

# *Biologie Moléculaire*

Intitulé du cours : Séquençage Haut Débit NGS

*Rédacteur : AIT BAHA Hajer*

*Ronéo n° : 3 (dernière ronéo de l'année youhou)*

**Corporation des Carabins Niçois**

UFR Médecine

28, av. de Valombrese

06107 Nice Cedex 2

<http://carabinsnicois.fr/>

[roneo.c2n@gmail.com](mailto:roneo.c2n@gmail.com)

*Partenaires*



Hello ! Le voici le voilà, le DERNIER cours de cette année est enfin arrivé #petitelarme. Le cours n'a pas changé par rapport à l'année dernière, du coup, j'ai repris ma ronéo de l'an dernier et ajouté, s'il y avait besoin, les précisions de la prof. Voilou, bonne lecture, bon courage à tous, pensez aux prochaines vacances qui vous attendent pour vous motiver, c'est la dernière ligne droite !! La bise 😊

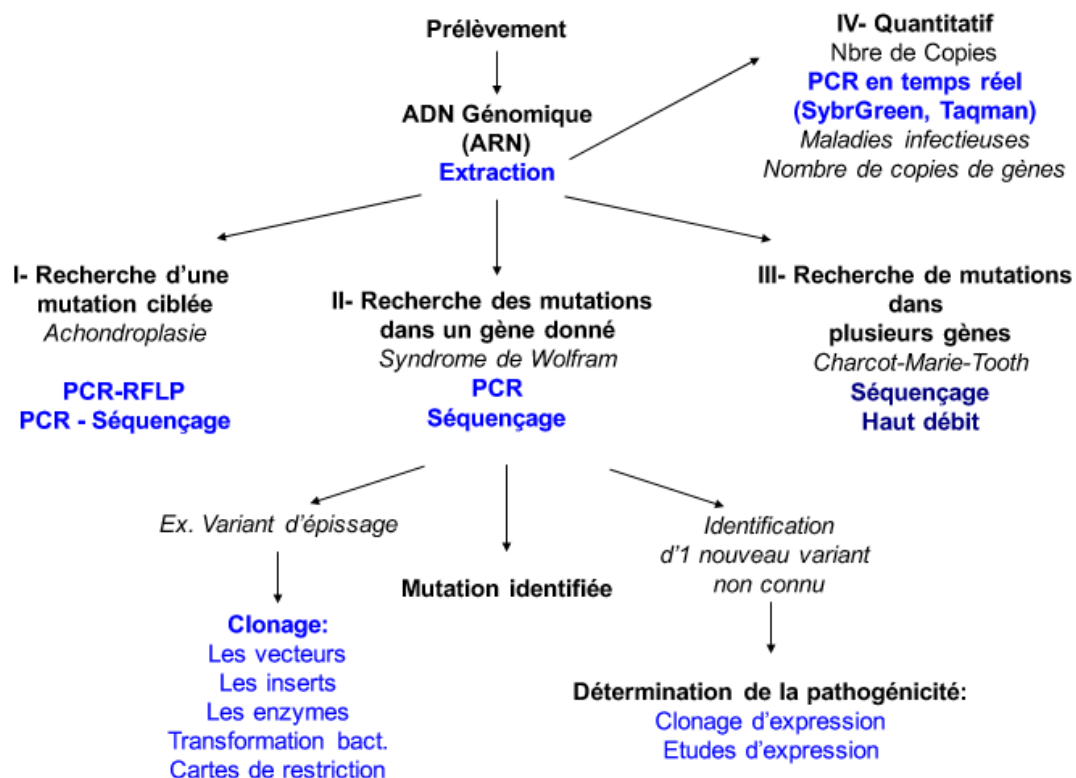
# Introduction

## A) Rappel des épisodes précédents

Après prélèvement biologique (sang ou tissus) et extraction du matériel génétique (ADN Génomique ++ ou ARN), on s'oriente selon ce que l'on souhaite chercher :

- Si on cherche une mutation ciblée (Achondroplasie), on utilise la **PCR-RFLP (PCR suivi d'une digestion enzymatique)** qui est toujours confirmée par une technique de séquençage ++.
- Si on cherche des mutations dans un gène donné au complet (Syndrome de Wolfram), on screene le gène dans son ensemble grâce à une PCR-Séquençage. Là en fonction des résultats :
  - Soit on a identifié notre mutation et on est tranquille, on ne va pas plus loin.
  - Soit lorsque les variants ne sont pas situés sur les exons par exemple, mais au niveau des introns avec des variants d'épissage, on effectue toute une série de manip (à savoir le clonage) pour identifier exactement le variant d'épissage (voir ronéo 2)
- ⇒ Là, si on identifie un variant non connu, on essaie de déterminer sa pathogénicité grâce aux clonages d'expression (clonage d'expression d'une protéine par exemple que l'on fusionne à un tag pour la localiser).
- Si on cherche des mutations dans plusieurs gènes (maladie de Charcot), on va utiliser le séquençage haut débit.
- Si on cherche à quantifier le nombre de copie d'un gène dans le cadre de maladies infectieuses par exemple, on utilise la PCR en temps réel (Sybr-Green, Taqman)

C'est important d'avoir ces différents cas de figure en tête pour les qcms ++ La prof veut que vous connaissiez les principes des techniques + dans quel cas on les utilise + « dans quel cas on utilise telle ou telle enzyme » ++



## B) Evolution des techniques

### 1977 : Méthode de référence : Méthode de SANGER = Méthode enzymatique des didésoxyribonucléotides (ddNTP)

Séquençage manuel utilisant des ddNTP radiomarqués A T C G (qui bloquent la polymérase) dans quatre réactions différentes. En fonction de la migration et de l'identité des pistes on en déduisait les séquences, on pouvait le faire pour **200 pdb maxi**.

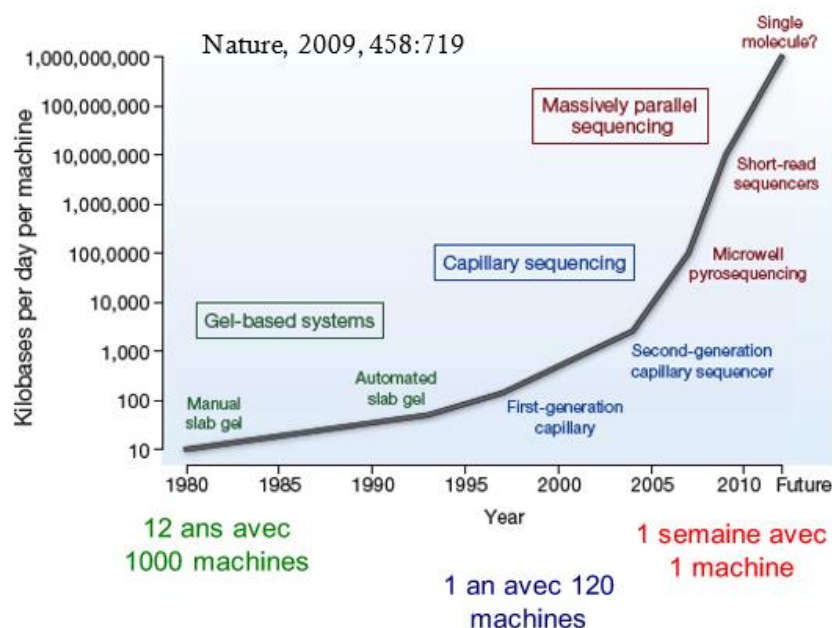
La résolution est pas top !

### 1993 : Séquenceurs automatiques= « séquençage capillaire »

On a une migration électrophorétique dans de petits capillaires remplis de gel d'agarose. On a le couplage de chaque ddNTP à un fluorochrome de couleur différente permettant l'ajout des 4 ddNTP dans un même tube réactionnel. On peut séquencer jusqu'à 16 réactions en simultanément en 1h. On peut le faire pour **800 pdb maxi**.

### 2007 : Séquençage haut débit (NGS)

Petits appareils de paillasse super puissants. MAIS un NGS doit toujours être confirmé par un SANGER car nous n'avons pas assez de recul à l'heure actuelle quant aux résultats obtenus par le séquençage au débit.



Le génome humain est composé de **3 milliards de paires de bases**. On note une évolution assez conséquente des capacités de séquençage du génome humain. Pour le séquencer il aurait fallu :

- 12 ans avec 1000 machines dans les années 80

- 1 an avec 120 machines en 1995

- 1 semaine avec 1 machine aujourd'hui.

En 2001, on a pu faire le séquençage de 95% du génome humain. Cela représente le séquençage de 30 000 gènes.

C'est en 1989 qu'on a commencé à séquencer le génome humain pour finalement l'achever en 2003.

## Le Séquençage haut débit (NGS)

### A) Définitions

Le NGS, pour Next-Generation DNA Sequencing, est le séquençage massif en parallèle de molécules d'ADN individuellement séparées et amplifiées sous forme de clones ou de molécules uniques.

Trois sociétés (avec trois méthodes différentes) ont pris en charge le développement de cette technique :

- **Illumina** (utilise des nucléotides fluorescents)
- **Roche** (utilise des nucléotides fluorescents, **mais plus trop sur le marché**)
- **Life technologie** (utilise les variations de pH) ++ Celle qu'on va développer

On note trois étapes communes à ces trois plateformes :

- **Extraction** de l'ADN génomique (on peut aussi partir directement d'un produit PCR) dans un premier temps
- **Fragmentation** de l'ADN dans un second temps

- Enfin, ajout **des adaptateurs + codes-barres** sur les fragments d'ADN.

L'étape d'après, qui est **l'amplification PCR-clonale** (on parle également de **PCR multiplex**) des fragments d'ADN, est cependant différente selon les sociétés :

Illumina	Roche	Life technologie (Ion Torrent++)
Capture les fragments d'ADN et amplification sur support solide (lame de verre)	Fixe les fragments sur <b>billes d'agarose</b> et les sépare par PCR en émulsion (eau/huile)	Fixe les fragments sur des <b>sphères</b> et les sépare par PCR en émulsion (Eau/huile)

Ce qu'il faut comprendre c'est qu'une fois les molécules d'ADN fragmentés, barre-codés et fixés à des adaptateurs, il faut les individualiser et à partir de là on amplifie chaque molécule pour pouvoir ensuite lancer notre séquençage.

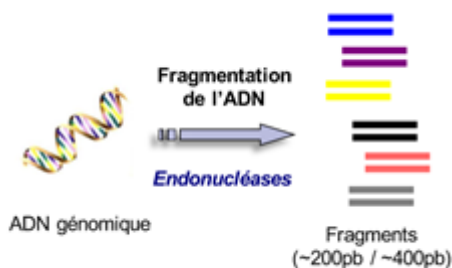
**NB : Dans la PCR multiplex on utilise un set de plusieurs couples d'amorces pour amplifier simultanément plusieurs cibles d'ADN en une seule réaction de PCR, plutôt qu'un seul couple d'amorces pour amplifier une cible d'ADN unique.**

## B) Préparation des échantillons

On va prendre l'exemple de la méthode par capture. On cible des régions exoniques des gènes dans le cadre d'une maladie multigénique. On pourrait faire un « whole génome », c'est-à-dire séquencer tout le génome mais ce n'est pas fait en routine, on ne capture en général qu'une région génomique d'intérêt.

Pour se faire on suit différentes étapes :

### a) Fragmentation de l'ADN.

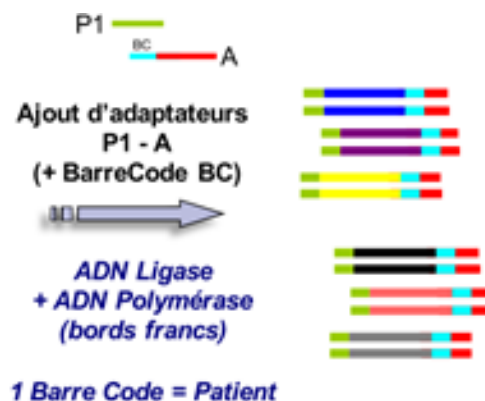


On extrait notre ADN (technique vue dans la ronéo1) puis on coupe notre molécule d'ADN en petits morceaux de **200 à 400 pb** grâce à des **endonucléases** qui coupent l'ADN double brin sans avoir de spécificité ++.

Attention, ce ne sont pas des enzymes de restrictions, on rappelle que celles-ci coupent l'ADN double brin en fonction d'une séquence spécifique++. Ici, les enzymes génèrent des fragments d'ADN aléatoirement.

### b) Ajout des adaptateurs + barre-codes

Après fragmentation, on ajoute **des adaptateurs** (P1 en 5' et A en 3' sur le schéma) pour que **les extrémités 3' et 5' soient les mêmes pour tous les fragments** ainsi que des **BarreCodes** spécifiques de chaque patient (1 Barre Code = 1 Patient)



- ✓ Grâce aux adaptateurs, on n'a plus besoin que d'un seul couple de primers pour l'amplification par PCR car tous les fragments auront les mêmes extrémités 5' ou 3' (adaptateurs P1 et A).
- ✓ Les codes-barres sont des petits primers accrochés à une extrémité du fragment spécifique d'un patient.

On peut alors séquencer plusieurs patients en même temps dans une même réaction. En fait, lors de la préparation des échantillons, nous avons un patient pour un tube, mais au moment du séquençage, on va pouvoir mélanger les échantillons grâce aux BarreCodes qui permettront, après analyse informatique, d'attribuer chaque séquence à un patient donné.

c) Ajout des ligases pour lier nos BC et nos adaptateurs + d'une ADN polymérase (pour être sûr d'avoir des bords francs car on n'est pas sûr que les extrémités coupées par nos endonucléases soient à bord francs).

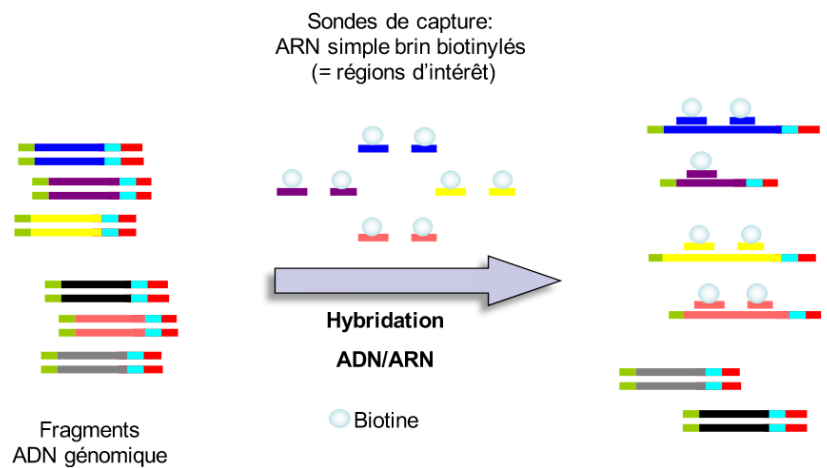
d) Sondes de capture

Une fois nos fragments d'ADN obtenus, on souhaite **isoler les régions exoniques d'intérêts** (car on ne veut pas tout séquencer #inutile).

Pour se faire, on ajoute **des sondes de capture composées d'ARN simple brin biotinylés++**.

(On a ajouté de la biotine à nos fragments d'ARN)

Chacun de ces ARNs est complémentaire à nos régions d'intérêt et va donc s'y fixer.

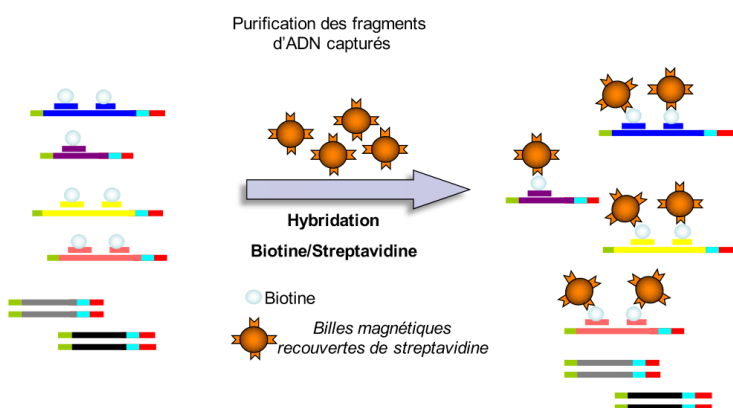


On dénature l'ADN (car l'ADN est double brin) puis il y a reconnaissance entre l'ARN biotinylé et notre région d'intérêt.

Si on fait de l'exome, on cible tous nos exons.

**Les régions d'intérêt à séquencer sont reconnues par les sondes de capture (Hybridation ADN/ARN) contrairement aux régions qui ne nous intéressent pas.**

e) Purification des fragments d'ADN capturés



Une fois l'hybridation faite, on veut capturer seulement les régions d'intérêts.

A notre mélange, on ajoute des **billes/nanobilles magnétiques** recouvertes de **streptavidine**. Ce qu'il faut comprendre c'est que la biotine et la streptavidine ont **une très grande affinité++**.

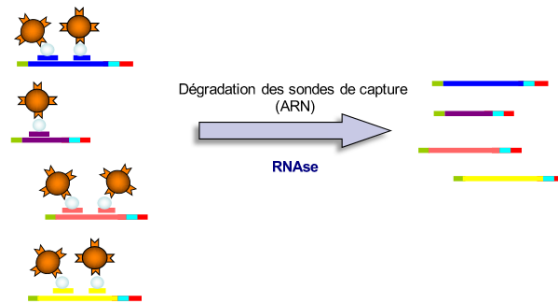
Ces billes vont donc accrocher l'ADN où l'ARN biotinylé s'est fixé.

Pour les capturer, on place les tubes dans un aimant. L'aimant contre la paroi permet d'attirer les billes magnétiques et de récupérer nos populations d'intérêt. Les autres fragments sans billes sont éliminés après plusieurs lavages et éluitions.

Aujourd'hui dès qu'on veut purifier quelque chose on utilise très souvent cette méthode avec les billes magnétiques, par exemple pour la purification/extraction de l'ADN, au lieu d'utiliser l'extraction au phénol chloroforme on peut utiliser ces billes ☺

### f) Dégradation des sondes de capture

Après ça, on dégrade les sondes de capture ARN grâce à une RNase.



## C) Amplification clonale par PCR en émulsion de nos fragments d'ADN

### a) Définitions

Les séquences d'intérêt ont été isolées ; il faut maintenant les séparer et les amplifier par PCR avant de les séquencer.

Notre ADN à séquencer est capturé par une sphère grâce à une émulsion.

En effet, on parle d'émulsion car c'est comme si on mélangeait de l'eau et de l'huile pour former des gouttelettes d'huile (= **microréacteur**). On se place de telle sorte qu'il y aura **dans chaque gouttelette, un seul fragment d'ADN++**. On aura donc un fragment d'ADN donné pour une sphère.

Dans chaque microréacteur on a donc :

- **1 sphère sur laquelle sont fixés des primers**
- **2 primers dont un est biotinylé ++** complémentaires des adaptateurs

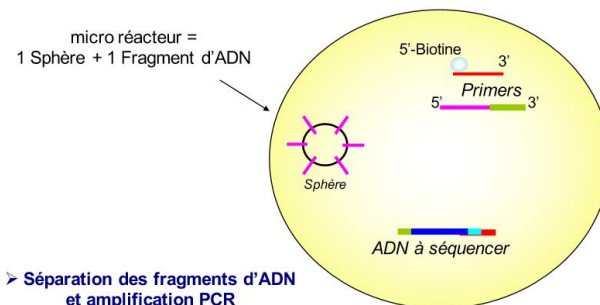
Rappel : Les adaptateurs sont les mêmes pour TOUS les fragments, donc quelque soit le fragment on utilise les mêmes primers pour tout le monde. *Ces petits adaptateurs sont l'équivalent de la borne d'amont et d'aval de la PCR classique.*

- **ADN polymérase**
- **Des dNTP**
- **Un tampon.**

#### 2- Amplification clonale par PCR en émulsion:

Formation de « micro réacteurs » (Emulsion)

= 1 Sphère + 1 fragment ADN + Primers + ADN Polymérase + dNTP + Tampon



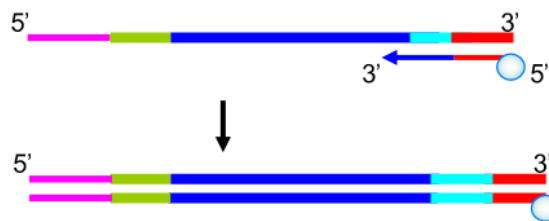
On retrouve les mêmes étapes que la PCR classique, à savoir : *Dénaturation, Hybridation, Elongation avec les différentes étapes de chaleur, pendant n cycle.*

**b) Amplification de l'ADN**

- 1) Dénaturation (par la chaleur) de l'ADN double brin
- 2) Hybridation de l'amorce P1 sur l'extrémité 5' du brin et synthèse (de 5' en 3') par l'ADN polymérase du premier brin complémentaire



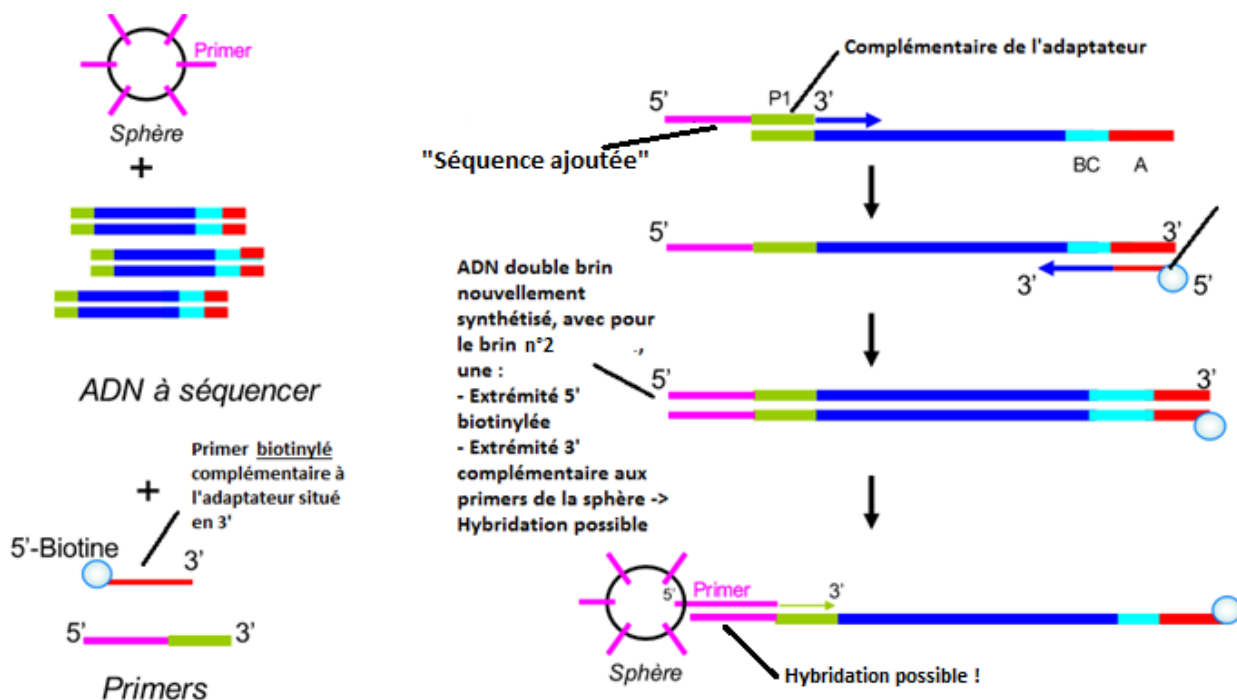
- 3) Dénaturation (rend l'ADN nouvellement synthétisé simple brin)
- 4) Hybridation de l'amorce A biotinylé sur l'extrémité 3' du brin nouvellement synthétisé et élongation par l'ADN polymérase. Au deuxième cycle, on a synthèse du brin complémentaire du brin nouvellement synthétisé.



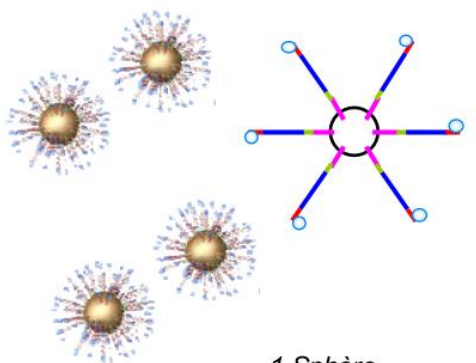
Contrairement au fragment d'ADN de base où on avait que la séquence de l'adaptateur (en vert sur la diapo), les brins fils sont composés d'une séquence supplémentaire (partie violette sur la diapo) complémentaire au primer qu'on a utilisé. Cette extrémité nouvellement synthétisée sur nos brins d'ADN est complémentaire aux primers fixés sur notre sphère.

⇒ Donc, après dénaturation de l'ADN nouvellement synthétisé, on peut l'hybrider au primer de la sphère par complémentarité. Après ça, on a synthèse du brin complémentaire par la polymérase de 5' en 3'.

**L'hybridation avec le primer situé sur la sphère ne peut se faire qu'avec le dernier brin nouvellement synthétisé. ++**



Lisez bien cette diapo, elle récapitule tout, j'ai ajouté des annotations qui devraient vous aider à comprendre, c'est pas franchement évident au début, mais lisez tranquillement et ça devrait le faire ! 😊



1 Sphère  
=  
1 fragment ADN amplifié

Après n cycle d'amplification, on se retrouve avec une sphère qui correspond à un fragment d'ADN qui a été amplifié n fois = amplification clonale. Il faut noter que **l'extrémité 5' de nos fragments fixés est biotinylée++**.

**1 Sphère = 1 fragment ADN amplifié**

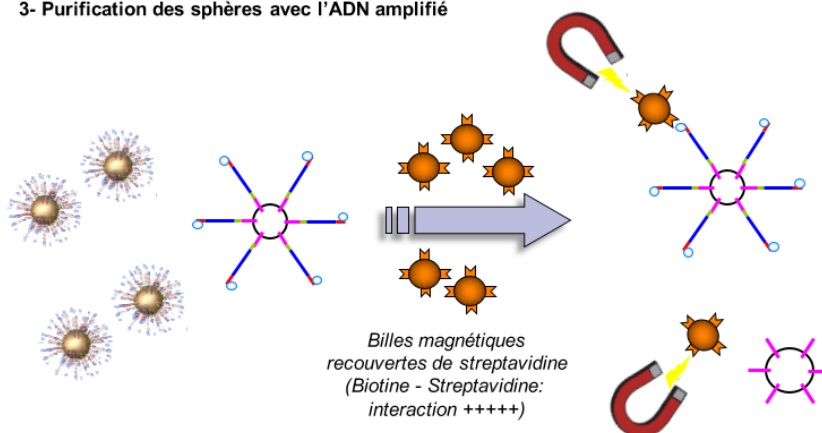
Après bien sûr dans certains de nos microréacteurs on aura certainement des ADN sans sphères ou des sphères avec plusieurs fragments d'ADN mais le traitement informatique détectera ces erreurs pour pas que ça fausse tous nos résultats.

### D) Purification des sphères avec l'ADN amplifié

Après amplification de notre fragment d'ADN, il faut purifier les sphères. On veut récupérer les sphères qui ont l'ADN amplifié et se débarrasser des sphères qui n'ont fixé aucun fragment d'ADN (= sphère sans extrémités biotinylées).

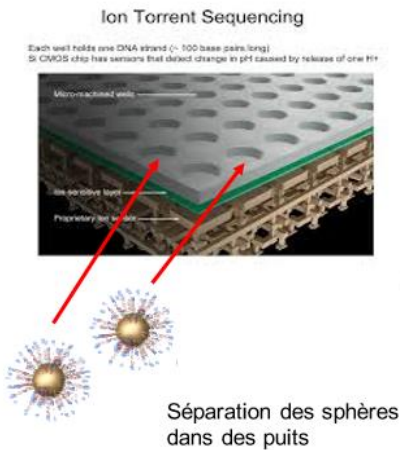
Pour se faire, on réutilise **l'interaction streptavidine/biotine**. On fixe nos sphères à **des billes magnétiques** recouvertes de streptavidine puis on les place dans un aimant. L'aimant contre la paroi permet d'attirer les billes magnétiques et de récupérer nos populations d'intérêt. Enfin, on se débarrasse des billes magnétiques non fixées, des sphères sans ADN amplifié et de l'ADN libre (non fixé).

#### 3- Purification des sphères avec l'ADN amplifié



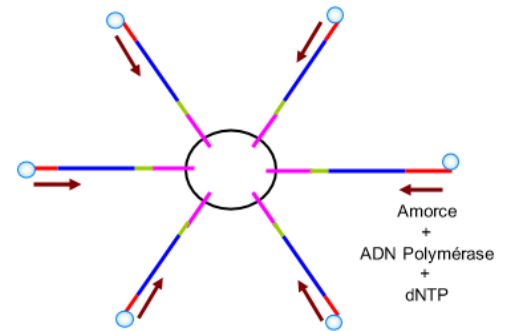
**E) Séquençage individuel des sphères**

**a) Séquençage**

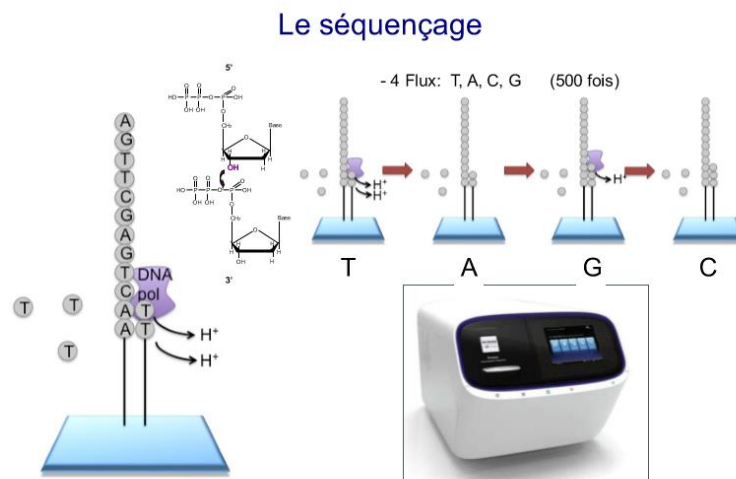


Pour le séquençage, on utilise **une puce** avec un support métallique composé d'un **grand nombre de puits (1 à 11 millions de 2 à 3 cm de côté** -> séquençage massif) dans lesquels on va placer les fragments d'ADN à séquencer. **Ces puces sont calibrées de telles sortes que chaque puit ne puisse accueillir qu'une seule sphère.**

- 1) Dénaturation de l'ADN (pour devenir simple brin) puis utilisation **d'une amorce** (qui est la même pour toutes les sphères, pour nos 11 millions de réactions on utilise qu'une seule amorce #adaptateurs++), **d'une ADN polymérase** qui synthétise et **de dNTP**.



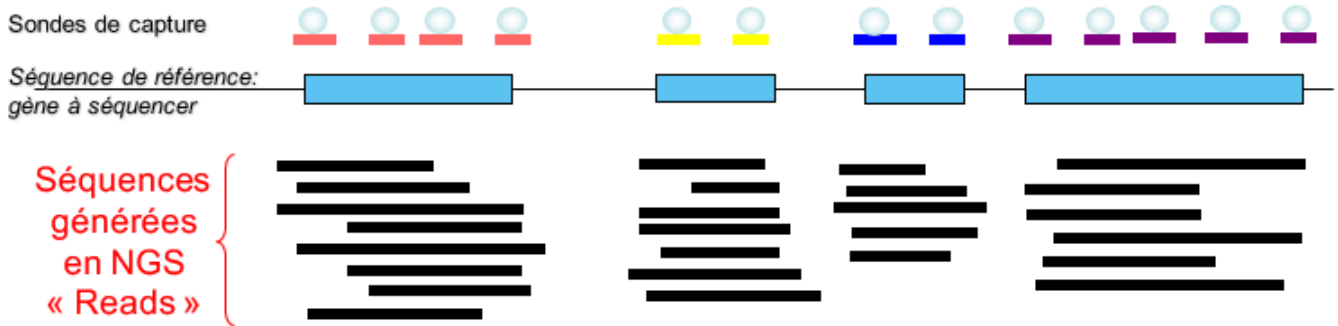
- 2) **Principe : Injection des dNTP** classiques les uns après les autres par la machine. **Pour Illumina, ils utilisent la fluorescence, à chaque incorporation d'un nucléotide il y a émission de fluorescence, un peu comme la méthode Sanger.** Ici on mesure le relargage des ions H<sup>+</sup> et donc la modification du pH.



- ✓ S'il y a complémentarité : Synthèse par la polymérase -> Libération d'un ion H<sup>+</sup> suite à la formation de la liaison entre les 2 bases -> Création d'une variation du pH détectée par la machine.
- ✓ S'il n'y a pas complémentarité -> Pas de variation du pH.

L'appareil mesure le pH en continu et envoie les dNTP un par un. De ce fait, en fonction de la variation du pH et du dNTP envoyé, on connaît au final l'ordre des nucléotides, c'est-à-dire la séquence de notre ADN génomique. Quand la polymérase incorpore un G par exemple, on a libération d'un proton -> modification du pH et donc on sait qu'il y a dans à cette position un C.

**b) Analyse informatique**



Toute la partie bioinformatique est très importante.

**Couverture** : Pourcentage des régions séquencées. Avoir une couverture de 100% signifie qu'on a réussi à séquencer 100% de notre région d'intérêt.

**Profondeur** : Nombre de séquence générée, c'est-à-dire le nombre de fois qu'une même région a été lue.

**Read** : Séquence générée.

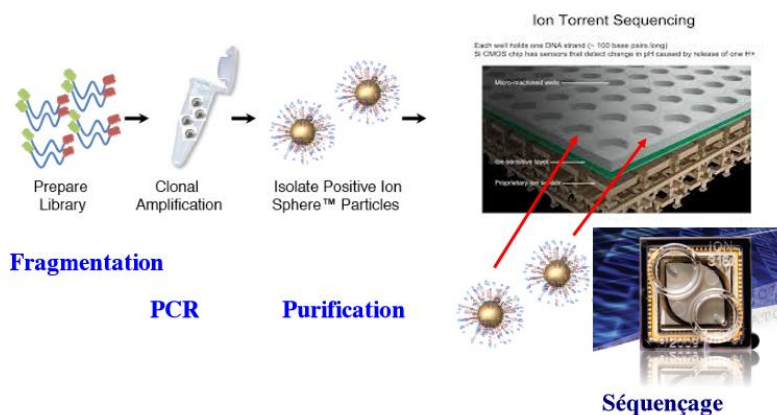
L'analyse informatique est très complexe : on a un tri et un réalignement sur la séquence de référence (ADN génomique « normal ») de toutes les séquences lues par la machine. C'est une technique extrêmement puissante ; une même position nucléotidique est lue des milliers de fois. Cela permet de **détecter les changements nucléotidiques (variants pathogènes potentiels)**.

La difficulté est d'identifier le variant et si variant nouveau il y a, il faut déterminer s'il est pathogène ou pas. Le problème aujourd'hui n'est pas la technique de séquençage en elle-même, mais l'interprétation qui suit, on ne connaît pas les conséquences de tous les variants. Est-ce que tel ou tel variant est pathogène ou pas ? Telle est la question !

Exemple : On séquence l'ADN d'un patient dans le cadre d'une pathologie (on fait un exome) et on découvre par hasard qu'il possède une mutation qui prédispose à tel ou tel cancer. Est-ce qu'il faut lui en faire part ou pas ? Ça soulève plein de questions éthiques comme ça.

En résumé, la prof veut vraiment que vous ayez ces étapes en tête !

La Technologie Ion Torrent



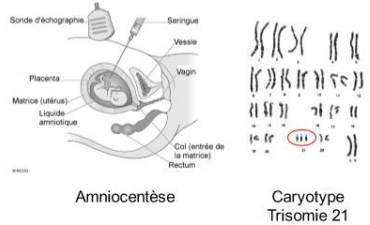

Remarque : Il est nécessaire de vérifier les résultats obtenus par la méthode SANGER ++

## Exemples d'applications du NGS

On utilise la technologie NGS pour le :

- Séquençage de plusieurs gènes (panel de gènes, plusieurs centaines de gènes)
- Séquençage d'exome (régions codantes d'un génome)
- Whole genome sequencing (séquençage de notre génome en entier, très peu utilisé mais très utile en cancéro dans le cadre de la médecine personnalisée, si le patient a telle mutation on pourra lui donner tel traitement)
- **Dépistage Prénatal non invasif : Pour la trisomie 13 18 21**

## Dépistage prénatal non invasif (DPNI)

Diagnostic Prénatal (DPN)	Dépistage prénatal non Invasif (DPNI)
<ul style="list-style-type: none"> <li>• Prélèvement <u>invasif</u> (amniocentèse, ponction du liquide amniotique)</li> <li>⇒ <b>Risque important de fausse couche ++</b></li> <li>• Méthode diagnostique : <u>Caryotype</u></li> </ul>  <p>Amniocentèse      Caryotype Trisomie 21</p>	<ul style="list-style-type: none"> <li>• Prélèvement <u>non invasif</u> (prise de sang de la maman dans des tubes Streck)</li> <li>• Méthode diagnostique : <b>NGS</b></li> </ul>  <p>Prise de sang (ADN foetal circulant)      NGS</p>

On ne parle pas de « diagnostic » mais bien de **dépistage** car on a pas assez de recul pour cette technique. Donc si on trouve quelque chose par le DPNI ou devra obligatoirement confirmer le résultat avec un caryotype.

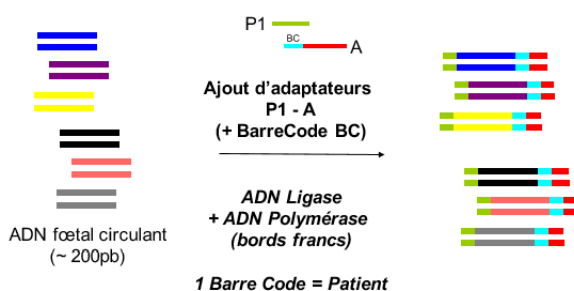
Dans le sang maternel, il y a **une petite quantité d'ADN foetal circulant** que l'on peut séquencer par NGS. Cet ADN vient des cellules trophoblastiques et est **d'apparition précoce** dans la circulation maternelle avec une **augmentation** de sa quantité au cours de la grossesse.

### Le Dépistage Prénatal Non Invasif (DPNI)

Les étapes:

1- Extraction ADN circulant à partir du plasma

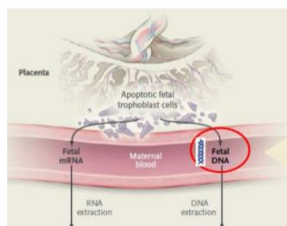
2- Préparation des échantillons:



Pour le séquençage, on reprend les mêmes principes du NGS classique, **à la différence que l'ADN foetal est déjà fragmenté** (donc pas besoin de le fragmenter et d'utiliser des endonucléases). On extrait notre ADN circulant à partir du plasma, on ajoute nos barre codes et adaptateurs, on séquence et on analyse informatiquement.

On n'a cependant pas assez de matériel pour séquencer plusieurs fois l'ADN foetal mais on a suffisamment de données pour quantifier le nombre de région de chacun de nos chromosomes. On recherche alors une surreprésentation chromosomique. C'est **une méthode quantitative**.

- L'analyse Bioinformatique:



- Analyse quantitative:  
Recherche d'une sur-représentation chromosomique

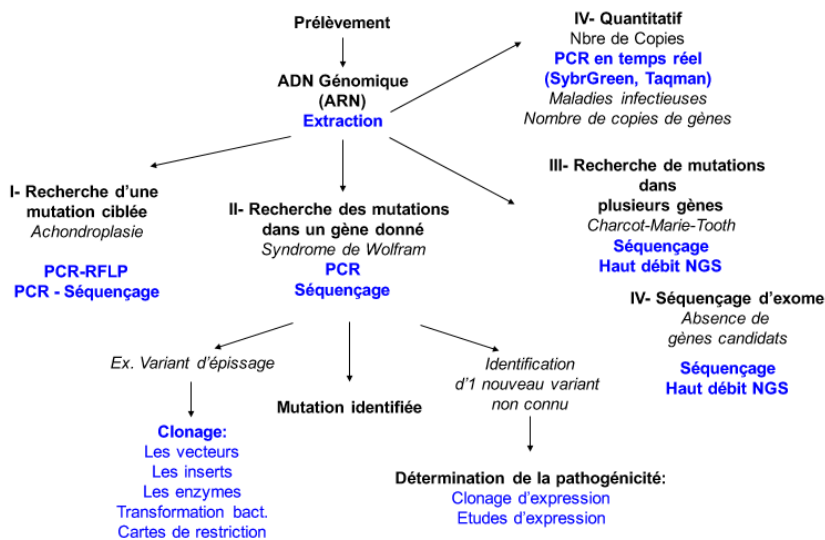
**Le DPNI est une analyse QUANTITATIVE car on recherche une sur-représentation chromosomique (chromosome surnuméraire), et non pas une variation nucléotidique. Elle ne s'applique que pour les trisomies 13 18 et 21 +++**

Mais si on a la présence d'un chromosome surnuméraire on est obligé de faire une amniocentèse pour confirmer. Le DPNI est en développement en ce moment.

**!/ Attention QCM :** Dans les QCMs d'annales où on demande à partir de quoi on peut extraire l'ADN, la prof comptait FAUX le plasma parce que le DPNI n'existait pas. En effet on n'extrait pas d'ADN génomique dans le plasma mais aujourd'hui on est capable d'extraire l'ADN circulant (foetal du coup) ++ Cependant la prof a dit qu'elle évitera ce genre de qcm un peu trop ambigu 😊

## Conclusion

En théorie, on peut utiliser le Séquençage haut débit pour tous types de recherches de mutations mais en pratique on ne l'utilise que pour la recherche de mutations dans plusieurs gènes. De ce fait, si dans un qcm on vous propose que le NGS est utilisé dans le cadre de la recherche d'une mutation ciblée par exemple ou de mutations dans un gène donné, il faut le compter faux !! Il faut vraiment garder le schéma suivant en tête ++ ! **Le NGS s'applique dans la recherche de mutations dans plusieurs gènes** (exemple de la maladie de Charcot Marie, maladie neurodégénérative dans laquelle plusieurs gènes sont impliqués). La prof explique qu'il y a un gros programme national de mis en place qui s'appelle France Médecine Génomique 20-25 qui a pour but d'ici 2025 de pouvoir séquencer tous les individus. Aujourd'hui, et de plus en plus, on essaie de classer les pathologies selon les gènes impliqués d'où l'intérêt du séquençage d'exome #introronéo1. Ce schéma c'est la base des QCMs, elle le répète 900



En résumé, le NGS c'est la base, on l'utilise dans pleins de domaines et on n'a pas fini d'en entendre parler. Des bisous à ma Audrey, t'es la meilleure !