

Les gènes et la cartographie du génome humain

Les progrès récents réalisés en génétique moléculaire ont déjà et auront des conséquences sur la compréhension et le traitement de nombreuses maladies. Ils bouleversent également les connaissances fondamentales en biologie, en particulier dans le domaine de la génétique des populations, une des activités de recherche de l'Ined. Ceci nous amène à faire le point sur les acquis et perspectives de la cartographie du génome humain.

Les bases de la vie

La cellule, unité de base de tous les êtres vivants, fonctionne toujours selon le même schéma. Dans le cas de la bactérie ou de l'amibe, elle assure, seule, sa survie et sa multiplication. Dans les autres cas, des cellules dont le nombre peut aller jusqu'à plusieurs dizaines de milliards s'associent, chacune d'elles se spécialisant dans une fonction, l'une d'entre elles étant la reproduction. Dans son extrême diversité, le monde du vivant recèle, sous ses formes végétales et animales et depuis son origine, une prodigieuse unité liée à la continuité de la vie [1].

Chaque cellule renferme un noyau, dont la molécule fondamentale est l'«acide désoxyribonucléique», plus connu sous l'abréviation d'«ADN». Cette molécule, qui porte toutes les instructions nécessaires à la vie, est universellement présente chez tous les organismes, animaux ou végétaux (cf encadré page 2).

Le patrimoine génétique de l'Homme, son génome, est constitué de la totalité de la molécule d'ADN, le constituant des chromosomes. On

estime que sur cette longue molécule, le nombre total de gènes est de l'ordre de 50 000 à 100 000. Et pourtant, ceux-ci ne constituent que 3 à 5 % du génome humain, le reste de l'ADN n'ayant pas encore de fonction connue ou élucidée.

Chez tout individu, telle fonction biologique, telle configuration organique, telle particularité physique dépend, sur une paire de chromosomes donnée, du gène transmis par le père et du gène transmis par la mère (1). L'information produite est directement liée à la «forme» (les généticiens disent l'«allèle») présentée par chacun de ces deux gènes. Le plus souvent, cette information permet, par le jeu des protéines, la production d'une fonction biologique ou d'une configuration organique normale. Cependant, si la protéine joue un rôle majeur dans l'organisme, une seule transformation génétique peut modifier ses aspects fonctionnels au point de compromettre la survie de l'individu. On emploie généralement le terme de mutations pour désigner ces modifications de l'ADN et le terme de maladies héréditaires pour désigner les dysfonctions qui en résultent sur le plan clinique.

Pour les populations humaines, le poids des maladies héréditaires n'est pas négligeable : près de 4 200 pathologies de ce type sont répertoriées actuellement (2), les plus connues étant la mucoviscidose (1/2 500 naissances), la phénylcétonurie (1/15 000 naissances), la myopathie de Duchenne

(1) Un caractère donné peut également résulter de la transmission de plusieurs gènes.

(2) On ne considère ici que le cas des maladies héréditaires « monogéniques » (liées aux mutations d'un seul gène).

Sommaire

Éditorial : Les gènes et la cartographie du génome humain

- Les bases de la vie 1
- L'exploration du génome 3

- Les perspectives du projet 3
- Les gènes et les populations 4
- Chromosomes, code génétique, protéines 2

Chromosomes, code génétique, protéines.

L'ADN est le principal constituant des chromosomes, qui ne sont visibles que dans les cellules en cours de division. Le nombre de chromosomes est constant, caractéristique d'une espèce : 1 pour la bactérie, 8 pour la mouche, 14 pour le petit pois, 18 pour le radis, 20 pour le maïs, 40 pour la souris, 78 pour le chien, 48 pour le chimpanzé. Chez l'homme, il y en a 46 (23 paires) (1). Pour chaque paire considérée, un chromosome est hérité du père et l'autre, de la mère. Chaque chromosome est donc un long filament d'ADN. Depuis la découverte de Watson et Crick en 1953, on sait que cette molécule se présente sous la forme d'un double brin, formé d'une succession d'unités fondamentales, les nucléotides (2). La partie essentielle de ces nucléotides, dite « base », n'existe que sous quatre formes différentes, désignées habituellement par leurs initiales : A (adénine), T (thymine), G (guanine), C (cytosine).

Les deux brins sont disposés face à face de telle sorte que toute base de type A se trouve associée, sur le brin opposé, à une base de type T et réciproquement. De même, toute base de type G se trouve associée à une base de type C et réciproquement. Cette molécule a la particularité de pouvoir se scinder en deux, par séparation des deux brins, puis de synthétiser, à partir de chacun d'eux, un nouveau brin complémentaire. Ainsi l'ADN peut-il se dupliquer indéfiniment. C'est ce qu'il advient dans le processus de division cellulaire.

C'est vers 1960 que Jacob, Monod et Lwoff élucident, à l'Institut Pasteur, la nature du gène. Ils découvrent qu'au long de l'ADN, certains segments ont une fonction particulière, celle de synthétiser des molécules indispensables à la vie de la cellule, les protéines. Ces segments d'ADN sont les « gènes ».

Les protéines sont, quant à elles, constituées par une succession de corps chimiques, les acides aminés, présents au nombre de 20 dans la nature. Comment passe-t-on de l'ADN, écrit dans un alphabet à

(1) Le chromosome n°2 humain semble résulter de la fusion de deux chromosomes du chimpanzé, ce qui explique la présence d'une paire supplémentaire chez ce dernier. On estime que le matériel génétique des deux espèces (homme, chimpanzé) est le même à 98 %.

(2) Sur la totalité de l'ADN humain, il y en a environ 3,5 milliards.

4 lettres (A, T, G, C) aux protéines, constituées par l'enchaînement des acides aminés ? Les biologistes ont découvert que chacun de ces 20 acides aminés était associé à un ou plusieurs groupes de trois bases successives (les « triplets » ou « codons ») choisies parmi les quatre de l'ADN. Il existe $4^3 = 64$ codons possibles. La correspondance entre ces 64 codons et les 20 acides aminés naturels est connue, elle représente le « code génétique » et se réalise par l'intermédiaire d'une autre famille de molécules proches de l'ADN, nommées ARN (molécules à simple brin où les bases « thymines », T, sont remplacées par des bases « uraciles », U). Par exemple, les triplets d'ADN : CAA, CAG, CAT, CAC codent pour un même acide aminé, la valine ; GTA et GTG codent pour l'histidine etc. Ce code est unique dans le monde du vivant (à quelques exceptions près). Il est le même pour les végétaux, les animaux, les mollusques marins ou les oiseaux. Parmi les 64 codons possibles, trois d'entre eux (ATT, ATC, ACT) ne codent pour aucun acide aminé. Ils sont dits « non sens », servent de signal stop pour arrêter la synthèse protéique et caractérisent la fin d'un gène. Un tronçon d'ADN peut être ainsi à l'origine d'une protéine dont la structure résulte de la séquence des bases de l'ADN (figure 2).

Figure 2 - Segment du gène (présent sur le chromosome 11) codant pour une des chaînes de l'hémoglobine (protéine du sang)

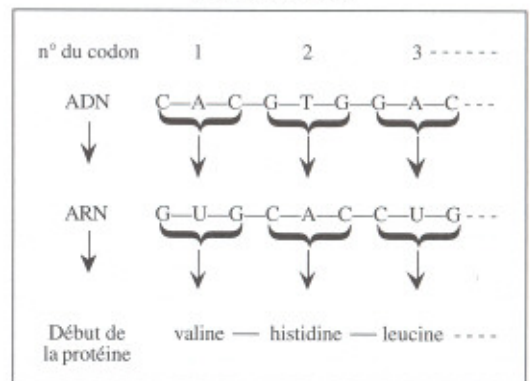
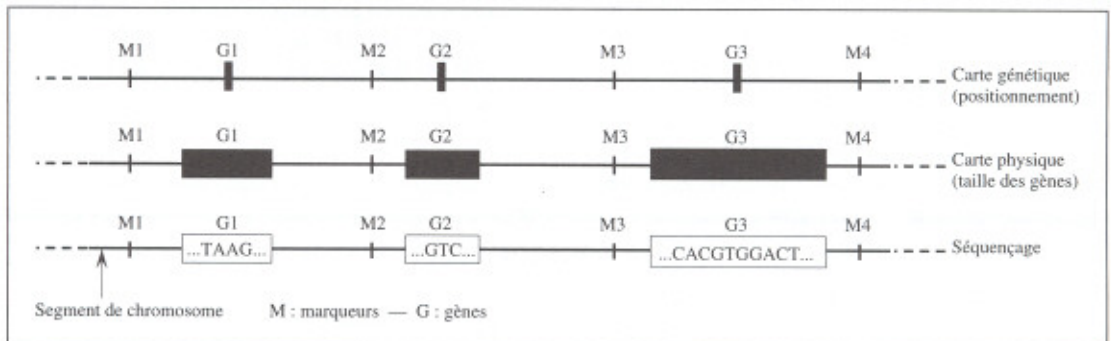


Figure 1 - Les différentes étapes de la cartographie des gènes sur un segment de chromosome



(1/3 500 naissances masculines), l'hémophilie (1/10 000 naissances masculines). En termes de morbidité, on estime que 10,5% des nouveau-nés vivants présentent une déficience à composante génétique: 0,4% des cas concerne des anomalies chromosomiques, 1,1% concerne des maladies héréditaires monogéniques et 9% représentent un groupe très hétérogène, les maladies multifactorielles, résultant d'une interaction entre un défaut génétique et l'environnement [2].

L'exploration du génome

Depuis longtemps, la recherche médicale s'est orientée vers la localisation, sur leurs chromosomes respectifs, des mutations responsables de maladies génétiques. En effet, identifier une mutation et sa composition moléculaire revient à décrypter l'anomalie du message et, en conséquence, à comprendre l'origine et la nature des mécanismes conduisant à la maladie.

Ces recherches sont longtemps restées le fait de laboratoires isolés. En 1986, un projet scientifique international, le *Projet sur le génome humain* (PGH) a été créé pour organiser l'ensemble de ces travaux et s'est donné pour objectif de déchiffrer l'intégralité du génome humain avant la fin de l'an 2002 (3). Il s'agissait d'un véritable défi compte tenu des techniques de travail dont on disposait alors, du budget et du temps nécessaires à une telle entreprise [3]. Pour accroître son efficacité, le PGH s'est doté, en 1989, d'une organisation (Hugo, *Human genome organization*) assurant la coordination des travaux effectués par les différents laboratoires et l'échange des outils et données utiles au projet.

Les pays engagés dans le PGH sont, pour la plupart, membres de l'*Organisation de coopération et de développement économiques* (OCDE). Au regard des engagements financiers, les principaux pays sont les États-Unis (170 millions de dollars investis en 1994), le Canada, le Royaume-Uni, le Japon. La France occupe une place singulière avec la participation d'organisations en partie privées: le *Centre d'étude du polymorphisme humain* (CEPH) et l'*Association française contre les myopathies* (AFM), celles-ci ayant contribué – grâce à l'apport du *Téléthon* – à l'installation et au financement du laboratoire du *Généthon* (près de 80 millions de francs de budget annuel).

Déchiffrer le génome humain requiert une haute technicité et l'information précise sur les gènes n'est obtenue qu'après une succession d'étapes. La première étape consiste à établir une «carte génétique». Celle-ci est réalisée à partir d'études familiales où l'on décrit la cotransmission d'une maladie héréditaire donnée et d'une série de mar-

queurs, petits segments d'ADN dont on connaît la localisation sur un chromosome et qui ont ici la fonction de repères. La probabilité de trouver une association entre cette maladie et certains marqueurs est liée à leur proximité: plus un marqueur est proche du gène responsable de la pathologie, plus il a de chance d'être hérité avec ce dernier. A ce stade, les études familiales portant sur différentes maladies ainsi que les associations décrites, permettent de situer la place occupée par les gènes sur les chromosomes et leurs positions respectives par rapport à une série de marqueurs. Quand l'emplacement et l'ordre des gènes sont établis, la deuxième étape donne lieu à une «carte physique». Il s'agit ici d'études moléculaires où différentes techniques, comme le clonage de l'ADN, peuvent être employées. Avec ces techniques, la mise en évidence d'un nombre élevé de points de repères identifiables sur l'ADN permet de mesurer la taille exacte des gènes ainsi que les distances qui les séparent. Cette topographie des régions chromosomiques aboutit enfin à la troisième étape, le séquençage du génome. Les procédés techniques utilisés à ce stade permettent d'obtenir la description complète de l'ADN, et en particulier l'agencement des 4 bases moléculaires A, T, G, C, constitutives des gènes (cf encadré et figure 1).

Le nombre de gènes séquencés s'accroît régulièrement. En 1981, on disposait de la séquence moléculaire complète de 16 gènes. Dix ans plus tard, plus de 2 300 gènes étaient entièrement décrits. A l'heure actuelle, c'est un peu plus de 16 000 gènes qui sont totalement séquencés. Cette progression spectaculaire a été permise grâce à des avancées technologiques qui ont révolutionné le savoir-faire des généticiens.

Les perspectives du projet

Ce décryptage du génome humain aura des conséquences multiples. Au plan médical, les retombées du projet suscitent beaucoup d'espoir. La connaissance des gènes eux-mêmes ouvre de nouvelles perspectives dans le cadre du diagnostic génétique et de la thérapie génique.

Le diagnostic génétique prénatal est une médecine prédictive qui permet, au stade fœtal, de révéler la présence d'une maladie héréditaire donnée. Cela concerne, le plus souvent, les couples déjà repérés comme étant «à risque», soit qu'ils aient déjà eu un enfant atteint, soit qu'ils aient subi un dépistage génétique révélant un risque important de donner naissance à un enfant malade. Ce diagnostic, lourd et coûteux à l'heure actuelle, ne s'applique qu'aux cas de maladies graves, dont le lien avec des marqueurs génétiques connus a été établi, ou mieux, dont le gène est ou sera identifié.

Si le résultat du diagnostic est positif, plusieurs conséquences sont envisageables. Il peut mener à la décision d'interruption volontaire de grossesse

(3) Parallèlement au génome humain, le projet propose de décrypter le génome d'organismes plus simples, à l'ADN plus court, tels que les bactéries, les levures etc.

dans le cas de maladies fortement invalidantes ou létales. Il peut conduire, à l'inverse, à la possibilité d'un suivi thérapeutique précoce dans le cas de maladies que l'on sait traiter, ou mieux, conduire à la prévention de la maladie par un traitement approprié. Il pourra aussi aboutir éventuellement à une décision de thérapie génique. Dans tous les cas, ces différentes possibilités permettront de diminuer de façon non négligeable la morbidité et la mortalité associées aux maladies génétiques.

Ce procédé pose néanmoins de nombreux problèmes éthiques. Que faire, par exemple, dans le cas d'une maladie de dégénérescence, létale, dont on sait qu'elle ne se manifesterait pas avant l'âge de 40 ou 50 ans (c'est le cas de la chorée de Huntington) ? Que décider dans le cas d'un gène muté dont on sait que la présence n'entraîne la manifestation de la maladie qu'avec une certaine probabilité et dans un contexte environnemental donné ? Le diagnostic prénatal est toujours établi par rapport à une maladie donnée, cependant le développement de ces techniques risque de favoriser la revendication d'un « droit à l'enfant sain ».

La thérapie génique, domaine encore incertain [4] mais cependant très réglementé sur le plan éthique et juridique (4), n'en est qu'au stade de l'expérimentation et ne s'applique qu'aux maladies graves pour lesquelles il n'existe aucune autre possibilité thérapeutique. Il s'agit de transférer, par l'intermédiaire d'un vecteur, le gène « sain » dans l'ADN du sujet atteint d'une maladie héréditaire, afin qu'il supplée au gène déficient. Cette nouvelle approche thérapeutique représente un immense espoir pour les patients. Si le principe en est simple, les obstacles biologiques et les difficultés sont nombreuses : l'insertion du gène sain doit permettre la synthèse d'une protéine normale en quantité suffisante et sur le long terme. La première thérapie génique, tentée en 1990 chez une petite fille de quatre ans, souffrant d'un déficit immunitaire grave [5] a donné des résultats prometteurs.

Les gènes et les populations

Outre le diagnostic génétique prénatal et la thérapie génique, la cartographie du génome humain aura pour conséquence de modifier la nature des données utilisables en génétique des populations. En effet, les études concernant la diversité biologique humaine ou celles concernant les causes et les modes de diffusion de certaines maladies héréditaires n'étaient possibles, dans le passé, qu'à

(4) Elle ne concerne notamment que les cellules somatiques, c'est-à-dire celles qui ne participent pas à la reproduction des individus. L'intervention porte sur le patient, mais n'empêche pas que la maladie puisse être transmise à sa descendance.

partir de caractères physiques extérieurs, facilement repérables. L'étude de Jean Sutter, décrivant dans les années 1960 la luxation congénitale de la hanche en Bretagne, dans ses aspects génétique, démographique et sanitaire, en est un exemple [6].

Par la suite, on a pu accéder à des entités biochimiques moins directement visibles telles que les enzymes ou les groupes sanguins. Ces caractères biologiques ont permis d'observer une grande variabilité individuelle et de décrire les particularités génétiques propres à certaines populations [7, 8].

Actuellement, la prise en compte simultanée d'un grand nombre de ces caractères, observés dans de multiples pays, permet d'analyser la répartition géographique des fréquences des allèles et, en conséquence, de proposer des processus évolutifs qui auraient joué dans l'histoire de l'Homme moderne [9].

Les travaux sur la cartographie et le séquençage des gènes concernent actuellement un nombre limité d'individus et fourniront le contenu d'un génome-type. Cette information sera d'une très grande utilité en biologie et en médecine. Pour le généticien des populations, il faudra cependant attendre une étape ultérieure pour que les données collectées portent sur des échantillons représentatifs et qu'elles fournissent une mesure de la variation individuelle. On pourra alors explorer le patrimoine génétique et sa diversité à son niveau le plus fin.

Gil BELLIS et Marie-Hélène CAZES

RÉFÉRENCES

- [1] Albert JACQUARD, *Éloge de la différence, la génétique et les hommes*, coll. Science ouverte, Paris, Seuil, 1978.
- [2] NATIONS UNIES, *Sources et effets des rayonnements ionisants*, rapport UNSCEAR 1977, New York, 1978.
- [3] OCDE (Organisation de coopération et de développement économiques), *Le grand programme sur le génome humain*, Paris, Éditions de l'OCDE, 1995.
- [4] Dossier spécial : « Faut-il avoir peur du génie génétique ? », *La Recherche*, n° 270, pp. 1106-1141, 1994.
- [5] French ANDERSON, « La thérapie génique », *Pour la Science*, n° 217, pp. 91-93, 1995.
- [6] Jean SUTTER, *La luxation congénitale de la hanche*, Travaux et Documents n°62, Paris, Ined/Puf, 1972.
- [7] Albert JACQUARD, *Génétique des populations humaines*, collection Sup, Paris, Puf, 1974.
- [8] Marie-Hélène CAZES (éd.), *Les Dogon de Boni, approche démo-génétique d'un isolat du Mali*, Travaux et Documents n°132, Paris, Ined/Puf, 1993.
- [9] Luca et Francesco CAVALLI-SFORZA, *Qui sommes-nous ? Une histoire de la diversité humaine*, Paris, Albin Michel, 1994.