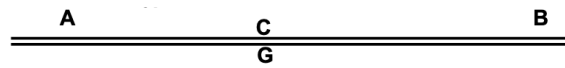


STRUCTURE DU GÉNOME

COMPOSITION DU GÉNOME	
Généralités	- Le génome est l'ensemble du matériel génétique d'un individu ou d'une espèce. Les gènes ne constituent qu'une partie du génome. Celui-ci est constitué de molécules d'acides nucléiques : l'ADN et l'ARN. - Chez l'homme, le génome nucléaire est réparti sur 46 chromosomes, soit 22 paires d'autosomes et deux gonosomes. - La taille du génome se mesure en nombre de nucléotides, ou bases ou paire de bases (pb), puisque la majorité des génomes est constituée de doubles brins. 1000 pb = Kpb (pour kilo-base) 1 000 000 bp = Mpb (méga-base) 1 pg représente environ 1 000 Mpb (978 Mpb précisément). - La taille du génome humain est de 3400 Mbp - Le nombre de gènes est de 25000
Le gène : unité fonctionnelle du génome	ex: exon ■ Portion non codante ■ Portion codante in: intron 1: base de début de la transcription; ATG: codon de début de traduction qui code pour une méthionine (Met); STOP: codon de fin de traduction; AATAAA: signal de polyadénylation Les portions codantes donneront la protéine (et en rouge des portions non codantes mais qui font partis de l'ARNm). - ATG = codon d'initiation à la traduction - Exon 1 = début de la transcription - En 3' du codon stop = séquences régulatrices (comme signal de polyadén)
Séquences codantes et non codantes	→ Les séquences codantes représentent 1-2 % du génome humain → Les séquences non codantes (98-99% du génome) sont : - Des introns des gènes - Des séquences régulatrices des gènes - Des séquences répétées (« repeats ») - Des séquences de fonction structurale (architecture chromosomique de l'ADN) - Des séquences de fonction inconnues (ont-elles une fonction ?) → Répartition très variable d'un génome à l'autre (différence entre drosophile et Homme par ex)
Séquences répétées ou « repeats »	→ Séquences d'ADN trouvées à plusieurs exemplaires dans le génome. - Origine : éléments mobiles (transposons, rétrovirus, etc...) - Chez l'homme, 2 types selon la taille : LINES (Long Interspersed Nuclear Elements) : x Kbp SINES (Short Interspersed Nuclear Elements) : x 100 bp - Alu : 10% du génome

Les variations de la séquence du génome sont très nombreuses (1 sur 2500 bp en moyenne).

On peut rencontrer 3 types de variations :



1- Remplacement d'un nucléotide, Single Nucleotide Polymorphism (SNP):

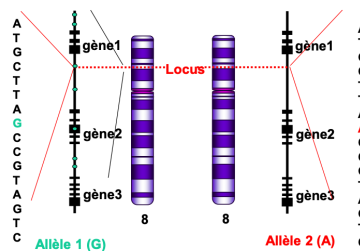


2- Insertion/délétion (ou inversion) d'un segment génomique



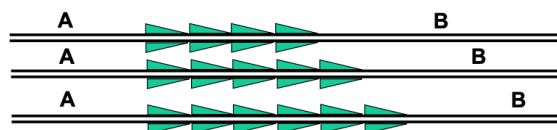
Polymorphisme

- **Polymorphisme** : Toute variation de séquence génomique entraînant l'existence au même locus d'au moins 2 formes différentes de la séquence, appelés allèles, dans la population
- **Le locus** : emplacement d'un segment d'ADN sur un chromosome, défini par son contenu informationnel (gène), ou sa séquence qu'elle soit ou non polymorphe.

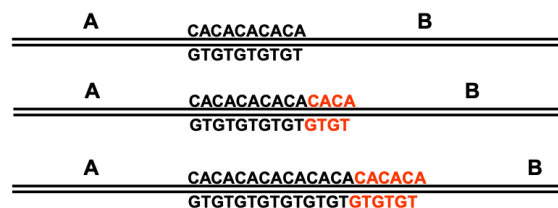


Séquences répétées en tandem et microsatellites

VNTR: variable number of tandem repeats, repeat > 10 pb



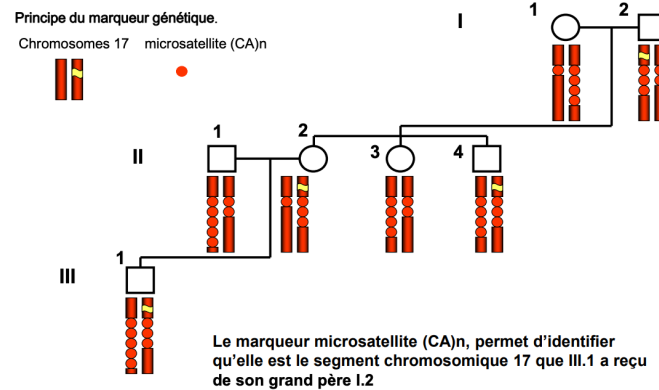
Microsatellites, repeat < 10 pb



Marqueurs en génétique

- **Le marqueur génétique** est un segment génomique de position connue sur le génome qui présente un polymorphisme
- Les marqueurs définissent le **génotype** (constitution génétique d'un individu)
- D'informativité estimée par le **taux d'hétérozygotie connue**.

EXEMPLE :



→ Un marqueur génétique est caractérisé :

- Par **ses allèles** correspondant à chaque variation possible de la séquence
- Son **taux d'hétérozygotie** qui est fonction du nombre et de la fréquence de chacun des allèles :

$$H = 1 - (\sum p_i^2)$$

- Un marqueur est informatif s'il a un **taux d'hétérozygotie élevé**.

Explication du taux d'hétérozygotie avec l'exemple précédent :

Pour faire simple, c'est 100% - tous les homozygotes. Imaginons qu'on ait 3 allèles de fréquence 0,2 / 0,3 / 0,5 → $H = 1 - (0,04 + 0,09 + 0,25) = 0,62$ → Cela signifie que 60% de la pop est hétérozygote pour ce marqueur.

Rq : Un marqueur biallélique ne pourra pas avoir un taux d'hétérozygotie sup à 0,5 qui sera atteint en cas d'allèles de fréquence égale.

PATHOLOGIES DE L'ADN

Généralités

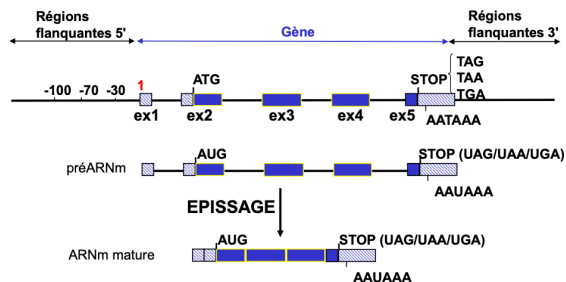
→ Chaque enfant naît avec **60 à 75 nouvelles variations ponctuelles de l'ADN**
: Mais la majorité dans les régions **non codantes**
→ On distingue :

- **Un variant somatique** : variant qui intervient dans une cellule d'un tissu ou d'un organe excepté la lignée germinale
- **Un variant germinale** : variant qui est intervenue dans une cellule de la lignée germinale et qui peut être transmise à la descendance

Types de variants	On distingue 2 types de variants selon leur nature et leurs propriétés : - Les variants simples (stables) : variants qui peuvent toucher toutes les séquences du génome avec une probabilité faible (de l'ordre de 10⁻⁵) - Les variants instables : variants qui surviennent sur des séquences instables formées par la répétition successive d'un motif de quelques nucléotides.
Variants simples	Ils sont de 3 types : remplacement d'une base, petits réarrangements, grands réarrangements.
(1) Remplacement d'une base	- Les transitions transforment la séquence CG ou CA représentent 40% des substitutions : le dinucléotide CpG est un point chaud de mutation. Le mécanisme responsable est la déamination des cytosines méthylées. Transition: C → A, G → T Transversion: C → G, A → T, C → T, G → A Examples: p.G1784A, p.Gly551Asp, p.G551D - A gauche, le C est remplacé par U, mais ça ce n'est pas possible dans l'ADN → donc restauration du CG - A droite, déamination donne un T, le système de surveillance de l'ADN ne le remarque pas → donc variation. Désamination de C → U-G → Restauration de G → C-G Désamination de methyl-C → T-G → Réplication → T-A - La méthylation des cytosines des doublets CpG participe au contrôle de l'expression des gènes par son rôle dans la compaction de la chromatine, empêchant l'accès des protéines (PN) du complexe de transcriptions à leur site de liaison sur l'ADN. En règle générale, la méthylation est donc associée à une répression transcriptionnelle (mais ce n'est pas une règle absolue). Rôle démontré dans le développement, l'empreinte parentale et le cancer. Maternal allele: DMD (methylated) → Repression of expression Paternal allele: DMD (unmethylated) → Expression Methylation and subsequent inactivation of one of the parental alleles of an imprinted gene DMD: domaine méthylé de façon différentielle entre la copie paternelle et maternelle; TC: facteurs de transcription

Cadre de lecture

Définition : le cadre de lecture est défini par le découpage en codon de trois nucléotides de la séquence de l'ARN messager, ce qui permet de déduire la séquence protéique correspondante. Il est déterminé par le **codon d'initiation de la traduction AUG** qui code pour une méthionine et par le **codon de terminaison (STOP)** :



1: début de la *transcription*
 ATG: codon initiateur de la *traduction*
 STOP: codon de terminaison de la *traduction*
 AAUAAA: signal de polyadénylation

- **Le cadre de lecture doit être connu** pour déduire les conséquences d'un variant au niveau de la séquence protéique.
- Chez les eucaryotes, les nucléotides entourant le triplet initiateur AUG sont bien conservés et constituent **une séquence nommée séquence consensus de Kozak** qui permet l'association de l'ARNm sur la petite sous-unité du ribosome.
- **La différenciation entre un triplet initiateur et un triplet codant une méthionine** au sein de la chaîne peptidique est favorisée par cette séquence.

NB : le code génétique est dégénéré. Donc, on a des mutations dites silencieuse ou synonyme (cf. exemple cadre de lecture sur les diapos)

- Ils correspondent à la **perte (délétions) ou ajout (insertions)** d'une à plusieurs bases
- Ils sont détectés dans **les exons ou les séquences introniques flanquantes**
- **La nomenclature :** délétion d'une paire CG en position 413 : 413delCG ou 413del2 / insertion d'un T après la base 121 : 121-122insT, OU insertion de 4 bases après la base 146 : 146ins4

- **Les délétions/ insertions de la séquence codante** qui ne sont pas des multiples de **3 bases** entraînant un décalage du cadre de lecture et conduisent en règle générale à l'apparition d'un codon STOP prématuré (PTC pour « premature termination codon »).

AUG G C C U C U A A C C A U G G C
 Met Ala Ser Asn His Gly ⇒ AUG C U C U A A
 Met Leu STOP

AUG A G C C U C U A A C C A U G G C
 Met Ala Ser Asn His Gly ⇒ AUG A G C C U C U A A
 Met Cys Leu STOP

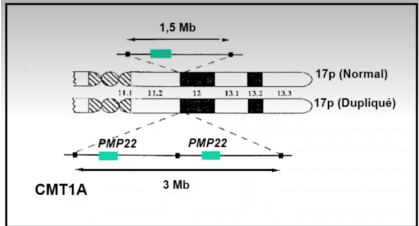
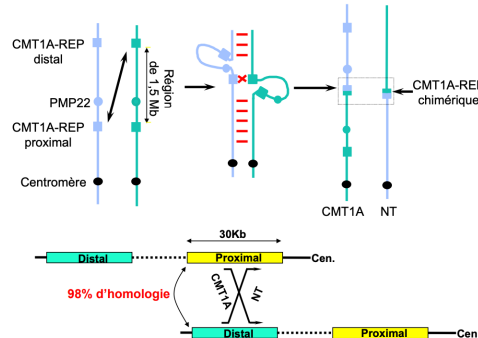
- **Les délétions/ insertions de la séquence codante** qui sont des multiples de 3 bases entraînent la **délétion ou l'insertion** d'un ou plusieurs acides aminés ou la production d'un codon STOP.

ACC
 AUG G C C U G U A A C C A U G G C
 Met Ala Thr Cys Asn His Gly

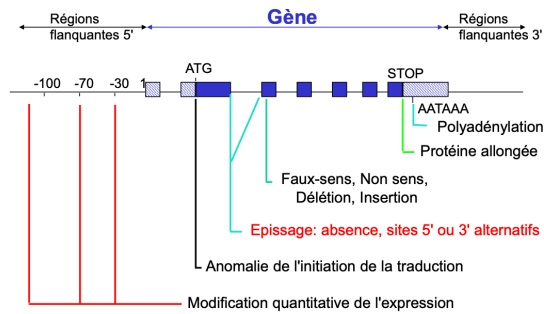
ACC
 AUG G C C U G U A A C C A U G G C
 Met Ala Cys Asn His Gly
 STOP

ACC
 AUG G C C U C U A A C C A U G G C
 Met Ala Ser Asn His Gly
 Lys Pro

(2) Les petits réarrangements

(3) Les grands réarrangements	- Ils correspondent à la délétion ou la duplication d'un segment génomique comportant soit un ou plusieurs exons d'un gène soit le gène en entier. - Généralement, les grands réarrangements ne sont pas caractérisés à la base près. On les identifie par la liste des exons du gène qui sont touchés : Exemple : délétions récurrentes des exons 45 à 47 dans le gène de la dystrophine - Pour les réarrangements qui touchent un ou plusieurs exons du gène, il est crucial de savoir si l'ajout ou la perte de séquence codante décale le cadre de lecture. Dans ce cas, il y a le plus souvent apparition d'un codon STOP prématuré.
Cas particuliers des réarrangements comportant le gène en entier	- La taille de ces réarrangements peut atteindre plusieurs millions de bases. - En pathologie, les délétions sont plus fréquemment observées que les duplications. - En général, pour une pathologie donnée, les réarrangements sont récurrents : la majorité des patients présentent la même région délétée ou dupliquée. - Le modèle : la maladie de Charcot et la neuropathie tomaculaire, 2 pathologies qui touchent la myéline du SNP • Dans le CMT, 30% des patients ont une duplication de la région 17p11.2 (CMT1A) • Dans la NT, 90% des patients ont une délétion de la région 17p11.2 EXEMPLE : Dans la région, 40 gènes ont été décrits : le gène PMP22 qui code pour une PN de la myéline du SNP, est responsable du phénotype. Chez les patients avec duplication, il y a 1,5 x ARNm et 1,5x PN PMP22. PMP22 est donc le seul gène de la région dont l'effet de dose est délétère. 
Recombinaison inégale entre les 2 chromosomes 17 homologues	 De part et d'autre de la région qui est dupliquée on a des repeats (longs 30 Kb). Dans certains cas, le REP distal ne reconnaît que le REP proximal car ils ont beaucoup d'homologies donc on obtient 2 chromosomes différents (le dupliqué et le délété). C'est pour ça qu'on a tjrs la même duplication/ délétion (de même taille).
Conclusion	- Les grands réarrangements contenant un gène en entier sont le plus souvent dus à des recombinaisons inégales par mauvais appariements entre des séquences homologues de grande taille (x10kb à x100kb) qui flanquent la région - Ces « repeats » sont concentrés dans les régions péri-centromériques - Ce mécanisme est symétrique mais pourtant plus de délétion identifiée que de duplications : <i>les duplications sont-elles létales ? Les duplications du gène ne donnent-elles pas de phénotype ? Les duplications du gène donnent-elles des phénotypes très variés difficile à définir au plan clinique ?</i>

Conséquences des variants en fonction de leur localisation



- Une mutation dans le promoteur peut modifier quantitativement l'expression de la protéine
- Une mutation ATG → ACG par ex → on n'a plus de méthionine, donc anomalie de l'initiation de la traduction
- Une mutation sur le codon STOP → protéine allongée
- Épissage...

Conséquences des mutations des sites d'épissage

Les séquences qui interviennent dans l'épissage:



Le mécanisme d'épissage:



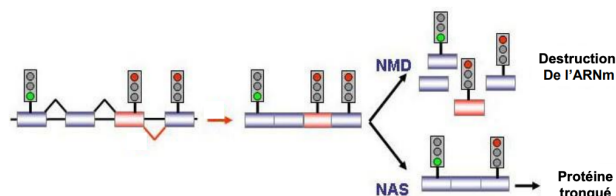
En conclusion, les conséquences d'un variant des sites d'épissage peuvent être variées :

- **Absence d'épissage de l'intron correspondant** : l'intron est donc retenu dans l'ARNm
- **Le démasquage d'un autre site d'épissage** (site donneur ou accepteur détruit, la machinerie cherche ailleurs)

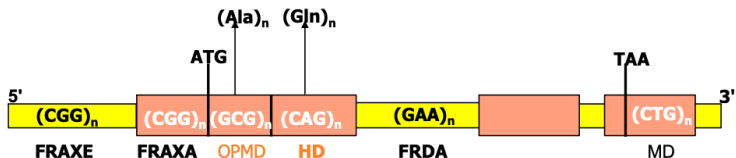
Dans les 2 cas, il y a le plus souvent un **décalage du cadre de lecture**.

Conséquences fonctionnelles des codons STOP prématurés

- **Les codons STOP prématurés** (càd présent sur l'ARNm avant le codon STOP légitime), sont reconnus sur l'ARNm par un système de surveillance « NMD » (pour nonsense-mediated mRNA decay) qui détruit l'ARNm qui en est porteur :



NMD: Non-sense mediated decay;
NAS: Non-sense mediated alternative splicing.

	Les conséquences d'un codon STOP prématuré (PTC) peuvent être de 2 ordres : → La synthèse d'une PN plus courte. Dans certains cours, elle a : - Une activité même partielle - <u>Des interactions (délétères) avec ses partenaires</u> - <u>Des interactions (délétères) avec la protéine normale</u> → La destruction par le système NMD de l'ARNm porteur : la PN sera donc synthétisée à partir de la copie normale du gène mais en moindre quantité : - Si une baisse de la Q de PN est délétère pour un ou plusieurs types cellulaires, alors il y a survenue d'une pathologie (« effet de dose ») - Si une baisse de la Q de protéine n'est délétère pour aucun type cellulaire, alors il n'y a pas de pathologie. Les études récentes montrent que la grande majorité des PTC sont soumis à la NMD, ce qui prévient la synthèse de PN délétères et l'apparition d'une pathologie quand le gène n'est pas soumis à un effet de dose.
Fréquence des différents types de variants	- Les variants de petites tailles sont en moyenne majoritaire : Substitutions, délétions/insertions de quelques pb. - Mais très variable d'un gène à l'autre : Dans la Neuropathie tacomulaire, la délétion de la région 17p11.2 correspond à 90% des mutations (les autres : petites délétions dans le gène PMP22) - Sur le chromosome X : 5-95% des mutations sont des grandes délétions : 2/3 des mutations sont des délétions d'au moins 1 exon dans la myopathie de Duchenne
Les variants instables	- Les séquences instables sont des répétitions successives et homogènes de trinucleotides comme (CGG)_n dans le syndrome de l'X fragile ou (CAG)_n dans la maladie de Huntington. - Plus de 25 maladies héréditaires sont dues à des expansions de triplets. - Il y a 2 types de pathologies, celles associées à des répétitions dans les séquences codantes (répercussion sur la structure de la PN) et celles avec des répétitions dans les séquences non codantes (Expression du gène) : Exemples à ne pas connaître  FRAXA: Syndrome de l'X fragile FRAXE: X linked mental retardation OPMD: Dystrophie musculaire oculopharyngée; HD: Maladie de Huntington FRDA: Ataxie de Friedreich MD: Dystrophie Myotonique de Steinert exons introns

Répétitions de triplets CAG dans les séquences codantes	- Elles entraînent des pathologies neurodégénératives par destruction sélective de certaines structures cérébrales par mort des neurones (Maladie de Huntington) - L'expansion du poly (CAG)_n va être traduite dans la protéine par une expansion de polyglutamines qui va induire une nouvelle conformation de la protéine. - Les malades sont hétérozygotes pour l'expansion du CAG. Leurs cellules contiennent des protéines normales et anormales :
Répétitions de triplets CAG dans les séquences NON codantes	- Ces expansions peuvent être de très grande taille, plusieurs centaines de répétition de triplets (jusqu'à 6000 répétitions pour moins de 200 pour les expansions de triplets dans les séquences codantes) - Elles sont associées à des modes de transmission variés ; AD, AR et lié à l'X - Leurs conséquences au niveau cellulaire sont variées
Modes de transmission	- Les modes de transmission découlent des conséquences fonctionnelles des variants - Ils sont déterminés par un caractère particulier des variants dont on classe les conséquences → En perte de fonction : le variant ne permet plus à la protéine d'assurer sa fonction dans les cellules. OU → Gain de fonction : le variant acquiert une (des) propriétés qui a (ont) une influence délétère sur la physiologie d'au moins 1 type cellulaire.
Mode autosomique récessif	- Le variant de chacune des copies du gène entraîne une perte de fonction. - Chez les hétérozygotes, la présence d'une seule copie normale du gène, donc de la moitié de la quantité de protéine observée chez un individu normal, est suffisante pour assurer la fonction au niveau cellulaire - La plupart des enzymopathies ont un mode de transmission autosomique récessif : Phénylcétonurie ; Le déficit en 21 hydroxylase ; La maladie de Gauche
Mode autosomique dominant	- Une maladie sera transmise sur le mode autosomique dominant si les conséquences du variant de l'une des deux copies du gène ne sont pas compensées par la copie normale du gène. - Le mode de transmission dominant peut être engendré par 2 mécanismes moléculaires distincts : Un déséquilibre quantitatif (haploinsuffisance, duplication) ; Mutation conférant des propriétés nouvelles à la protéine mutée.

Déséquilibre quantitatif	- <u>Le déséquilibre quantitatif (DQ) peut être rencontré :</u> Pour toutes les délétions (perte de fonction) et duplications (gain de fonction) de gènes ; Pour les variants qui entraînent un codon Stop prématuré soumis à la NMD (perte de fonction) - <u>Cas particulier des enzymopathies :</u> <i>on peut citer les porphyries par déficit d'un des enzymes de biosynthèse de l'hème qui sont dus le plus souvent à une haploinsuffisance et sont de transmission AD (perte de fonction).</i> <u>Haploinsuffisance :</u> il y a haploinsuffisance quand une seule copie n'est pas suffisante pour assurer sa fonction normale. - En conclusion, la majorité des maladies AD sont dues à un déséquilibre quantitatif.				
Variant conférant des propriétés nouvelles à la protéine	- Le variant de l'une des deux copies du gène ne l'inactive pas mais introduit <u>un caractère nouveau</u> à l'origine de la maladie (gain de fonction) : - Variant conférant à la protéine une toxicité - Variant auto actif - Variant dominant négatif				
Conclusion	<table border="1"> <thead> <tr> <th data-bbox="402 981 831 1021">Gain de fonction</th><th data-bbox="831 981 1248 1021">Perte de fonction</th></tr> </thead> <tbody> <tr> <td data-bbox="402 1021 831 1227"> - <u>Dominant :</u> gène surnuméraire, variant toxique, variant activateur (oncogène, récepteur), variant à effet dominant négatif </td><td data-bbox="831 1021 1248 1227"> - <u>Dominant :</u> haplo-insuffisance - <u>Récessif :</u> perte de fonction </td></tr> </tbody> </table>	Gain de fonction	Perte de fonction	- <u>Dominant :</u> gène surnuméraire, variant toxique, variant activateur (oncogène, récepteur), variant à effet dominant négatif	- <u>Dominant :</u> haplo-insuffisance - <u>Récessif :</u> perte de fonction
Gain de fonction	Perte de fonction				
- <u>Dominant :</u> gène surnuméraire, variant toxique, variant activateur (oncogène, récepteur), variant à effet dominant négatif	- <u>Dominant :</u> haplo-insuffisance - <u>Récessif :</u> perte de fonction				