



# Génétique

## COURS 3 : Séquençage et NGS

### AVANT PROPOS :

Coucouuuu, c'est Roxygène et comme d'habitude, mes annotations seront dans cette police/couleur ! On se retrouve pour le 3e (et dernier cours !!) en Génétique, je sais que vous allez triste de ne plus lire mes fiches... Bref, trêve de plaisanteries, après avoir lu cette fiche, vous aurez vu TOUT le programme de Génétique !

Ce dernier cours sur le séquençage de l'ADN est peut-être le plus dur en Génétique (selon moi) car il nécessite pas mal de compréhension mais on se détend, on y va en plusieurs fois si besoin et on essaie de comprendre AVANT d'apprendre ! Surtout, si vous avez la moindre question, vous savez quoi faire → hop hop le forum <3

Je vous laisse, bonne lecture !

### SOMMAIRE

#### I - Séquençage de l'ADN

1. Principe
2. Etapes
3. Cycles successifs
4. Histoire
5. Méthode automatisé
6. Retour sur l'exemple de l'achondroplasie

#### II - Exemple d'analyse d'un gène par PCR et séquençage Sanger : le syndrome de Wolfram

1. Tableau clinique
2. Rappels : organisation, expression des gènes et nomenclature
3. Recherche de mutations dans un gène

#### III - Séquençage Haut Débit ou NGS (Next Generation DNA Sequencing).

1. Etapes globales du NGS
2. Préparation des échantillons
3. Enrichissement par PCR clonale avec Illumina (PCR en pont, clusters)
4. Séquençage avec la plateforme Illumina
5. Enrichissement par PCR clonale avec ThermoFischer
6. Séquençage individuel des sphères avec ThermoFischer
7. Analyse bio-informatique

# I - Séquençage de l'ADN

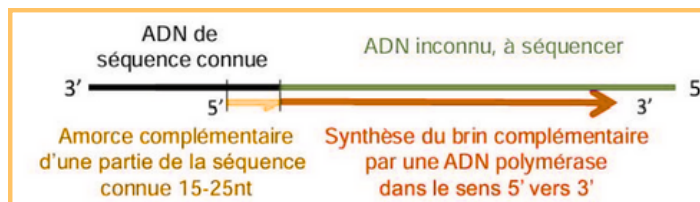
**BUT** : Le séquençage d'un fragment d'ADN est une technique qui permet de **déterminer la succession** des nucléotides qui le compose.

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, entre les nucléotides), avec ces deux brins antiparallèles, liés entre eux par des liaisons hydrogènes (entre les deux brins). Le principe du séquençage d'ADN (aussi appelée méthode de Sanger) est basé sur les di-désoxynucléotides.

C'est une **méthode enzymatique**, aujourd'hui 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

Globalement le principe est le même que la PCR, les étapes sont très proches cependant l'amplification sera un peu différente.



On va utiliser :

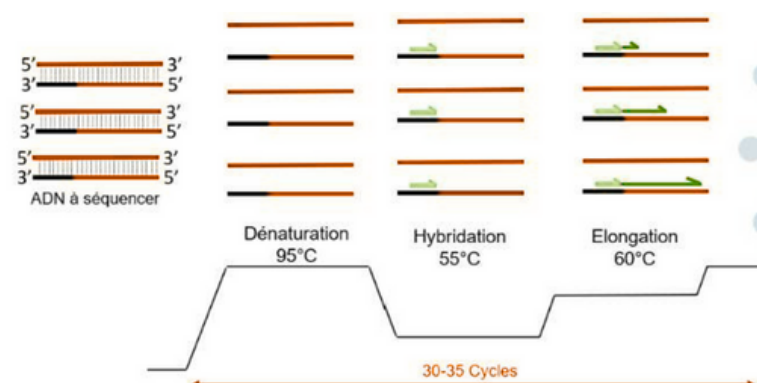
- Le brin d'**ADN inconnu à séquencer**
- Une **seule amorce complémentaire** (= primer)

*On ne veut lire qu'un seul brin d'ADN à la fois.*

*Dans le séquençage ce sont exactement les mêmes amorces qu'on a vu dans la PCR, sauf que là au lieu d'en mettre une en amont et une en aval, il n'y en a qu'une !*

- Une **ADN polymérase** (pour la synthèse du brin d'ADN complémentaire) qui synthétise de 5' en 3' dans un seul sens en incorporant les nucléotides (de la même façon que la PCR).

## 2. Etapes



**Mêmes étapes que la PCR** (même jeu sur la température) :

- Dénaturation à **95°**
- Hybridation à **55°** (avec un seul primer)
- Élongation à **60°**

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.

*En séquençage, on ne cherche pas à amplifier les deux brins comme en PCR. On choisit un seul des deux brins et on le lit de bout en bout grâce à une amorce qui sert de point de départ → ça suffit pour obtenir la séquence.*

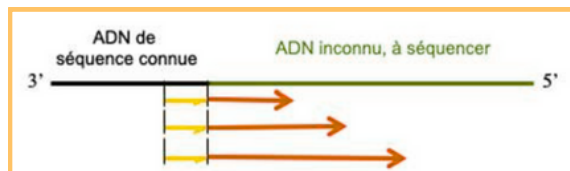
- Le matériel nécessaire à la Taq polymérase et à la synthèse du nouveau brin, à savoir les **nucléotides (dNTPs)**
- Les **tampons**
- Des **ddNTPs** (spécificité du séquençage)

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. **Élonger notre brin**, à **60°C** : le brin complémentaire au brin que l'on veut séquencer se forme.

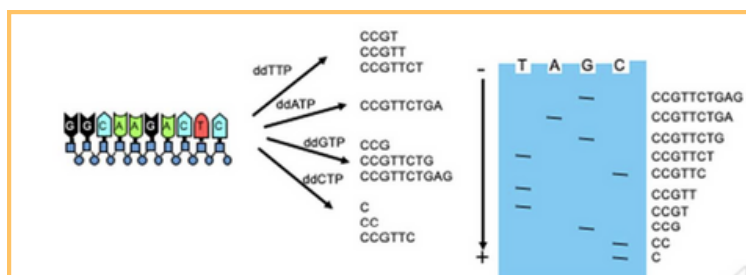
**L'autre particularité pour l'étape de séquençage est que l'on va utiliser un mélange de dNTPs (désoxyribonucléotides ATCG), et de ddNTPs (didésoxyribonucléotides).**

### 3. Cycles successifs



L'introduction d'un dNTP ou d'un ddNTP se fait au hasard. On va réaliser des cycles successifs de dénaturation des brins – hybridation de l'amorce – élongation du brin complémentaire. On se retrouve à la fin avec de très nombreux fragments d'ADN, de tailles différentes, complémentaires de l'ADN qu'on veut séquencer.

**ATTENTION** : Le brin d'ADN copié par la polymérase par complémentarité des bases est 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 !



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.

Le séquençage Sanger repose sur l'incorporation aléatoire de nucléotides classiques (**dNTPs**) et de nucléotides modifiés appelés **ddNTPs**.

- Les **dNTPs** permettent la poursuite de la synthèse.
- Les **ddNTPs**, eux, bloquent la synthèse (car ils manquent du groupement OH nécessaire à la liaison suivante).

☛ **Résultat** : chaque fois qu'un ddNTP est inséré, la synthèse s'arrête, ce qui génère des fragments d'ADN de tailles variables.

### Mise en place expérimentale :

---

Historiquement, on préparait 4 tubes de réaction où chaque tube contenait :

- le fragment d'ADN à séquencer
- une ADN polymérase
- une amorce (pour initier la synthèse)
- les 4 dNTPs (A, T, C, G)
- un seul type de ddNTP (soit ddATP, soit ddTTP, soit ddCTP, soit ddGTP) donc sachant que chaque tube possède un seul de ddNTP → chaque tube produit des fragments stoppés spécifiquement sur une base donnée

### Synthèse et arrêt aléatoire

---

Lors des cycles (dénaturation – hybridation de l'amorce – élongation) : la polymérase insère des nucléotides de manière complémentaire, mais la réaction peut s'arrêter à n'importe quelle position où un ddNTP est incorporé.

☛ On obtient donc une série de fragments de toutes les longueurs possibles, chacun stoppé à une base bien précise.

### Séparation et lecture

---

Les fragments produits sont séparés par électrophorèse sur gel :

- Les petits fragments (moins lourds) migrent plus vite (en bas du gel).
- Les grands fragments (plus lourds) migrent plus lentement (en haut du gel).

Chaque piste du gel correspond à un nucléotide (A, T, C ou G).

☛ En lisant de bas en haut, on reconstitue la séquence du brin complémentaire.

### Exemple :

---

Si le brin à séquencer est GGCAAGACTC comme sur l'exemple, on lit sur le gel le brin complémentaire : CCGTTCTGAG.

Le fragment le plus court correspond au premier nucléotide complémentaire ajouté (ici un C). En remontant, on lit progressivement toute la séquence.

### Pour résumer :

4 tubes → incorporation aléatoire des ddNTPs → fragments de tailles variées → séparation par électrophorèse → lecture de bas en haut = séquence du brin complémentaire.

## 5. Méthode automatisée

La méthode Sanger a, par la suite, été automatisée et simplifiée grâce à l'utilisation de ddNTPs fluorescents.

### Principe général :

- Chaque ddNTP est couplé à un fluorochrome de couleur spécifique : A = vert, T = rouge, C = bleu, G = noir
- Les 4 ddNTPs fluorescents sont mélangés dans un même tube réactionnel.
- ☞ Grâce au code couleur, on peut identifier quel nucléotide a été incorporé.

### 1 Mise en place expérimentale :

Le tube de réaction contient :

- le fragment d'ADN à séquencer,
- une ADN polymérase,
- une amorce (unique, complémentaire du brin à lire),
- les dNTPs (qui permettent la poursuite de la synthèse),
- les 4 ddNTPs fluorescents (qui stoppent la synthèse et marquent la position).

### 2 Synthèse et arrêt :

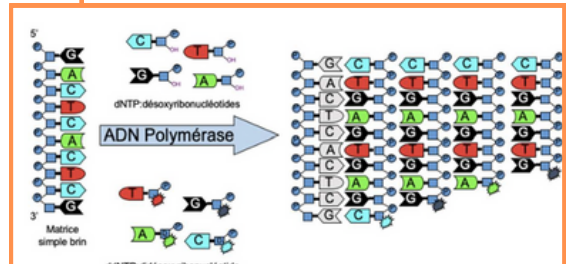
L'ADN polymérase synthétise le brin complémentaire à partir de l'amorce.

Elle incorpore **aléatoirement** :

- soit un dNTP (la synthèse continue),
- soit un ddNTP fluorescent (la synthèse s'arrête).

Les fragments obtenus sont donc de tailles variables :

- un fragment peut être complet si aucun ddNTP n'est incorporé,
- un fragment peut être très court si un ddNTP est incorporé tôt,
- toutes les combinaisons intermédiaires sont possibles.



### 3 Séparation et migration :

Les fragments produits sont séparés en fonction de leur taille par électrophorèse capillaire. Chaque capillaire correspond à un gel d'agarose miniaturisé.

La migration se fait grâce à un champ électrique :

- la machine plonge chaque extrémité dans un tampon avec 2 électrodes (cathode et anode),
- la migration se fait **de la cathode (-) vers l'anode (+)**.

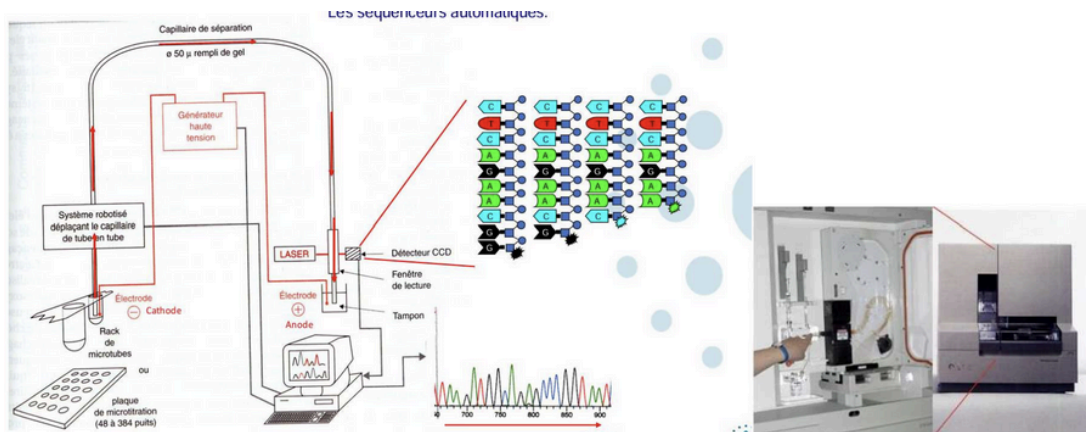
Le capillaire est rempli d'un gel de polymère spécial qui joue le même rôle que l'agarose et permet de séparer l'ADN par taille : les fragments petits migrent vite (lus plus tôt), et les fragments grands migrent lentement (lus plus tard).

## Détection

Au fur et à mesure de la migration, les fragments passent devant une caméra/laser.

Le fluorophore du ddNTP incorporé est lu : vert = A, rouge = T, bleu = C, noir = G.

Chaque tube de séquençage correspond à un capillaire, et chaque fragment qui passe est identifié par sa taille (position) et sa couleur (nucléotide).



## Traitement informatique

Le signal est converti en électrophorégramme, un tracé de pics colorés.

- Position (taille du fragment) → indique l'ordre des bases.
- Couleur (fluorophore du ddNTP) → indique l'identité de la base.

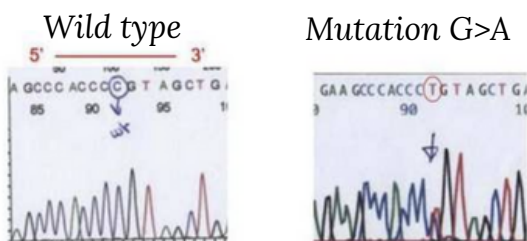
☞ On obtient ainsi directement la séquence nucléotidique.

## 6. Retour sur l'exemple de l'achondroplasie

Après la PCR-RFLP (= PCR suivi d'une digestion enzymatique), on séquence de cette manière cette même PCR et on va lire sur l'électrophorégramme l'enchaînement de nucléotides du patient en fonction de la couleur affichée..

☞ On pourra alors connaître le nucléotide pour chacune des positions sur l'ADN et donc savoir s'il y a un mauvais nucléotide donc une mutation.

*On utilise ici le séquençage pour vérifier notre digestion enzymatique : on a défini qu'on avait une mutation G>A en position 1138, on vérifie en séquence qu'on a bien G>A (représenté par un double pic au niveau de la position 1138.)*



Lorsqu'on arrive à une position normale d'un C (par complémentarité un G), il y a un **deuxième pic qui se superpose, pic correspondant à un T** et par complémentarité à un A : la mutation est donc un G → A.

## II - Exemple d'analyse d'un gène par PCR et séquençage Sanger : le syndrome de Wolfram

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.

**C'est-à-dire ?** On ne sait pas quelle mutation est présente, donc on va séquencer l'intégralité des exons de ce gène pour identifier le variant pathogène.

Pour l'achondroplasie, on cherchait une seule mutation spécifique. Ici, on cherche une mutation mais sachant qu'il en existe plusieurs sur ce même gène, on les cherche toutes.

*En résumé : on sait quel gène est impliqué → on séquence tous les exons car on ne connaît pas la mutation précise → on identifie la mutation pathogène compatible avec le phénotype du patient parmi tous les variants trouvés.*

### 1. Tableau clinique

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

- Diabète
- Surdité
- Atrophie optique
- Troubles neurologiques



Tableau clinique du Syndrome de Wolfram

👉 C'est une maladie à **transmission autosomique récessive**, ce que signifie que les individus atteints ont forcément les 2 allèles mutés pour que les signes apparaissent (les parents sont porteurs mais non atteints).

Le gène **WFS1** est un gène constitué de **8 exons**, et l'ATG (codon permettant le début de la traduction de l'ARN messager) se trouve dans le deuxième exon du gène. Cela signifie donc que le **premier** exon est **non codant**.

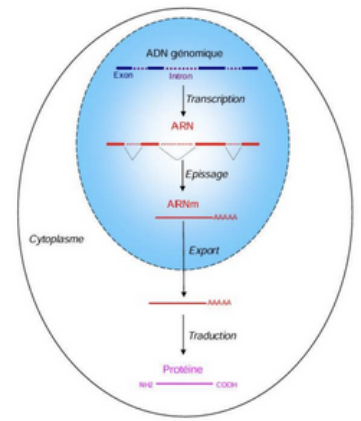
La **Wolframine** (protéine codée par le gène WFS1) a une fonction encore peu connue. Elle jouerait un rôle dans le flux calcique.

## 2. Rappels : organisation, expression des gènes et nomenclature

Dans le noyau se trouve l'ADN génomique, organisé en chromosomes (contenant les gènes constitués d'exons et d'introns).

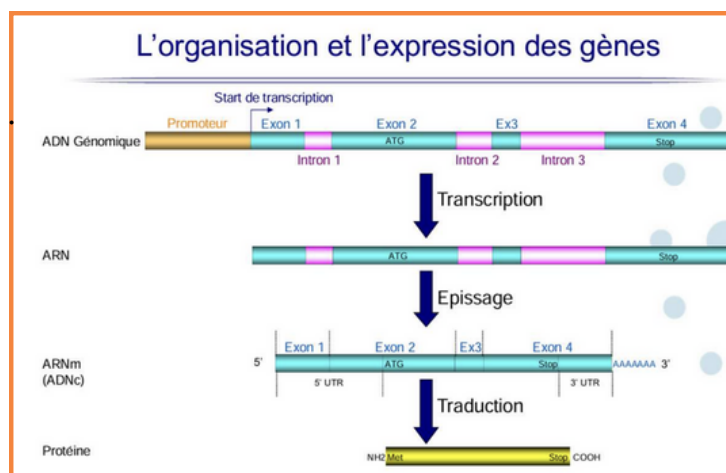
Ces gènes vont être transcrits en ARN, et après les mécanismes d'épissage, il ne restera plus que les régions codantes.

☞ Les introns seront donc éliminés pour obtenir un ARN messager mature avec, à leur extrémité 3', une queue Poly A. Ces ARN messagers vont ensuite être exportés du noyau vers le cytoplasme où ils seront traduits en protéines.



L'unité fonctionnelle du gène, c'est la protéine, un gène en tant que tel, tout seul, si on ne sait pas à quoi correspond la protéine pour lequel il code, ne va pas nous permettre d'en déduire grand-chose. On retient que ce qui nous intéresse fondamentalement, c'est la protéine.

### 🔍 Zoom sur l'organisation et l'expression des gènes :



Ici, on représente sur la 1ère ligne 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.

Le **start de transcription** définit le cadre de lecture. On retrouvera ainsi les **régions codantes (exons)**, séparées par les **régions non codantes (introns)**.

⚠ **La traduction ne commence pas forcément aux premiers nucléotides du premier exon !** ⚠

Elle commence à partir du codon ATG (sur l'exon 2 dans cet exemple : le premier n'est donc pas traduit).

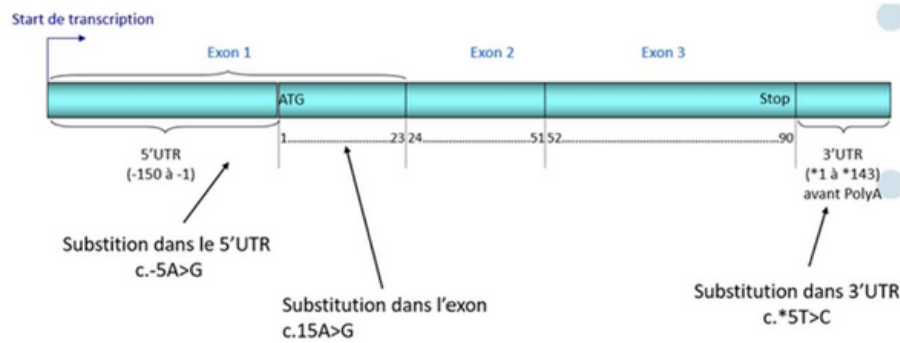
Ceci sous-entend que si le codon ATG se trouve dans l'exon 2, tout ce qui est en amont de cet exon ne sera pas traduit, même si la séquence est présente dans l'ARN messager.

☞ Un **exon** est sur l'ARNm mais n'est **pas forcément codant** pour la protéine. La région en amont du codon ATG (codon START) est appelée 5' UTR (non codante) et la région en aval du codon STOP est appelée 3' UTR, cette région est également non codante.

Les régions 5' UTR et 3' UTR sont donc des **régions transcrites mais non codantes** donc **non traduites**.

En général ces régions jouent un rôle dans la stabilité de l'ARNm, son exportation ou autre (**elles ne servent pas à rien**). Ainsi, la protéine commence au niveau de l'ATG (= méthionine, premier codon codant = c.1) et se termine au niveau du codon stop.

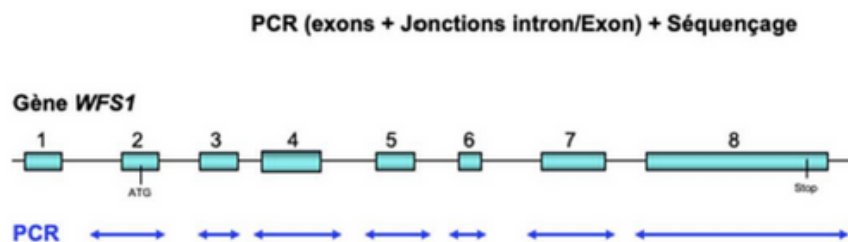
## Terminologie :



Les substitutions situées en 5' UTR seront numérotées « c. - » en partant de l'ATG et en remontant (donc de droite à gauche).

Les substitutions situées en 3' UTR seront numérotées « c. \* » en partant du codon stop (donc de gauche à droite).

### 3. 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 **8 exons**, et l'ATG se situe dans l'exon 2.

L'exon 1 est donc transcrit mais non traduit.

Pour diagnostiquer cette pathologie, il faut **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. 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 la conséquence de cette modification sur la protéine.

On va réaliser **7 PCR pour amplifier les régions codantes** : les PCR seront faites sur les exons 2 / 3 / 4 / 5 / 6 / 7 / 8.

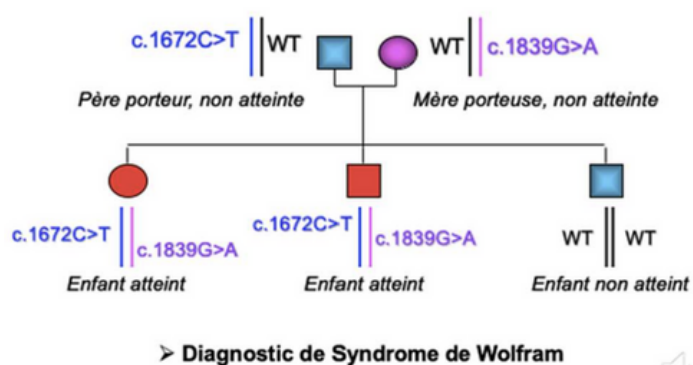
☞ On n'amplifie pas l'exon 1 un car il est non codant. S'il y avait une mutation dans cet exon, il serait très difficile de l'interpréter et de savoir quelles conséquences seraient engendrées puisqu'il n'est pas exprimé.

☞ Aujourd'hui, on n'a pas encore les outils pour prédire ce qu'il va se passer lorsque le variant nucléotidique est dans une région non codante. Malgré tout, on sait qu'un **variant sur les tous premiers nucléotides de l'intron** a généralement un **effet sur l'épissage** (et donc sur la protéine finale) donc on prend des amorces pour la PCR qui amplifient l'exon mais aussi une partie des introns qui l'encadrent (à 50 nucléotides de cette jonction intron/exon).

## On récap' :

- On **amplifie** : les **exons codants** (on interprète facilement les changements nucléotidiques dans les parties codantes) + **une partie des introns** qui les encadrent (ont potentiellement des effets sur l'épissage)
- On **n'amplifie PAS** : les **exons non codants** (car on n'a pas encore les outils pour prédire ce qu'il va se passer lorsque le variant nucléotidique est dans une région non codante)

## Prenons l'exemple de cette famille :



On amplifie les 7 exons par PCR, on les séquence, et on identifie :

Chez le père un variant en position 1672 (qui modifie un C en 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.

## Pourquoi on fait un séquençage ici ?

On connaît le **gène responsable (WFS1)** → donc pas besoin de faire un exome entier ou un génome, on cible uniquement ce gène.

Cependant, **on ne connaît pas à l'avance quelle mutation est présente**. Contrairement à l'achondroplasie (toujours la même mutation FGFR3), ici il existe plein de mutations différentes possibles sur WFS1.

Donc on doit explorer tous les exons du gène → c'est pourquoi on fait une PCR sur les 7 exons, puis un séquençage.

Le séquençage permet de retrouver le ou les variants pathogènes. On voit la position respective de la mutation chez le père et chez la mère.

Les **enfants atteints** → ils héritent des deux mutations, donc ils sont **hétérozygotes composites** (= chaque allèle porte une mutation différente, mais au final aucun allèle n'est fonctionnel).

## Conclusion :

Les parents sont porteurs sains → pas de maladie car ils ont encore un allèle normal.

Les enfants atteints ont deux allèles mutés → maladie exprimée.

Le séquençage permet donc de poser le diagnostic moléculaire précis en identifiant quelles mutations exactes sont responsables chez cette famille.

### III - Séquençage Haut Débit ou NGS (Next Generation DNA Sequencing)

Aujourd'hui, le **NGS** est de plus en plus utilisé car c'est une technique extrêmement puissante qui coûte de moins en moins cher et est de plus en plus accessible.

#### Rappel historique :

- Le « **Séquençage Manuel** » (dit Sanger) inventé en **1977**
- Le « **Séquençage Automatique** » avec ses DDNTPs fluorescents, datant de **1993**
- Depuis **2007**, on a le « **Séquençage Haut Débit** » ou NGS qui a révolutionné la biologie moléculaire

En 2001, 95% du génome humain avait été séquencé. C'est un énorme projet mondial débuté en 1989 et achevé en 2003, réalisé à l'aide de séquenceurs capillaires, car initié bien avant la fabrication des machines NGS.

Le **génomome humain** correspond à environ **3 milliards de pdb**.

On y trouve **30 000 gènes** et les régions codantes (**exons**) ne représentent que **1 à 2%** du génome.

Il nous reste donc **encore beaucoup de choses à comprendre** car les 98% restants ne servent certainement pas à rien mais aujourd'hui on n'est pas capable de comprendre exactement à quoi servent toutes ces régions non codantes.

On aurait d'ailleurs pu les trouver moins intéressantes car non codantes. On s'attendait bien sûr en séquençant le génome à retrouver plus de régions codantes que 1 ou 2 %.

Les régions non codantes ont **certainement un rôle au niveau de la régulation de l'expression des gènes** voire dans d'autres mécanismes encore incompris. La difficulté aujourd'hui, quand on fait un séquençage complet du génome pour retrouver le variant responsable, n'est donc plus au niveau du séquençage mais dans l'interprétation des variants.

Enfin, pour réaliser le séquençage complet d'un génome humain il aurait fallu :

- 12 ans avec 1000 machines PCR avec la méthode Sanger en 1980
- 1 an avec 120 machines avec le séquenceur automatisé dans les années 2000
- Il faut 1 semaine (voire moins : 2-3 jours) avec 1 machine, avec le NGS, aujourd'hui pour réaliser un séquençage du génome humain de plusieurs dizaines d'individus en même temps.

**Les capacités de séquençage ont donc augmenté de manière exponentielle ces dernières années.**

## LEXIQUE :

**Exome** : Ensemble des exons

**Génome** : Ensemble des gènes (comprenant les exons et les introns)

**Séquençage** : Le séquençage haut débit est un séquençage massif en parallèle de molécules d'ADN individuellement séparées et amplifiées sous forme de clones ou de molécules uniques.

La définition peut paraître un peu compliquée, mais une fois que vous aurez vu les différentes étapes vous comprendrez que ce qu'on va faire c'est : **fragmenter notre ADN, l'isoler et le séquencer.**

*Vous visualiserez mieux à la fin du cours à quoi correspond le NGS. Je vous conseille de faire une première lecture du cours avec comme but de comprendre et vous apprendrez ensuite (conseil de ma vieille que j'approuve à 100%).*

### Pourquoi faire un NGS ?

---

On fait un NGS lorsqu'on veut **séquencer énormément de gènes** voire l'exome ou le génome au complet.

Le NGS sera utilisé pour **rechercher les mutations dans les maladies génétiques** qui peuvent être dues à un très grand nombre de gènes.

On pourrait utiliser cette technique pour diagnostiquer l'achondroplasie, mais elle ne sera pas en adéquation avec l'utilité de cette dernière car l'achondroplasie est une maladie qui atteint un gène bien précis.

☞ **On fera donc une PCR-RFLP (= PCR suivie d'une digestion enzymatique), plutôt qu'un NGS pour diagnostiquer l'achondroplasie.**

*Je reprends l'explication de ma vieille : le NGS donnerait bien sûr le diagnostic donc en théorie on pourrait l'utiliser mais on ne le fait pas car ce serait inutile de séquencer tout le génome alors qu'on recherche un seul gène, c'est comme si vous lisiez tout le dictionnaire alors que vous cherchez un mot en p. (Elle a géré la métaphore, pas vrai ?)*

Les **outils de biologie moléculaire** utilisés pour réaliser un NGS sont **pratiquement les mêmes que pour les techniques vues auparavant** : polymérase, ligase etc....

Ce qui a permis l'arrivée du séquençage massif c'est : **l'amélioration des technologies, la miniaturisation des systèmes, et la bio-informatique qui s'est développé et se développe encore aujourd'hui.** (*Attention je préviens, c'est une phrase importante du cours, je dis ça, je dis rien...*)

Actuellement, il existe deux plateformes (2 sociétés) qui commercialisent deux technologies différentes au niveau des étapes d'amplification et de séquençage les sociétés :



Illumina est la plateforme la plus fréquemment utilisée.

La professeure vous conseille de voir 2 vidéos YouTube (qui explique chacune le déroulement de ces 2 méthodes) dont vous avez le lien ci-contre. Ces liens sont directement issus du diaporama du Pr. Bannwarth.

## 1. Etapes globales du NGS :

A titre informatif, voici ce qu'a dit le Pr. Bannwarth lors du présentiel de l'année dernière :

“Les étapes sont les mêmes pour les deux systèmes (Illumina et ThermoFisher) : ce qu'il faut que vous compreniez c'est comment on fait un NGS et quelles sont les grandes étapes. Je ne vous demande pas de retenir les détails de ces étapes qu'on présentera après.”

# 1

La préparation des échantillons/librairies : Cette étape est identique quelle que soit la plateforme utilisée (illumina ou ThermoFischer).

Elle consiste à **fragmenter notre ADN** (ADN génomique ou des fragments PCR). Les techniques de NGS permettent aujourd'hui de séquencer des fragments d'environ 200 à 400 pb en fonction des kits et des plateformes utilisés.

A ces petits fragments d'ADN, vont être ajoutés **des adaptateurs** (= **primers**) et des **BC** (= **code-barres**), c'est-à-dire des oligonucléotides dont la séquence est connue.

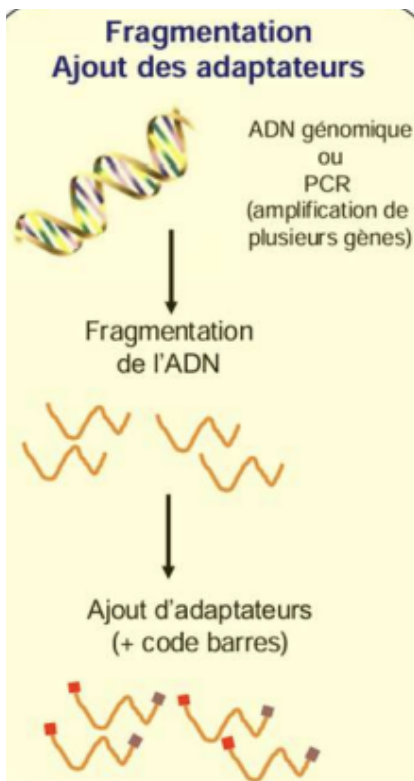
Ils vont permettre :

- **L'amplification PCR pour les adaptateurs** (identiques pour toutes les séquences et donc tous les patients)
- **L'identification des patients pour les BC** (donc différents/spécifiques pour chaque patient)

En effet ces BC ont une séquence bien déterminée et chaque BC va être attribué à un patient, soit **une séquence de BC = un patient**.

Dans les étapes de séquençage on va donc pouvoir mélanger nos patients dans un même run, dans une même réaction. Lors de l'analyse bio-informatique, ces séquences vont être triées grâce à ces BC ce qui va permettre d'identifier et d'attribuer chacune des séquences à chacun des patients.

Les adaptateurs sont **ajoutés aux extrémités 5' et 3' des fragments d'ADN** et ils servent à ce que **toutes les extrémités 5' et 3' soient identiques** et donc qu'on puisse **amplifier l'ensemble de nos fragments d'ADN avec uniquement un seul couple de primer** (1 primer sens + 1 primer reverse).



2

### Enrichissement : PCR Clonale

Fixation et amplification  
des fragments d'ADN  
sur support solide

llumina :  
Lame de verre (Flow cell)  
PCR « Bridge » en clusters

ThermoFisher :  
Sphères métalliques  
PCR en émulsion



3

### Séquençage

llumina:  
Séquençage par synthèse  
Détection de fluorescence

ThermoFisher:  
Semi-conducteur  
Séquençage par synthèse  
Détection de variation de pH



4

### Analyse Bioinformatique

- Détermination des bases
- Contrôle qualité
- Alignement au génome de référence
- Détermination des variants
- Annotation
- Interprétation

**La PCR : Enrichissement de nos fragments d'ADN** par PCR clonale, (car on part d'un seul fragment et on l'amplifie x fois) il va y avoir une étape de fixation et d'amplification de chacun des fragments d'ADN sur un support solide, différents selon les plateformes :

- Technologie **llumina** = **Lame de verre**
- Technologie **Thermo Fischer** = **sphères métalliques**

L'idée c'est de **séparer** tous les fragments d'ADN (en les fixant) de manière à pouvoir les **amplifier** de manière clonale (à en avoir des copies, des clones). Il faut imaginer qu'on le fait pour nos milliards de séquence en même temps.

**Le séquençage :** Ces supports solides (lame de verre, sphère métallique) vont être **séquencés individuellement**. On retrouve un peu ce qu'on avait avec le séquençage Sanger mais avec des détections qui vont être différentes en fonction des plateformes :

- **llumina** : détection de **fluorescence** (*mémo de ma vieille : illumina → lumière → fluorescence*)
- **ThermoFischer** : des variations de **pH**

Mots du Pr Bannwarth « Peu importe ce qu'il faut que vous compreniez c'est **qu'on fragmente, on isole, on amplifie et on séquence** »

**L'analyse bio-informatique :** C'est une étape **commune** et **extrêmement importante**, on va générer des données massives et il va falloir ensuite transformer ces signaux lumineux (détection fluorescence → Illumina) ou transformer ces variations de pH (ThermoFischer) en « nucléotides », soit en données de séquences. Cette partie représente **la plus grosse difficulté**, car elle nécessite beaucoup de systèmes informatiques avec des algorithmes de plus en plus puissants.

En effet, on arrive à séquencer de plus en plus avec des fichiers de plus en plus gros qui engendrent des difficultés de stockage de données. Cela engendre des **problèmes d'annotations ou d'interprétations**.

**Par exemple**, vous avez séquencé vos 3 milliards de nucléotides sur votre génome, vous cherchez la faute d'orthographe, le nucléotide variant pathogène parmi tous les polymorphismes (car on a tous des variants nucléotidiques d'un individu à l'autre mais qui n'entraînent aucun caractère pathogènes : yeux marrons/verts, cheveux bruns ou blonds qui sont des caractères physiologiques).

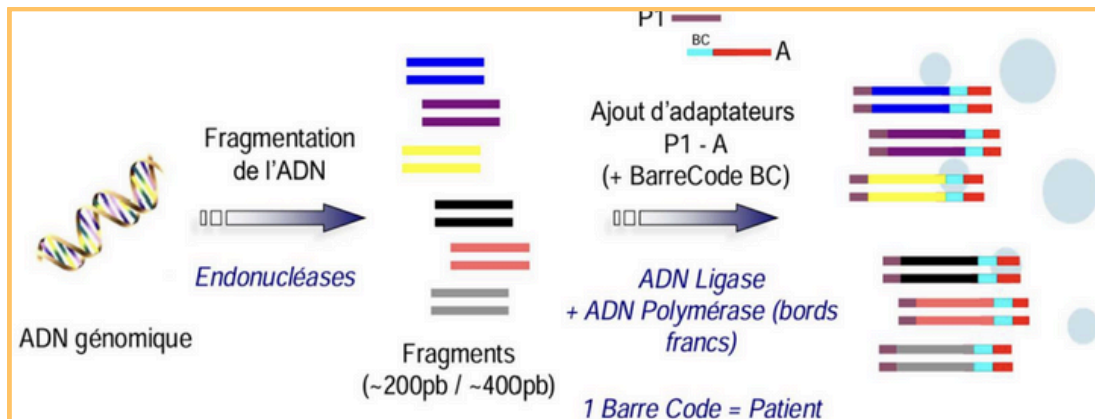
## 2. Etape 1 : Préparation des échantillons (= préparation des librairies)

*Ici, il va falloir essayer de visualiser au maximum, même si ça paraît un peu compliquée !*

### Comment peut-on fragmenter et ajouter les adaptateurs sur ces échantillons (ADN génomique) ?

On est en mesure de récupérer uniquement les régions qui nous intéressent (des régions exoniques en général, pas toujours les mêmes) grâce à un système/technique de capture.

☛ On capture/extrait du milieu réactionnel uniquement les régions d'ADN génomique qui nous intéressent.



Tout d'abord, cet ADN génomique va être **coupé en fragments de 200 à 400 pb** grâce à des endonucléases qui ont comme propriété de couper l'ADN double brin en son milieu.

☛ On fragmente avec **plusieurs endonucléases** qui vont générer des fragments **aléatoirement** sur l'ADN génomique.

Ensuite à ces fragments d'ADN, on va rajouter des **adaptateurs** (P1 et A) et des **barres codes (BC)**.

☛ Pour se faire, on retrouve les outils vus précédemment combinés différemment :

1) Des **ADN ligases** pour lier les fragments d'ADN aux adaptateurs et aux barres codes (qui sont aussi des séquences d'ADN)

2) Des **ADN polymérases**, en fonction de la digestion réalisée par les endonucléases, de leur site de restriction les extrémités ne sont pas forcément toutes à bords francs, alors elles peuvent préparer les extrémités pour que les ADN ligases puissent ensuite faire leur travail de liaison proprement.

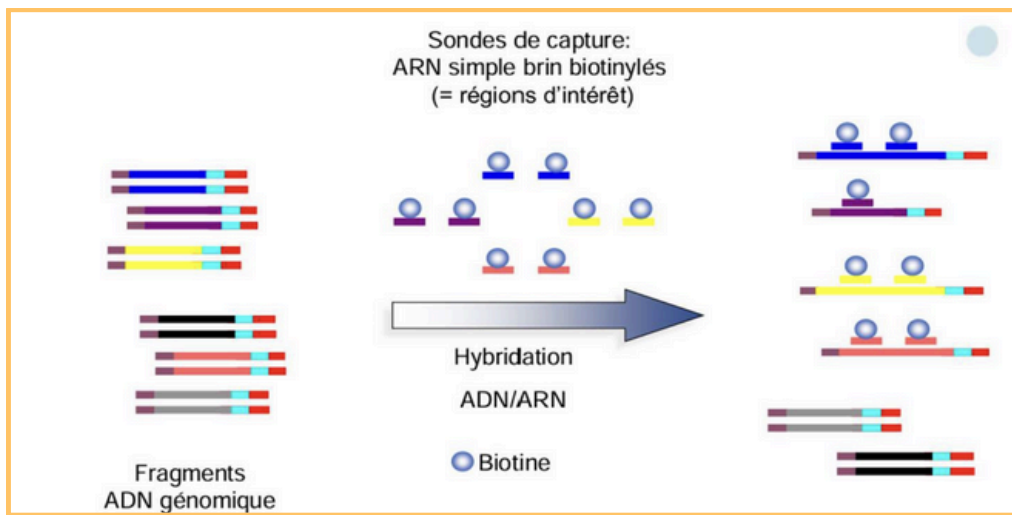
Grâce à ces outils on va pouvoir couper et lier nos fragments d'ADN qui auront alors les mêmes extrémités 5' et les mêmes extrémités 3' :

**⚠ Toutes les extrémités 5' sont identiques, toutes les extrémités 3' sont identiques mais les extrémités 5' et 3' sont différentes entre elles ⚠**

Mots du Pr Bannwarth : "C'est ça que vous devez savoir ! On n'ira pas vous demander si l'adaptateur en 5' c'est le P1 ou le A"

A présent, **on peut mélanger plusieurs patients**, car ils seront reconnus par la séquence des BC. C'est aussi pour ça qu'on est capable de séquencer énormément, on peut mélanger nos patients ! Par contre attention, il ne faut **pas les mélanger avant d'avoir mis le BC !**

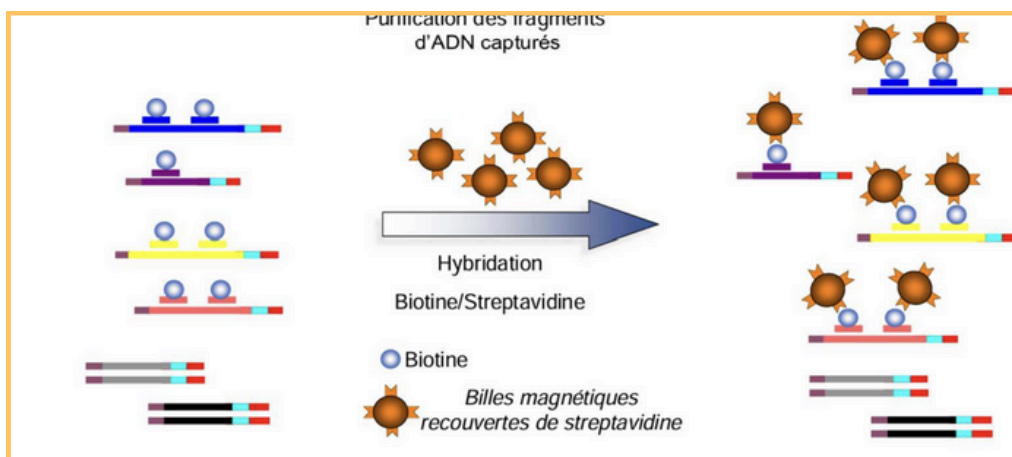
A présent, **on va chercher la séquence spécifique qui nous intéresse.**



Pour trouver la région d'ADN spécifique qui nous intéresse, on ajoute à ces fragments **des sondes de capture = sondes ARN simple brin**. Ce sont des petits fragments d'ARN complémentaires de la séquence d'ADN que l'on veut séquencer.

On a un mélange de différentes sondes ARN spécifiques aux fragments d'ADN ciblés. Ces sondes sont biotinylées, c'est-à-dire que de la biotine est fixée sur ce fragment d'ARN. Comme pour la PCR, on dénature, on casse les liaisons hydrogènes qui maintiennent les deux brins d'ADN entre eux.

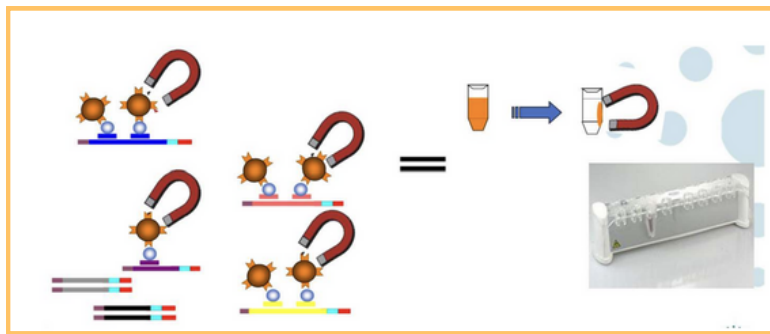
☞ Grâce à des variations de température, on **aura la formation d'hybrides ADN/ARN**, avec ces ARN biotinylés qui vont se fixer sur les fragments d'ADN qui leur sont spécifiques et qui nous intéressent.



Nous allons ensuite **purifier ces fragments d'ADN capturés en ajoutant des sondes**, soit des billes magnétiques recouvertes de **streptavidine**.

L'interaction « biotine-streptavidine » est une interaction très forte (protéine-protéine) qui est utilisée dans beaucoup de techniques différentes. Lorsqu'on rajoute ces billes recouvertes de streptavidine, **elles vont se fixer sur la biotine** donc sur les hybrides ADN/ARN.

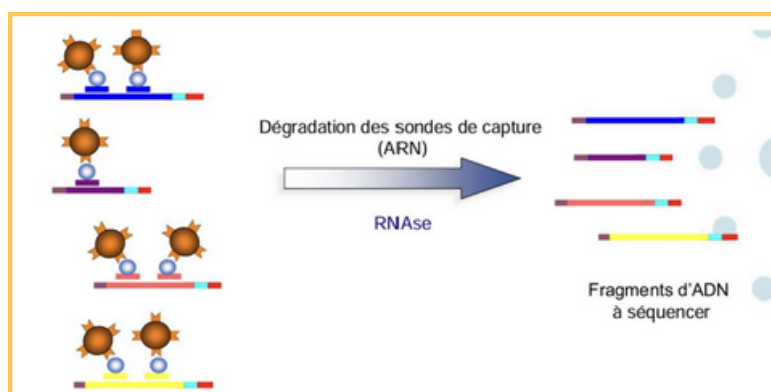
☞ **Tous les fragments d'ADN reconnus vont être piégés sur ces billes magnétiques.**



Les billes recouvertes de streptavidine, étant des billes magnétiques, si on ajoute des aimants sur le bord du tube, ces billes vont être attirées sur les côtés, se coller sur le bord du tube.

☛ Cela donne une coloration marron sur le bord du tube.

On va pouvoir réaliser des **lavages** (ajouter des liquides, des tampons), aspirer, retirer le surnageant de notre tube qui ne nous intéresse pas, **de façon à nettoyer et ne récupérer que les billes qui auront fixées ces sondes ARN, elles-mêmes hybridées sur nos régions d'intérêt** (= fragments d'ADN qu'on veut séquencer).



Une fois qu'on a purifié les régions d'intérêt, **on va dégrader les sondes ARN hybridées à l'ADN d'intérêt**. Pour cela on va ajouter des RNases, enzymes qui dégradent l'ARN hybridé dans le complexe ADN/ARN, pour qu'il ne reste plus que l'ADN à séquencer.

☛ **On obtient donc nos régions d'ADN ciblées qui sont barre codées et avec leurs adaptateurs.**

On rappelle que quelle que soit la plateforme utilisée, la préparation des échantillons avec cette étape de capture est identique.

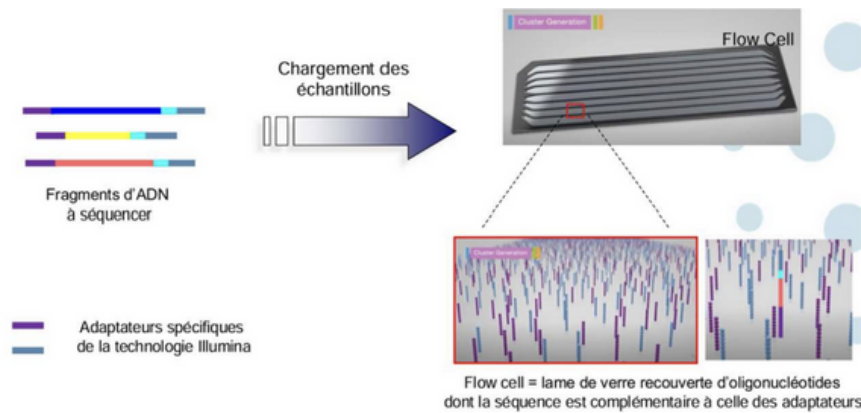
**Récap' de ma vieille (il est juste génial <3), n'essayez pas d'apprendre par coeur sans rien comprendre, vous devez visualiser :**

- On a TOUS nos ADN fragmentés mais **on ne veut que ceux qui nous intéressent**
- **Comment on peut les récupérer ?** Avec des sondes ARN (car on peut facilement prendre des ARN avec la séquence qu'on veut et hop ça s'hybride)
- Oui mais **après faut les récupérer ces fragments ARN/ADN !?**
- Alors on n'utilise pas que des ARN mais des **ARN biotinylés !**
- **Pourquoi de la biotine ?** Par ce qu'elle a une très **forte interaction avec la streptavidine**
- Alors **on recouvre des billes magnétiques de streptavidine**, qu'on ajoute au mélange
- Finalement, **un aimant permet de récupérer les billes magnétiques**, recouvertes de streptavidine ce qui récupère donc l'ARN biotinylé, et par conséquent nos fragments d'ADN que l'on voulait séquencer
- Ensuite on lave tout et **on dégrade les ARN** (avec RNase), car on en a plus besoin !

**RECAP :** Aimant > Billes magnétiques > Streptavidine > Biotine > ARN > ADN (à séquencer)

### 3. Enrichissement par PCR clonale avec Illumina (PCR en pont, clusters)

Une fois qu'on a nos fragments d'ADN à séquencer, on veut les amplifier par PCR. Pour la technologie Illumina, l'enrichissement/la PCR clonale est réalisée sur une **lame de verre, également appelée Flow Cell**.

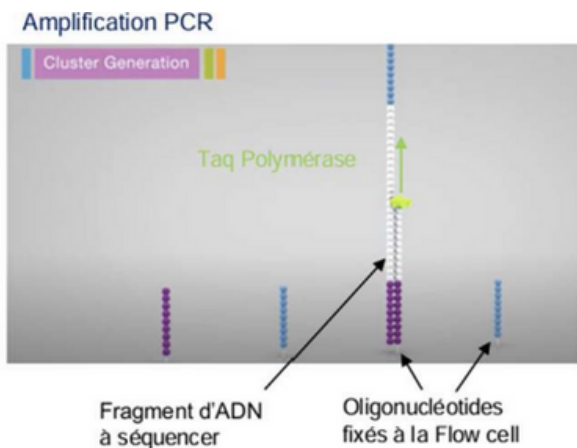


Cette lame est recouverte de **2 types d'oligonucléotides** (= amorces) fixés :

- **Oligonucléotides spécifiques des adaptateurs ajoutés en 5'** = leur séquence est complémentaire aux adaptateurs ajoutés en 5'
- **Oligonucléotides spécifiques des adaptateurs placés en 3'** = avec une séquence complémentaire aux adaptateurs placés en 3'.

Les échantillons préparés précédemment sont déposés sur cette lame de verre et, par complémentarité des bases, les fragments d'ADN préparés vont s'hybrider sur les différents oligonucléotides.

Il faut donc imaginer que l'ensemble de nos échantillons vont être hybridés à différents endroits, sur cette lame de verre de sorte à générer des clusters.



Ici, on voit que notre fragment d'ADN s'hybride à partir de son extrémité (5' ou 3') sur un oligonucléotide fixé à la lame de verre (ou Flow Cell). Ainsi, à partir de ces fragments fixés sur cette lame de verre, on réalise une amplification de chacun de ces fragments par PCR.

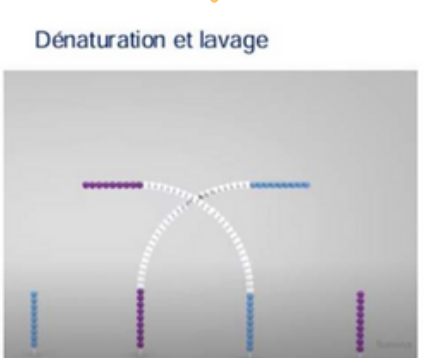
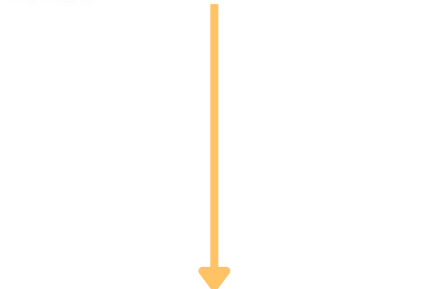
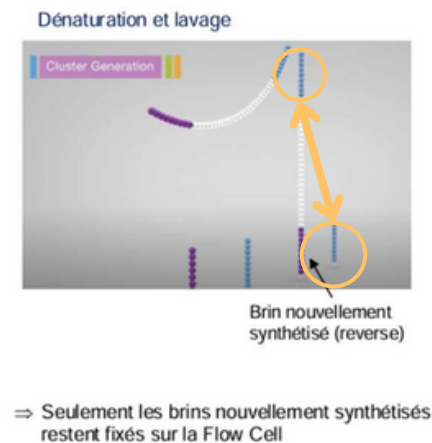
On va donc retrouver tout ce qu'il faut pour amplifier un fragment d'ADN, notamment **une Taq polymérase**. Cette dernière va synthétiser de 5' en 3' un brin complémentaire au brin hybridé sur notre lame de verre.

On a ensuite une étape de **dénaturation** et de **lavage** qui va permettre de libérer, de notre lame, le premier fragment qui a été hybridé sur celle-ci.

☛ Les **fragments nouvellement synthétisés, eux, ne sont pas enlevés par les lavages** puisque l'oligonucléotide qui a permis d'amorcer cette synthèse est fixé sur la lame de verre.

Par contre, le fragment que nous avons ajouté et qui provenait de notre échantillon va être enlevé de cette lame lors du lavage son amorce n'est pas fixée de manière covalente sur la lame de verre (il était simplement hybridé par complémentarité des bases à l'aide de l'oligonucléotide).

**Seul le brin nouvellement synthétisé est attaché à la lame de verre puisque celui-ci a été formé à partir des oligonucléotides fixés sur cette lame.**



On a ensuite une **étape de ré amplification de ces fragments isolés = PCR bridge ou PCR en ponts.**

Lors de cette ré amplification, on va avoir la formation de ponts puisque l'adaptateur non fixé de notre fragment s'hybride à l'oligonucléotide complémentaire à cette séquence, localisé juste à côté.

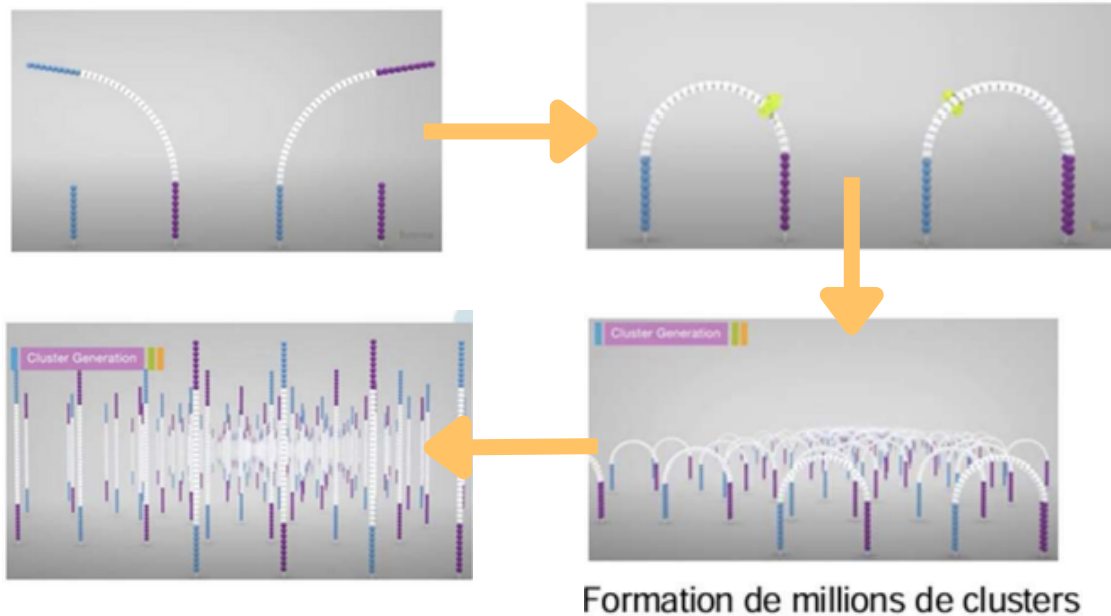
*Vous pouvez voir en haut et à gauche, que notre adaptateur bleu non fixé sur le fragment nouvellement synthétisé s'est hybridé avec l'oligonucléotide bleu juste à côté, afin de former un pont.*

C'est pour cette raison qu'on parle de PCR en pont : notre molécule d'ADN nouvellement synthétisée forme des ponts grâce à l'hybridation de chacun de ses adaptateurs en 5' et en 3'.

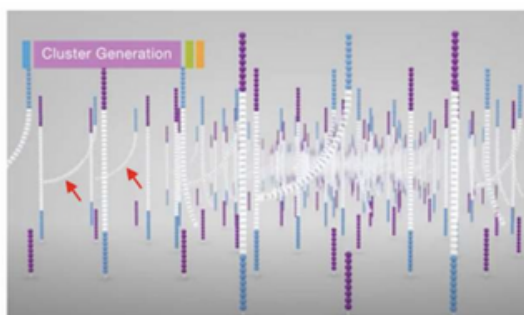
Suite à la formation de ce pont, **la Taq polymérase** va pouvoir synthétiser le brin complémentaire de 5' en 3'.

Ce pont est à son tour dénaturé, ce qui va permettre d'obtenir deux fragments d'ADN simple brin, fixés sur notre lame de verre.

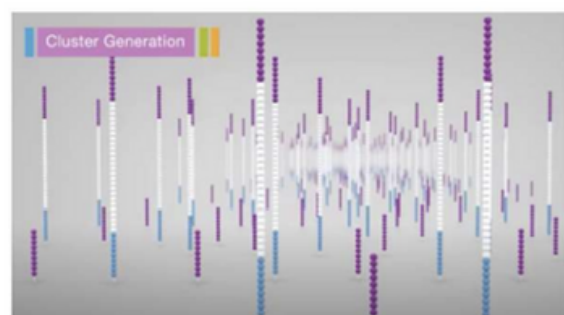
Nous obtenons donc 2 nouvelles molécules d'ADN nouvellement synthétisées (un brin sens et un brin reverse).



Ces **étapes de formation de ponts vont se répéter**, ainsi que la formation du brin complémentaire grâce à la Taq polymérase. Cette répétition **se fait jusqu'à la formation d'un grand nombre de brins : c'est la formation d'un cluster d'amplification** (formation de groupes d'amplification de produits PCR). Sur ce support solide, à la fin des cycles successifs de PCR, **on obtient des millions de clusters qui correspondent initialement à un seul fragment d'ADN.**



Clivage des brins reverses



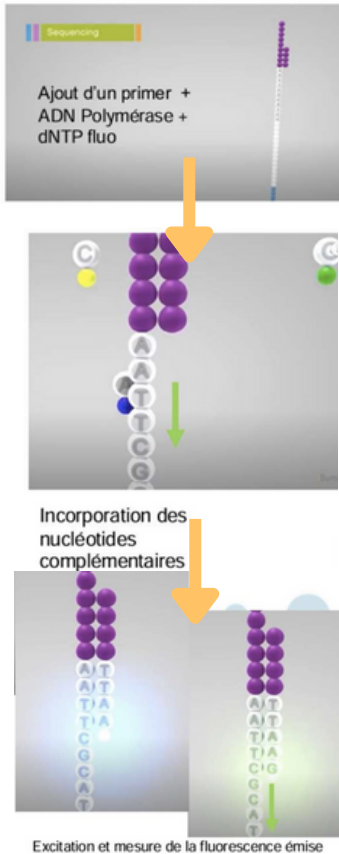
Elimination des brins reverses par lavage  
Conservation des brins sens uniquement

À la fin de nos différents cycles d'amplification, on a une **étape qui va permettre de cliver les brins reverse**, qui sont les brins complémentaires du brin que nous avons mis initialement sur la lame.

☞ **On ne garde donc que les brins sens**, c'est-à-dire **les brins qui ont la même séquence que notre fragment d'ADN original**, initialement hybridés sur la lame. Ce sont ces brins sens que l'on va séquencer.

*Cette partie est compliquée à visualiser, regardez bien les diapos, n'hésitez surtout pas à regarder la vidéo de la prof (perso c'est ce qui m'a aidé à comprendre durant ma P1) et ne vous prenez pas trop la tête sur les détails ! Ce qu'il faut surtout retenir , on a isolé chaque fragment d'ADN en le fixant à une lame de verre et ensuite on a amplifié = formation de cluster. (C'est déjà très bien !)*

## 4. Séquençage avec la plateforme Illumina



Le séquençage est réalisé directement sur notre lame de verre.

Pour cela on rajoute un primer, qui correspond à un oligonucléotide complémentaire de **l'adaptateur**. On ajoute une **polymérase** pour la synthèse du brin de 5' en 3', et des nucléotides fluorescents (= dNTPs fluorescents).

☞ Lorsque la polymérase introduit un nucléotide par complémentarité des bases, une fluorescence d'une couleur différente selon le dNTP est émise.

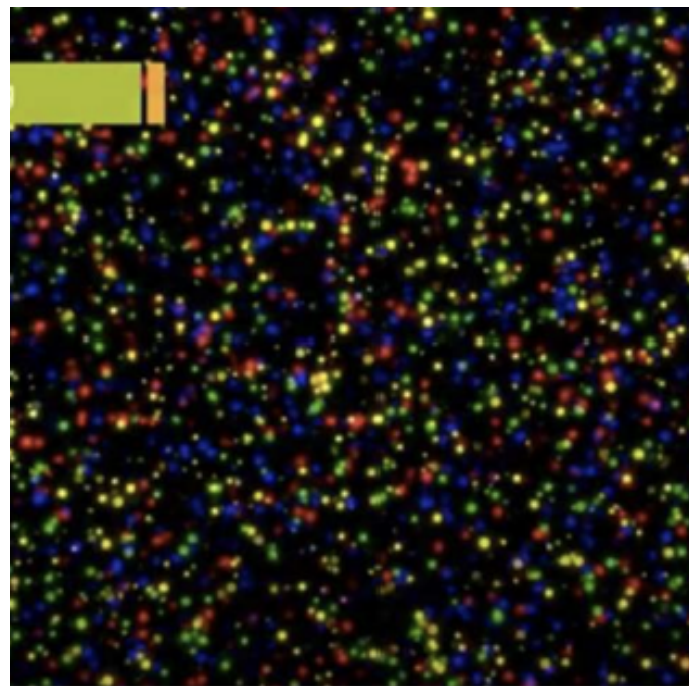
☞ L'automate va lire/mesurer en direct la fluorescence émise par le dNTP ajouté sur chacun des fragments et dans les différents clusters au cours de la synthèse. L'appareil lit la couleur émise et donc on saura par déduction quel nucléotide est incorporé.

On obtient ce type d'image. Nous avons ici une toute petite partie d'une Flow Cell.

☞ On voit bien les signaux lumineux émis chaque fois qu'on incorpore un nouveau nucléotide.

☞ Le système arrive à les lire au fur et à mesure et à les transformer en données de séquence.

On parler de **Séquençage Massif**, puisque chaque cluster peut représenter un tube de séquençage. Ils sont séquencés et lus en parallèle. Avant, en séquençage Sanger, on avait besoin d'un tube par réaction de séquençage.



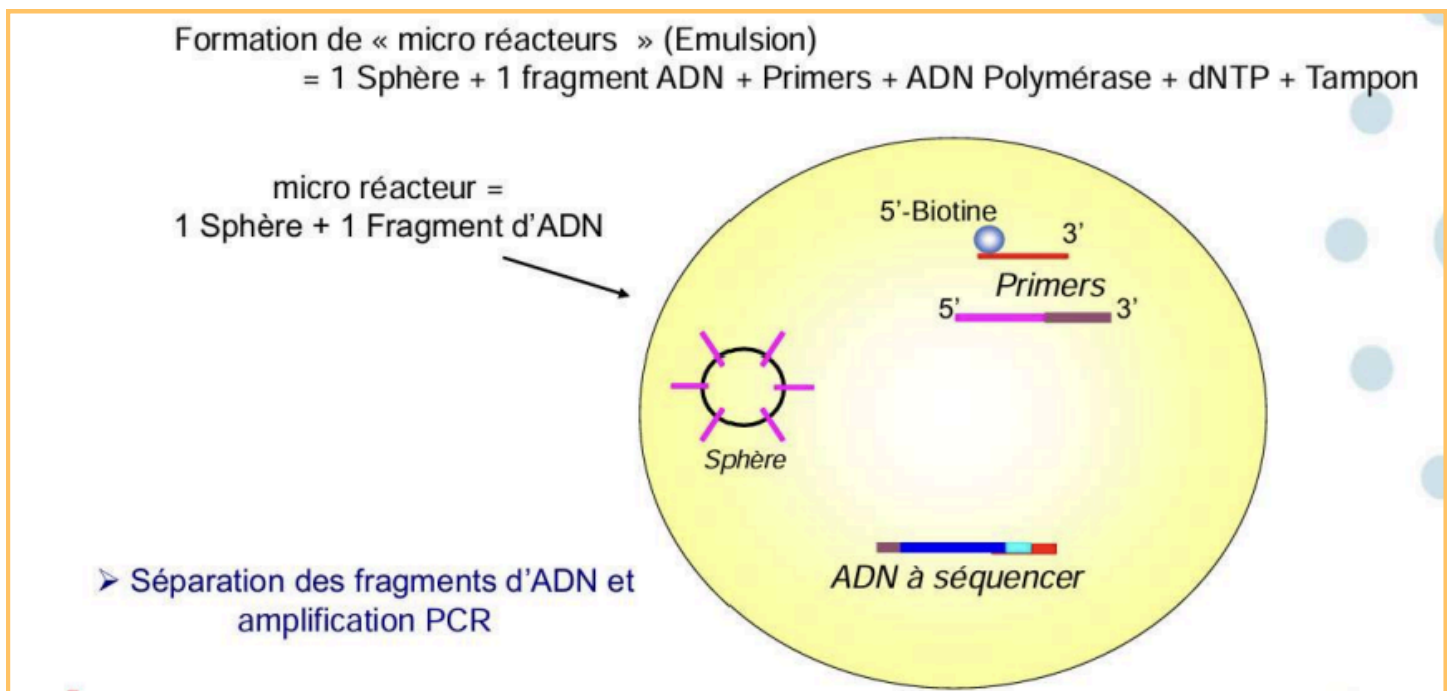
*On a vu comment on effectue la PCR et le séquençage avec la plateforme Illumina, on va faire de même pour ThermoFischer ! Courage, on s'approche de la fin <3*

## 5. Enrichissement par PCR clonale avec ThermoFischer

Avec les techniques d'amplification clonale sur la plateforme ThermoFischer, l'ADN capturé n'est pas fixé sur une lame de verre, mais **sur une sphère métallique**.

➡ Néanmoins, le principe reste le même (amplification clonale puis séquençage). Contrairement à Illumina, l'amplification clonale **ne va pas se faire par cluster mais dans un microréacteur créé par un système d'émulsion**.

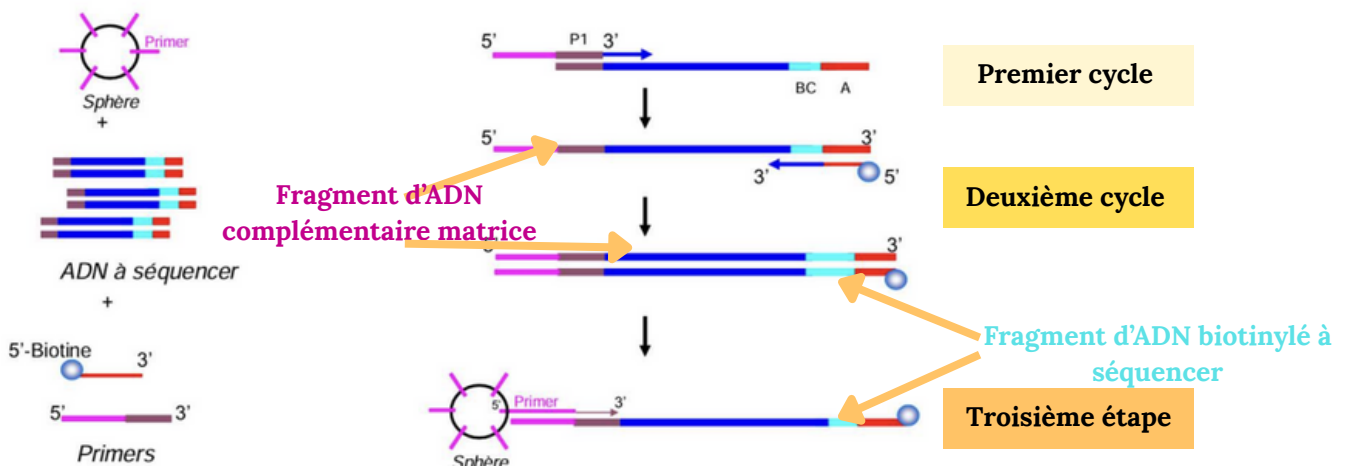
➡ Lorsqu'on mélange une substance grasse (ex : huile) avec une solution aqueuse, il y a la formation de microgouttelettes. Ce sont ces **microgouttelettes** qui vont former nos microréacteurs.



Le système est fait de sorte que chaque **microréacteur soit l'équivalent d'un tube contenant :**

- Une **sphère** avec des **oligonucléotides spécifiques et complémentaires** des adaptateurs de nos fragments d'ADN à séquencer
- Un **fragment d'ADN** à séquencer
- **Deux primers** : un biotinylé, l'autre non
- **L'ADN polymérase**
- Les **dNTPs**
- Un **tampon** (pour que la Taq Polymérase fonctionne correctement).

On réalise ensuite la PCR clonale avec les mêmes étapes de dénaturation, hybridation, élongation que la PCR classique.



### Premier cycle :

- Le **1er priméur** s'hybride au niveau de l'extrémité 3' de notre fragment d'ADN à séquencer.
- On a l'élongation du **brin complémentaire** à notre fragment d'ADN d'intérêt de 5' en 3'.
- On obtient donc un ADN double brin.
- On dénature.

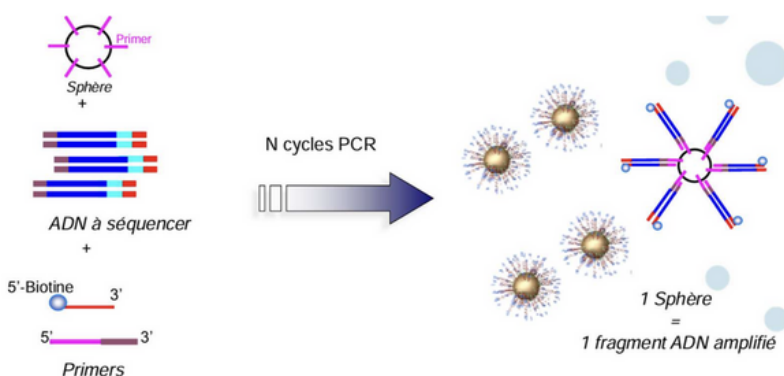
### Deuxième cycle :

- A présent, le **brin complémentaire** créé précédemment sert de **matrice** pour le 2e cycle.
- Le **2e priméur (biotinylé en 5')** s'hybride en complémentarité de l'adaptateur.
- La polymérase allonge, on fabrique **un nouveau brin, qui est donc le complémentaire du brin complémentaire**. Autrement dit, on revient à une **copie du brin original (cible)**, mais cette fois biotinylée.
- On dénature pour libérer le fragment d'ADN biotinylé à séquencer

### Troisième étape :

- Notre **fragment d'ADN biotinylé à séquencer** peut se fixer par complémentarité des bases sur le priméur localisé sur une petite sphère magnétique
- De cette manière, on peut piéger nos fragments d'amplification sur un support solide.

À chaque étape, nous avons amplification de chacun des fragments et fixation de ces fragments sur une sphère magnétique.



Après **n cycles de PCR**, on se retrouve avec cette **sphère recouverte d'une multitude de fragments d'ADN**, correspondant à notre fragment initial présent dans notre microréacteur et qu'on a amplifié.

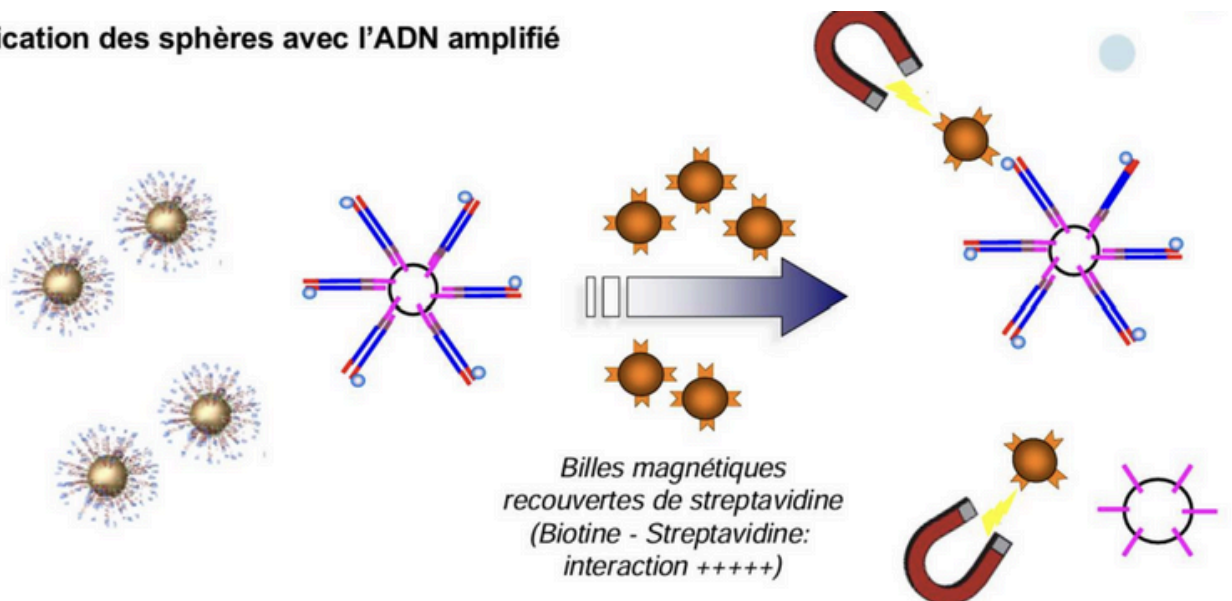
Regardez bien les couleurs de mon paragraphe et des images, elles sont là pour vous aider à comprendre <3

On parle de PCR clonale, parce que c'est réellement un clone d'une molécule d'ADN amplifiée dans chaque microréacteur avec, sur chaque sphère différente, des produits PCR uniques.

→ 1 sphère contient de multiples copies identiques d'un même fragment d'ADN (sphères « pures »)

→ ⚠ D'une sphère à l'autre les fragments d'ADN diffèrent. ⚠

### Purification des sphères avec l'ADN amplifié



### Quatrième étape :

→ Ensuite, il faut **purifier ces sphères** recouvertes d'ADN pour les récupérer.

→ Pour cela on va utiliser la propriété d'un des primers qui avait été recouvert de biotine : on va ajouter nos billes magnétiques recouvertes de streptavidine.

→ Encore une fois, nous avons une **interaction biotine / streptavidine**, sur laquelle, si nous appliquons un aimant, nous allons **pouvoir récupérer et purifier uniquement les sphères recouvertes d'ADN**.

→ Pourquoi faire ?

Il faut garder à l'esprit que dans le microréacteur nous pouvons avoir tout type de situation, **comme des microréacteurs n'ayant pas de fragment d'ADN amplifié**.

Il n'y aura donc pas de fixation de la streptavidine (car pas d'ADN fixé = pas de biotine = pas d'interaction biotine/streptavidine) : **ces sphères ne nous intéressent pas**. Grâce à cette interaction, nous allons donc pouvoir purifier uniquement les sphères qui nous intéressent.

*Je sais que cette partie est un peu dure, ne vous en faites pas, essayez de bien visualiser sans apprendre par coeur, si vous comprenez bien, vous allez retenir ! Posez vous tranquillement, et ça va le faire <3*

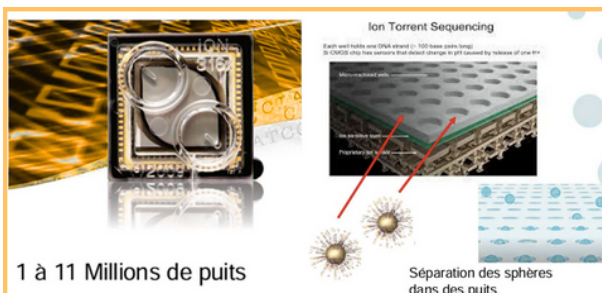
## 6. Séquençage individuel des sphères avec ThermoFischer

Nous allons ensuite récupérer nos sphères et les séparer physiquement en les plaçant individuellement dans des puits qui leur seront propres.

☞ Une seule sphère sera placée dans un seul puit.

☞ On a des toutes petites puces (quelques cm) contenant 1 à 11 millions de puits, dans lesquelles nous allons donc pouvoir séquençer 1 à 11 millions de sphères en parallèle.

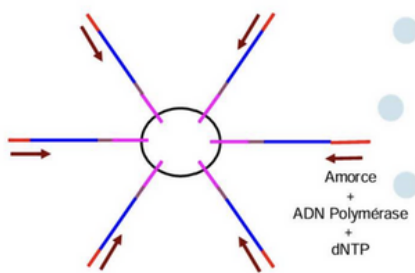
C'est donc un système extrêmement puissant, qui génère énormément de données de séquençage. Le séquençage se fait directement dans ces puces.



Si on reprend la définition du NGS, on comprend bien maintenant pourquoi ce séquençage est dit :

- **Massif** → 11 millions de réactions en même temps.
- **Clonal** → isolation puis amplification d'un même fragment identique de très nombreuses fois.

C'est ce qui fait la puissance du système



On va ajouter dans notre milieu réactionnel ce qu'il faut pour réaliser le séquençage :

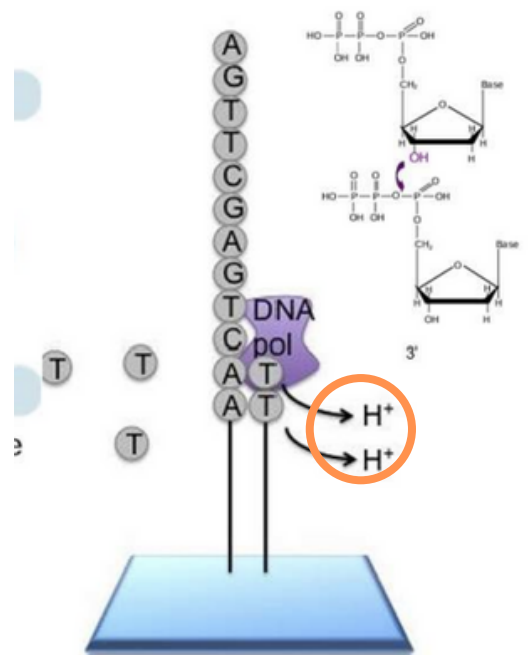
- Une **ADN polymérase**
- Des **dNTPs**
- Une **amorce** (pour s'hybrider à l'un des **adaptateurs** et à partir de laquelle la polymérase synthétisera le brin complémentaire)

Lorsque l'**ADN polymérase** ajoute un nouveau **nucléotide** pour synthétiser le brin complémentaire, il y a formation d'une liaison phosphodiester entre les nucléotides.

☞ La formation de cette liaison phosphodiester libère un ion **H+** dans le milieu réactionnel, entraînant des variations de pH.

Il faut visualiser ce séquençage comme :

- Chaque puits = un mini laboratoire.
- Chaque sphère = un clone d'un fragment unique.
- Le séquençage se fait sur la sphère dans le puits, et le signal pH indique la séquence.



L'automate va **lire/mesurer ces variations de pH dans tous ces puits en direct.**

Comment ça marche ?

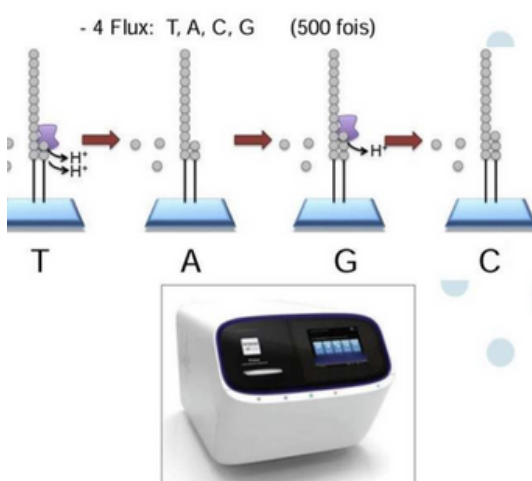
On **ajoute les nucléotides un par un dans un ordre précis** : T → A → G → C → puis on recommence (environ 500 cycles).

Quand un **nucléotide est incorporé** par la polymérase dans le brin en cours de synthèse :

- Il libère un ion H<sup>+</sup>.
- Cet ion H<sup>+</sup> entraîne une variation locale de pH détectée par un capteur sensible.

Si le **nucléotide n'est pas complémentaire**, il n'est pas incorporé → **pas de libération d'H<sup>+</sup>** → pas de signal.

Si plusieurs nucléotides identiques s'incorporent d'affilée (ex. AAA), la polymérase libère plusieurs H<sup>+</sup> → le signal est proportionnel au nombre de bases ajoutées.



### Pour résumer :

☞ C'est la succession contrôlée d'injections (T, A, G, C, ...) qui permet de savoir quel nucléotide a été incorporé à chaque cycle.

*En gros, si on injecte un nucléotide T et qu'il y a une variation de pH, cela signifie que T se trouve à cet endroit dans notre séquence.*

*En revanche, si on injecte un nucléotide A et qu'il n'y a aucune variation de pH, cela signifie que le A ne se trouve pas dans cet endroit précis de la séquence.*

## 7. Analyse bio-informatique

Il va falloir transformer nos données obtenues (pH, fluorescence) en données de séquence.

Le travail bio-informatique réalisé à la fin de notre séquençage est énorme.

**C'est grâce à ces outils que nous avons pu développer nos nouvelles techniques de NGS : à la fois par la miniaturisation des systèmes (milieux réactionnels, automates...) mais également par le développement des outils bio-informatiques.**

Nous allons donc détailler les différentes étapes lors de cette analyse bio-informatique.

*Courage, tu es quasiment sur la fin, je suis trop fière de toi !! <333*

## Étapes :

1

### Détermination des bases (base calling) :

→ **Transformation** des signaux générés (pH, fluorescence) en données de séquence

2

### Étape de contrôle qualité

→ **Élimination** des lectures de séquences de mauvaise qualité (données générées : des millions.

Il y a forcément des puits dans lesquels la mesure ne s'est pas faite correctement car pas de produit PCR ou alors 2 produits PCR présents initialement donc l'amplification n'a pas été clonale, etc.).

→ On ne garde que les lectures de bonne qualité.

3

### Étapes d'alignement :

→ **Positionnement des séquences d'ADN générées sur la séquence de référence** : on va essayer d'aligner nos différents « reads » / lectures sur notre séquence de référence.

→ Après avoir tout fragmenté et tout séquencé individuellement, on « reconstruit le puzzle » grâce à l'informatique.

→ On prend pour référence le génome humain puisqu'il est entièrement connu. On pourra ensuite chercher les variants / les différences

4

### Étape d'analyse :

→ On regarde si les gènes que l'on voulait étudier, ont été séquencés correctement : Est qu'il a été séquencé dans sa totalité ? Quel a été le nombre de lecture pour chacune des bases ?

## 2 notions très importantes :

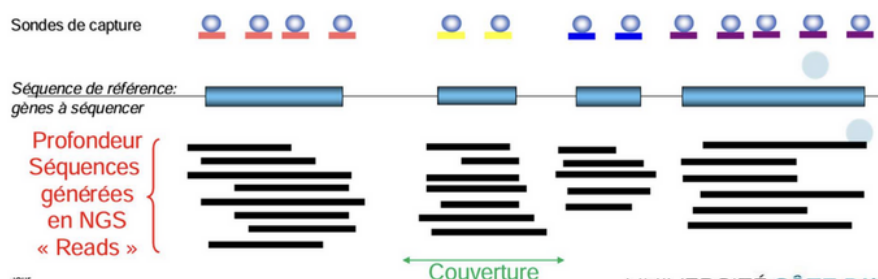
→ **Profondeur de lecture** : nombre de lectures « reads » indépendantes d'une base.

C'est le nombre de fois où chacune des bases d'intérêt a été lue sur des reads différents.

→ **Couverture** = est ce que je lis bien toute ma séquence d'intérêt ?

$$\frac{\% \text{ de bases séquencées par lecture indépendante}}{\text{Nombre total de bases de la région d'intérêt}}$$

C'est une valeur donnée pour une profondeur de lecture donnée.



☛ On vérifie donc que tout l'exon a été lu correctement et si la profondeur est suffisante.

☛ On considère un séquençage de bonne qualité lorsque chacune des bases de l'ensemble des exons a été lue au moins 20 à 30 fois (au moins 20 à 30 séquences indépendantes).

## 5

### Étape d'annotation des variants

→ Consiste à **rechercher les localisations chromosomiques des variants sur la séquence**, en confrontant la séquence de notre patient et celle de référence.

→ Les annotations, permettant une classification, vont être :

- **Localisation** sur le chromosome ? Sur quel gène ? À quel endroit du gène ?
- **Variants** nucléotidiques exoniques ? introniques ?
- **Conséquences** sur la protéine après traduction ? Caractère pathogène ? Si on a une mutation qui donne un codon stop très tôt dans le gène, on n'aura pas la protéine.
- **Fréquence** de ce variant ? Dans quelles populations ? (Caucasiennes, asiatiques...)  
Si votre variant est retrouvé chez 70 % de la population on ne va pas pouvoir le tenir responsable d'une maladie rare.

Le gros de ce tri et de l'annotation (et quelques prédictions) est heureusement fait par l'informatique, par des algorithmes de plus en plus performants. Cela aide à l'interprétation du variant, et à créer des bases de données.

On classe donc les variants en 5 classes :

1. **Bénin** → Polymorphisme, variante génétique fréquente dans la population, sans conséquence pathologique.
2. **Probablement bénin** → Variante a priori sans impact clinique, mais pour laquelle il existe encore un doute. En général, on ne la prend pas en compte dans le diagnostic.
3. **Variants de signification inconnues** → Variante dont on ne sait pas encore si elle est pathogène ou bénigne. Ce sont les plus problématiques aujourd'hui, car on ne peut pas les classer avec certitude.
4. **Probablement pathogène** → Variante suspectée d'être responsable de la maladie, mais sans preuve complète (il manque certains arguments).
5. **Pathogènes** → Variante clairement démontrée comme causale, décrite dans la littérature et associée à une maladie.

## 6

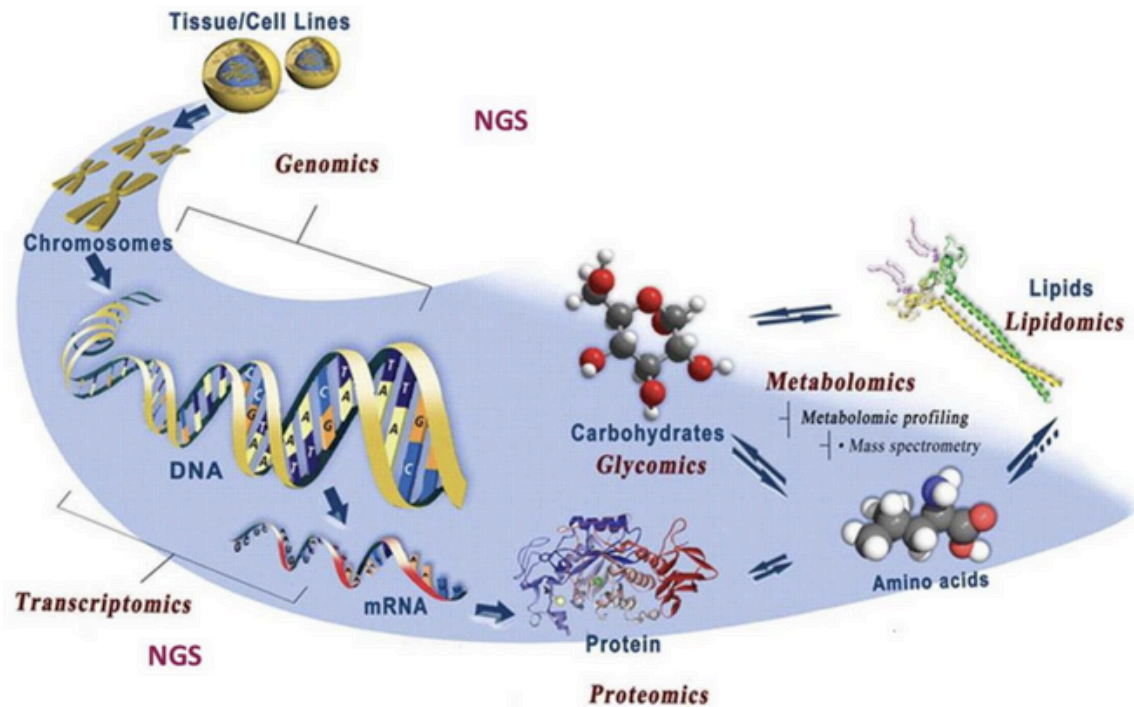
### Étape d'interprétation des variants = expertise clinico-biologique

→ Étape **la plus importante et la plus difficile**, séquencer est devenue facile mais la difficulté se trouve dans l'interprétation.

→ Nous avons une **confrontation** biologistes/cliniciens, car pour bien interpréter un variant, il faut l'interpréter dans son contexte biologique propre mais aussi en regardant le dossier du patient et notamment en prenant en compte les signes cliniques du patient, l'histoire familiale...

→ Il nous faut répondre à la question suivante : **est-ce que les variants identifiés par NGS permettent d'expliquer le phénotype et la clinique que le patient présente et de lui donner un diagnostic ?** Il faut trouver le variant responsable de la pathologie parmi de nombreux variants.

## L'ère des Multi-OMICS



☛ Le NGS est l'une des briques du multi-OMICS : génomique (génome, exome), transcriptome. D'autres techniques s'intéressent aux protéines (protéomique), au métabolisme (métabolomique).

☛ Ce n'est pas tout de séquencer le génome mais il faut savoir l'interpréter. Pour une mutation déjà connue, un syndrome bien défini, l'identification et l'interprétation d'un séquençage peut aller très vite, c'est simple. Cependant, dans beaucoup de maladies génétiques et notamment les maladies rares, tous les gènes ne sont pas connus. On connaît la séquence complète du génome mais pas encore le rôle de tous les gènes.

### Utilité du NGS

**Séquençage de plusieurs gènes**  
(= panel de gènes puisqu'on s'intéresse à plusieurs centaines de gènes)

**Séquençage de l'exome en entier** (= toutes les régions codantes d'un génome) (**WES**)

« **Whole génome sequencing** » (on séquence tout le génome de manière intégrale) (**WGS**)

**DPNI**, dépistage prénatal non invasif, qui permet de rechercher les trisomies 13, 18 et 21



**Analyse des ARN : RNA-Seq (expression, épissage alternatif...) (Transcriptome)**

## 🔍 ZOOM sur le rôle du séquençage dans l'analyse de l'ARN :

👉 On va séquencer les ARN, pour avoir une **idée sur l'expression des gènes (quantitatif)** mais aussi pour **appréhender les variants d'épissage (qualitatif)** dans tel ou tel tissu.

👉 On va ainsi pouvoir apercevoir un peu plus précisément **la fonctionnalité d'un gène** car ce sont les ARN qui seront traduits en protéines. On s'aide d'un aspect plus fonctionnel pour comprendre.

Au niveau de l'ARN on va voir les **modifications au niveau de l'épissage** mais ça ne nous donnera pas la mutation responsable.

👉 On **peut donc croiser les données** (génomique + transcriptome par ex.) pour avoir plus d'informations. Si on trouve un problème au niveau de l'ARN, on revient à notre séquence d'ADN pour voir si on peut trouver la variation nucléotidique responsable de cette anomalie dans notre ARNm.

---

Les **multi-omics** sont vraiment de **nouvelles technologies** dont on entendra beaucoup parler dans les années à venir, puisque l'avenir n'est pas dans le séquençage (déjà fait), mais dans l'interprétation et l'application clinique.

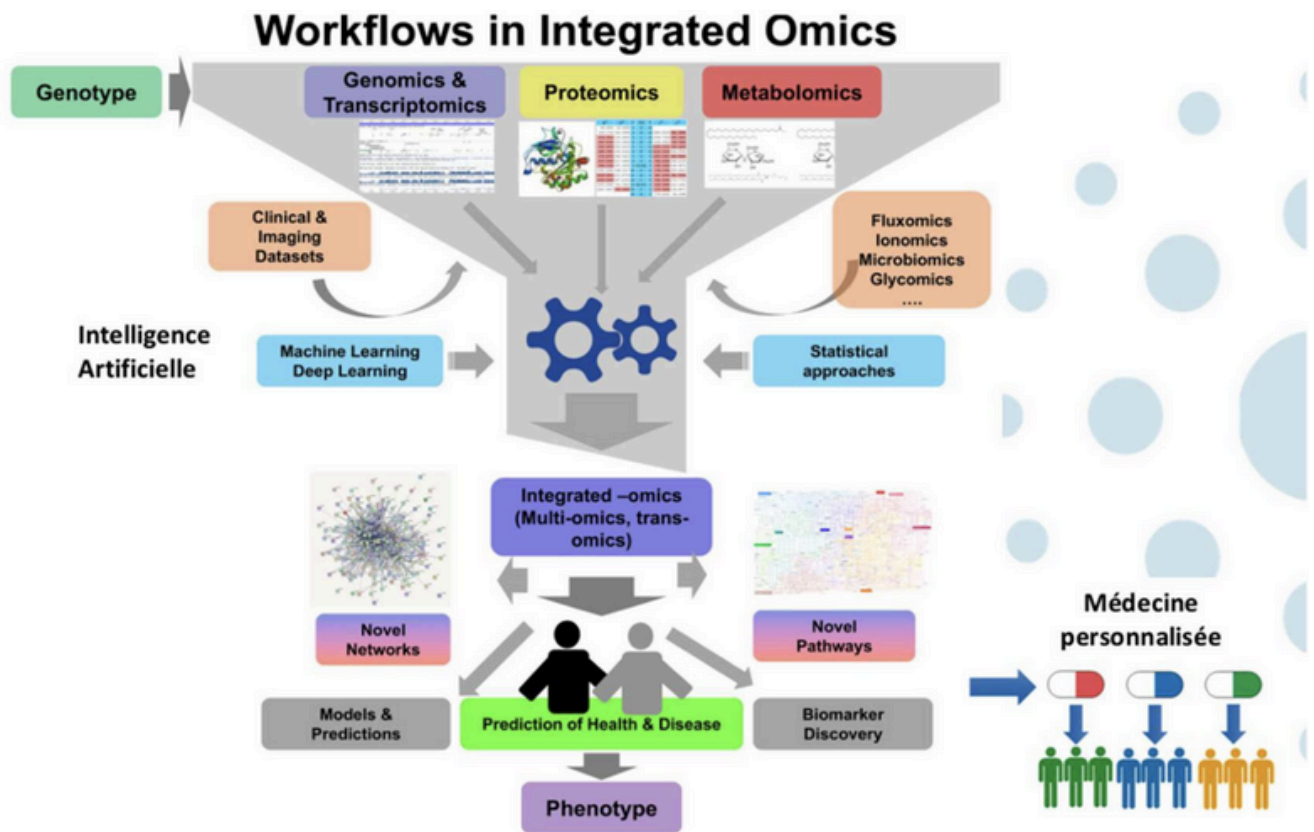
👉 Il va falloir obtenir la génomique d'un individu dans son intégralité, pour comparer avec sa transcriptomique, sa protéomique, sa métabolomique.

Grâce aux **progrès de l'informatique et de l'intelligence artificielle**, il devient possible de **croiser l'ensemble des données issues du séquençage** afin d'obtenir une vision globale du métabolisme et de ses altérations. Ces approches ouvrent la voie à une meilleure compréhension et utilisation des réseaux métaboliques.

Le séquençage NGS contribue également au **développement de thérapies ciblées**, à la base de la « médecine personnalisée » : adapter le traitement à chaque patient en fonction de ses voies métaboliques, de ses marqueurs et de ses caractéristiques propres, plutôt que de prescrire un même médicament pour tous.

Aujourd'hui, on combine déjà le génome et le transcriptome, et on y ajoute progressivement les données issues de la protéomique et de la métabolomique, ce qui conduit à des approches multi-omiques.

En France, **deux plateformes nationales de séquençage** du génome prennent en charge les échantillons lorsque les analyses génétiques classiques ne permettent pas d'identifier la mutation responsable. L'intelligence artificielle est de plus en plus mobilisée pour analyser et intégrer ces masses de données issues du séquençage (génomique, transcriptome, protéome, métabolome, etc.).



### CONCLUSION (ça y est les gars, c'est la fin 😊😊):

La génétique moléculaire est une discipline qui **évolue extrêmement rapidement**.

Il faut toujours être **à l'affut des nouvelles technologies**, de ce qui sort dans la bibliographie, des derniers résultats (on peut être amenés à reprendre un dossier 10 ans plus tard car une nouvelle publication est sortie).

Par exemple, ThermoFischer et Illumina voient déjà une nouvelle génération de séquenceurs arriver alors que ceux actuels sont mis en routine dans les laboratoires depuis à peine une dizaine d'années.

On est très dépendants de ces connaissances, et ces connaissances sont dépendantes de ces nouvelles technologies. Ce n'est pas parce qu'on **séquence** le génome/l'exome/une grande quantité de gènes en 2 jours qu'on a les **résultats** d'un test génétique en 3 jours due à la **grande difficulté d'interprétation** des variants et la **nécessité de concertations pluri/interdisciplinaires** entre plusieurs spécialistes dont des cliniciens et des biologistes (expertise clinico-biologique) pour l'interprétation des variants identifiés en NGS.

Selon les variants qu'on va trouver, on peut mettre 4 à 5 ans voire plus à donner un diagnostic car il a fallu trouver la fonction de la protéine, comprendre, faire le lien avec la clinique (rendre des variants pour rendre des variants n'a pas de sens).

Enfin pour finir, le meilleur moment de la fiche (la dernière olala), les dédis :

Pour commencer, je vais faire la plus grosse dédidace du monde au TUUUUUUUTORAT NICOIS <333

J'ai fait tellement de rencontres en si peu de temps, et je peux vous dire que vos tuteurs sont vraiment des êtres exquis ! Si vous avez l'occasion de le faire l'année prochaine, sautez le pas, c'est une expérience fantastique : vous ne le regretterez pas ;)

Dédi aux lieux qui ont accompagné ma P1 à savoir la BU de Valrose (avec NOTRE TABLE) et le campus de Carlone (un labyrinthe dans lequel j'ai quelques bons souvenirs)

Dédi aux cours de Lettres en P1 avec Lucacide et Lauraorte surtout ceux de Mr. Biscéré (j'aimerais y retourner aled)

Dédi à Isaure dont le courage ne cessera jamais de m'impressionner, à Pierre et cette soirée où on a récupéré ce crapaud sous le canapé, et à Marine et Faustine qui sont des cousines de cœur pour nous

Dédi à Laurine et à nos jeux interminables dans les champs accompagné de Monsieur Hiboux

Dédi à Lina, qui fera médecine l'année prochaine, je n'ai aucun doute sur ta réussite, tu as toujours été une jeune fille brillante et persévérante <3

Dédi à mes adorables vieux : Naomi, Yacine et ma marraine de P2 Maëva :')

Dédi à tout ce qui rend cette vie un peu meilleure tel que Disney, Ghibli, et Patrick Swayze (cette liste n'a aucun sens)

Dédi à mes meilleurs souvenirs de P1 pour que vous vous rendiez que OUI c'est possible de passer une bonne année de P1 :

- La Tut'Rentrée du S1 où on a dansé nos premiers Make You Move
- La sortie à St Jean Cap Ferrat avec Marine, ses fillots, Lucie, Iris et moi après le 1er EB
- Les sorties cinémas avec ma sœur pour aller voir le film de Inoxtag (c'était sa volonté mais ça reste un bon souvenir), et avec Aimie pour aller voir L'Amour Ouf
- Les cours de Biophysique, d'Anatomie, et de Génétique en amphitheâtre
- Les appels hebdomadaires avec Nono le dimanche soir
- Le mois de décembre post concours avec le voyage à Paris pour Noël (oui c'est possible de partir quelques jours entre vos 2 semestres, promis ça régénère <3)
- Les sorties après les EB au 2e semestre qui ont été ma carotte pour tenir la fin d'année
- Le voyage au ski où j'ai pas fait de ski (mdrr l'objectif du coup?) mais le simple fait d'être à la montagne et de manger du fromage, c'était génial
- Notre sortie à Eze la veille du concours du S2 avec Iris et Sara

Bref tout ça pour vous dire que ces petites choses peuvent transformer les choses pour vous, autorisez vous des petits plaisirs, ça vous sourira !

Pour finir, dédi à la moi d'il y a un an terrifiée à l'idée de rater, spoiler : t'as réussi 😊

Et dédi à toi, parce que tu es toujours là à lire les cours pour atteindre ton objectif, tu peux être fière de toi ! J'espère lire tes dédis l'année prochaine parce que je SAIS que tu en es capable ;)

Anti dédi à l'administration de la FAC qui nous prévient 3 jours avant pour tout et n'importe quoi et aux machines à café qui servent de l'eau