

Génétique 2023/2024



Cours 3 : Séquençage Haut Débit (NGS)

Le séquençage haut débit est utilisé en **diagnostique pour les maladies génétiques** pour séquencer un **très grand nombre de gènes**, voir un exome au complet (=toutes les régions codantes).


C'est un séquençage massif : on l'utilise lorsque l'on veut **séquencer énormément de gènes**.

On pourra bien sur 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**+++.

Il est donc plus utile de faire une **PCR-RFLP, plutôt qu'un NGS** qui va séquencer **tout le génome**.

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

Ce qui a permis l'arrivé 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.

 Petit rappel historique :

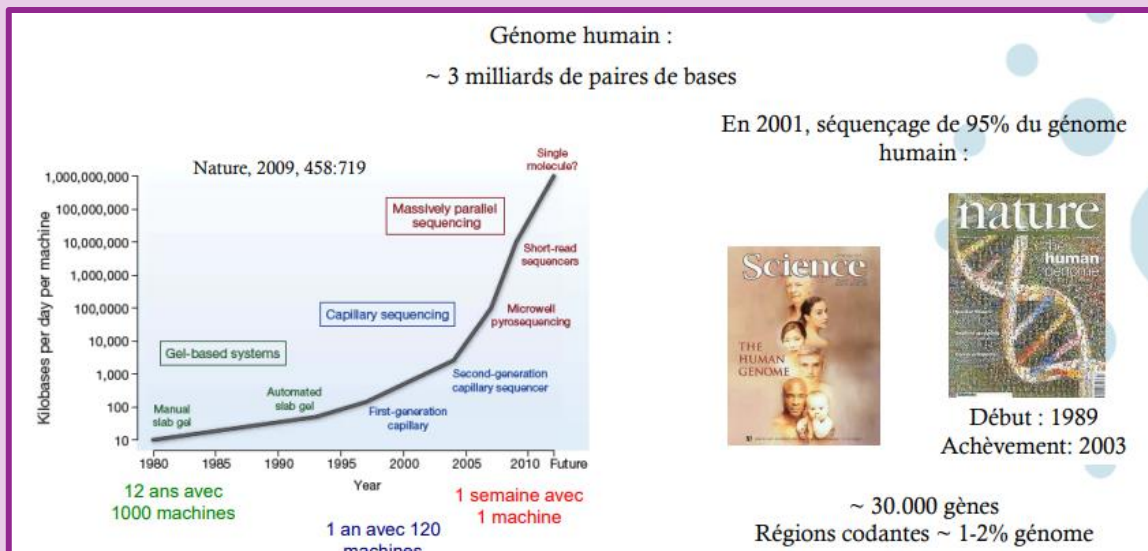
Le « **Séquençage Manuel** »
(=Sanger) inventé

En **1977** (avec ses DDNTPS radiomarqués)

Le « **Séquençage Automatique** » avec ses
DDNTPs

Fluorescents, datant de
1993

Depuis **2007**, on a le « **Séquençage Haut Débit** »
qui révolutionne la biologie
moléculaire
(Next Generation Sequencing)



Pour vous donner un ordre de grandeur, le génome humain correspond à **3 milliards de paires de base (pb)**.

En **1980**, si on avait utilisé la méthode Sanger, il aurait fallu **12 ans avec 1000 machines PCR** pour réaliser le séquençage complet d'un génome humain.

Dans les années **2000** avec le séquenceur automatisé, il aurait fallu **1 an avec 120 machines** pour séquencer le génome complet.

Aujourd'hui, avec le NGS, on est capable, en **1 semaine avec 1 machine**, de réaliser un séquençage du génome humain de plusieurs dizaines d'individus en même temps.

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

Pour mémoire, dans le génome humain, on trouve 30 000 gènes et les **régions codantes (exons) ne représentent que 1 à 2% du génome**.

Par définition le séquençage au débit est un séquençage massif en parallèle de molécules d'ADN individuellement séparées et amplifiées sous forme de clones ou de molécules uniques +++

Actuellement il existe **deux plateformes** (2 sociétés) qui commercialisent deux technologies, qui sont différentes au niveau des étapes d'amplifications et de séquençage. Il s'agit des sociétés **Illumina et ThermoFisher** (ou LifeTechnology) avec la technologie Ion Torrent.

Il faut noter qu'il existe une troisième société un peu en retrait ces dernières années : la société « **Roche** » qui, comme les deux autres, s'est lancée dans cette nouvelle technologie qu'est le NGS.

Les deux plateformes à retenir sont Illumina et ThermoFisher / Life Technologie+++

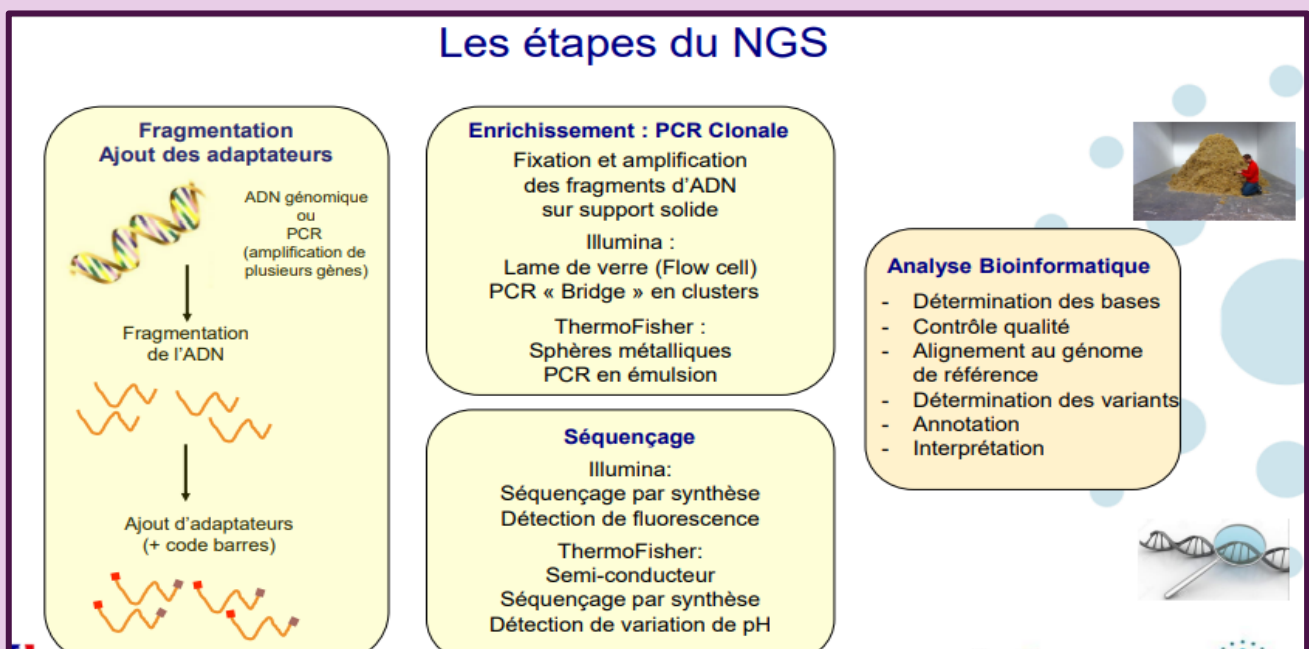


Voici les liens vers les différentes vidéos YouTube. Je vous conseille de les regarder ++++ car ça aide grandement à visualiser les étapes.

1. Les étapes du NGS

Quelle que soit la plateforme (Illumina ou ThermoFisher), on retrouve 4 étapes principales dans le NGS :

1. La **préparation des échantillons** ou préparation des **libraires** qui consiste à **fragmenter l'ADN** et ajouter des **adaptateurs** et des **barres-codes (BC)** au séquenceur d'ADN génomique.
- 2- La **PCR clonale** qui consiste à **amplifier** chacune de ses molécules d'ADN génomique d'intérêt.
- 3- Le **séquençage de ces molécules** d'ADN génomique d'intérêt.
- 4- **L'analyse bio-informatique** qui est nécessaire au traitement des données générés par les séquenceurs de façon à identifier les variants et à rechercher d'éventuelles mutations (par ex)



La première étape est identique quel que soit la plateforme utilisée (illumina ou ThermoFisher).

Elle consiste à **fragmenter notre ADN génomique** ou à fragmenter des produits PCR dont les tailles seraient supérieures à 200 ou 400 Pb. En effet, le NGS permet de séquencer de fragments dont la taille est **d'environ 200 à 400 pdb** en fonction des kits et des plateformes utilisés.

Lors de cette première étape, on va **couper notre ADN** en petit fragments à la taille nécessaire au séquençage.

A ces petits fragments d'ADN vont être ajoutés des **adaptateurs et des BC**, c'est-à-dire des **oligonucléotides** dont la séquence est **connue** qui vont permettre :

 **L'amplification PCR pour les adaptateurs**

 **L'identification des patients pour les BC**

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 pouvoir **mélanger nos patients** dans un même run, dans une même réaction. Il va donc falloir ensuite, **identifier et attribuer chacune des séquences à chacun des patients**.

C'est toute l'analyse bio-informatique qui va trier ces séquences et qui va permettre en reconnaissant le BC numéro 1 d'attribuer l'ensemble des séquences à ce patient numéro 1.

Les **adaptateurs** sont ajoutés **aux extrémités 5' et 3'** des fragments d'ADN et on verra qu'ils vont permettre **d'amplifier l'ensemble de nos fragments d'ADN avec uniquement un seul couple de primer** (1 primer sens + 1 primer reverse).

La seconde étape est un enrichissement de nos fragments d'ADN par PCR clonale.

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 :

- **Technologies illumina = lame de verre**

(PCR en pont, en cluster qui permettent d'amplifier chacun des fragments)

- **Technologie Thermo Fischer = sphères métalliques** et on verra que la PCR clonale se fait dans un système de PCR en émulsion.

Ces supports solides (lame de verre, sphère métallique) vont être séquencés individuellement avec des technologies un petit peu différente en fonction des plateformes :

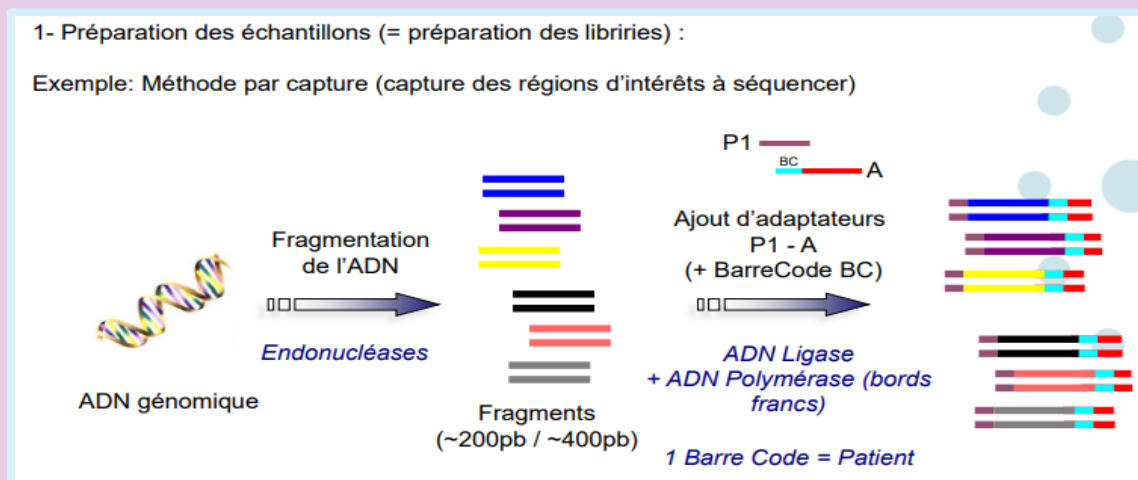
- Illumina, il s'agit du séquençage par synthèse avec détection de **fluorescence**

- ThermoFischer utilise la technologie des **semi-conducteurs**, on verra que les étapes de séquençage, engendrent des **variations de pH** vu par ses automates

La quatrième étape commune et extrêmement importante, l'analyse bio-informatique.

On va générer des données massives et il va falloir ensuite transformer ces **signaux lumineux** (détection fluorescence à illumina) ou transformer **ces variations de pH** (thermofischer) en « nucléotides », en **données de séquences**.

1) Préparation des échantillons



On va voir comment on peut **fragmenter** et **ajouter les adaptateurs** sur ces échantillons (ADN génomique).

Dans cet exemple on va voir le système de capture, c'est-à-dire qu'à partir **d'une extraction d'ADN génomique** on va être en mesure de récupérer **uniquement les régions qui nous intéressent** (régions exoniques en fonction des sondes possibles).

Cet ADN génomique va être tout d'abord **fragmenté**, en morceaux de 200, 400 pb. Pour cela, on va utiliser des **endonucléases** qui ont comme propriété **de couper l'ADN double brin en son milieu**. Il y a une combinaison bien particulière de reconnaissance. On fragmente donc avec plusieurs endonucléases qui vont générer des fragments **aléatoirement sur l'ADN génomique**.

Ensuite à ces fragments d'ADN on va **rajouter des adaptateurs et des barres codes**.

L'adaptateur P1, se place en 5' et les **adaptateurs A** qui vont se placer **en 3'**.

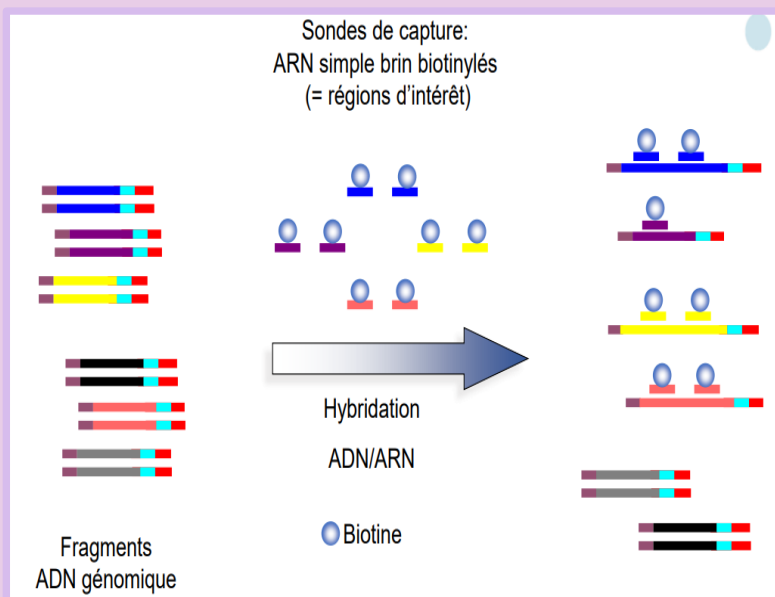
Entre les adaptateurs A et notre fragment d'ADN on aura notre BC qui sera spécifique de notre patient.

Donc la séquence de **P1 et de A** sont **identiques quel que soit le patient**. C'est la séquence du **barre-code** (une dizaine de nucléotides) qui **diffère** et qui est **spécifique** de tel ou tel patient.

Pour tout cela on retrouve les outils vus précédemment :

- Des **ADN ligases** (qui vont lier les fragments d'ADN, les adaptateurs, les barre codes)
- Des **ADN polymérases** car en fonction de la digestion réalisée par les endonucléases, les extrémités ne sont pas forcément toutes à bords francs.

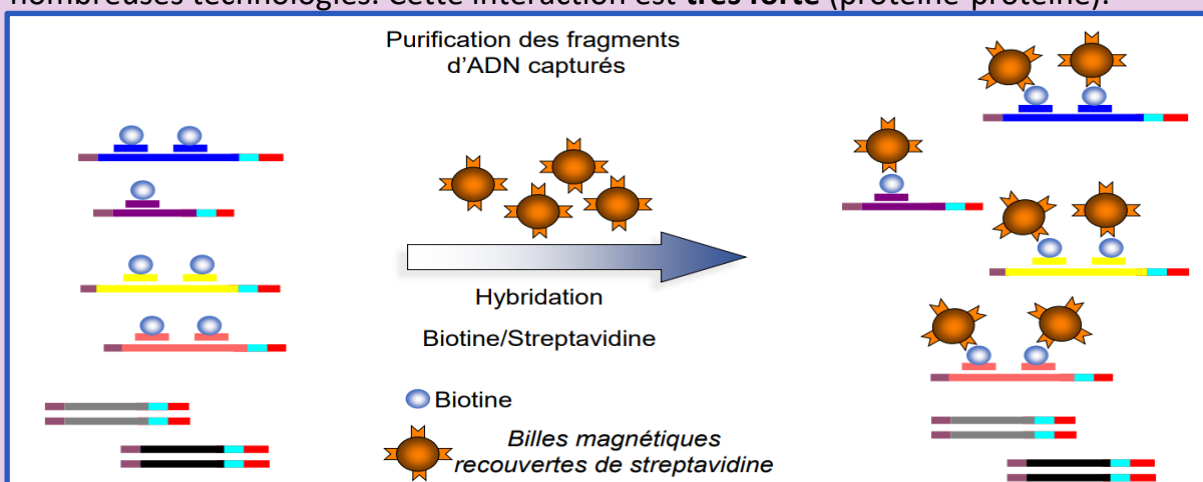
Cette combinaison d'enzyme va permettre à l'adaptateur P1 d'être intégré en 5' et au BC et à l'adaptateur A d'être ajouté au fragment d'ADN en 3'. Maintenant qu'on a obtenu nos fragments d'ADN qui sont munis d'un BC et des adaptateurs, on peut mélanger plusieurs patients, car ils seront reconnus par la séquence des BC.



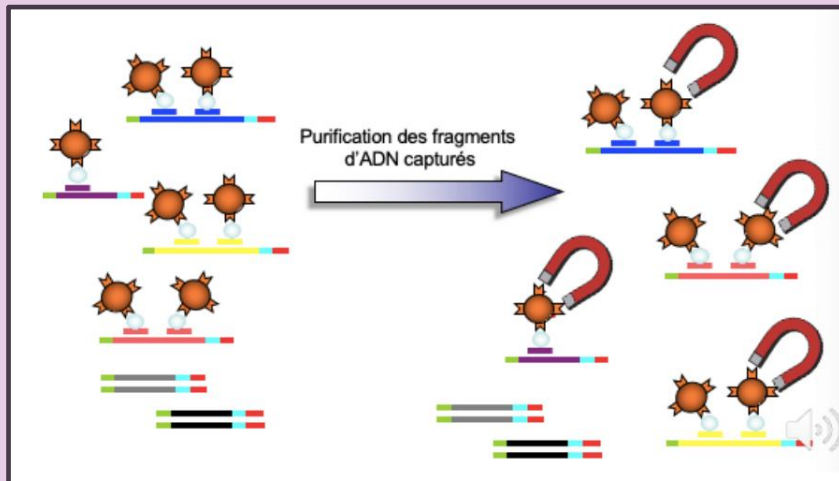
On ajoute à ces fragments des **sondes de capture (ARN simple brin biotinylés)** qui correspondent au niveau **des séquences à des complémentaires** de la séquence d'ADN que l'on veut séquencer.

On aura la formation **d'hybride ADN/ARN**, avec ces **ARN biotinylés** qui vont se localiser sur les fragments d'ADN qui nous intéresse.

Nous allons ensuite **purifier ces fragments d'ADN** capturés grâce à des **sondes** et pour cela on va rajouter des **billes magnétiques** (recouverte de **streptavidine** qui forme l'interaction « **biotine-STREP** ». Cette interaction est utilisée dans de nombreuses technologies. Cette interaction est **très forte** (protéine-protéine).



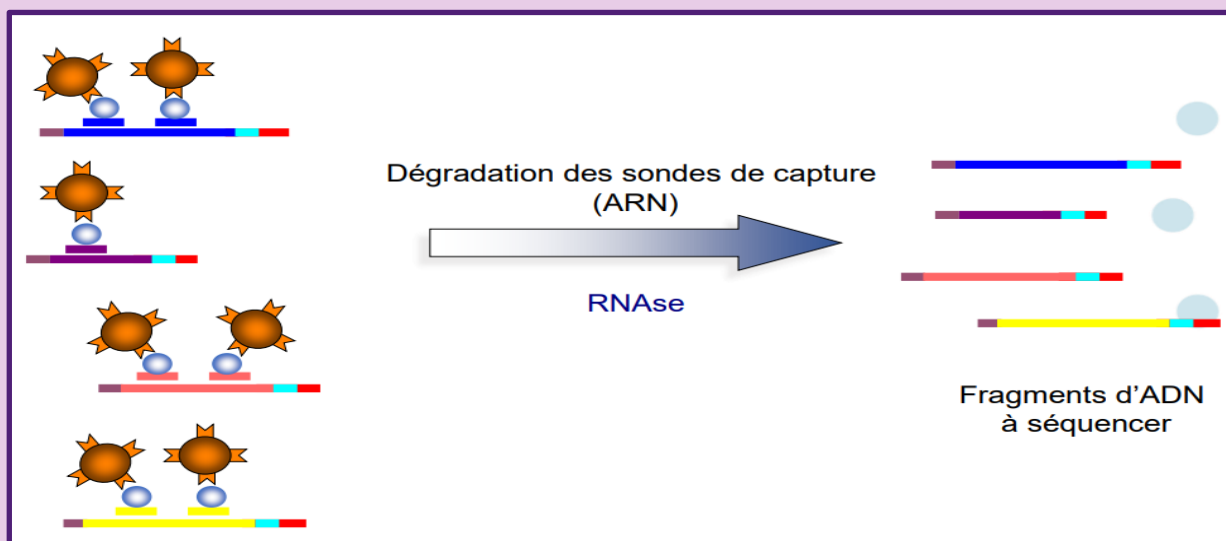
Lorsqu'on rajoute ces billes recouvertes de streptavidine, l'interaction va permettre à tous les fragments d'ADN reconnus d'être piégés sur les billes. Les régions qui nous intéressent (celles que l'on veut séquencer) vont donc être **capturées et récupérées**.



Les **billes recouvertes de streptavidine** étant des billes **magnétiques**, si on ajoute des **aimants** sur le bord du tube, les billes vont être **attirées sur les côtés** et on va pouvoir réaliser des **lavages**. (=ajouter des liquides, des tampons) Ainsi on nettoie et on ne récupère que les billes qui auront fixés ces **ARN sondes** elles-mêmes

hybridées sur nos **régions d'intérêt**. Les billes donnent **une coloration marronne** à notre tube.

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**, qui dégradent l'ARN hybridé dans un duplex ADN/ARN. Il ne restera ainsi que l'ADN à séquencer. On obtient donc les régions **d'ADN désirés** avec **leur BC**, et **les adaptateurs**.



Quelle que soit la plateforme utilisée, les échantillons seront toujours préparés de cette manière.

A présent nous allons détailler la technique de séquençage des plateformes Illumina et ThermoFisher.

2. Illumina

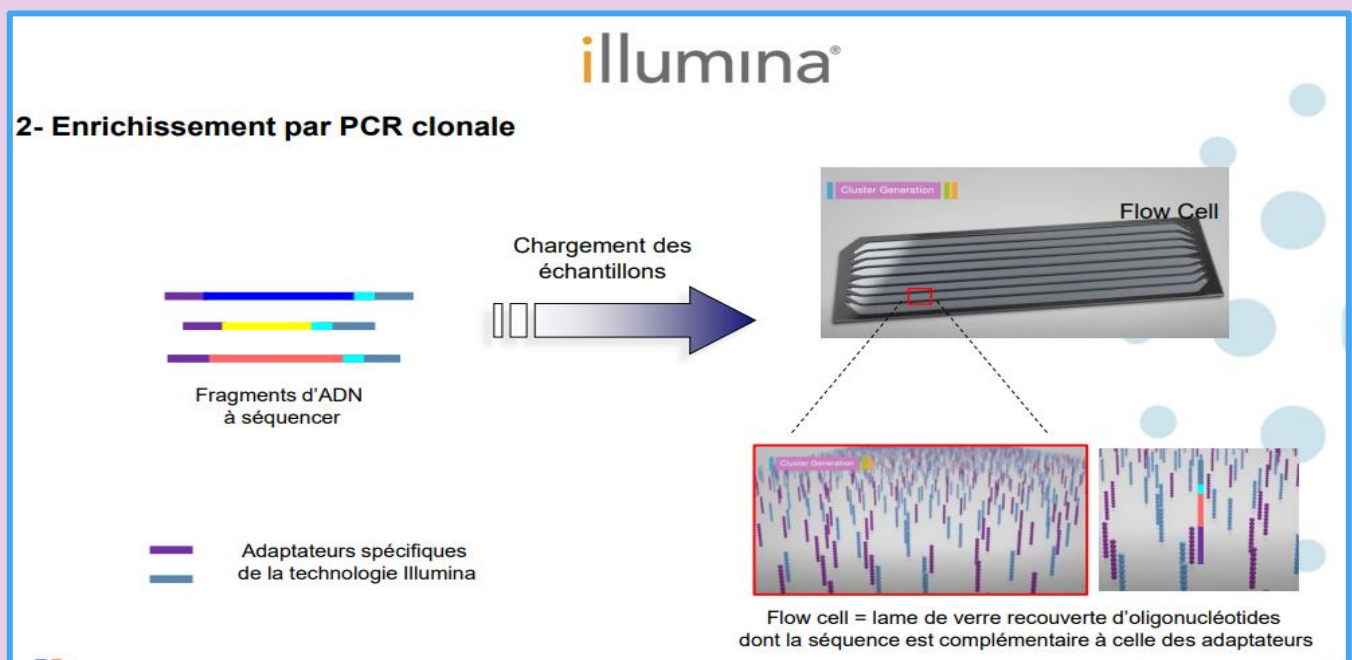
1) Lame de verre

Pour la technologie Illumina, l'Enrichissement ou la **PCR clonale** est réalisée sur une **lame de verre**, également appelée **Flow Cell**.

Cette lame est recouverte de **2 types d'oligonucléotides fixés** :

🧪 Des **oligonucléotides spécifiques des adaptateurs** ajoutés en 5' (ont la séquence est complémentaires aux adaptateurs en 5'),

🧪 **Spécifiques**, qui présentent une **séquence complémentaire aux adaptateurs placés en 3'** de nos fragments d'ADN.



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

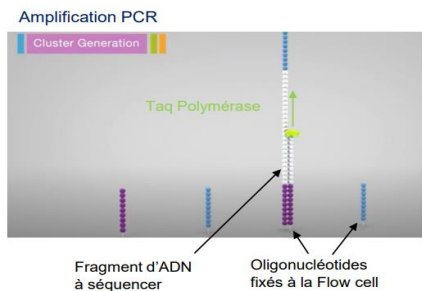
Il faut donc imaginer que l'ensemble de nos échantillons vont être **hybridés à différents endroits, sur cette lame de verre**.

À partir de ces **fragments fixés sur cette lame de verre**, on va réaliser une **amplification** de chacun de ses fragments.

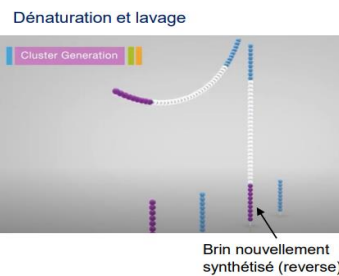
Cette amplification va se faire par **PCR** : on va donc retrouver les caractéristiques de cette technique, et tout ce qu'il faut pour amplifier un fragment d'ADN. On retrouve donc notamment une **Taq polymérase**, qui va à synthétiser **de 5' en 3'** un brin complémentaire au brin hybridé sur notre lame de verre.



2- Enrichissement par PCR clonale



Amplification PCR des fragments d'ADN fixés sur la Flow Cell



⇒ Seulement les brins nouvellement synthétisés restent fixés sur la Flow Cell



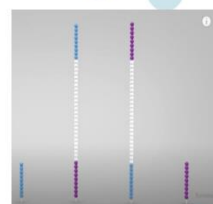
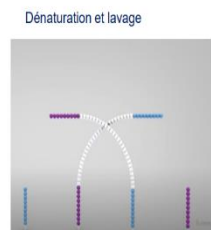
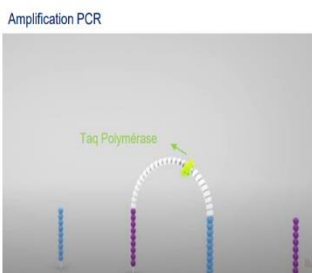
Hybridation de l'extrémité 3' du brin d'ADN nouvellement synthétisé avec les 2èmes oligonucléotides (séquence complémentaire à celle de l'adaptateur)

<https://www.youtube.com/watch?v=fCd6B5HRaZ8>

2) Dénaturation et lavage

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

2- Enrichissement par PCR clonale – PCR « Bridge » en clusters



⇒ Seulement les brins nouvellement synthétisés restent fixés sur la Flow Cell

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

En revanche, le fragment que nous avons ajouté et qui provenait de notre échantillon va pouvoir être enlevé de cette lame. Il


subit l'étape de dénaturation et de lavage puisque son amorce **n'est pas fixée de manière covalente** à la lame de verre (*il était simplement hybridé par complémentarité des bases à l'aide de l'oligonucléotide violet sur le diaporama*). **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.

3) Formation des ponts

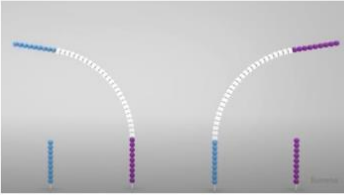
Vient ensuite l'étape de la **formation des ponts (amplification)** : l'adaptateur placé en 3' de notre fragment va être capable de **s'hybrider à l'oligonucléotide complémentaire** à cette séquence, localisé juste à côté. C'est pour cette raison qu'on

parle de **PCR en pont**, puisqu'après cette étape de dénaturation et de lavage, notre molécule d'ADN nouvellement synthétisé va former **des ponts avec ses adaptateurs en 5' et en 3'**.

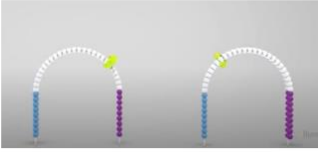
À la suite de la formation de ce pont, la **Taq polymérase** va pouvoir synthétiser le **brin nouveau complémentaire de 5' en 3'**. Puis, on a une étape de **dénaturation** et de **lavage** qui va permettre d'obtenir **deux fragments d'ADN simple brin**, fixé sur notre lame de verre. Nous obtenons donc **2 nouvelles molécules d'ADN nouvellement synthétisées (un brin sens et un brin reverse)**.




2- Enrichissement par PCR clonale – PCR « Bridge » en clusters

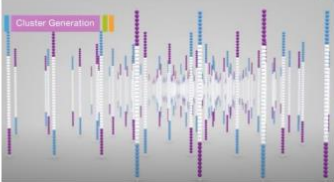


Cycles successifs d'amplification PCR



Formation de millions de clusters





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 brin** : c'est la formation d'un **cluster** (= *formation de groupes d'amplification de produits PCR*).

Le point de départ de chacun des clusters est **un brin d'ADN hybridé initialement en un point localisé**. C'est le même principe que les colonies bactériennes.

On obtient, par ses **multiples étapes de dénaturation**, plein de molécules **d'ADN simple brin**. 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 ADN original** et initialement hybridé sur la lame. Ce sont ces brins sens que l'on va séquencer.

4) Séquençage

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 (*violet sur la diapo*). On ajoute une **polymérase** pour la synthèse du brin de **5' en 3'**, et des **nucléotides fluorescents**. Le système est fait de telle sorte que la polymérase va introduire les nucléotides en complémentarité des bases les **uns après les autres**.

Cette étape est suivie d'une étape **d'excitation et de mesure de la fluorescence** émise par le **dNTPs ajouté**. Ici, on voit que c'est un A qui émet une fluorescence, en complémentarité d'un T. C'est ensuite un G...

illumina®

3- Séquençage

Ajout d'un primer +
ADN Polymérase +
dNTP fluo

Incorporation des
nucléotides
complémentaires

Excitation et mesure de la fluorescence émise

L'automate va donc mesurer cette **émission de fluorescence** suivant l'ajout des **différents nucléotides**, sur chacun des fragments et dans les différents clusters.

Sur ces Flow Cell, nous avons des centaines de milliers de clusters / de fragments d'ADN **séquencés en même temps**. On peut donc parler de **Séquençage Massif**, puisque chaque cluster peut représenter un tube de séquençage. Ils sont séquencés et lus en parallèles.

illumina®

Image représentative d'une petite fraction de la Flow Cell

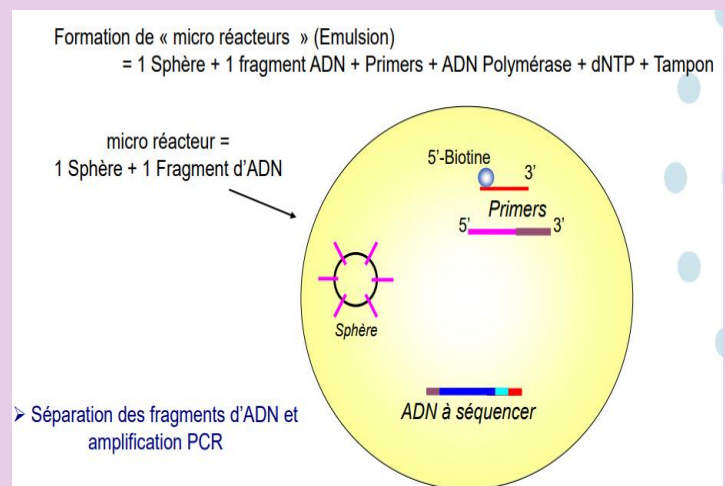
Flow Cell = Centaines de milliers de Clusters

3. ThermoFisher

Concernant les techniques d'amplification clonale et de séquençage sur les plateformes ThermoFisher, on remarque que cette fois-ci, **notre ADN est bien capturé** mais il n'est pas fixé sur une lame de verre, mais sur une **sphère métallique**. ++ Contrairement à Illumina, l'amplification clonale ne va pas se faire par cluster mais dans un **micro-réacteur** créé par un système **d'émulsion**.

Lorsqu'on mélange de l'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 à ce que chaque microréacteur soit l'équivalent d'un tube contenant :

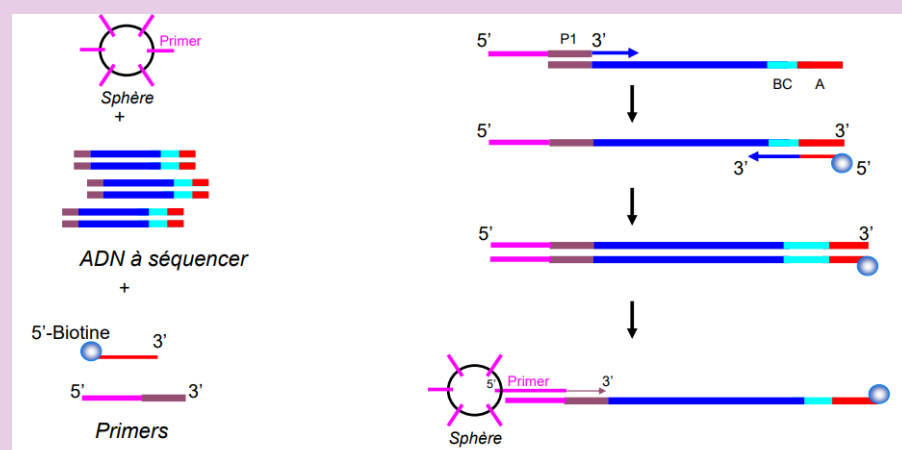
- 🧪 Une **sphère**,
- 🧪 Un **fragment d'ADN** à séquençer,
- 🧪 Deux **primers**,
- 🧪 L'ADN **polymérase**,
- 🧪 Les **dNTPs**,
- 🧪 Un **tampon**.



On réalise ensuite la PCR clonale : ce sont les mêmes étapes que la PCR classique (dénaturation / hybridation / élongation).

1. Hybridation du primer

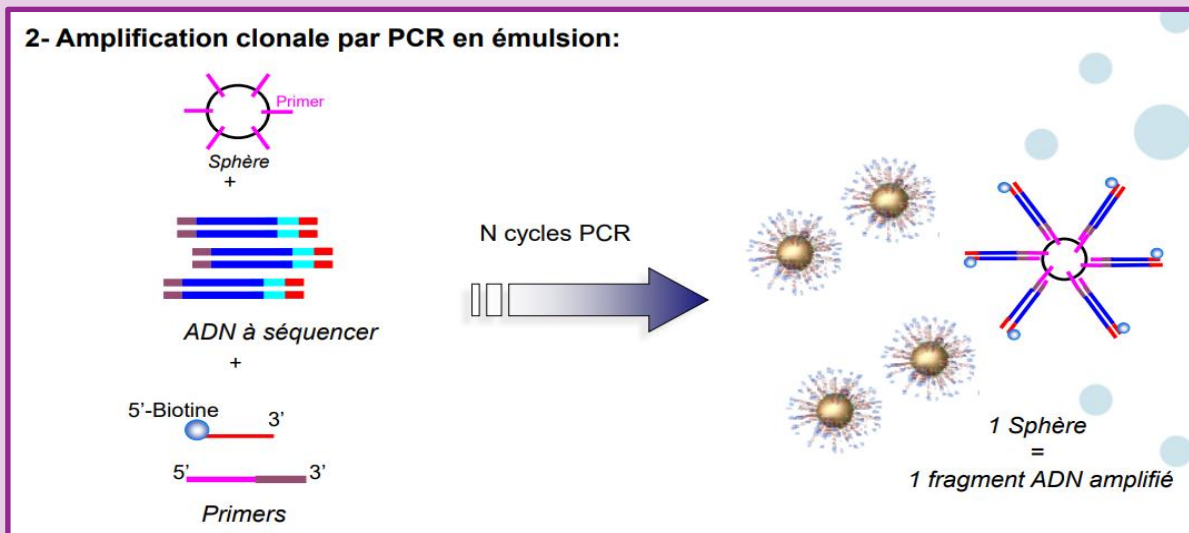
Il s'hybride grâce à la séquence P1. Lors du **deuxième cycle**, c'est le deuxième primer qui **s'hybride en complémentarité de l'adaptateur**. Cette fois-ci, le primer est **biotynilé en 5'**.



2.Elongation

Le brin complémentaire est créé. Lors **du second cycle**, le fait d'avoir synthétisé ce brin complémentaire va permettre à ce nouveau brin **de pouvoir se fixer toujours par complémentarité** des bases sur le **primer localiser sur une petite sphère magnétique**.

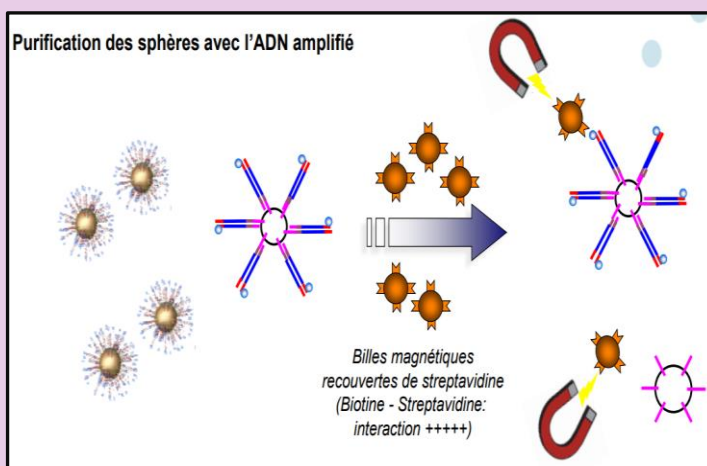
Donc à chaque étape, nous **avons amplification de chacun des fragments et fixation de ces fragments sur une sphère magnétique**. Au final, on se retrouve avec des sphères **recouvertes de cet ADN** que l'on veut séquencer, et que l'on a déjà **amplifier x fois et fixer sur nos sphères**.



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**.

On parle **de PCR clonale** parce que c'est réellement un clone d'une molécule d'ADN qui aura été amplifiée dans chaque microréacteur avec, sur chaque sphère différente, des **fragments d'ADN différents**.

3.Purification des sphères avec l'ADN amplifié



On va ensuite **purifier ces sphères** recouvertes d'ADN. 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 **recupérer et purifier** uniquement les sphères recouvertes d'ADN.

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 : ces sphères ne nous intéressent pas.

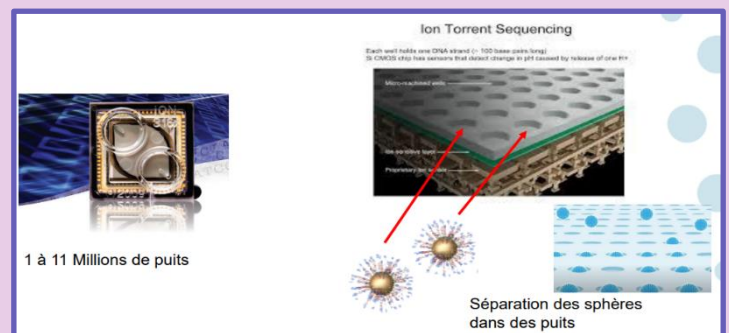
Grâce à cette interaction, nous allons donc pouvoir **purifier uniquement les sphères qui nous intéressent**.

4. Séquençage individuel des sphères

Nous allons **ensuite séparer physiquement nos sphères**, et les placer **individuellement** dans des **puits** qui leur seront propres.

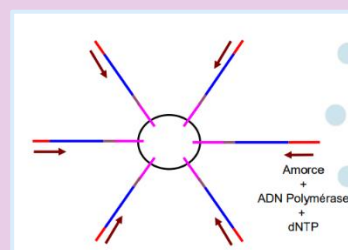
On a des toutes petites puces contenant 1 à 11 millions de puits, dans lesquelles nous allons pouvoir séquencer 1 à 11 millions de sphères. **Une seule sphère va pouvoir être placée dans un seul puit.**

La réaction de séquence va se faire directement **dans ces puces**.



On va ajouter dans notre milieu réactionnel ce qu'il faut pour réaliser la séquence :

- Une **amorce**,
- Une **ADN polymérase**,
- Des **dNTPs**.

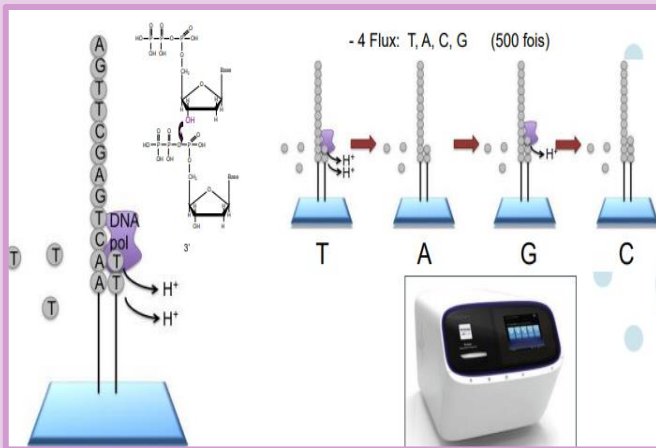


Ainsi, la polymérase pourra synthétiser le brin complémentaire à partir de **l'amorce hybridée sur l'un des deux adaptateurs**.

Lorsque la polymérase synthétise le brin complémentaire et qu'elle ajoute un nouveau nucléotide, il y a formation **d'une liaison phosphodiester** entre les nucléotides. Cette formation de liaison phosphodiester libère en **ions H+** dans le milieu réactionnel, entraînant des **variations de pH que les automates vont pouvoir mesurer**.

Les nucléotides étant **ajoutés séquentiellement** en fonction de ce qu'envoie l'automate, on va pouvoir retrouver la séquence de notre gène. Entre chaque nucléotide envoyé, un lavage est réalisé.

Petite explication : si l'automate envoie un T et que le pH ne varie pas, alors



l'automate ne détectera rien. En revanche, s'il envoie un A et que le pH diminue, c'est qu'il y a eu libération d'un H⁺. Ceci est dû à la création d'une liaison phosphodiester, qui signifie que le nucléotide suivant était un A. L'automate retranscrit le nucléotide envoyé lors de la variation de pH.

Cette succession de nucléotides se fait **environ 500 fois dans un run de séquençage.**

Petit Récap du love :

Les deux principales plateformes de NGS sont **illumina** et **ThermoFisher**, dont le point de départ est identique :

- Préparation de **librairie**,
- Préparation de nos **échantillons** en petits fragments, auxquels sont ajoutés **ces adaptateurs en 5' et en 3' +++**

Nos échantillons peuvent également provenir de **PCR multiplexes** : ce sont de **très grand nombre de PCR** dans la taille est déjà de 200 ou 400 pb.

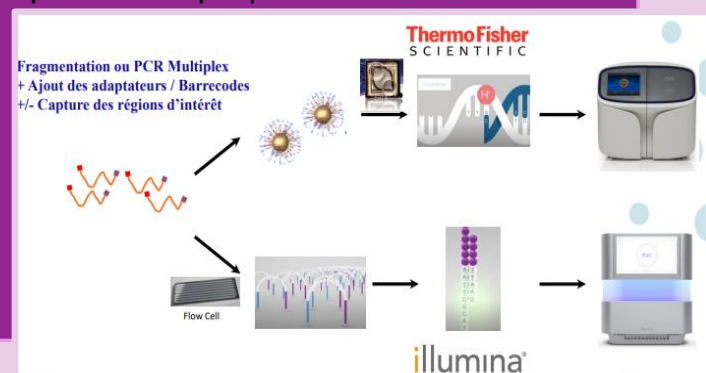
Dans ce cas, on peut mélanger les PCR directement, sans étape de fragmentation. La fragmentation n'est nécessaire **qu'à partir du moment où on utilise de l'ADN génomique ou de PCR grands fragments.**

Chacune des plates-formes a ses spécificités, une amplification et donc un enrichissement des fragments qui nous intéressent :

- Soit en PCR en **émulsion**, avec des **microréacteurs** : **ThermoFisher**,
- Soit fixation de nos fragments d'ADN sur **lame de verre**, et d'effectuer cette PCR en **ponts** pour amplifier nos ADN par **cluster** : **illumina**

Les techniques de séquençage restent spécifiques de chaque plate-forme :

- Soit par **variation de pH** : **ThermoFisher**,
- Soit par **fluorescence** : **illumina**.



IV) 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**.


Même si les outils peuvent être **différents** en fonction des plateformes, l'analyse informatique est **primordiale** pour **déterminer les variants**, permettre les **annotations** et les **interprétations**.


Des **échanges** entre **cliniciens, biologistes et techniciens** sont nécessaires pour ces interprétations (ce n'est pas tout d'avoir la mutation).

Les interprétations doivent être faites en fonction de **la fonction du gène, de la protéine, des répercussions des variants...** On cherche si le caractère pathogène d'une protéine peut **expliquer la clinique d'un patient**.

1) Les étapes

 **Détermination des bases (base calling)** : Transformation des **signaux générés** (pH, fluorescence) en **données de séquence**.

 **Contrôle de la qualité** : Nettoyage / Élimination des **lectures de séquences de mauvaise qualité** (données générées : des millions à il y a forcément des puits dans lesquels la mesure ne s'est pas faite correctement). On ne garde que les lectures de bonne qualité.

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



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 ?

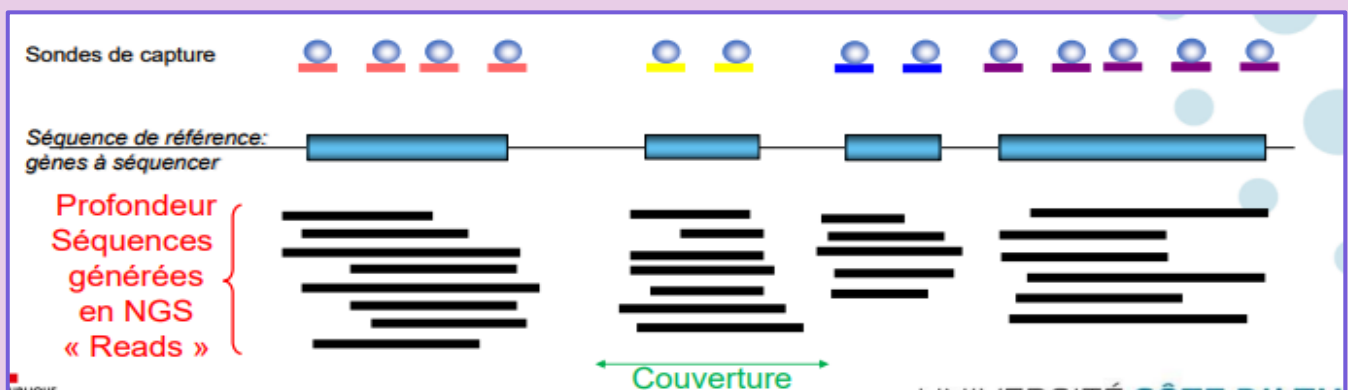
Deux notions très importantes :

- **Profondeur de lecture** : nombre de lecture **indépendante** d'une base. C'est le **nombre de fois où chacune des bases d'intérêt a été lue** sur des reads différents.

- **Couverture** :

(% de bases séquencées par lecture indépendantes) / (Nombre total de base de la région d'intérêt).

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



On vérifie donc que chacun de cet **exon a été lu correctement**. Par exemple, dans un exome, on s'attend à un séquençage de bonne qualité, où chacune des bases de l'ensemble des exons ont été **lues au moins 20 fois** (au moins 20 séquences indépendantes).

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 chromosome ?

Sur quel gène ?

À quel endroit du gène ?

Variants exoniques ?

Introniques ?


Classification bénin / malin ?

Conséquences sur la protéine après traduction ?

Caractères pathogènes ?

Fréquence de ce variant ?

Dans quelles populations ? (Caucasiennes, asiatiques...)
 Ça aide à **l'interprétation du variant, et à créer des bases de données.**

 **Interprétation des variants = expertise clinico-biologique :**

Etape la plus IMPORTANTE

Nous avons une confrontation **biologistes / cliniciens**, puisque pour bien interpréter un variant, **il faut l'interpréter dans son contexte biologique** qui lui est propre. Il va falloir prendre 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 de donner la clinique que le patient présente ?**

C'est l'étape aujourd'hui, **la plus difficile, mais aussi la plus importante** : séquencer est devenu facile grâce au NGS, mais l'écueil se trouve au niveau de l'interprétation. Il faut trouver **LE variant responsable de la pathologie parmi de nombreux variants.**



2) Exemple d'application du NGS

- Séquençage de **plusieurs gènes** (on parle de panel de gènes puisqu'on s'intéresse à plusieurs centaines de gènes),
- Séquençage de **l'exome en entier** (ce sont toutes les régions **codantes** d'un génome) (WES, ou Whole Exome),
- « **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**).
 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** puisque ce sont les ARN qui seront traduits en **protéines**.

Le NGS est également l'une des briques de ce que l'on appelle le **multi-OMICS** (tout ce qui est génomique, donc exome, génome...). Il va nous permettre de très nombreuses approches qui permettent de s'intéresser aux protéines (**protéomique**), au métabolisme (**métabolomique**).

Ce sont vraiment des 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**. Ceci nous permettrait d'obtenir un panel complet du métabolisme modifié. Les voies métaboliques vont, à l'avenir, **être de plus en plus utilisées**.

La NGS pourra également nous aider à **développer des thérapeutiques ciblées**, notion dont découle la « Médecine Personnalisée » : à chaque patient son traitement, développé à partir de ses voies métaboliques, de ses marqueurs, qui caractérisent un individu.

« Le NGS est une petite brique de ce gros domaine qu'est le multi-OMICS »

V) Conclusion

En conclusion, le séquençage haut débit est très utile **par sa puissance pour trouver des mutations dans plusieurs gènes** : cela concerne donc les maladies multigéniques, les séquençages d'exomes (régions codantes d'un individu) ...

Cette technique est donc plus utilisée dans les pathologies où le gène en cause **n'est pas connu**.

L'identification de ces séquences, que ce soit par séquençages Sanger ou par NGS, nécessite ensuite de **comprendre quel est l'impact de ce variant** sur la fonction de la protéine. Pour cela, nous verrons plus tard quelques techniques qui permettent de **déterminer la pathogénicité d'un variant**.

Il faut toujours garder en tête qu'on identifie des **variants au niveau de l'ADN**, mais que le **phénotype / les symptômes présentés par le patient viennent de la modification de la fonction de la protéine**. Il faut toujours comprendre ce que les / les variant(s) identifiés ont comme effets sur **les protéines et les voies métaboliques**.
Et voilà !! C'est la fin de ce cours <3 Dédi à vous !

