Elles ne sont pas ainsi sensibles aux mêmes hormones, par exemple.

Par exemple, l'adrénaline induit le resserrement de certains vaisseaux sanguins chez l'Homme (vasoconstriction), du fait de la présence de récepteurs $\alpha 1$ à l'adrénaline, alors qu'elle a un effet contraire (vasodilatateur) sur d'autres vaisseaux sanguins, exprimant le récepteur $\beta 2$ de l'adrénaline.

Etude d'un exemple : Modification du protéome et transition épithélium-mésenchyme. Voir figure 1.

B. Différents facteurs contrôlent l'expression génétique

L'expression d'un gène peut être sous le **contrôle de nombreux facteurs**, ce qui permet d'aboutir à un contrôle fin de l'expression génétique, en fonction des cellules et des conditions dans lesquelles elles se trouvent.

1. Communication intercellulaire et expression génétique

Une cellule est capable de réagir à la présence de molécules de signalisation extracellulaire, grâce à la présence de **récepteurs**. Dans le cas où ces molécules informatives sont hydrophiles, les récepteurs sont des protéines transmembranaires de la membrane plasmique.

Les molécules informatives peuvent être des hormones, mais aussi des molécules intervenant dans une communication interspécifique (Ex : communication par flavonoïdes et facteurs Nod entre les bactéries rhizobium et les tissus racinaires des Fabacées).

La fixation de la molécule informative sur son récepteur active une voie de transduction, qui aboutit à l'activation de facteurs de transcription, et ainsi à une modulation de l'expression génétique.

Dans le cas d'hormones liposolubles, comme les hormones stéroïdes, le récepteur est une protéine intracellulaire, qui agit comme facteur de transcription suite à la fixation de l'hormone.

Dans l'exemple étudié précédemment (figure 1), la modification de l'expression génétique des cellules, pendant une transition épithélium-mésenchyme, est induite par la présence de protéines TGFβ dans le milieu.

2. Facteurs internes et expression génétique

Diverses protéines sont capables de lier à l'ADN et de modifier l'expression génétique : Elles constituent un contrôle interne de l'expression.

C'est le cas des **facteurs de transcription**, mais aussi de protéines de **remodelage de la chromatine**.

Il existe ainsi des systèmes de contrôle génétique en cascade : un premier gène produit une protéine, qui régule l'expression d'autres gènes, dont les protéines régulent elles-mêmes d'autres gènes.

De même, un gène peut plusieurs séquences régulatrices différentes au niveau de sa région promotrices : L'expression du gène peut ainsi être sous le contrôle de plusieurs autres gènes, ou d'une combinaison de paramètres intra et extracellulaires. Cette multiplicité des contrôles permet d'aboutir à une régulation très fine de l'expression génétique.

3. Effet de l'environnement

Les conditions environnementales peuvent moduler l'expression génétique.

Chez la bactérie *E. coli*, par exemple, la présence de lactose et/ou de glucose dans le milieu permet une modulation de l'expression de **l'opéron lactose**. L'expression des gènes permettant le métabolisme du glucose est ainsi maximale :

- En présence de lactose (qui permet l'inhibition de la protéine répresseur) : la transcription est possible ;
- En absence de glucose (qui permet la stimulation du complexe AMPc-CAP) : la transcription est stimulée.

Un exemple d'intervention de l'environnement est le processus de **vernalisation** chez certaines Angiospermes : Les bourgeons floraux formé à l'automne ne sont pas capables de germer, ce qui évite une floraison précoce lors d'un redoux hivernal. La capacité à fleurir n'est acquise que par suite d'une exposition prolongée au froid.

La vernalisation est, ainsi, la modification de l'expression du génome de la plante suite à une exposition prolongée au froid. Elle peut inhiber l'expression de gènes, tout comme elle peut stimuler l'expression d'autres gènes. Parmi les gènes dont l'expression est inhibée, se trouve le gène **FLC** (cf. II.C ci-après).

C. Différentes méthodes d'étude de l'expression génétique

Voir TP F3

L'étude de l'expression génétique peut mettre en jeu diverses méthodes utilisant des techniques de biologie moléculaire.

1. Différents types de modification du génotype

Le lien entre le génotype et le phénotype peut être étudié en utilisant les techniques de modification de l'information génétique d'une cellule ou d'un organisme, et en observant le phénotype induit par la modification du génotype ainsi produite :

- Transgénèse : ajout d'une séquence nucléotidique dans le génome ; il s'agit donc d'un ajout au génotype. L'intégration du transgène se réalise de manière aléatoire, ce qui peut jouer sur son expression.
- Mutagenèse aléatoire : mutation induite de manière non dirigée, par exemple grâce à des agents mutagènes. L'altération du génotype n'est pas prévisible, et doit donc être déterminée afin de pouvoir la relier au phénotype observé.

- Mutagenèse dirigée : mutation induite de manière contrôlée. La démarche est ici de modifier une séquence précise afin d'analyser le rôle de cette séquence dans la réalisation du phénotype.
- Recombinaison homologue : cette technique permet de réaliser une mutagenèse dirigée, consistant à remplacer un fragment d'ADN génomique par une construction génétique. La construction peut être un gène rapporteur, permettant d'analyser le profil d'expression du gène remplacé par le gène rapporteur. La construction peut aussi permettre de remplacer une copie sauvage d'un gène par une copie légèrement modifiée, afin d'étudier les conséquences éventuelles de cette modification.

2. Etude du génome, du transcriptome et du protéome

Voir aussi le TP F1

Le génome peut être étudié par différentes techniques, dont le séquençage, complet ou partiel. La présence de séquences connues peut aussi être évaluée par PCR, ainsi que par électrophorèse d'ADN et Southern blot. Des hybridations *in situ* peuvent être réalisées sur des chromosomes afin de localiser des séquences précises (technique de FISH).

Le transcriptome peut être étudié par RT-PCR (et donc aussi qRT-PCR), par puces à ADN, électrophorèse d'ARN et Northern blot, etc. L'hybridation *in situ* permet d'analyser les patrons d'expression de gènes.

Ces études au niveau de l'ADN et des ARN mettent souvent en jeu des sondes (ADN ou ARN), qui permettent des interactions avec des séquences spécifiques.

L'étude des protéomes peut être réalisée par des électrophorèses de protéines et Western blot. L'identification de protéines spécifiques peut être réalisée grâce à l'utilisation d'anticorps.

Un protéome peut être étudié de manière globale par diverses techniques.

Par exemple, il est possible d'extraire l'ensemble des protéines d'une cellule ou d'un tissu, de les fragmenter, puis de séparer les fragments obtenus par spectrométrie de masse (séparation en fonction du rapport charge/masse) : la comparaison avec des bases de données permet alors d'identifier les protéines présentes.

II. Le contrôle transcriptionnel de l'expression du génome

Dans une cellule donnée, seule une partie du génome s'exprime. Cette expression différentielle peut être contrôlée au niveau de la transcription (contrôle transcriptionnel) ou après la réalisation de la transcription (contrôle post-transcriptionnel).

A. Facteurs de transcription et expression génétique

Le contrôle de la transcription peut mettre en jeu des facteurs de transcription spécifiques. Ces protéines agissent en modulant le recrutement de l'ARN polymérase :

- L'ADN présente, à plus ou moins grande proximité de la séquence transcrite, une séquence consensus reconnue par le facteur de transcription. On parle d'élément cis-régulateur, car la séquence a une action sur la transcription d'une autre séquence située à proximité sur l'ADN. En fonction de l'effet sur la transcription, on parle d'enhancer (stimulant la transcription) ou de silencer (inhibant la transcription).
- Le facteur de transcription se lit à l'ADN de manière spécifique et module le recrutement de l'ARN polymérase. On parle d'élément trans-régulateur, car le gène codant la protéine peut être situé n'importe où dans le génome (pas de proximité génomique, au contraire des éléments cis-régulateurs).

Les facteurs de transcription sont des protéines caractérisées par la présence d'un domaine de liaison à l'ADN. D'un point de vue structural, les domaines de liaison à l'ADN correspondent à quelques structures tridimensionnelles usuelles, dont, par exemple, le motif bHLH.

Le motif bHLH correspond à la succession de deux hélices α , séparées par une boucle : On parle de motif hélice-boucle-hélice (*basic helix-loop-helix*). Ce motif protéique « s'insère » dans le grand sillon de l'ADN, ce qui permet une association avec un ADN de séquence spécifique. On retrouve ce motif en particulier dans les facteurs de transcription myogénique, qui interviennent lors de la différenciation des cellules musculaires au cours du développement embryonnaire.

B. Intervention de facteurs de remodelage de la chromatine

La réalisation de la transcription nécessite que les facteurs de transcription et l'ARN polymérase aient accès à l'ADN. Les eucaryotes possèdent ainsi des **complexes de remodelage de la chromatine**, comme SWI/SNF.

Ces complexes sont capables :

- De faire **glisser** les nucléosomes le long de la molécule d'ADN.
- De **retirer** ou **ajouter** des nucléosomes en un site.

- De **modifier** les **interactions** entre l'ADN et les nucléosomes, pour contrôler plus globalement la compaction de la chromatine

Ces modifications de l'état de la chromatine sont transmissibles (au moins partiellement) aux cellules filles lors des divisions cellulaires. De telles modifications correspondent à des modifications épigénétiques.

C. L'épigénétique : modifications chimiques de la chromatine

1. Définition

Le terme **d'épigénétique** correspond aux altérations de la chromatine, qui modifient l'expression du génome, et peuvent se transmettre au travers de divisions cellulaires.

L'épigénétique correspond donc à des modifications moléculaires du génome, jouant sur son expression.

De manière très générale, cela fait donc partie des mécanismes d'expression différentielle des gènes au sein d'un même organisme, conduisant à la différenciation cellulaire.

La particularité ici réside dans la possibilité de transmission de ce profil d'expression d'une génération cellulaire à une autre.

2. Epigénome et modifications épigénétiques

Le passage de la chromatine entre un état condensé (hétérochromatine ; pas de transcription) et un état décondensé (euchromatine ; transcription possible) peut être contrôlé grâce à des modifications chimiques de la chromatine. Ces modifications peuvent toucher l'ADN ou les histones.

a. La méthylation de l'ADN

Un nucléotide sur 100, en moyenne, peut porter un groupement **méthyle** supplémentaire, toujours fixé à une **cytosine**. (-CH₃ ajouté)

C'est un facteur important conduisant à l'**inactivation des gènes** présentant des cytosines méthylées. La méthylation de l'ADN entraîne en effet sa condensation sous forme d'hétérochromatine.

Cette méthylation est réalisée par des enzymes spécifiques (méthyl-transférases). Cette inactivation du gène est réversible (grâce à une déméthylation).

La méthylation est reproduite à l'identique lors de la réplication de l'ADN, ce qui permet de la maintenir au fil des divisions cellulaires.

Les séquences méthylées ne sont pas réparties au hasard et se trouvent souvent dans les **ilots CpG** impliqués dans la régulation des gènes.

b. Modifications des histones et état de la chromatine

La partie N-terminale des histones H2A, H2B, H3 et H4 s'étendent hors du cœur nucléosomique. Les acides aminés de ces chaînes peuvent être modifiés chimiquement de manière réversible, par méthylation, acétylation, phosphorylation, ubiquitinylation (ajout d'ubiquitine, petite protéine de 76 acides aminés), etc.

Ces modifications chimiques peuvent être mises en place ou retirées grâce à diverses enzymes.

L'ensemble des modifications épigénétiques des histones constituent le code histone : la chromatine est condensée ou décondensée en fonction de ces modifications chimiques.

Exemple : marque épigénétique sur l'histone H3 au niveau de promoteurs.

- H3K9ac = acétylation de la Lysine 9 de l'histone 3 → marque activatrice
- H3K9me3 = triple méthylation de la Lysine 9 de l'histone 3 → marque inhibitrice ; chromatine condensée (hétérochromatine)

3. Un exemple de gène contrôlé par épigénétique : FLC et le contrôle de la floraison des Angiospermes

La floraison de certaines plantes (blé d'hiver, betterave sucrière, certaines variétés d'*Arabidopsis thaliana*) nécessite impérativement une exposition prolongée au froid lors des stades précoces du développement ; ce processus est la **vernalisation**. Durant cette période, des modifications affectent notamment les profils des histones, remodelant la chromatine et inhibant l'expression du gène **FLOWERING LOCUS C (FLC)** qui code un facteur de transcription empêchant la floraison.

Le processus de vernalisation est une adaptation aux climats avec un hiver marqué.

Chez *Arabidopsis*, la répression de FLC augmente quantitativement avec la durée du traitement par le froid, elle est mitotiquement stable même après le retour de températures plus clémentes mais se réinitialise à la génération suivante.

Le déroulement de la vernalisation se produit en 4 phases.

Avant la vernalisation, l'expression de FLC est élevée dans la plante à l'état végétatif. La chromatine du locus FLC est alors associée à des marques histones en lien avec une transcription active (notamment des méthylations d'histones).

La répression de FLC est induite par le froid **au cours de la vernalisation**. Elle résulte de l'association de 2 complexes protéiques (PRC2 et PHD) activés par le froid qui permet d'ajouter sur une partie du locus FLC une marque réprimant la transcription et de retirer la marque activatrice.

De plus, dès le début de la période de froid, un ARN non-codant antisens de FLC est produit et participe à la répression de FLC.

Après la vernalisation, lors du retour à des températures ambiantes et jusqu'à la floraison, le complexe PRC2-PHD étend son action (ajout d'une marque répressive) sur tout le locus FLC. Cette action stabilise alors la chromatine dans un état répressif et stable au cours des mitoses.

Lors de la **fécondation**, l'expression de FLC semble consécutive à l'action d'un facteur de transcription embryonnaire capable d'activer la transcription de gènes portant des marques répressives de la chromatine. Puis des facteurs de transcription embryonnaires stabilisent l'expression de FLC à un niveau élevé, ce qui empêche une floraison trop précoce.

III. Le contrôle post-transcriptionnel

A. L'existence d'un contrôle post-transcriptionnel de l'expression génétique

Le contrôle post-transcriptionnel de l'expression génétique correspond aux processus permettant de moduler l'expression génétique, au-delà de la transcription.

Ce contrôle comprend l'épissage alternatif (permettant l'obtention de transcrits matures différents en fonction des types cellulaires), ou encore le contrôle de la traduction.

Le contrôle post-transcriptionnel comprend aussi le contrôle des ARNm cytosoliques par interférence ARN.

B. Interférence ARN et expression génétique

L'interférence ARN est un mécanisme de contrôle des ARNm cytosoliques, grâce à de petits ARN non codants inhibiteurs.

Il s'agit en particulier des **siARN** (petits ARN interférents) et des **miARN** (micro ARN interférents), dont le fonctionnement est assez similaire. Ces petits ARN interférents ont une longueur de 21 à 24 nucléotides, et sont codés par des gènes spécifiques (par exemple, le génome humain contiendrait environ 6000 gènes de miARN).

La transcription du gène de miARN ou siARN conduit à la formation d'un ARN double brin. Après action de plusieurs enzymes, dont la ribonucléase **DICER**, on obtient un petit ARN simple brin, associés à des protéines (dont **Argonaute**) au sein d'un complexe ribonucléoprotéique : **RISC**.

Le complexe RISC s'associe de manière spécifique à des ARNm cibles.

Cette association peut induire la dégradation de l'ARNm, ou à l'inhibition de la traduction en empêchant l'action du ribosome.