# 

|  |  |
| --- | --- |
| BLOC SANTE | Auteur : RonéoAI |
| UE-1 Constitution et transformation de la matière | Date : 2025-08-17 |

# 

# **Architecture du génome humain**

# **I. Présentation du module et plan du cours**

Ce module porte sur l'organisation du génome humain et sur les méthodes biotechnologiques permettant de le caractériser et de l'analyser. Il vise à donner une vision intégrée de la composition, de l'architecture et de la dynamique des génomes, ainsi que des conséquences fonctionnelles et médicales des variations génomiques.

Le plan du cours présente successivement une introduction aux concepts évolutifs et aux organismes modèles, les notions fondamentales de génétique moléculaire, la description du génome humain (nucléaire et mitochondrial), le projet de séquençage, l'architecture globale du génome, la description des éléments constitutifs (gènes codants et non codants, pseudogènes, éléments répétés), la comparaison intergénomique, les différents types de variations et leurs conséquences, la variabilité du génome humain et, enfin, les notions essentielles et mécanismes de l'évolution des génomes.

Les objectifs pédagogiques incluent la compréhension des éléments structuraux et fonctionnels du génome, la capacité à interpréter les résultats de cartographie et de séquençage, et l'aptitude à appréhender les mécanismes évolutifs et les types de variations qui façonnent la diversité génomique. Le cours associera présentations conceptuelles et introduction aux outils d'exploration des données génomiques.

# **II. Introduction au contexte évolutif et organismes modèles**

L'étude du génome s'inscrit dans un cadre évolutif où les concepts de temps, d'adaptation et de sélection expliquent la conservation et la divergence des caractères moléculaires et structuraux entre espèces. Comprendre ces notions est indispensable pour interpréter la distribution des gènes, des éléments répétitifs et des variations observées au sein des populations et entre espèces.

L'approche comparative et l'utilisation d'organismes modèles constituent des outils méthodologiques majeurs pour déduire des fonctions biologiques, tester des hypothèses mécanistiques et reconstituer des histoires évolutives à partir d'observations expérimentales et de données génomiques.

## **1. Arbre de la vie et concept de LUCA**

L'arbre de la vie représente les relations phylogénétiques entre les êtres vivants et illustre comment la variation accumulée au fil du temps a conduit à la diversification des lignées. Il fournit le cadre conceptuel pour comprendre la parenté entre espèces et la distribution des caractères hérités d'ancêtres communs.

Le concept de LUCA (Last Universal Common Ancestor) renvoie à l'ancêtre commun le plus récent à l'origine de l'ensemble des organismes actuels ; il constitue une reconstruction phylogénétique permettant d'identifier des traits ancestraux partagés au niveau moléculaire et métabolique, et d'envisager les étapes initiales de l'évolution de la vie.

## **2. Arbre des eucaryotes et choix d'organismes modèles**

L'arbre des eucaryotes organise les grands groupes (levures, plantes, protistes, métazoaires, etc.) et sert de référence pour sélectionner des organismes modèles représentatifs des processus biologiques étudiés. La diversité des eucaryotes permet de comparer des mécanismes conservés et des innovations spécifiques à certaines lignées.

Le choix d'un organisme modèle repose sur des critères pratiques et scientifiques : intérêts économique et technique, facilité d'hébergement et de manipulation, durée du développement embryonnaire et du temps de génération, ainsi que le partage d'éléments moléculaires et cellulaires fondamentaux (machinerie de transcription, ADN, voies métaboliques) qui rendent les résultats transférables à d'autres systèmes.

# **III. Notions fondamentales et définitions clés**

Parmi les notions fondamentales figurent la définition de l'homologie, la distinction avec l'analogie, et les ramifications conceptuelles associées telles que l'orthologie et la paralogie. Ces concepts permettent d'interpréter les similitudes entre caractères anatomiques, moléculaires ou génétiques à la lumière de leur histoire évolutive.

La maîtrise de ces définitions et des méthodes de comparaison des séquences est essentielle pour inférer des fonctions, reconstituer des arbres phylogénétiques et comprendre l'origine des duplications et des divergences au sein des génomes.

## **1. Homologie: définition, orthologie, paralogie et méthodes de détection**

L'homologie désigne une origine évolutive commune pour un caractère observé dans plusieurs espèces ; il convient de la distinguer de l'analogie, qui traduit une similarité fonctionnelle apparue indépendamment par convergence évolutive. Au niveau génétique, on distingue deux cas fréquents : les gènes orthologues, issus d'une spéciation et souvent conservant une fonction similaire entre espèces, et les gènes paralogues, résultant d'une duplication et susceptibles d'évoluer vers des fonctions différentes.

La détection de l'homologie repose essentiellement sur l'alignement des séquences d'ADN ou de protéines et sur des mesures quantitatives telles que le pourcentage d'identité (résidus strictement identiques) et le pourcentage de similarité (résidus identiques ou physico‑chimiquement proches). Ces mesures fournissent des indications de proximité mais leur interprétation dépend de la famille protéique considérée ; des seuils empiriques (par exemple une identité souvent jugée significative au-delà d'un certain pourcentage) peuvent orienter l'analyse, sans constituer une règle absolue.

Plusieurs précautions méthodologiques sont nécessaires : un faible pourcentage d'identité n'exclut pas l'homologie, car des séquences peuvent diverger fortement tout en conservant une relation évolutive, et des similarités peuvent résulter de régions de faible complexité ou de convergence fonctionnelle. Par conséquent, l'interprétation doit intégrer le contexte évolutif et analytique.

Des approches complémentaires renforcent les inférences d'homologie : alignements multiples, analyses phylogénétiques, conservation de la syntenie, études structurales, et méthodes statistiques basées sur des profils ou modèles de Markov cachés (HMM) sont couramment employées pour distinguer relations homologues et similitudes fortuites.

## **2. Biologie du gène, dogme central et omiques**

Le gène est défini comme un segment d'ADN utilisé comme matrice pour synthétiser un ARN simple brin au cours de la transcription. La distinction fondamentale entre ARN codants (ARNm, porteurs de l'information pour la synthèse protéique) et ARN non codants (ARNt, ARNr, miARN, snARN, snoARN, lncARN, etc.) souligne la diversité des fonctions moléculaires qui ne se résument pas à la production de protéines.

Le génome constitue une référence structurelle et fonctionnelle, mais il doit être considéré en relation avec d'autres couches d'information : le transcriptome (ensemble des ARN exprimés), le protéome (ensemble des protéines), le méthylome (marquages épigénétiques) et d'autres jeux de données omiques. L'intégration de ces niveaux permet de relier la séquence à l'expression, à la régulation et aux phénotypes observables.

# **IV. Le génome humain: composition et organisation générale**

Le génome humain se compose d'un génome nucléaire, organisé en chromosomes linéaires abondamment compactés et régulés, et d'un génome mitochondrial, de structure circulaire et d'une taille nettement plus réduite. La connaissance de la composition et de l'organisation de ces deux compartiments est essentielle pour comprendre l'héritage et la variabilité génétique.

L'organisation chromosomique varie selon l'état cellulaire (par exemple condensation lors de la métaphase versus organisation plus lâche en interphase) et reflète à la fois des niveaux d'expression différents et des structures répétitives qui posent des défis techniques pour l'assemblage et l'annotation.

## **1. Composition générale du génome: haploïde et diploïde**

Chez l'Homo sapiens, le génome nucléaire est réparti en 22 paires d'autosomes et une paire de chromosomes sexuels (X et Y) ; la plupart des cellules somatiques sont diploïdes et contiennent donc deux exemplaires (origines maternelle et paternelle) de chaque chromosome. La condensation des chromosomes pendant la mitose contraste avec leur organisation en territoires chromosomiques plus diffus en interphase.

Le génome mitochondrial est une molécule circulaire compacte, présente en de nombreuses copies par cellule et transmise le plus souvent selon un mode maternel. Sa compacité et son mode d'hérédité entraînent des dynamiques de variation et d'expression distinctes de celles du génome nucléaire.

Les notions de génome haploïde (un seul jeu chromosomique, par exemple dans les gamètes) et diploïde (deux jeux, caractéristiques des cellules somatiques) ont des conséquences immédiates pour la génétique mendélienne, la cartographie et l'interprétation des variations alléliques chez l'individu et au sein des populations.

# **V. Détermination et séquençage du génome humain**

La détermination de la séquence du génome humain a été conduite dans le cadre d'un projet international d'envergure lancé au début des années 1990, d'une durée d'environ quinze ans et d'un coût global évalué à plusieurs milliards de dollars. Un objectif prioritaire de ce projet était la production d'une séquence de référence pour la région euchromatique, riche en gènes et plus accessible aux approches d'assemblage initiales.

Ce travail visait à fournir une ressource de base pour la découverte des gènes, l'analyse des variations interindividuelles et le développement d'outils diagnostics et de recherche. Il a nécessité le développement et la combinaison de stratégies de cartographie, de séquençage et d'assemblage adaptées aux contraintes liées aux régions répétées et à la complexité chromosomique.

## **1. Historique, objectifs et ciblage de l’assemblage initial**

L'assemblage initial du génome humain a ciblé prioritairement l'euchromatine en raison de son contenu en gènes et de sa plus grande tractabilité pour l'assemblage séquentiel. Cette stratégie permettait d'établir rapidement une référence fonctionnelle pour la majorité des régions codantes et régulatrices.

La mise en place d'une carte de référence a nécessité des étapes successives d'amélioration afin de traiter ensuite les zones plus riches en répétitions (hétérochromatine), qui exigent des approches méthodologiques et technologiques spécifiques pour être résolues correctement.

## **2. Cartographie génétique basée sur fréquences de recombinaison**

La cartographie génétique s'appuie sur les fréquences de recombinaison observées lors de la méiose et s'exprime en centimorgans (cM). Les distances en cM reflètent la probabilité d'un événement de recombinaison entre deux loci et permettent d'ordonner et d'estimer des distances relatives le long des chromosomes.

Les cartes génétiques sont un outil essentiel pour localiser des gènes responsables de phénotypes, pour concevoir des croisements expérimentaux et pour ancrer des fragments lors de l'assemblage du génome ; toutefois, elles n'indiquent pas la distance physique exacte en nucléotides et sont soumises aux variations locales du taux de recombinaison.

## **3. Cartographie physique: distances en kilobases et mégabases**

La cartographie physique mesure les distances moléculaires en paires de bases et s'exprime ordinairement en kilobases (kb) ou en mégabases (Mb). Elle renseigne sur la distance réelle entre séquences et constitue la base de l'assemblage physique et de l'annotation structurale du génome.

La combinaison des cartes physiques et génétiques fournit une vision complémentaire de l'organisation génomique : la carte physique précise l'espacement réel en nucléotides, tandis que la carte génétique informe sur le comportement des régions au cours de la recombinaison, information utile pour étudier la structure chromosomique et les mécanismes évolutifs.