

Chapitre 1

Structure et diversité du génome humain

I. Préambule

Le projet de séquençage du génome humain est certainement, après la conquête de la lune, l'une des aventures humaines récentes les plus passionnantes et dont les retombées scientifiques et industrielles sont majeures. Dans le cadre de ce chapitre nous ferons d'abord un bref rappel historique afin de comprendre l'évolution des concepts et connaissances ayant finalement abouti à la réalisation de ce projet de collaboration international. Nous aborderons le génome dans sa globalité et découvrirons qu'il est à la fois nucléaire et mitochondrial, et définirons quelques termes essentiels en génétique humaine. Nous découvrirons la plasticité du génome qui est en constante évolution sous l'influence de forces antagonistes : d'une part les recombinaisons et les mutations qui ont tendance à le modifier et d'autre part, la sélection naturelle qui a tendance à moduler ces évolutions afin de s'adapter au milieu.

Nous aborderons ensuite les étapes de cartographie génétique et physique, laborieux processus sans lesquels le génome n'aurait pas pu être assemblé. Enfin nous ferons le point sur les connaissances liées à la réalisation de ce projet et les découvertes qui y sont associées.

Enfin nous verrons comment nous sommes passés en quelques années du projet de séquençage « du génome humain » qui a coûté près de 3 milliards de dollars au séquençage « du génome individuel » pour quelques milliers de dollars ouvrant ainsi la porte à la médecine personnalisée.

II. Repères historiques

Bien avant la découverte de l'ADN vecteur de transmission de l'information génétique, l'Homme a tenté d'expliquer la transmission de caractères entre les parents et les enfants par différentes théories. Ainsi, de l'antiquité au XIX^e siècle ont été proposées :

Le **préformisme** qui correspond à la doctrine de la préexistence et de l'emboîtement des germes : Dieu étant le créateur de toute chose, il a, dès le commencement, créé tous les animaux, toutes les plantes et tous les Hommes amenés à peupler le monde jusqu'à la fin des temps. Les enfants à naître existent donc déjà, minuscules mais totalement formés, dans leurs géniteurs ; ces enfants eux-mêmes abritent, dans cet état minuscule, leurs enfants et, par emboîtements successifs, toutes les générations suivantes...

La découverte du spermatozoïde, par Antoni van Leeuwenhoek en 1677, vient populariser la théorie alternative de l'**animalculisme** selon laquelle l'embryon préexiste dans le spermatozoïde, l'œuf féminin servant uniquement à le nourrir. Le **patroclinisme** est une théorie assez voisine puisqu'elle considère que les caractères de l'enfant sont transmis par la lignée paternelle uniquement. À l'inverse, pour l'**ovisme** l'embryon est

préformé dans la femelle, le sperme apporte juste une « *aura seminalis* » qui est une essence vitale qui animera l'embryon. Avec l'avancée des connaissances scientifiques apparaît **l'épigénisme** qui stipule que les organes apparaissent progressivement au cours de la croissance embryonnaire sous l'influence de forces extérieures. On considère traditionnellement que la théorie de l'épigénèse l'a emporté à la fin du XIX^e siècle avec l'émergence de l'embryologie expérimentale.

C'est en 1865 que Gregor Mendel jette les bases de la génétique moderne avec l'étude des petits pois et la découverte des allèles, des phénomènes de dominance entre allèles et de la notion d'hétérozygotie (cf. IV.1.a). Quelques années plus tard, en 1910 Thomas Morgan découvre la recombinaison méiotique et publie ses résultats qui sont les bases de la théorie chromosomique de l'hérédité. Enfin en 1913 est publiée la première carte génétique. Il faut attendre une quarantaine d'années pour qu'en 1953 James Watson, Francis Crick, Maurice Wilkins et Rosalind Franklin décryptent des clichés de diffraction et publient la structure en double hélice de l'ADN. Douze ans plus tard, en 1965, Jacques Monod, François Jacob et André Lwoff font une autre découverte clé qui leur vaudra également un prix Nobel. Ils découvrent ainsi les ARN (Acides RiboNucléiques) et surtout la régulation de l'expression génique chez les procaryotes. Autre date importante, 1973 puisque c'est à cette époque qu'est mis en place un moratoire sur le clonage et le génie génétique.

Plus proche de nous, les années 1980 correspondent au début des grands progrès du génie génétique avec notamment la découverte de marqueurs polymorphes, des techniques de RFLP (polymorphisme de fragments de restriction) et un peu plus tard des microsatellites (cf. IV.1.a), qui vont permettre la réalisation de projets de cartographie intensive.

Tout cela va être véritablement révolutionné par la découverte de la PCR (réaction de polymérisation en chaîne) par Kary Mullis et vers la fin des années 1980 se met en place une organisation internationale la « *HUman Genome Organization* » afin de colliger l'ensemble des données de séquençage.

Cependant, avant de pouvoir véritablement séquencer le génome humain, il fallait disposer de cartes génétiques de référence, et la France a joué un rôle majeur avec l'établissement des cartes dites du Généthon.

En 2000, Alain Fisher et son équipe, publient les premiers résultats positifs de thérapie génique sur l'Homme chez des enfants touchés par une déficience grave du système immunitaire, appelée Déficit Immunitaire Combiné Sévère lié au chromosome (X-SCID). Pour leur survie, ces enfants sont placés dès leur naissance dans une atmosphère complètement stérile sous une bulle de plastique. La correction du défaut génétique a ainsi permis à cette époque à deux enfants de réintégrer le domicile familial, libérés de tout traitement.

Enfin dernières dates importantes : 2001 quand est publiée la première carte du génome humain, qui est finalement considérée comme achevée en 2003.

III. Structure et plasticité du génome humain

1. Génome nucléaire et mitochondrial

Alors que savons-nous aujourd'hui du génome humain ?

Nous savons que l'information génétique est contenue dans l'ADN (Acide DésoxyriboNucléique) et qu'il existe plusieurs molécules d'ADN dans une cellule humaine. Ces molécules sont localisées soit dans le noyau, et peuvent être visualisées lors de la division cellulaire sous la forme de chromosomes, soit dans les mitochondries qui comportent chacune une molécule d'ADN circulaire.

a. Génome nucléaire

Le génome nucléaire comprend plusieurs molécules d'ADN. Chez l'Homme, nous en distinguons ainsi 46 qui se répartissent en 22 paires d'autosomes (chromosomes n'intervenant pas dans la détermination du sexe) et 2 chromosomes sexuels.

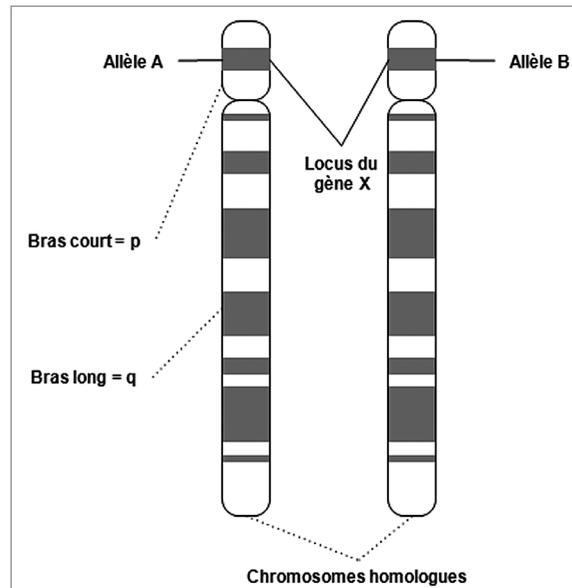
Si l'on tient compte de la taille du génome humain, mises bout à bout ces molécules ont une taille voisine de 2 m.

La question est donc : comment faire tenir ces molécules dans le noyau d'une cellule qui a une taille moyenne de 50 à 100 μm ? Tout simplement en le compactant. Tout d'abord en compactant la double hélice décrite par Watson et Crick qui a un diamètre de 2 nm sous la forme de nucléosomes : des histones (protéines basiques) permettent à l'ADN de s'enrouler en chapelets d'un diamètre de 11 nm ; puis en compactant ces nucléosomes sous la forme d'une fibre de chromatine d'un diamètre de 30 nm ; cette fibre de chromatine peut elle-même se condenser, faisant apparaître des structures de 300 nm, puis s'hyper-condenser en structures de 700 nm pour enfin aboutir à la forme la plus condensée qui correspond aux chromosomes métaphasiques qui ont un diamètre de l'ordre de 1400 nm. Il faut se souvenir qu'à cette étape de la division cellulaire la cellule vient de dupliquer son ADN et elle en contient donc l'équivalent de 4 m !

Un chromosome (Figure 1) est classiquement représenté avec une structure centrale le centromère ; un bras court ou bras p et un bras long aussi nommé bras q. Lors de la division cellulaire, chaque chromosome est dupliqué et ces deux molécules sœur

sont nommées chromatides ou chromatides sœurs, ces deux chromatides donneront naissance à 2 chromosomes dans les cellules filles.

Figure 1 : Représentation schématique d'un chromosome



b. Génome mitochondrial

Le génome mitochondrial est lui composé d'une molécule d'ADN circulaire de petite taille puisque celle-ci a été séquencée dès 1981 et comprend 16,568 paires de bases. Il contient 37 gènes qui codent pour 22 ARN de transfert, 2 ARN ribosomiques et 13 protéines membranaires appartenant aux différents complexes impliqués dans la production d'énergie sous forme d'ATP (adénosine triphosphate).

2. Définitions

Comme nous l'indiquons dans l'introduction, il est important, avant d'aller plus loin, de définir certains termes spécifiques à la génétique et indispensables à la bonne compréhension des notions qui seront abordées ensuite.

Ainsi, le terme **locus** désigne un emplacement précis et invariable sur un chromosome. Il peut faire référence à un gène ou à une séquence quelconque.

Un **allèle** désigne l'une des formes sous lesquelles un gène ou un segment d'ADN se présente après mutation. Vous entendrez souvent parler d'allèle sauvage et d'allèle muté. Chacun a une séquence différente que l'on peut étudier par de nombreuses techniques.

Le **génotype** correspond à l'ensemble des allèles d'un ou plusieurs locus dans un organisme. Nous pouvons le représenter sous la forme d'un tableau avec pour chaque locus, le ou les allèles dont est composé l'ADN d'un individu. Dans le cas du génotype, les allèles sont énoncés de façon brute. Ainsi nous ne savons pas si l'allèle

A du locus D1S100 est localisé sur la même molécule d'ADN que l'allèle C du locus D1S300 ou bien sur la molécule d'ADN porteuse de l'allèle D de ce même locus.

C'est l'**haplotype** qui permet de décrire l'arrangement linéaire ordonné des allèles sur un chromosome. Ainsi, grâce à l'haplotype, nous pouvons décrire précisément les 2 chromosomes homologues 1 et ainsi savoir quels haplotypes correspondent au génotype présenté dans le tableau 1.

Locus	Génotype
D1S100	A, B
D1S200	C, C
D1S300	C, D
D1S400	A, A

Comme vous le voyez, au locus D1S200, l'individu présente 2 allèles identiques nommés C. Les 2 chromosomes sont donc porteurs de la même séquence à ce locus. Il en va de même pour le locus D1S400 avec l'allèle A. En revanche, lorsque nous regardons les locus D1S100 et D1S300, cela ne va pas de soi. Pour le premier, le génotype est A, B. Les 2 chromosomes ont donc 2 allèles différents. Il en est de même pour le locus D1S300 avec les allèles C et D. Nous avons donc 2 possibilités : soit l'un des chromosomes possède pour ces 4 locus les allèles A, C, C et A et l'autre B, C, D et A ; soit ces 2 chromosomes possèdent les combinaisons A, C, D et A et B, C, C et A. L'analyse des ascendants ou des descendants permet de trancher.

La détermination des haplotypes revêt un intérêt majeur en génétique afin de suivre au cours des générations la transmission de caractères liés à un locus donné en observant quel chromosome a été reçu de chaque parent.

3. Plasticité du génome humain ; recombinaisons et mutations

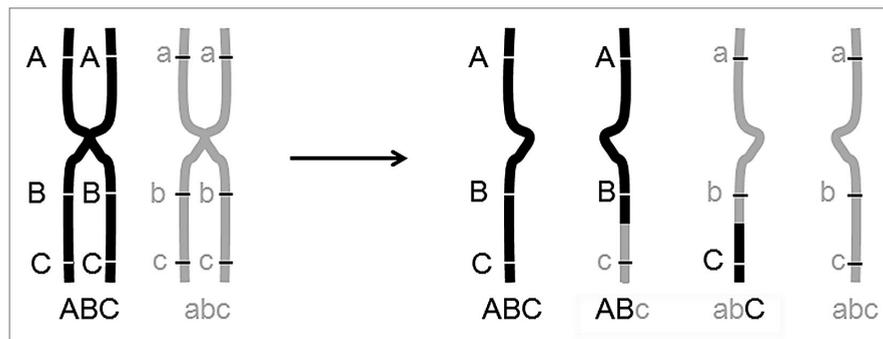
Nous pouvons maintenant aborder l'évolution du génome humain et notamment les recombinaisons et les mutations.

Il est logique de concevoir que le génome doit être suffisamment stable pour que les espèces puissent exister mais également suffisamment dynamique pour qu'il y ait une évolution possible. L'un des moteurs de cette évolution est la capacité des molécules d'ADN à recombiner par un phénomène dit de **recombinaison** ou « *crossing over* ». Ce phénomène correspond à l'échange de matériel génétique entre les chromatides sœurs pendant la prophase de la 1^{re} division méiotique. Pour que le caractère soit transmis à la descendance, il faut en effet toucher la lignée germinale et non les tissus somatiques.

Sur la Figure 2, vous reconnaissez en noir et gris une paire de chromosomes homologues avec leurs chromatides sœurs. Nous avons choisi un exemple simple avec 3 locus A, B et C possédant 2 allèles, l'un en minuscule et l'autre en majuscule. S'il se produit un croisement de 2 chromatides (une rouge et une bleue) entre les locus B et C, il peut alors y avoir cassure puis recollement des molécules d'ADN. Nous

aboutissons alors à 2 chromatides nouvelles qui ont échangé du matériel génétique. L'haplotype sera différent. Alors que nous attendions 2 haplotypes ABC et abc, nous avons maintenant 4 haplotypes ABC, abc, mais aussi ABc et abC. Dans le cas présent les chromatides résultant de cette recombinaison sont de même taille. Dans d'autres situations, nous pouvons observer une recombinaison inégale qui aboutit à une molécule d'ADN de taille supérieure, l'autre étant de taille inférieure. Ces phénomènes de recombinaison jouent un rôle majeur dans l'évolution du génome.

Figure 2 : Phénomène de recombinaison homologue



Le second mécanisme à l'origine de la diversité du génome correspond aux **mutations** (cf. Chapitre 2).

Les mutations sont souvent associées aux maladies génétiques mais ce n'est pas toujours le cas. Elles peuvent en effet être associées à de nombreux caractères phénotypiques variables tels que la taille ou la couleur des yeux.

4. Complexité du génome humain

Nous venons de citer les phénomènes qui jouent un rôle dans la plasticité du génome : recombinaisons et mutations, il est maintenant logique de regarder de plus près la structure du génome humain.

L'une des caractéristiques physiques fondamentales de la double hélice d'ADN est sa capacité de dissociation à chaud (dénaturation) et de réassociation spontanée lors du refroidissement (renaturation). En effet la chaleur va entraîner la rupture des liaisons hydrogènes qui maintiennent la structure en double hélice. Si les simples brins formés sont laissés tels quels, le refroidissement progressif va permettre la renaturation de la double hélice. Historiquement, la connaissance des génomes a d'abord débuté par celle des procaryotes (organismes unicellulaires dépourvus de noyau). Ces organismes possèdent un génome simple composé presque uniquement de gènes. Ainsi plus l'organisme est complexe, plus la taille de son génome est importante. La Figure 3 présente la courbe de vitesse de renaturation de l'ADN d'un procaryote type et du génome humain. Nous constatons que ces génomes se comportent de façon

très différente. En effet, Pour un procaryote, l'ensemble des séquences d'ADN va posséder une vitesse homogène de renaturation que nous pouvons arbitrairement définir fixer à 1. À l'inverse, pour l'ADN humain, la courbe a un profil qui fait apparaître 3 types de séquences :

- Un premier type (1) ayant une vitesse de réassociation rapide = 10^{-2} ;
- Un second type (2) ayant une vitesse de réassociation moyenne = 1 ;
- Et un troisième type (3) ayant une vitesse de réassociation lente = 10^2 .

Il aura fallu de nombreuses recherches pour finalement comprendre à quoi ces différents types de séquence correspondent :

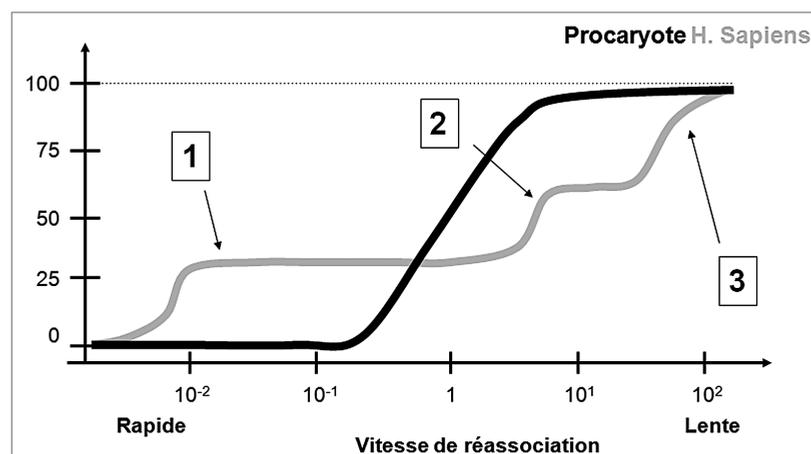
Le type « rapide » correspond à des séquences fortement répétées qui constituent 10 à 15 % du génome. Ce sont des séquences localisées au niveau des centromères, des télomères (extrémités des chromosomes) et des séquences de type mégasatellites, minisatellites et microsatellites, pouvant différer par la taille de la répétition ce qui explique leurs noms.

Le deuxième type ayant une vitesse de réassociation « moyenne » sont des séquences moyennement répétées. Elles correspondent à 20 à 40 % du génome. Elles sont constituées de transposons tels que les séquences SINE « *Short INterpersed Elements* » et LINE « *Long INterpersed Elements* », de rétrovirus endogènes ainsi que de quelques gènes codant des ARN ribosomiques, des ARN de transfert ainsi que les sous-unités 5S et 7S du ribosome.

Enfin le troisième type, qui a une vitesse de réassociation lente, correspond à des séquences uniques. Elles représentent environ 50 % du génome et contiennent la plupart des gènes codant pour des protéines ainsi qu'une majorité de séquences de nature inconnue.

Ces éléments montrent bien que la structure du génome humain est plus complexe que celle des procaryotes.

Figure 3 : Courbe de vitesse de renaturation de l'ADN



IV. De la cartographie à la diversité du génome humain

Ayant mis en évidence que le génome humain possédait une structure complexe et afin de mieux comprendre les différences associées à ce génome entre individus, il était important d'en déterminer sa structure le plus finement possible jusqu'à en réaliser le séquençage complet. Pour cela plusieurs étapes ont été nécessaires comme nous allons le voir.

1. La Cartographie

De nombreuses équipes de recherche ont décidé d'entreprendre la cartographie du génome humain afin de répondre à différentes questions notamment :

- Puisque le génome humain est plus complexe que celui des procaryotes, alors quelle est sa composition ?
- Contient-il plus de gènes que celui des autres espèces ?
- Comment peut-on expliquer les différences de niveau d'évolution entre les espèces en relation avec leur génome ?
- L'hypothèse selon laquelle plus une espèce est évoluée, plus elle a de gènes est-elle vraie ?

Bien sûr, en parallèle, cette approche devait également permettre d'identifier les gènes responsables des maladies génétiques et pour cela, au préalable, pouvoir placer les gènes sur les différents chromosomes, ce que nous appelons la cartographie. Il existe en réalité deux types de cartographie : la cartographie génétique et la cartographie physique.

a. Outils

Quels sont les outils utilisés pour réaliser cette cartographie du génome humain ?

Le premier consiste à mettre en évidence les phénomènes de **recombinaison méiotique**. Ces phénomènes vont, comme nous l'avons vu, séparer des allèles qui étaient liés (rappel : pour cela il est nécessaire d'étudier les haplotypes). Pour cela, nous devons étudier des **familles**. C'est en effet le seul moyen de mettre en évidence les phénomènes de recombinaison ;

Le deuxième outil nécessaire correspond aux **marqueurs polymorphes**. Un marqueur polymorphe est un marqueur pour lequel il existe plusieurs allèles et qui permet donc de distinguer les chromosomes en suivant leur ségrégation au cours des générations.

Enfin le troisième outil fait appel aux **statistiques**. Ainsi, l'étude des génotypes et des haplotypes dans des familles utilise les probabilités afin de déterminer si ce qui est observé peut être lié au simple hasard ou au contraire s'il existe une très forte probabilité d'être lié à une maladie par exemple. Ces études statistiques sont connues sous le terme de « *Lodscore* » qui correspond à un test de liaison génétique entre des marqueurs et un phénotype.