

# Le Génome Humain: Principes fondamentaux sur l'ADN

E. G. Berger, T. Hennet

## Introduction

Qui est HUGO? Cet acronyme simple à mémoriser qui s'est fait connaître au public en février dernier par la publication de la structure du génome humain, correspond à **H**uman **G**enome **O**rganisation. Alors que beaucoup de gros titres de Recherche et Technique aussi spectaculaires soient-ils parfois, ne sont en fait que des coups d'éclat, ce que HUGO présente au public du monde entier ainsi qu'au corps médical est une véritable révolution comparable à l'exploration anatomique du corps humain par Vesale en 1543 ou à l'établissement de la classification périodique des éléments chimiques. Il vaudrait d'ailleurs certainement la peine, dans quelques années, de jeter un regard rétrospectif afin de mesurer ce que HUGO a déclenché en février 2001. Il serait également intéressant de tirer des parallèles sur l'effet de la première dissection anatomique et celui de la découverte de la structure du génome sur le progrès médical. Ils ont tous deux en commun la création d'une langue nouvelle, autrefois le latin, appelé de nos jours en anglais le «genome speak». Entre le passé et l'époque actuelle, la différence majeure devrait résider dans la rapidité d'application des connaissances récentes. La communauté de la recherche biomédicale est d'ores et déjà consciente de la source inépuisable de connaissances nouvelles que constituent les bases de données du génome. Nous croyons que nos collègues praticiens devraient pouvoir en bénéficier, ne serait-ce qu'afin d'apprécier cette contribution culturelle ou de trouver accès à la biologie cellulaire et moléculaire moderne dont est de plus en plus empreinte la médecine au quotidien. Nous avons l'intention de tenir compte de cette révolution dans le cadre de la formation de nos jeunes étudiants en leur proposant sous peu, et ceci dès la première année d'étude, une filière de formation intégrée en biologie cellulaire et moléculaire basée sur un cycle de cours interdisciplinaire systématique comprenant des travaux pratiques, sur des tutorats orientés sur des projets de médecine moléculaire, ainsi que par l'utilisation de multimédias comme internet. Cette dernière sera

donc enseignée en lieu et place de la zoologie, de la morphologie comparée des vertébrés et de la botanique actuellement encore requises par le règlement.

Dans les brefs compte-rendus sur le génome humain traités de façon non systématique, nous posons tout d'abord les bases puis rappelons de quelques propriétés élémentaires du porteur de l'information génétique: l'ADN. Par la suite nous essayerons de décrire le potentiel médical contenu dans l'abondante information actuellement disponible par l'accès à de nombreuses banques de données\*. A ce propos, une série d'articles, parus il y a quelques années dans le *New England Journal of Medicine* et qui avaient trouvé bon écho, nous a également servi de parrain.

## Pourquoi publier maintenant le génome humain?

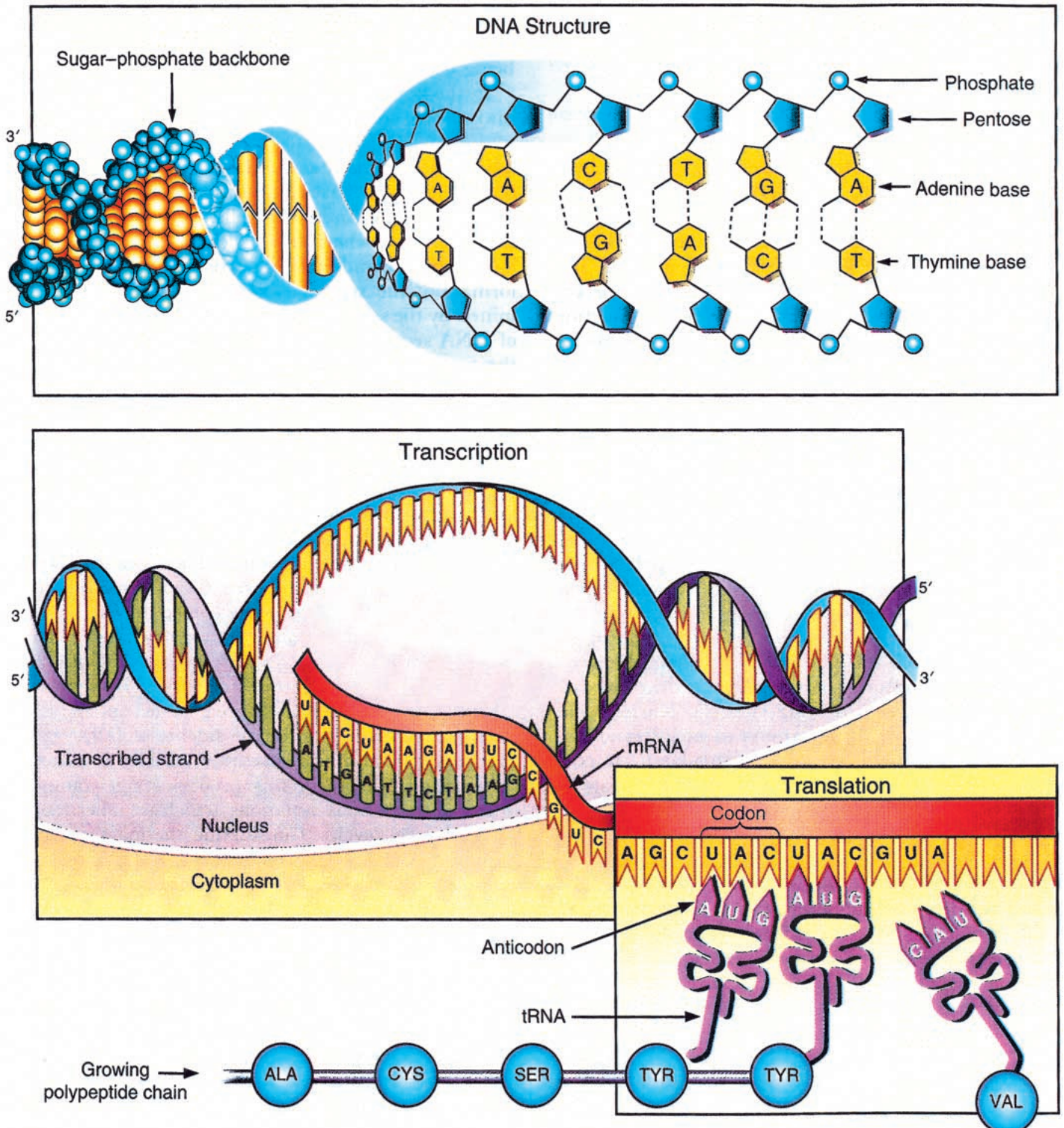
L'énoncé de quelques faits marquants dans le domaine de la recherche génétique moderne (voir l'encadré) comprend également celui, récompensé du prix Nobel, de la méthode de séquençage de l'ADN par Sanger. Cette méthode s'est révélée si solide qu'une automatisation fut rendue possible. La faisabilité de l'élucidation de la structure du génome humain se dessinait déjà au milieu des années quatre-vingt: à Montreux, en septembre 1988, un consortium soutenu par des subventions publiques sous la direction de V. McKusik fut constitué sous le nom de HUGO ([www.gene.ucl.ac.uk/hugo/history.htm](http://www.gene.ucl.ac.uk/hugo/history.htm)). L'organisation est inscrite au registre de Genève et comprend 220 membres provenant de 23 pays. L'organisation s'est clairement fixé comme but l'élucidation du génome humain et a coordonné les efforts entre les laboratoires impliqués. Il y a quatre ans, une puissante concurrence s'opposa aux efforts de progression plutôt lents de HUGO, par le biais d'une entreprise (Celera) sous la direction de Craig Venter qui, bien qu'ayant une stratégie différente, s'était fixé le même but à la différence près que cette dernière envisageait la protection par un droit de brevet. Dès lors, subsistait le danger d'une rétention de l'information sur la séquence obtenue dans le domaine privé, qui n'a pu être écarté qu'en l'an 2000 lors d'une déclaration commune de Clinton et Blair, chefs d'Etat des deux nations dirigeantes dans ce domaine. La pression exercée par Celera a permis d'accélérer considérablement l'élucidation de la structure, si bien qu'à l'heure actuelle 94% du génome humain sont séquencés. Les revues *Nature* et *Science* se sont d'ailleurs partagé les différentes stratégies: *Nature* rapporte les résultats du projet public financé par les contribuables, *Science* ceux de la méthode utilisée par Celera. Cette information est désormais ac-

**Figure 1.**

**En haut:** Représentation de la structure de la double hélice d'ADN. L'appariement se fait par les ponts hydrogène (en traitillés).

**Au milieu:** Transcription: l'utilisation de l'information génétique par la cellule se fait par la copie de la partie à lire. La copie (ARN) est légèrement différente du point de vue chimique (cf texte) ce qui empêche un appariement stable.

**En bas:** L'ARNm passe du noyau dans le cytoplasme où l'information est traduite en une séquence colinéaire d'acides aminés. L'ARN de transfert qui sert d'adaptateur porte d'une part un triplet de bases complémentaire et, d'autre part, l'acide aminé correspondant au code génétique. La liaison des acides aminés en protéine se déroule dans le ribosome qui consiste en ARN ribosomal et en un grand nombre de protéines. Modifié après Rosenthal N, N Engl J Med 1994;331(1):39-41. Avec la permission des éditions.



### Les grandes étapes de la découverte de l'ADN.

1869	<b>Miescher</b> découvre les acides nucléiques contenus dans les leucocytes
1944	Démonstration par <b>Avery</b> du rôle de l'ADN dans l'hérédité
1953	Elucidation de la structure en double hélice par <b>Watson</b> et <b>Crick</b>
1961	Le code génétique est élucidé par <b>Nirenberg</b> et <b>Ochoa</b> . Le code détermine la transposition de l'information génétique dans les protéines
1962	Découverte des endonucléases de restriction par <b>Dussoix</b> et <b>Arber</b>
1973	Premier clonage d'un plasmide bactérien par <b>Boyer</b> et <b>Cohen</b>
1975	Méthode de séquençage par <b>Sanger</b> et <b>Barrell</b>
1977	Découverte des introns par <b>Robert</b> et <b>Sharp</b>
2000	Fin de la cartographie et du séquençage provisoires du génome humain
2001	Les revues Nature et Science publient une carte génétique du génome humain séquencée à 94%

cessible à tous, mais son application, elle, constitue un grand défi: (le fait pour un cuisinier de posséder les meilleures recettes ne fait pas encore de lui un grand chef)?

### L'ADN en tant que porteur de l'information génétique

L'ADN (acide désoxyribonucléique) est la base chimique de l'information génétique de tout être vivant (microorganismes, végétaux, animaux et humains). Il est formé par la séquence des différentes bases A,G,C,T (fig.1). La «langue génétique» formée de ces quatre lettres qui en résulte détermine la diversité de la création. Il n'y a pas de différences fondamentales du comportement chimique entre l'ADN végétal, animal ou humain. C'est pourquoi il est sans autre possible d'implanter, par exemple, des fragments d'ADN humain dans celui d'un végétal ou vice-versa.

L'ADN des cellules humaines est contenu dans le noyau d'une part et d'autre part, en moindre mesure, dans les mitochondries. Un grand chromosome humain, c'est-à-dire un fil d'ADN linéaire mesure, dans sa forme déliée, presque 10 cm de long. Comment est-ce qu'alors 46 chromosomes peuvent-ils être contenus dans un noyau cellulaire dont le diamètre mesure environ 5–10 µm? La nature a trouvé une réponse sous la forme d'un impressionnant emballage dans des nucléosomes formant une sorte de mini-collier de perles qui, lui-même, est à son tour emballé de façon compacte. La partie contenue dans les mitochondries connue dans sa totalité depuis 25 ans déjà, contient l'information d'une partie des protéines mitochondriales et de leur biosynthèse. L'ADN mitochondrial n'est pas seulement important pour la biologie et la médecine en raison de l'infor-

mation nécessaire aux protéines mitochondriales qu'il contient, mais également en raison de son hérédité maternelle, de sa signification dans les troubles héréditaires liés à la phosphorylation oxydative (respiration cellulaire), de son «vieillesse» et de sa signification comme marqueur de l'évolution et de la migration des populations.

Fait important pour la **pratique**: le génome complet est représenté dans chaque cellule. Il est ainsi possible de pratiquer des analyses génétiques sur des échantillons sanguins, à partir de frottis des muqueuses, de racines de cheveux, etc. L'ADN diffère peu d'un être humain à l'autre formant ainsi la base de l'individualité biochimique. Afin que les données sur les séquences du génome humain soient représentatives, Celera a utilisé l'ADN de 21 donneurs choisis au hasard, de sexes, âges et ethnies différents. Le consortium HUGO a agi de façon similaire. Toutes les données personnelles des donneurs d'ADN furent ensuite, bien entendu, détruites.

Les articles qui suivent traitent de médecine moléculaires développent des thèmes concernant les différences interindividuelles de l'ADN dans le cadre de polymorphismes (ce sont les mutations sans conséquences pathophysiologiques évidentes) et de maladies liées à l'hérédité (modifications des gènes dont les propriétés sont prouvées engendrer la maladie).

### Fonctions de l'ADN

L'ADN ne sert de porteur de l'information que lorsque que celle-ci puisse d'une part être transmise et que d'autre part elle puisse être également utilisée. La transmission aux cellules-filles se fait par **réplication**: Lors de la découverte de la double hélice, Watson et Crick avaient déjà attiré l'attention sur le fait que la complémentarité des paires de bases se prêtait de façon excellente à la duplication: Dans la citation originale il est dit avec un soupçon d'euphémisme: «it has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material». Le processus de réplication n'est pas exempt d'erreurs. Des mécanismes de réparation étudiés avec subtilité préviennent une accumulation de copies erronées (donc de mutations). On compte une mutation sur trois divisions cellulaires; un gène de taille moyenne sera donc affecté sur une durée de temps de 200 000 ans par une mutation. Afin qu'une mutation ait des conséquences sur la descendance il faut qu'elle ait lieu lors de la germination. Cette fréquence de mutation est à la base de l'évolution. Nous constatons non sans étonnement que les comparaisons systématiques des séquences des protéines du règne vi-

vant ayant la même fonction conduisent à de nouvelles relations ontogénétiques. Les générations de médecins précédentes ont encore reçu, lors du premier examen propédeutique, un aperçu de l'évolution par l'étude de la morphologie comparée des vertébrés (Portmann); cette ontogénèse fondée sur l'étude morphologique est dès lors remplacée par la comparaison des séquences d'ADN.

La lecture de l'information dans le temps et dans l'espace se déroule dans l'organisme de façon hautement spécifique: ce processus, la **transcription** (fig.1) conduit à une image complémentaire du segment d'ADN copié sous forme d'un monobrin d'ARNm (ARN messenger). L'ARN contient de l'uracile à la place de la thymine et du ribose à la place du désoxyribose. Ceci permet de prévenir la formation d'une double hélice stable et de former des molécules de courte durée de vie à un débit élevé. Ceci permet d'obtenir une modification rapide de l'information transmise à la cellule et, par la même, une capacité rapide d'adaptation aux signaux extérieurs.

#### Données quantitatives

- Chaque noyau cellulaire contient 46 chromosomes (jeu diploïde).
- Chaque jeu haploïde contient 23 chromosomes, dont 22 autosomes et un chromosome sexuel; X (femelle) ou Y (mâle). Dans le jeu diploïde XX correspond au sexe féminin et XY au sexe masculin.
- Chromosome = double hélice d'ADN de longueur variable, mesurant dans sa forme déliée environ 1-10 cm.
- $2,91 \times 10^9$  paires de bases (haploïdes) au total, dont 5% transcrites.
- 30 000-35 000 gènes.

#### Propriétés

- Structure composée de chaînes désoxyribose (molécule saccharinique dotée de 5 atomes de C) reliées par ponts phosphodiester de 5' à 3'. Une base purique (adénine ou guanine) ou pyrimidique (cytosine ou thymine) est liée à l'atome C<sub>1</sub> du ribose.
- Cette chaîne forme un monobrin qui se forme par l'appariement des bases A et T, respectivement C et G, se transforme en double brin. L'appariement se fait par la liaison de ponts hydrogène dont l'énergie de liaison est mesurable, livrant ainsi la base de l'hybridation. Le chauffage casse les ponts H (dénaturation), le refroidissement permet la renaturation à l'occasion de laquelle les segments complémentaires se réappariaient.

Enfin, l'information de l'ARN est traduite au moyen du code génétique en une séquence colinéaire d'acides aminés. La traduction de l'information stockée sous forme de séquences de paires de bases dans la séquence d'acides aminés correspondante porte le nom de **translation**. Chaque groupe de trois lettres de l'ADN (triplet) code pour un des vingt acides aminés. Quatre lettres différentes étant à disposition, 64 triplets différents peuvent être formés. La cellule dispose d'interfaces (adaptateurs), comme le montre l'illustration, entre l'ARNm et les acides aminés à intégrer. Ces interfaces portent le nom d'ARN de transfert (ARNt), lesquels disposent également de gènes spécifiques. Certains acides aminés sont codés par un seul triplet, d'autres par six triplets différents; il existe en outre une sorte de ponctuation: le triplet AUG détermine le début et les triplets UAG, UAA et UGA la fin de la translation et ainsi donc la taille de la molécule protéique. Le code génétique, abstraction faite de petites exceptions pour les mitochondries et les bactéries, est universel. On le trouve fréquemment représenté, en raison de son importance, sur la reliure des manuels. La **bioinformatique** l'utilise afin de traduire automatiquement les nouvelles séquences d'ADN déchiffrées en séquences d'acides aminés y correspondant. La structure primaire d'une protéine (donc sa séquence d'acides aminés) n'est plus, depuis vingt ans au moins, déchiffrée que grâce à l'information génétique correspondante. Cependant, seule une petite partie de l'ADN sert à la traduction en protéines (voir le paragraphe: Qu'est-ce qu'un gène?). Le produit d'un gène est tout d'abord une protéine unique qui est ensuite remaniée par la cellule jusqu'à ce qu'elle obtienne sa structure définitive (comme glycoprotéine polymère p. ex.) et qu'elle ait trouvé son site d'action cellulaire ou systémique. Les modifications suivant la translation de la structure protéique portent le nom de «modifications post-translationnelles». Ces dernières sont fréquemment déterminantes dans la capacité de fonctionnement de la protéine (exemple: l'hormone érythropoïétine est inefficace sans transmission post-translationnelle des résidus saccharose, de même que l'insuline doit être dédoublée pour passer de sa forme première inactive à sa forme active. Ces deux processus de «saccharification» et de dédoublement se font après la translation et sont donc «post-translationnels»).

#### Qu'est-ce qu'un gène?

En génétique classique, un gène était décrit comme un facteur héréditaire contrôlant un caractère héréditaire précis. Aujourd'hui, un gène correspond à un segment d'un brin d'ADN

chimiquement défini qui est transcrit, entraînant la synthèse d'un ARN (partie codante) comprenant les segments régulant la formulation du gène (partie non codante). Le transcrit conduit le plus souvent à la formation d'une protéine; dans de rares cas, cependant, également à l'ARN structural des ribosomes. Il devint rapidement évident que plus d'ADN que nécessaire requis pour la production des séquences protéiques était empaqueté dans les chromosomes. Étrangement, certains végétaux et organismes amphibiens contiennent encore davantage d'ADN que le génome humain. La fonction de l'ADN n'étant pas utilisée directement pour les produits génétiques, soit environ 90%, n'est par contre pas encore élucidée. Une partie de cet ADN non codant (introns) est insérée parmi les segments de l'ADN codant (exons). Entre deux intervient, après la transcription, un processus d'épissage (splicing) des introns prenant place dans le noyau, conduisant ainsi à la formation de l'ARNm apte à la translation. L'épissage est un processus apportant une variabilité intracellulaire supplémen-

taire du modèle protéique qui, dans certains cas, peut également être défectueux. La structure même du gène ne peut cependant pas être déduite à partir de son produit génique: ceci tout d'abord car le code génétique ne peut pas être traduit à l'envers et ensuite parce qu'une part variable d'ADN non-codant est contenue dans le gène lui-même. Le gène de l'insuline, par exemple, ne contient que deux introns alors que celui responsable de la dystrophie musculaire (maladie de Duchenne), la dystrophine, contient plus de 50 introns. Il n'existe pas de réponse simple à la question: *qu'est-ce qu'un gène?* On ne peut déduire de la séquence des bases la structure du gène, car contrairement à la translation, il n'existe pas de séquence indiquant le début ou la fin de la transcription. Le postulat déclarant que tout ce qui est traduit en ARN appartient à un gène pourrait être émis. Ce dernier, bien que nécessaire n'en est pas suffisant: il existe, en effet, des indices indiquant que des segments du génome sont transcrits mais que ces transcrits ne sont pas dévolus à une utilisation ultérieure et

## Glossaire

<b>Bioinformatique:</b>	Terme d'ensemble regroupant toute l'informatique traitant les données génétiques.
<b>Chromosome:</b>	Structure composée d'une double hélice d'ADN et de plusieurs protéines associées qui porte l'information héréditaire d'un organisme.
<b>ADN:</b>	Acide désoxyribonucléique: polynucléotide formé d'unités désoxyribonucléotidiques liées de façon covalente consistant en deux brins antiparallèles.
<b>Exon:</b>	Partie codante d'un gène.
<b>Code génétique:</b>	Un triplet de nucléotides (codon) déterminant un des vingt acides aminés. 64 triplets différents peuvent donc en partie coder le même acide aminé.
<b>Génome:</b>	Totalité de l'information génétique d'une cellule ou d'un organisme.
<b>HUGO:</b>	Human Genome Organization. Société multinationale de recherche sur le génome humain siégeant à Genève.
<b>Intron:</b>	Partie non codante d'un gène.
<b>Mitochondrie:</b>	Organelle cellulaire ayant son propre génome. Assure par la phosphorylation oxydative avec l'aide de la chaîne respiratoire la production d'énergie sous forme d'ATP.
<b>Mutation:</b>	ADN: Modification transmissible et durable de la séquence nucléotidique d'un gène avec ou sans conséquences sur la fonction de son produit. Dans les cellules corporelles: mutation somatique non héréditaire avec ou sans conséquences sur la fonction de son produit génétique.
<b>Nucléosome:</b>	d'une courte longueur d'ADN enroulée autour d'un noyau d'histones. Unité structurale composée
<b>Polymorphisme:</b>	Différences ponctuelles et inter-individuelles dans la structure de l'ADN sans conséquences sur la fonction du gène.
<b>Réplication:</b>	Procédé servant à dupliquer l'ADN lors de la division cellulaire. A cet effet, les deux brins sont séparés afin qu'un ADN complémentaire pour chaque brin puisse être formé. Chacun des brins originaux est ensuite lié à un des brins nouvellement formé dans une cellule-fille.
<b>Séquençage:</b>	Détermination de l'ordre des nucléotides dans une molécule d'ADN.
<b>Epissage (splicing):</b>	Une des grandes surprises fut la découverte vers la fin des années soixante-dix de séquences au sein des gènes non traduites dans la protéine. Les parties non codantes sont les introns et celles codantes exons. Il s'agit d'un processus par lequel des séquences introniques sont excisées de molécules d'ARN dans le noyau pendant la formation d'ARN messenger.
<b>Translation:</b>	Processus se produisant sur un ribosome par lequel la séquence de nucléotides d'une molécule d'ARN messenger dirige l'incorporation d'acides aminés dans une protéine.
<b>Transcription:</b>	Lecture de l'information génétique par la formation d'un monobrin complémentaire d'acides nucléiques, également appelé ARN nucléaire hétérogène. Ce dernier est transformé par l'excision (épissage) de séquences non fonctionnelles (introns) en ARNm fonctionnel. Les produits de la transcription sont l'ARNm (ARN messenger) pour la fabrication de protéines, l'ARNr pour la formation de ribosomes et l'ARNt (ARN de transfert) comme interface entre les acides aminés et l'ARNm dans la synthèse protéique.

qu'ils sont dégradés. Un autre aspect complexe réside dans le fait que des gènes comportant beaucoup d'introns peuvent mener, par ce que l'on appelle l'épissage alternatif, à différents produits géniques. Pour cette dernière et différentes raisons les données sur le nombre probable de gènes varient considérablement. Depuis la publication de 95% de la séquence du génome humain, un ordre de grandeur d'environ 30 000 est communément accepté. On reconnaît cependant de façon unanime que la détermination exacte prendra encore quelques années.

### Références

- Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J. Molecular Cell Biology. WH Freeman & Co. 2000, New.
- The human genome, Nature 409 (6822), 2001.
- The human genome, Science 291 (5507), 2001.

### Conclusion

La pédagogie actuelle met l'accent sur la nécessité d'un perpétuel apprentissage. Durant les études de médecine d'un des auteurs, pratiquement rien de ce qui précède n'était connu (E. G. Berger). La double hélice était publiée mais ne figurait pas encore dans les notes de cours. En revanche, la pédagogie actuelle tait le bon adage selon lequel: ce que l'on n'apprend pas étant petit on ne l'apprendra jamais. Rien ne permet d'y échapper.