

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE BADJI MOKHTAR-ANNABA
FACULTE DE MEDECINE
Département de médecine

1^{ère} Année médecine
Cours de Génétique

LE GENE

Elaboré par :
Dr KEBIR .S
Maître assistant
Histologie-Embryologie et Génétique

Année universitaire : 2015-2016

LE GENE

I. GENERALITES-DEFINITIONS :

- Le gène est l'**unité de l'information génétique**, il est constitué d'une séquence d'acides nucléiques codant pour la synthèse **d'un polypeptide spécifique ou d'un ARN**.
- Jusqu'à la fin des années 70, le gène était visualisé comme **un segment de la molécule d'ADN** contenant l'information de la séquence des acides aminés de la chaîne polypeptidique et ses séquences régulatrices nécessaires pour son expression.
- Nous savons maintenant que c'est une **description incomplète** des gènes du génome humain. En fait, **un très petit nombre** de gènes existe sous la forme d'une séquence **codante continue** ; au contraire **la grande majorité** des gènes sont **interrompus par une ou plusieurs régions non codantes**. Ces séquences sont appelées **introns** sont d'abord transcrites en ARN dans le noyau mais ne sont pas présentes dans l'ARNm mature présent dans le cytoplasme, et par conséquent **ne sont pas représentées dans le produit final**.
- Les introns alternent avec des séquences codantes ou exons, qui déterminent la séquence en acides aminés du produit final.

II. STRUCTURE:

- Chez l'homme, on estime qu'il y a entre **50.000 et 100.000 gènes sur 23 chromosomes**. L'information génétique est déterminée par l'ordre dans lequel sont disposées les bases qui forment la chaîne d'ADN. Chaque gène **comporte toutes les instructions nécessaires pour la synthèse protéique**.
- **Le gène s'exprime par une copie de l'ADN en ARN**. L'ARN possède une séquence complémentaire au brin d'ADN, ayant servi de matrice: le processus de **transcription**. L'ARN permet la synthèse d'un polypeptide, dont la séquence en AA est déterminée par la séquence en bases de l'ADN: **la traduction**.
- **Seul l'un des deux brins de l'ADN est porteur de l'information**, c'est le **brin matrice** ou bien **codant** ou **brin sens** utilisé pour la transcription de l'ARN.

Les gènes sont de tailles variables de **quelques dizaines de paire de bases à plusieurs millions de paires de bases.**

- Le gène d'un **eucaryote est composé de de plusieurs segments codants: exons**, séparés par des séquences non codants: **les introns**.
- Bien qu'un petit nombre de gènes dans le génome humain ne possèdent pas d'introns, **la plupart en contiennent au moins un**, habituellement plusieurs.
- Dans un grand nombre de gènes, la longueur totale des **introns a une proportion plus grande que la longueur totale des exons**. Alors que certains gènes ont **moins d'1 kb de longueur**, d'autres tel que le gène du facteur VIII s'étendent sur **plusieurs centaines de kb**.
- Un gène exceptionnel, le gène de **la dystrophine sur le chr X** (dont les mutations conduisent à la dystrophie musculaire de Duchenne) s'étend sur **plus de 2000 kb**.

A. Les éléments de contrôle du gène:

1) Le promoteur:

- Dans les organismes eucaryotes, **l'expression des gènes est régulée**. Tous **les gènes présents** sur l'ADN d'une cellule **ne sont pas exprimés**, ceux **exprimés** par une cellule dépendent du type cellulaire considéré, ce qui détermine les caractères morphologiques et fonctionnels spécifiques de la cellule dans l'organisme. Par exemple, certains gènes, spécifiquement activés dans le muscle ou le foie, sont totalement inactifs dans la cellule nerveuse, laquelle en exprime d'autres.
- **L'expression du gène est régulée** par une séquence ADN appelée **promoteur**. Celui-ci est une séquence de 800 pb à 1000 pb, située en 5' en amont du gène, **adjacente au site d'initiation** de transcription du gène. Le promoteur **définit le site d'initiation de la transcription et sa direction**.
- Le promoteur comporte plusieurs de ces **séquences régulatrices**, dont les séquences consensus. Celles-ci sont soit spécifiques, soit ubiquitaires, retrouvées partout dans l'organisme. **Des facteurs de transcription**, interagissent avec les séquences régulatrices pour moduler l'expression du gène.

2) Les séquences consensus

- Ce sont des **séquences nucléotidiques situées dans une région particulière** du gène et qui **remplissent des fonctions semblables**. Les séquences

consensus sont retrouvées dans de nombreux gènes, mais en comparant l'ordre des nucléotides, on retrouve quelques variations mineures, autrement dit les séquences consensus sont homologues, sans être rigoureusement identiques.

- Souvent on donne à une séquence un nom en notant le monomère le plus fréquemment retrouvé dans la séquence. La boîte TATA ou TATA box, retrouvée dans le site d'initiation de transcription de gènes « tissus spécifiques », se situe à une trentaine de pb en amont du site d'initiation de la transcription. Elle a été appelée ainsi à cause des motifs retrouvés dans sa séquence consensus : 5' TATA (A ou T) A (A ou T) 3'.
- Certains gènes possèdent un autre élément d'initiation à la place de la boîte TATA comme **la CAT box** (séquence : CCAAT), une autre séquence consensus jouant un rôle dans l'initiation de la transcription.
- Dans les gènes qui s'expriment dans tous les tissus (« gènes de ménage » ou « house keeping gènes »), la TATA box est remplacée par une séquence d'initiation riche en motifs GC: **GC box**, formée de 100 à 200 pb d'éléments GC répétés en amont du site de démarrage de la transcription.
- Enfin d'autres gènes n'ont aucun élément d'initiation.

3) Enhancers

Les enhancers stimulent l'expression des gènes. Ce sont des séquences ADN situées en 5' ou en 3' du gène, ou même à l'intérieur du gène, dans un exon ou dans un intron. La fonction de l'enhancer ne dépend ni de sa position ni de son orientation par rapport au gène. Un enhancer contient un ou plusieurs sites de fixation pour les facteurs de transcription.

4) Silencers

Ce sont des séquences ADN ayant une action contraire à celle des enhancers, elles diminuent ou inhibent l'expression d'un gène. Elles se localisent souvent entre le promoteur et l'enhancer.

5) Les facteurs agissant « en trans »

Ce sont des séquences régulatrices situées loin du gène, généralement sur un autre chromosome. Ils agissent sur la transcription par leurs produits protéiques. Ces protéines régulatrices agissent soit en activant (Activateur/Enhancer) ou en inhibant partiellement ou totalement la transcription (Répresseur/Silencer).

B. Les autres éléments constitutifs du gène:

1) Le capping:

Immédiatement en 3' du promoteur on rencontre une séquence qui **n'est pas traduite** mais qui sert à l'initiation de la transcription en ARN. Cette séquence initiatrice permet aussi la fixation d'un cap (chapeau) sur l'ARN messager. Le capping est nécessaire pour la traduction ultérieure de l'ARN.

2) Les exons:

Chez les eucaryotes, l'information codante des gènes est souvent morcelée, répartie sur une série de séquences codantes, appelées **exons**.

Les exons débutent en 3' du promoteur, **suite à la zone de cap**. Les exons sont séparés les uns des autres par des séquences non codantes, les introns. A cause de l'alternance entre exons et introns, les gènes eucaryotes sont aussi appelés « gènes mosaïques » en comparaison avec les gènes procaryotes qui ne possèdent que des séquences codantes. Ce sont les exons qui sont transcrits sur l'ARN puis traduits en protéines.

3) Les introns:

Les introns, sont également transcrits sur l'ARN mais ils sont éliminés avant la traduction en protéines. Le nombre et la longueur des introns et des exons sont très variables. Les introns sont en général plus longs que les exons, constituant parfois la majeure partie du gène. Les introns n'existent pas dans les gènes de procaryotes.

4) Le codon initiation:

Entre la séquence d'initiation et le premier exon, c'est à dire au début de la partie codante du gène se trouve le codon initiation de traduction, ATG, qui code pour la méthionine.

Toutes les protéines débutent leur synthèse par l'acide aminé méthionine, celui-ci est ensuite éliminé s'il ne fait pas partie de la séquence polypeptidique.

5) Le codon stop:

A l'extrémité 3' du dernier exon, on retrouve l'un des trois codons stop, **TAA, TAG ou TGA** qui indique la **fin de la traduction** sur le ribosome.

6) La séquence de polyadénylation (poly A):

En 3' du codon stop, le dernier exon comporte une séquence de polyadénylation, AATAAA, dont la transcription, sert à **stabiliser l'ARN primaire**. Un gène peut comporter plusieurs sites de polyadénylation.

III. FAMILLES MULTIGENIQUES :

- Chez les eucaryotes **certains gènes** sont groupés en familles multigéniques. **Les gènes d'une famille** sont identiques ou presque, mais **ne sont pas régulés ensemble**.
- Certaines familles multigéniques sont constituées de gènes **semblables dispersés sur plusieurs chromosomes**, ceci est probablement dû aux réarrangements de l'ADN au cours de l'évolution. Dans les familles **multigéniques simples** tous les gènes sont identiques.
- Les familles **multigéniques complexes**, contiennent des gènes très semblables mais non identiques. Les gènes des globines constituent un exemple de famille multigénique complexe, leurs produits diffèrent par quelques acides aminés.
- Les gènes de **l' α -globine et la β -globine**, respectivement sur les chromosomes 16 et 11, dérivent vraisemblablement par duplication il y a environ 500 millions d'années d'un gène ancestral.
- Ces deux régions contiennent des gènes codant pour les chaînes de globines très semblables, et exprimées à différents moments au cours du développement, de la vie embryonnaire à la vie adulte.
- **La distribution** des exons et des introns des gènes de globine apparaît remarquablement conservée au cours de l'évolution ; chacun des gènes fonctionnels de la globine possède deux introns situés à des endroits semblables.

IV. EXEMPLE: LE GENE DE LA BETA GLOBULINE:

- Le flux d'informations peut être illustré en utilisant comme exemple un gène particulièrement bien étudié, le gène de la β -globine occupe approximativement **2 kb** sur le bras court du **chr 11** et code pour un polypeptide de **146 AA**.
- Le gène a **trois exons et deux introns**.
- Le gène de la β -globine tout comme les autres gènes plus complexes de globines, est transcrit dans une direction **centromère-télomère**. Cette orientation n'est pas la même pour d'autres gènes dans le génome, et pour un gène donné, elle dépend du brin de la double hélice sur lequel le message est lu.
- Les séquences d'ADN requises pour une initiation exacte de la transcription du gène de la β -globine sont localisées dans le promoteur à environ **200 paires** de bases en amont du site du début de la transcription.

- C'est le brin 3'-5' de l'ADN qui est en fait transcrit, mais la séquence 5'-3' de l'ARNm est directement comparable avec le brin 5'-3' de l'ADN (et en fait y est identique, excepté que U est substitué pour T).
- En raison de ceci, c'est le brin 5'-3' du gène (c'est-à-dire le brin qui n'est pas transcrit qui est habituellement donné dans la littérature scientifique.
- En accord avec cette convention, la séquence complète s'étendant sur environ 2 kb du chr 11 incluant le gène de la β -globine est donné à la figure.
- A l'intérieur de ces 2 kb sont contenus la plupart, si ce n'est la totalité, des séquences requises pour coder et réguler l'expression de ce gène.
- Le promoteur de la β -globine est formé d'une série de petits éléments fonctionnels courts qui interagissent avec des protéines spécifiques qui régulent la transcription.
- Une séquence importante de la boîte TATA, une région conservée, riche en adenine et thymine, qui est localisée environ à 25 à 30 paires de bases en amont du site du début de la transcription.
- La boîte TATA est importante pour déterminer la position de départ de la transcription, qui est localisé à environ 50 paires de bases en amont du site de la traduction.
- La 2^{ème} région conservée la boîte CAT (CCAAT), est localisée à quelques dizaines de paires de bases encore plus en amont.

V. GENES DE LA MITOCHONDRIE:

Le génome de la mitochondrie est un ADN extrachromosomique. C'est un ADN circulaire de 16 000 pb, codant pour les ARN de structure, les ARNr et les ARNt, nécessaires aux synthèses mitochondriales. L'ADN mitochondrial est bicaténaire comme l'ADN nucléaire, mais ses deux brins sont codants. La réplication, la transcription et la traduction de l'ADN mitochondrial sont indépendantes de celles de l'ADN nucléaire.

- L'ADN mitochondrial se caractérise par :
 - L'absence d'introns ;
 - L'absence d'histones le protégeant ;
 - L'absence de recombinaison et de système de réparation ;
 - Un taux de mutations, dix fois plus élevé que l'ADN nucléaire ;
 - Un grand polymorphisme génétique
 - L'ADN mitochondrial est transmis par la mère.

- Le code génétique de la mitochondrie diffère légèrement de celui du noyau. Certains codons de l'ADN nucléaire ne correspondent pas aux mêmes acides aminés dans la mitochondrie. Par exemple les codons AGA et AGG qui codent l'arginine dans le noyau ont la signification de codons stop dans la mitochondrie.

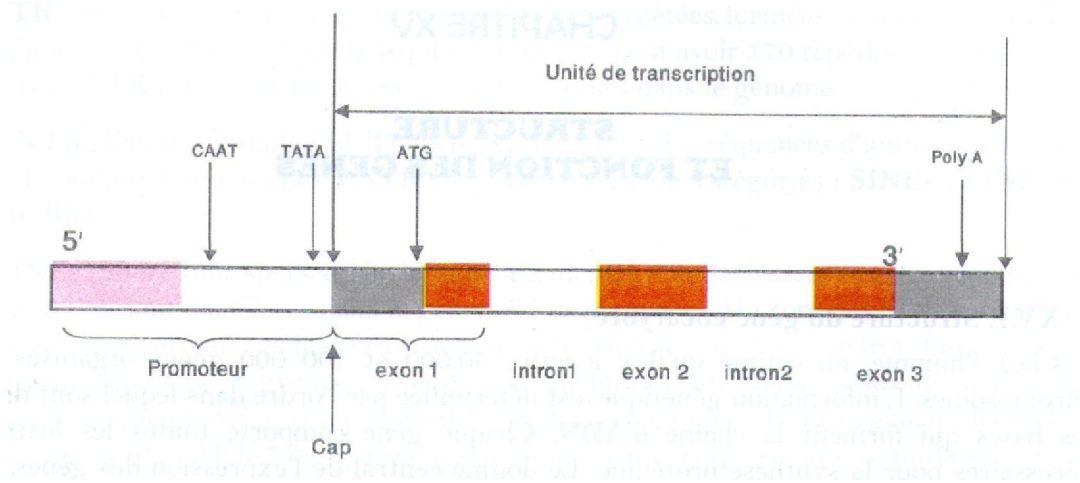


Fig 01 : Structure d'un gène codant pour une protéine.

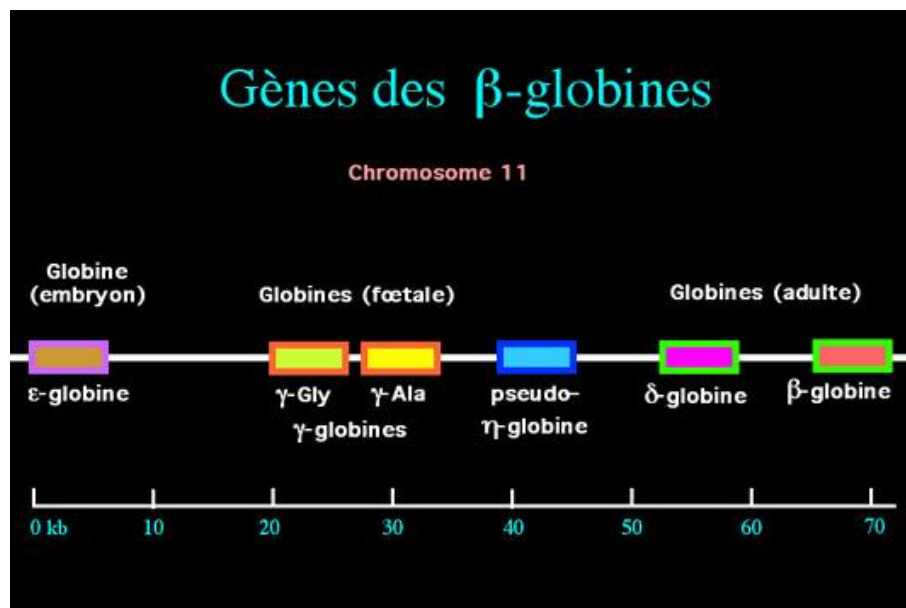


Fig 02 : Gènes de la β-globine.

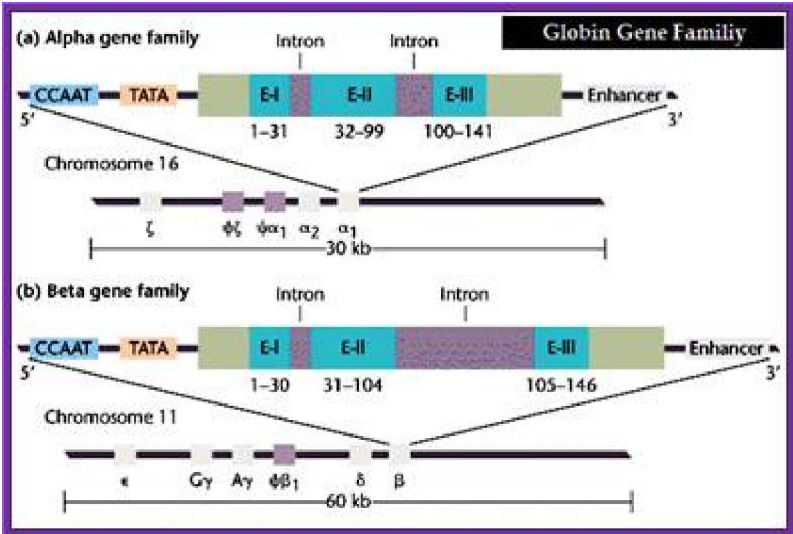


Fig 03: Gènes des globines.