

Génétique 2023/2024



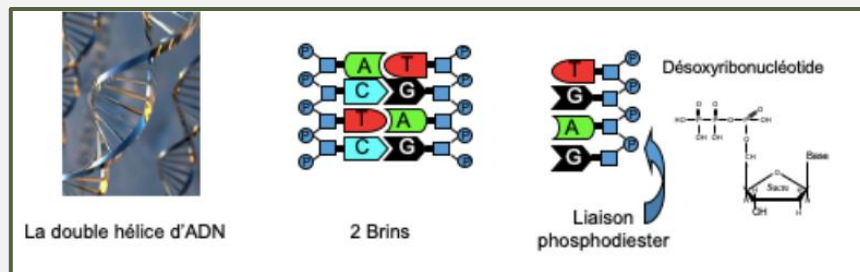
Cours 4 : Séquençage et PCR

(recherche des mutations dans un gène donné)

I) Le séquençage ADN

Le séquençage d'un fragment d'ADN est une technique qui permet de déterminer la succession des nucléotides qui le compose.

Pour rappel, l'ADN est cette **double hélice composée des 4 nucléotides A, T, C, G**, reliés par des **liaisons phosphodiester** (dans un même brin), avec ces deux brins **antiparallèles**, liés entre eux par des **liaisons hydrogènes** (entre les deux brins).

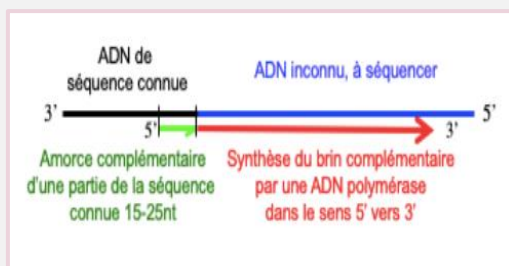


Le principe du séquençage d'ADN basé sur les **di-désoxynucléotides** (aussi appelée **méthode de Sanger**). C'est une méthode **enzymatique**, aujourd'hui la **méthode de référence** pour séquencer l'ADN.

Malgré l'apparition de nouvelles technologies comme le Séquençage Haut Débit, la **méthode de Sanger reste la méthode de référence**.

1) Principe

Nous allons voir les différentes étapes de ce séquençage qui sont très proches de la méthode PCR.



En effet, nous allons utiliser une **ADN polymérase**. Les **polymérases sont des enzymes qui permettent de synthétiser un brin complémentaire d'ADN à partir d'une amorce** (petit brin d'ADN simple brin complémentaire

d'une séquence d'ADN donnée). Ce sont les **mêmes types d'amorces** que celles utilisées pour la PCR.

Il va y avoir synthèse d'un brin d'ADN complémentaire grâce à l'ADN polymérase qui va synthétiser, dans **le sens 5' 3'**, un **brin complémentaire du brin que l'on veut séquencer**. La reproduction est fidèle à la séquence d'ADN utilisée. C'est la même chose que la PCR mais **on a un seul primer**.

2) Les étapes

« L'ADN polymérase synthétise, à partir d'une amorce, un brin complémentaire, fidèle à la séquence d'ADN à étudier, grâce à des cycles successifs (3à à 35 cycles) de 3 étapes : Dénaturation - Hybridation – Élongation ».

Explications : Dans un tube, on va mettre :

- L'ADN à séquencer, qui peut être le résultat d'un produit PCR,

- La **Taq Polymérase**,

- L'**amorce** (1 seule) qui doit être **spécifique** du brin que l'on veut séquencer (si on veut en séquencer deux, alors on utilise une autre amorce dans une autre tube : ici, une seule amorce),

- Le matériel nécessaire à la Taq polymérase et à la synthèse du nouveau brin, à savoir les **nucléotides** (DNTPs),

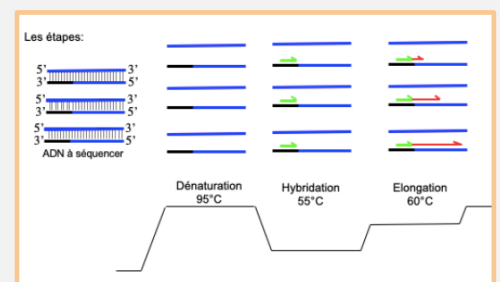
- Les **tampons**.

Nous allons ensuite avoir des **variations successives de températures** de façon à :

1. Dénaturer à 95°C notre double brin d'ADN : les liaisons hydrogènes qui lient les deux brins cassent à cette température.

2. Hybrider selon la complémentarité notre amorce sur les ADN devenus simples brins, à 55°C

/ ! \ **ATTENTION** : ici il n'y a **qu'une seule amorce** complémentaire à l'un des deux brins. (Dans la PCR, vous aviez 2 amorces : la complémentaire et la reverse)



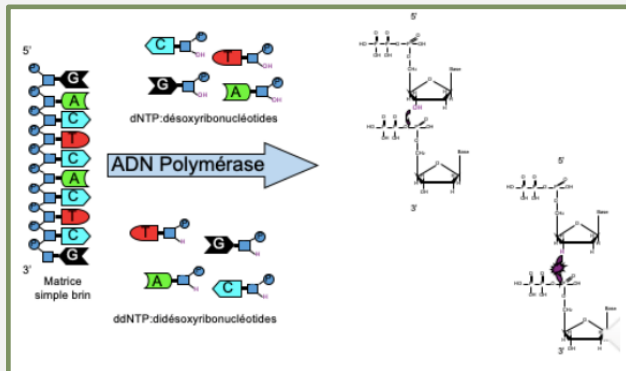
3. Élongation du brin, à 60°C (température optimale pour la Taq polymérase).

Le brin complémentaire au brin que l'on veut séquencer se forme.

/ ! \ ATTENTION : L'élongation est à 60°C ici et pas 72°C comme pour la PCR.

(Vous voyez les pièges venir.....)

L'autre particularité pour l'étape de séquençage est que l'on va utiliser un **mélange de dNTPs** (désoxyribonucléotides ATCG), et de **ddNTPs** (didésoxyribonucléotides) +++.

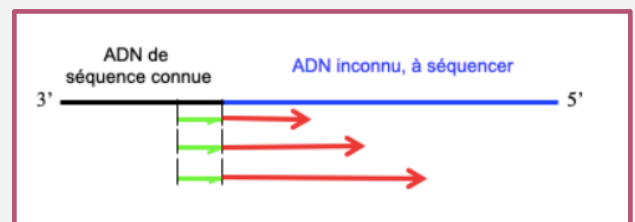


Lors de la synthèse, la polymérase va utiliser **aléatoirement**, en fonction de la complémentarité des bases, **des dNTPs ou des ddNTPs**. L'utilisation des dNTPs permet la création des **liaisons phosphodiester**, entre un **nucléotide** et le **suivant**, et donc de **continuer** la **synthèse** du brin. En revanche, un **ddNTP** est un **dNTP** mais avec un

groupement **H** au lieu d'un groupement **OH** : la **liaison** avec le nucléotide suivant **ne peut pas se faire**. À partir du moment où un **ddNTP** est **introduit**, la **synthèse s'arrête**, à la différence de l'introduction de dNTPs.

3) Cycles successifs

Comme l'introduction des dNTPs et des ddNTPs se fait **au hasard**, et on va réaliser des **cycles successifs de dénaturation des brins – hybridation des amorces – élongation du brin complémentaire**. On se retrouve à la fin avec de très nombreux fragments d'ADN, de **tailles différentes**, complémentaire de l'ADN qu'on veut séquencer.



ATTENTION : vous remarquerez qu'on obtient **le brin complémentaire**. À la fin, il ne faudra pas oublier que ce n'est pas le brin que l'on veut séquencer que l'on a mais bien son brin complémentaire !!

4) Histoire

Historiquement, nous utilisons la **Méthode Sanger** avec laquelle nous faisons **4 réactions distinctes**, avec dans chacune des réactions, un **seul type de ddNTP** (soit A, soit C, soit T, soit G), et tous les **dNTPs**.

Chacun des tubes contenaient tout ce qu'il fallait pour que la polymérase puisse séquencer :

- Le fragment d'ADN,
- La polymérase,
- L'amorce,
- Les dNTPs (ne stoppant pas la synthèse),
- **Un seul type de ddNTPs (stoppant la synthèse).**

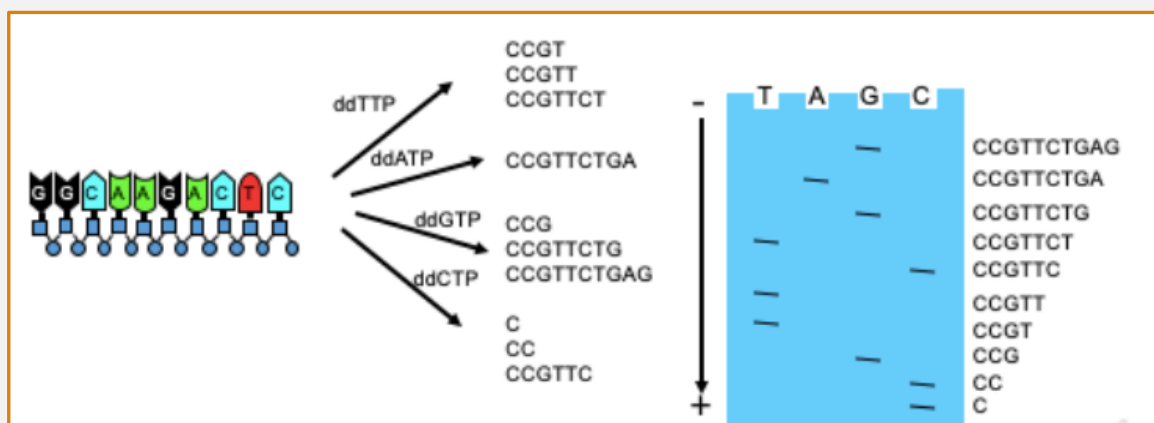
On avait donc **4 tubes** comme on peut le voir sur ce schéma. Ces fragments obtenus étaient séparés sur un **gel** par **électrophorèse**, puisqu'on a vu que si l'on veut séparer l'ADN en fonction de sa taille, on réalise une migration électrophorétique.

Nous avons donc ce genre de résultats, avec sur les différentes pistes, différents morceaux d'ADN migrants **plus ou moins loin en fonction de l'endroit où le ddNTP était introduit** (on avait donc pleins de brins de tailles totalement différentes).

L'identité des nucléotides est apportée **en fonction des pistes dans laquelle on va lire le fragment** (T, A, C ou G) et l'ordre des nucléotides se fait par rapport à la **taille des fragments**.

De plus, il faut savoir que plus **un morceau d'ADN est lourd, plus il a de difficultés à migrer. Donc on lit de haut en bas** (du plus léger / plus court au plus lourd / plus long).

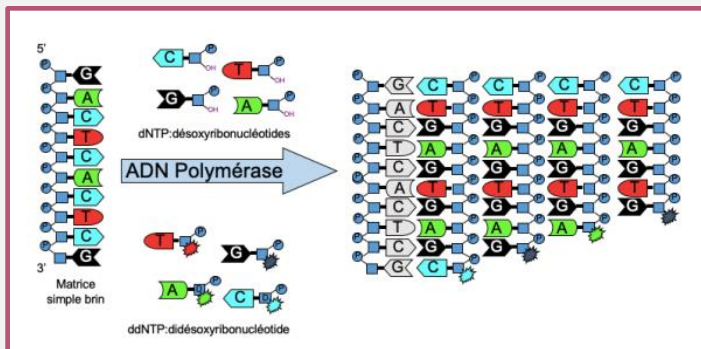
C'est ainsi que, dans cet exemple, le brin à séquencer étant GGCAAGACTC, nous obtenons du C (le plus bas). **CCGTTCTGAG (soit son brin complémentaire +++).** La première synthèse à s'être arrêtée, correspondant au ddNTPs complémentaire au premier DNTP de l'ADN à séquencer, donne le fragment qui ira le plus loin, soit le premier nucléotide complémentaire.



5) Méthode automatisée

La méthode Sanger a, par la suite, été automatisée et s'est nettement **simplifiée** grâce à l'utilisation de **ddNTPs fluorescents**. Chaque ddNTP est couplé à un **fluorochrome** de couleur différente. **Les 4 ddNTPs sont ajoutés dans le même tube**

réactionnel. L'ADN polymérase synthétise, à partir d'une amorce, un brin complémentaire fidèle à la séquence d'ADN à étudier en incorporant au hasard des dNTPs ou des ddNTPs.



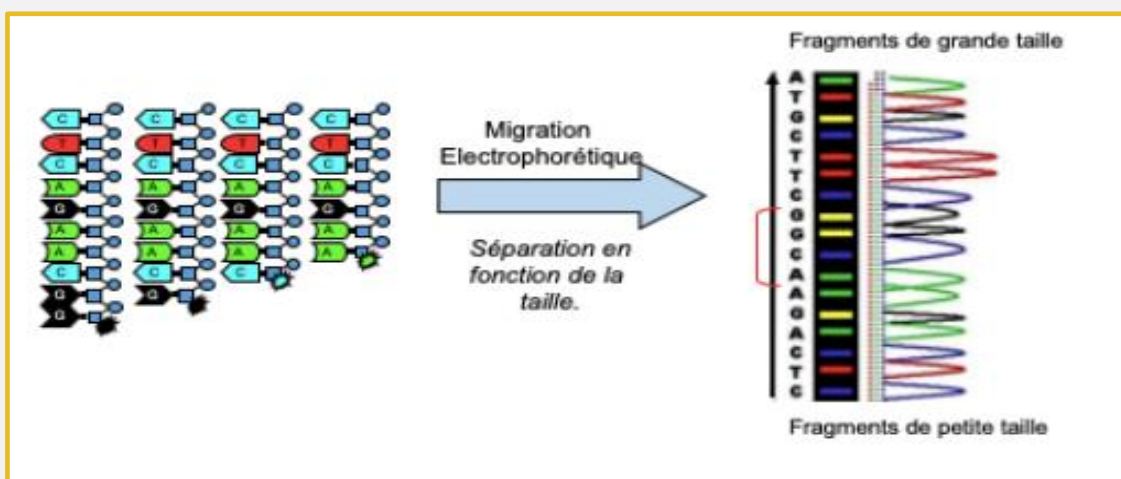
Chaque ddNTP est couplé à un fluorochrome de couleurs différentes. L'incorporation d'un ddNTP stoppe la synthèse.

Le code est : T → Rouge / A → Vert / G → Noir / C → Bleu (sachant que si on obtient un vert, ça veut dire que c'est un A, et par complémentarité, que le

brin que l'on veut étudier contient un T).

Il se passe donc **exactement la même chose que dans la méthode Sanger** mais dans **un seul et même tube** puisqu'on **mélange les 4 ddNTPs**. C'est grâce à un **code couleur** que l'on va pouvoir savoir quel nucléotide va être incorporé (et non plus selon le puits).

Les produits synthétisés sont séparés, en fonction de leur **taille par migration électrophorétique**. L'enchainement des nucléotides est apporté en fonction de la **séparation des fragments synthétisés** et l'identification des nucléotides se fait grâce à ces **fluorochromes de couleur différente en fonction du nucléotide incorporé**.



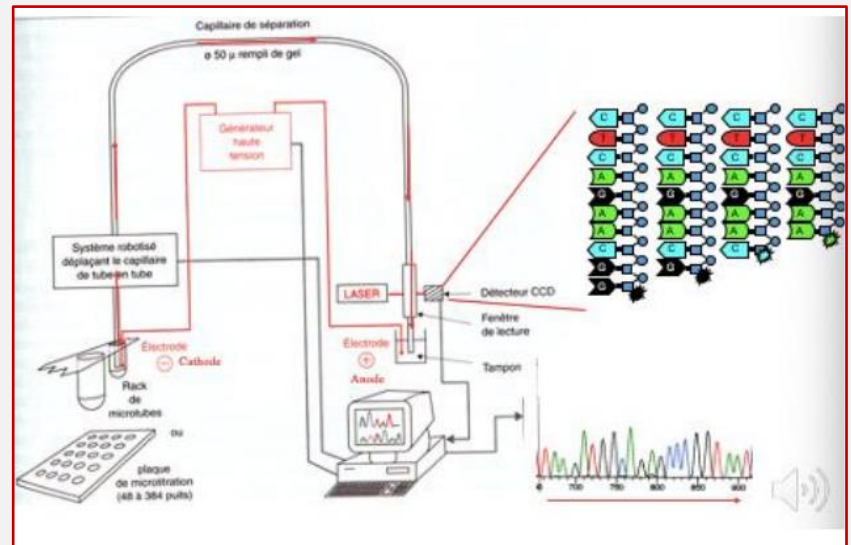
6) Les séquenceurs automatiques

Cette migration électrophorétique et l'identification des nucléotides se fait dans ce que l'on appelle les « **Séquenceurs Automatiques** » dans lesquels on retrouve des **petits capillaires**. Chaque capillaire correspond à un **gel d'agarose** et donc à une **migration électrophorétique**.

Les fragments d'ADN vont ensuite passer devant une **caméra** qui va lire le **fluorophore** qui a été incorporé. Plus en détails, la machine trempe chacune de ses **extrémités dans un tampon avec deux électrodes** : d'un côté une **cathode** et de

l'autre **l'anode** de façon que la migration puisse faire de la **cathode vers l'anode**. Ce capillaire est également rempli d'un **gel de polymère particulier** (équivalent à de l'agarose vu précédemment). Ceci va permettre de séparer **l'ADN en fonction de sa taille** pour donner suite à la **migration dans un champs électrophorétique**.

À gauche nous avons les tubes de séquences, injectés dans ces capillaires. Les préparations vont migrer dans un champ électrique. Les fragments vont ensuite passer devant une caméra qui va, grâce à un laser, identifier les fragments et la couleur des fragments qui passent.



Donc en fonction du ddNTPs, la caméra va voir plutôt un A incorporé si c'est du vert...

Par traitement informatique, on a ensuite la séquence qui apparaît sous forme **d'électrophorégramme**. L'enchaînement des nucléotides est apporté par migration électrophorétique qui va séparer les fragments d'ADN en fonction de leur taille et l'identité des nucléotides est apportée par la couleur des différents ddNTPs incorporés.

COULEUR → Nom du ddNTPS / POIDS SUR ÉLECTROPHORÈSE OU MIGRATION → Position.

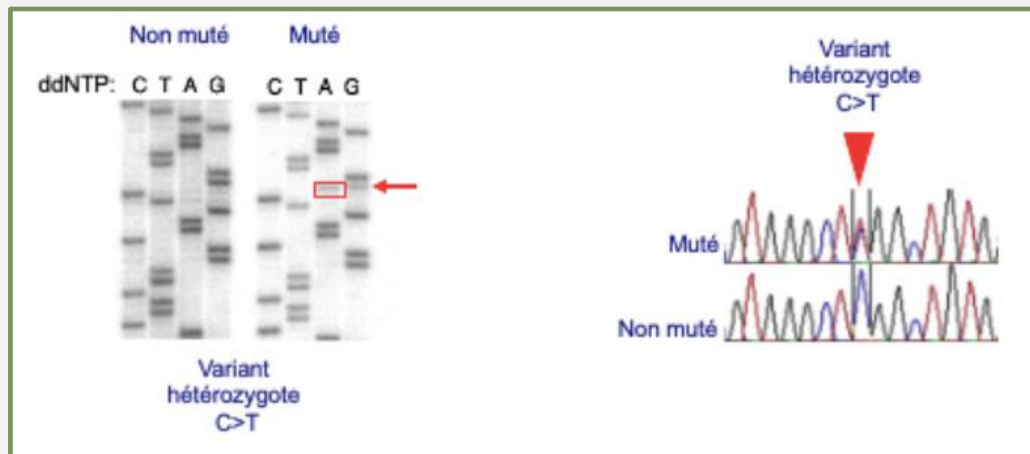
7) Exemple de résultats

Nous avons ici quelques résultats :

Sur la gauche, nous avons des résultats de la techniques utilisés en 77, avec des **ddNTPs radiomarqués dans 4 réactions différentes**. Nous avons 4 tubes par patients, avec le patient de gauche sain et le patient de droite muté. On note une mutation au niveau de la flèche est présente : nous avons un fragment arrêté avec un ddNTP A, et un autre arrêté avec un ddNTP G. Cela signifie que pour cette position-là, nous avons un **hétérozygote C → T** (puisque'on parle en complémentarité : si c'est un A qui a stoppé la synthèse, c'est que nous avons un T).

Sur la droite, la **méthode automatisée**. Aujourd'hui, cette même mutation serait lue avec cette façon. C'est à dire que nous avons des **électrophorégramme** avec des couleurs pour chaque nucléotide.

Chez le patient muté, on a une **superposition de 2 pics puisqu'il est hétérozygote** : le fragment a la même taille, mais comme nous avons 2 allèles différents, nous avons **deux possibilités d'incorporation de nucléotides**. On lit donc un C et un T, soit la superposition des 2 pics.



8) Résumé

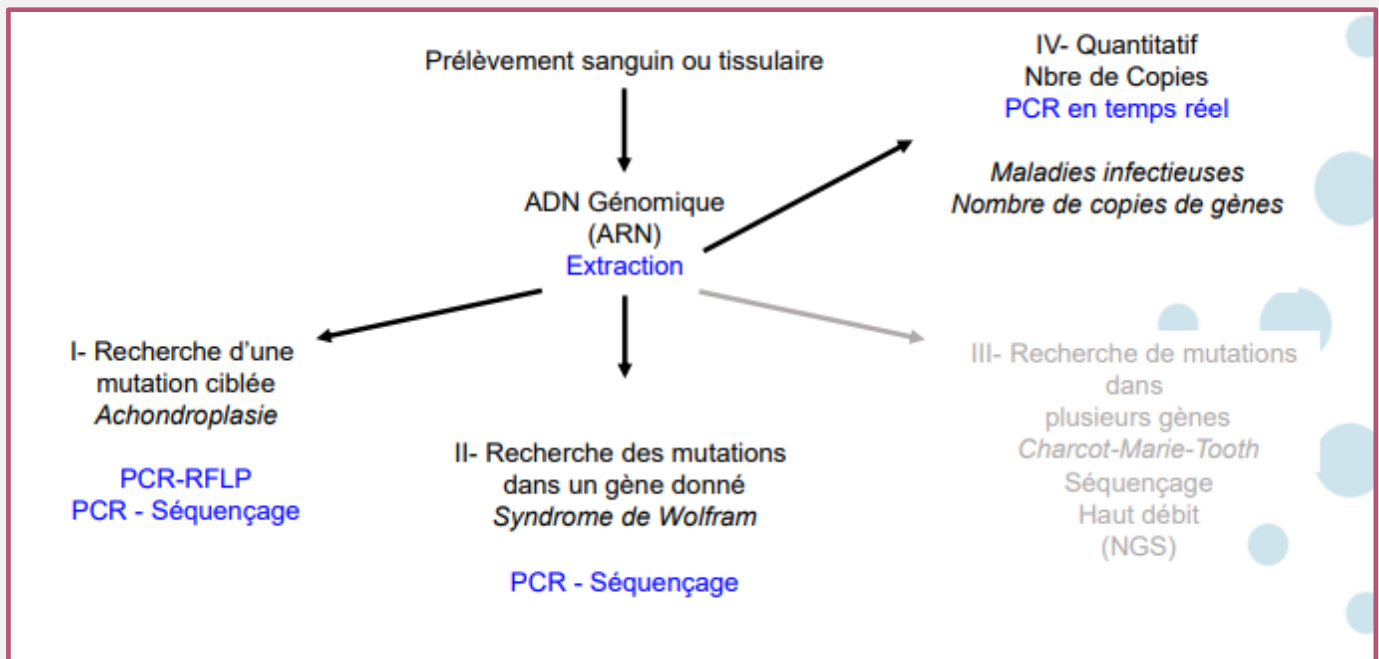
Jusqu'à présent, on a montré comment à partir d'un prélèvement, notamment à partir des **globules blancs**, on est capable d'extraire **l'ADN génomique** (on peut travailler sur l'ARN mais moins fréquemment).

On extrait cet ADN génomique, et dans le cas de **l'achondroplasie** par exemple (lorsque l'on va rechercher une mutation en général) : on **sait quelle mutation** rechercher, dans quel gène, quel changement nucléotidique attendu, on utilise la **PCR-RFLP** (PCR + Enzyme de restriction), dont le résultat peut être confirmé par une deuxième technique telle que la **PCR suivie d'un séquençage**.

Cette stratégie est appliquée pour l'achondroplasie, mais elle peut être appliquée pour **toute recherche d'une mutation ciblée**. C'est une technique qui va **rapidement** pouvoir nous donner **une réponse**.

Nous avons également vu la manière dont d'autres techniques de biologie moléculaire sont appliquées et sont développées, lorsque les questions sont un peu plus compliquées (screen de la totalité d'un gène, recherche de mutation dans plusieurs gènes, voire comment on peut quantifier un nombre de copies, en particulier lors des maladies infectieuses par exemple).

Ainsi, en fonction de la question posée, nous savons quelle technique utiliser (la plus adaptée).

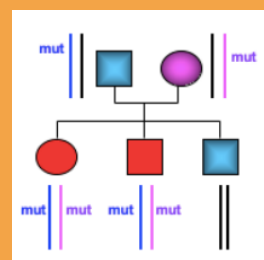


II) Analyse d'un gène par PCR et séquençage Sanger

Nous allons voir maintenant une application de l'utilisation de la PCR et du séquençage, mais cette fois-ci dans **le but de screener la totalité d'un gène**. On ne sait **pas quelle mutation est présente**, donc on va séquencer l'intégralité du gène pour identifier le variant pathogène.

Le syndrome de Wolfram associe d'un point de vue clinique :

- ✚ Diabète,
- ✚ Surdit ,
- ✚ Atrophie optique,
- ✚ Troubles neurologiques.



C'est une maladie à **transmission autosomique r cessive**, ce que signifie que les individus atteints ont forc ment **les 2 all les mut s** pour que les signes apparaissent. Le g ne WFS1 est un g ne constitu  de **8 exons**, et l'ATG (codon permettant le d but de la traduction de l'ARN messager) se trouve dans **le deuxi me exon du g ne**. Cela signifie donc que le premier exon est **non codant**. La **Wolframine**

(protéine codée par le gène WFS1) a une fonction encore peu connue, très étudiée. Elle jouerait un rôle dans le **flux calcique**.

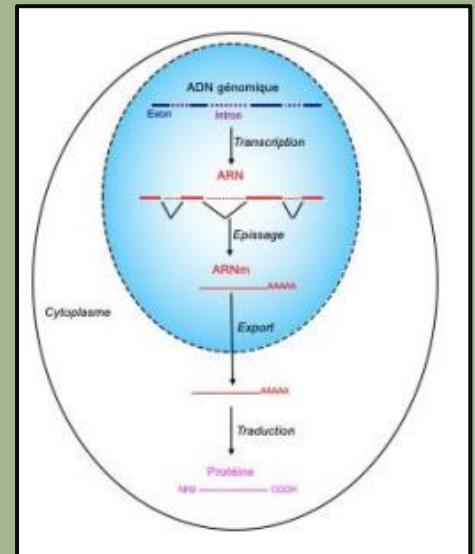
Petit Rappel :

Dans le noyau on va retrouver l'**ADN génomique organisé en chromosomes** (contenant les gènes constitués d'exons et d'introns).

Ces gènes vont être **transcrits en ARN**, et après les mécanismes **d'épissage**, il ne restera plus que les **régions codantes**.

Les **introns** seront donc **éliminés** pour obtenir un **ARN messager mature** avec, à leur extrémités 3', une queue Poly A.

Ces ARN messagers vont ensuite être **exportés du noyau vers le cytoplasme** où ils seront **traduits en protéines**.



Il faut bien retenir la terminologie : traduction, épissage, transduction...

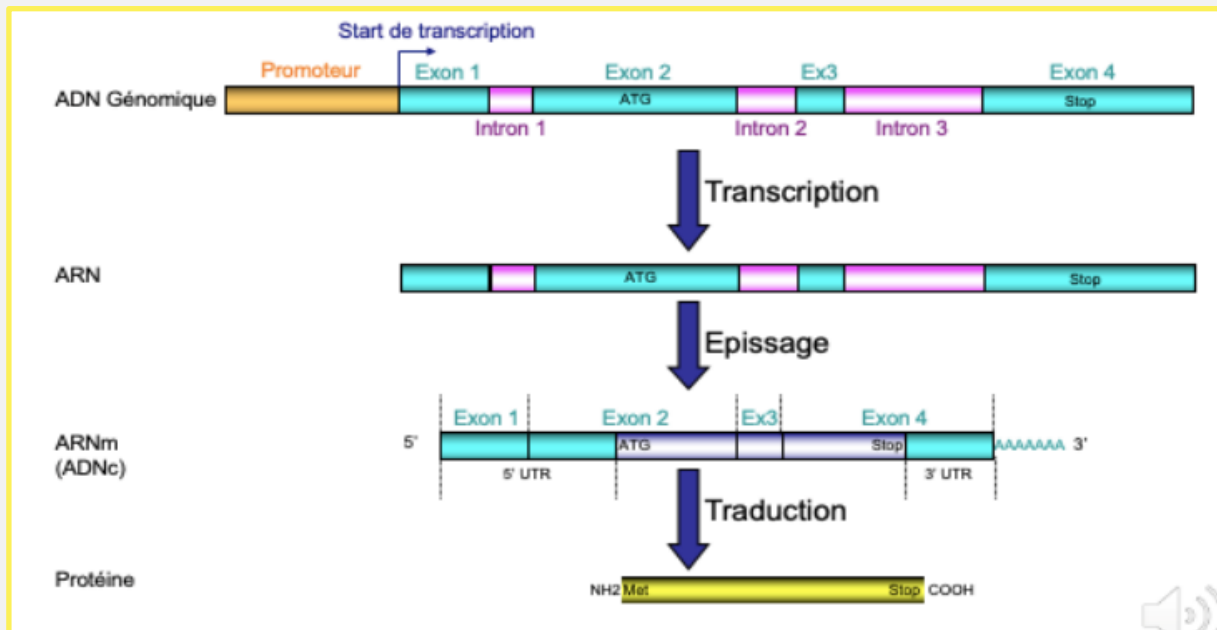
1) L'organisation et l'expression des gènes

Si on regarde de plus près l'organisation et l'expression des gènes que l'on représente sur la première ligne, on a l'**organisation génomique avec en 5' la région promotrice**. Ce 5' est le lieu où vont **se fixer les protéines** qui vont permettre le **démarrage de la transcription de ce gène en ARN**. C'est le start d'inscription qui va permettre de définir le **cadre de lecture**. On retrouvera ainsi les **régions codantes** (exons), séparées par les **régions non codantes** (introns).

Après transcription, on obtient **la copie en ARN de cet ADN génomique**. Cet ARN va ensuite être **épissé** (on va retirer par épissage les introns). Cet ARN messager sera **ensuite traduit en protéine**.

Cependant, il faut se rappeler que **la traduction ne commence pas forcément aux premiers nucléotides du premier exon**. Elle commence à partir de du codon **ATG** (dans cet exemple, il se trouve sur le deuxième exon : le premier n'est donc pas traduit). Ceci sous-entend que si le codons ATG se trouve dans l'exon deux, tout ce qui est en amont de cet **exon ne sera pas traduit, même si la séquence est présente dans l'ARN messager**. La région en amont du codons ATG (codon START) est appelée **5' UTR** (non codantes) et la région en aval du codon STOP est appelée **3' UTR**, cette région est également non codante.

Dans cet exemple, le codon stop est situé dans l'exon 4. La protéine commence donc au niveau de l'ATG (méthionine, premier codon codant), et se termine au niveau du codon stop.

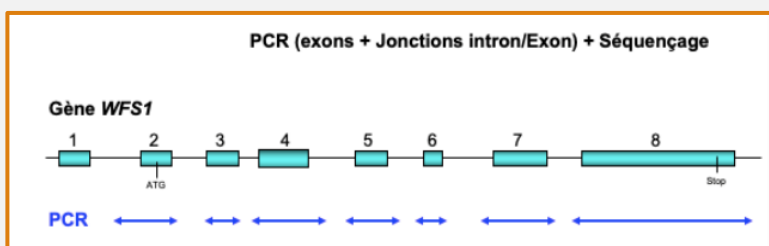


2) Recherche de mutations dans un gène

Si on reprend notre exemple de départ, à savoir le screening du gène WFS1, on remarque que ce gène est organisé en huit exons, et l'ATG se situe dans l'exon 2. Notre codon 1 doit donc être non traduit.

Pour le diagnostic de cette pathologie, il faut pouvoir **identifier** le variant nucléotidique responsable d'un dysfonctionnement de la Wolframine, qui va modifier la fonction de cette protéine.

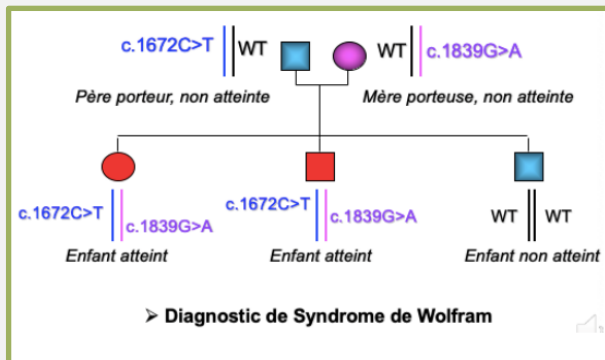
De ce fait, on va particulièrement s'intéresser aux **régions codantes**. C'est-à-dire que l'on va amplifier les exons codants parce qu'on sait facilement **interpréter des variants nucléotidiques situés dans les exons**. On va facilement pouvoir savoir si tel ou tel variant nucléotidique modifie un acide aminé et quelle va t'être la conséquence de cette modification d'acide aminé sur la **protéine**.



C'est pourquoi on va réaliser **7 PCR** qui vont amplifier les régions codantes : les PCR seront faites sur les exons 2 / 3 / 4 / 5 / 6 / 7 / 8. On n'amplifie pas l'exon 1 un **car il est non codant**. S'il y avait une mutation dans cet exon, il

serait très difficile de l'interpréter et de savoir quelles conséquences seraient engendrées puisqu'il n'est pas exprimé.

Prenons l'exemple de cette famille : on a amplifié les 7 exons, on les a séquencés, et on a **identifié chez le père un variant en position 1672** (qui modifie un C et T) à l'état **hétérozygote**.



Chez la mère, **un deuxième variant apparaît**, toujours sur le gène WFS1, mais en position 1839.

On voit donc que les **deux enfants atteints** ont hérité des **deux gènes mutés** : celui du **père** et celui de la **mère**.

En revanche le **troisième enfant** a hérité des **gènes non mutés**.

Le résultat permet de poser le diagnostic de Syndrome de Wolfram. Cette pathologie étant **autosomique récessive**, les parents ne présentent donc aucun signe. En revanche les **deux enfants atteints développeront la maladie cliniquement**.

Et devinez quoi ? C'est la fin !!

Le programme de Génétique..... est terminé.....ne soyez pas triste, séchez vos larmes je vous ai préparé des DMs, des EBs et pleins d'autres surprises afin que vous faisiez face à ce traumatisme. N'oubliez pas que la génétique vous aime !

PS : on impasse pas la Génétique c'est une petit matière toute mimi et surtout très très accessible ! Suri vous a à l'œil



Révissez bien, des bisous sur vos nucléotides mutés <3