

# UE11 - Méthodes d'étude et d'analyse du génome



## Séquençage Haut Débit Next Generation Sequencing

### I. Introduction

Nous avons vu dans les deux derniers cours comment effectuer un **prélèvement biologique** dans le but d'en **extraire du matériel génétique** (surtout de l'**ADN génomique**, mais aussi de l'**ARN**). Il est très important de retenir qu'on oriente la **démarche diagnostique** selon ce qu'on veut chercher :

- Si on cherche une **mutation ciblée** (*achondroplasie*), on utilise la **PCR-RFLP** qui est **toujours** confirmée par une technique de **séquençage** !
- Si on cherche **plusieurs mutations dans UN gène donné** complet (*syndrome de Wolfram*), on étudie le gène dans son ensemble grâce à une **PCR-Séquençage**. En fonction des résultats de ce séquençage :
  - Soit on **identifie toutes les mutations** et on pose le diagnostic
  - Soit les mutations ne sont **pas toutes exoniques** et on est en présence de **variants d'épissage introniques**, on va donc devoir **séparer les allèles** pour les séquencer à part en utilisant le **clonage moléculaire**
  - Dans les deux cas précédents, si le **variant identifié est inconnu**, on essaie de déterminer sa **pathogénicité** grâce à des études d'expression et au **clonage d'expression** d'une **protéine de fusion**
- Si on cherche à diagnostiquer une **maladie infectieuse** (déterminer la **charge virale**) ou qu'on veut **quantifier** le nombre de copies d'un gène amplifié on utilise la **PCR en temps réel** (SYBR Green, sonde TaqMan)
- Si on cherche **plusieurs mutations dans plusieurs gènes** (*syndrome de Charcot-Marie-Tooth*) on va utiliser le **séquençage haut débit** !

### II. Evolution des techniques

**1977 : séquençage « manuel » (Sanger)** = méthode de **référence**  
méthode de Sanger = méthode enzymatique des ddNTP  
**Séquençage manuel** utilisant des **ddNTP radiomarqués** dans 4 réactions différentes (A, T, C et G). En fonction de la **migration** et de l'identité des pistes on en déduisait les séquences, on pouvait le faire pour **maximum 200 pb**.



**1993 : séquenceurs « automatiques » (Sanger)**  
**Séquençage capillaire** basé sur une migration électrophorétique dans de petits capillaires remplis de gel d'agarose. Chaque ddNTP est couplé 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 simultané en 1h**, pour des fragments de maximum **800 pb** !

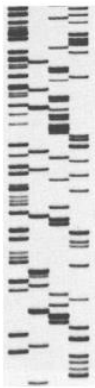


**2001 : séquençage de 95% du génome** (~ 30 000 gènes)

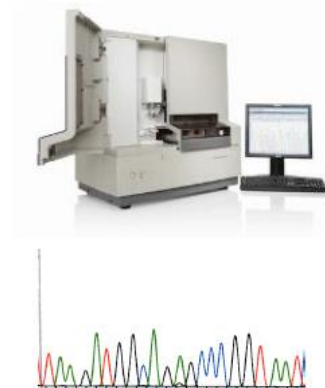


**2007 : séquençage haut débit (NGS)**  
Un NGS doit toujours être **confirmé** par un Sanger car nous n'avons pas suffisamment de recul à l'heure actuelle quant aux résultats obtenus par le séquençage haut débit pour nous affranchir de cette vérification.

## Un NGS doit toujours être confirmé par un Sanger qui est LA méthode de référence



1977  
Séquençage  
« manuel »  
Sanger

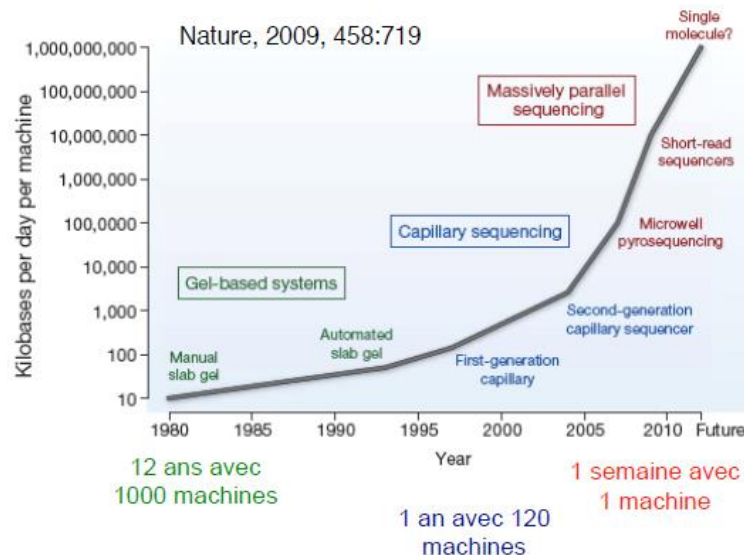


1993  
Séquenceurs  
« automatiques »  
Sanger



2007  
Séquençage  
haut débit  
« NGS »

**Remarque :** Aujourd'hui, grâce au NGS, il suffit d'une semaine pour séquencer tout le génome humain au lieu de plusieurs années !



## Le génome humain est composé de 3 milliards de pb

Il y a une évolution assez conséquente des **capacités de séquençage** du génome humain. Pour le séquencer entièrement, il aurait fallu :

- 12 ans avec 1000 machines dans les années 80
- 1 an avec 120 machines en 1995
- 1 semaine avec une seule machine aujourd'hui

On a réussi à séquencer **95% du génome humain** en 2001, ce qui représente le séquençage de **30 000 gènes** (soit **3 milliards de pb**) ! Ce séquençage a commencé en **1989** pour s'achever en **2003**.

## III. Définitions

Le NGS, signifiant **Next-Generation DNA Sequencing**, correspond au séquençage massif **en parallèle** de molécules d'ADN **individuellement séparées et amplifiées** sous forme de **clones** ou de **molécules uniques**.

Trois sociétés ont pris en charge le développement du NGS, chacune avec une méthode différente :

- o La société **Illumina** utilise des **nucléotides fluorescents**
- o La société **Roche** utilise des **nucléotides fluorescents**
- o La société **Life Technologies** utilise des **variations de pH**

**Remarque :** Dans le cadre de ce cours nous allons **uniquement** nous intéresser à la méthode des **variations de pH** (appelée la technologie **Ion Torrent**) développée par Life Technologies !

Il est capital de comprendre que les **étapes du NGS** sont **presque toutes les mêmes** pour ces trois sociétés, les deux seules choses qui vont changer concernent **l'amplification PCR-clonale** (que l'on va voir juste après) ainsi que **la manière dont la séquence va être lue : nucléotides fluorescents ou variations de pH** (Ion Torrent) !




Il existe donc **trois étapes communes** à ces trois sociétés :

- **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
- Ajout des **adaptateurs** et des **codes-barres** dans un dernier temps

Ces **trois étapes** se dérouleront **exactement de la même façon** et seront **strictement identiques** que l'on utilise la méthode de Roche, Illumina ou Life Technologies (peu importe donc que l'on utilise des nucléotides fluorescents ou des variations de pH à ce stade).

L'étape suivante (la quatrième si vous avez tout suivi) est ce qu'on appelle l'**amplification PCR-clonale** des fragments d'ADN.

Cette étape va être l'une des deux façons de **différencier la technique utilisée** par chacune des trois sociétés (avec aussi la manière dont sera lue la séquence, même si Roche et Illumina utilisent toutes deux des nucléotides fluorescents).

	ILLUMINA	ROCHE	LIFE TECHNOLOGIES
	<b>Extraction</b> de l'ADN génomique ou produit PCR (amplification de plusieurs gènes)		
	<b>Fragmentation</b> de l'ADN		
	Ajout des <b>adaptateurs</b> et des <b>code-barres</b>		
<b>Amplification PCR-clonale</b>	Capture des fragments d'ADN et <b>amplification sur support solide</b> (lame en verre)	Fixation sur billes d'agarose et séparation des billes (Eau/huile) <b>PCR en émulsion</b>	- Fixation sur des <b>sphères</b> - Séparation des billes (Eau/huile) <b>PCR en émulsion</b>
<b>Lecture de la séquence</b>	Nucléotides fluorescents	Nucléotides fluorescents	Variations de pH (Ion Torrent)

## IV. Préparation des échantillons

Nous allons à présent détailler les **trois étapes communes** évoquées précédemment, qui constituent la **préparation de nos échantillons**, étape préalable et nécessaire à l'**amplification PCR-clonale** et au **séquençage** à proprement parler !

### L'extraction, la fragmentation de l'ADN et l'ajout des adaptateurs et code-barres constituent la préparation des échantillons

Nous allons prendre l'exemple de la **méthode par capture** (qui peut être utilisée par les trois sociétés puisqu'encore une fois ces trois étapes de préparation des échantillons sont communes).

Nous allons grâce à cette méthode pouvoir **cibler les régions exoniques** des gènes qui nous intéressent dans le cadre d'une maladie **multigénique**, c'est-à-dire impliquant **plusieurs mutations sur plusieurs gènes** (comme le syndrome de Charcot-Marie-Tooth).

De manière générale nous avons **deux manières** d'utiliser le NGS :

- Soit il n'y a **pas de gènes candidats**, on ne sait pas lesquels peuvent causer un syndrome, on va donc **séquencer tout l'exome/le génome du patient** : on fait ce qu'on appelle un *whole genome* (le NGS est suffisamment puissant pour cela). Ce n'est cependant pas fait en routine.
- Soit on **connaît les gènes impliqués**, on va alors s'arranger pour **cibler** les régions génomiques correspondantes qui sont les seules à nous intéresser. Et comment les cible-t-on ? Justement grâce aux **sondes de capture** de cette méthode !

Nous sommes donc dans le cadre de ce cours dans la deuxième configuration, celle où l'on **connaît les gènes d'intérêt** et on veut donc **séquencer les régions exoniques** qui leur correspondent pour diagnostiquer une pathologie **multigénique** comme le **syndrome de Charcot-Marie-Tooth**. Nous allons donc voir les étapes permettant entre autres de **capturer nos régions génomiques d'intérêt**.

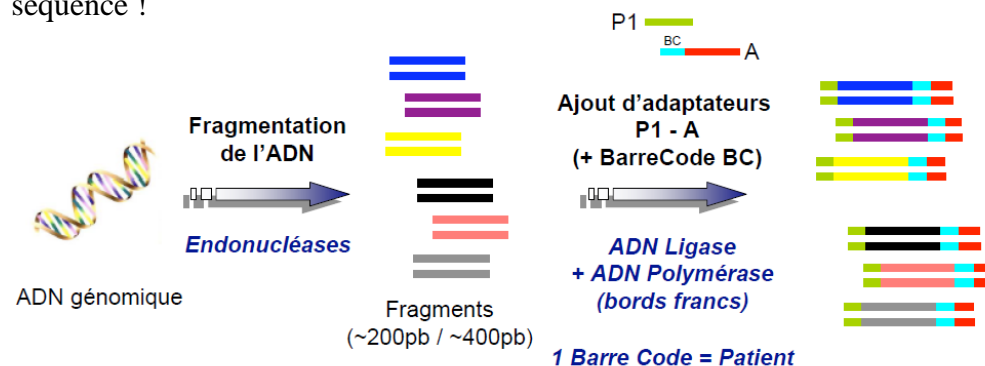
## A. Fragmentation de l'ADN

On ne revient pas sur les méthodes d'extraction du matériel génétique, cf. cours 1 pour plus de détails.

La molécule d'ADN que l'on a extrait du patient est **coupée en petits fragments de 200 à 400 pb** par des **endonucléases** qui coupent **l'ADN double brin sans avoir de spécificité !**

**Attention ce ne sont pas des enzymes de restriction**

Pour rappel, une enzyme de restriction coupe l'ADN double brin en fonction d'une **séquence spécifique** ! Or ici les endonucléases utilisées génèrent des fragments d'ADN de 200 à 400 pb **aléatoirement** sans reconnaître de séquence !



## B. Ajout des adaptateurs et des barre-codes

Une fois que l'ADN est fragmenté, on ajoute des **adaptateurs** (P1 et A sur le schéma ci-dessus) pour que les **extrémités 5' et 3'** soient **strictement identiques** pour tous les fragments !

On ajoute également des **barre-codes** (BC sur le schéma) qui eux sont **spécifiques du patient (1 BC = 1 patient) !**

Il ne faut pas confondre adaptateurs et barre-codes, les deux sont nécessaires au bon déroulement du NGS mais ils ont des rôles pratiquement opposés !

- **Les adaptateurs** nous permettent de n'utiliser plus qu'**un seul couple de primers/d'amorces** pour l'amplification par **PCR** qui va suivre car tous les fragments auront les **mêmes extrémités 5' et 3' !**
- **Les barre-codes** sont de petites séquences d'une dizaine de nucléotides bien définies, accrochées à une extrémité du fragment d'ADN et **spécifiques du patient**. On peut ainsi **séquencer plusieurs patients en même temps** dans une **même réaction** (dans un seul tube à essai). En effet, lors de la préparation des échantillons nous avons **un seul tube par patient** le temps d'ajouter à la séquence les barre-codes, mais au moment du **séquençage** les échantillons de tous les patients seront **mélangés** dans un seul tube. Ce sont les barre-code qui permettront après **analyse informatique** d'attribuer chaque séquence à un patient donné : le séquenceur reconnaît la séquence du BC et attribue l'ADN qu'il est en train de séquencer au bon patient !

**Remarque :** on ajoute également des **ligases** pour lier nos **BC** et nos **adaptateurs** aux fragments d'ADN de chaque patient ainsi qu'une **ADN polymérase** pour être sûr d'avoir des **bords francs** (on n'est pas sûr que les endonucléases aient générés des extrémités franches et non cohésives).

**Attention à ce stade l'ADN est toujours double brin !**

Nous avons donc obtenu les **multiples fragments** des gènes d'intérêt de nos patients, ceux qui sont concernés par de **multiples mutations** dans le cadre du syndrome de Charcot-Marie-Tooth (ou de n'importe quelle autre **pathologie multigénique** dont on connaît les gènes).

Le problème est que nous avons **fragmenté plusieurs gènes**, y compris donc des **régions exoniques** qui ne nous intéressent pas car on sait qu'elles ne seront **pas porteuses de mutations** : il faut donc trouver un moyen de les **éliminer** et de ne conserver que les autres.



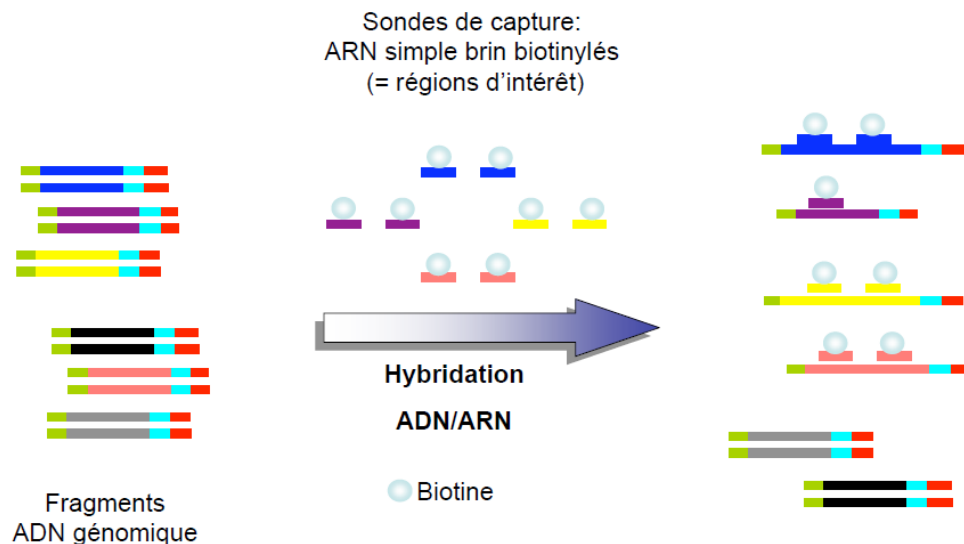
### C. Sondes de capture

Une fois que nos fragments d'ADN sont dotés de **deux adaptateurs** et d'un **BC**, on souhaite ensuite **isoler les régions exoniques d'intérêt**.

Nous allons donc utiliser des **sondes de capture biotinylées** composées d'**ARN simple brin** !

**Remarque :** une sonde biotinylée est tout simplement une **molécule d'ARN simple brin** à laquelle on a ajouté de la **biotine** (vitamine B8). Le fait de coupler nos sondes à de la biotine a un intérêt majeur qui sera expliqué lors de la prochaine étape de préparation des échantillons !

L'ADN est **toujours double brin**, il faut donc le **dénaturer** pour permettre une reconnaissance et un **appariement spécifique** entre la séquence de la **sonde d'ARN biotinylée** (= sonde de capture) qui est **complémentaire** d'un bout de séquence d'une de nos **régions d'intérêt** que l'on veut séquencer !



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

### D. Purification des fragments d'ADN capturés

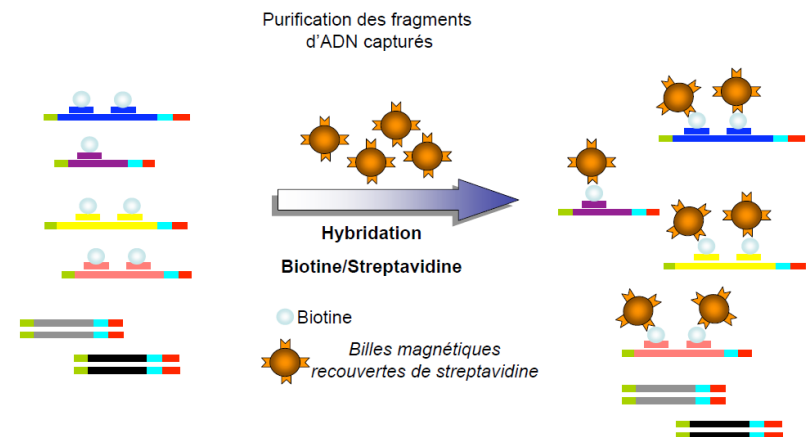
On arrive à obtenir des séquences d'ARN complémentaires d'un fragment de la région exonique qui nous intéresse avant même de séquencer cette dernière. *Comment est-ce possible ?*

Tout simplement parce qu'on a réussi à séquencer tout le génome humain (et oui on n'oublie pas le début du cours), et même en tenant compte du **polymorphisme génétique**, la grande majorité des séquences de chaque gène est **relativement identique** d'un individu à l'autre (que ce gène soit muté ou non). Ainsi, on arrive à obtenir de **courtes séquences nucléotidiques typiques/caractéristiques** d'un gène donné, qui sont **suffisantes** pour **permettre de l'identifier/de le capturer** par complémentarité tout en étant sûrs de ne pas se tromper !

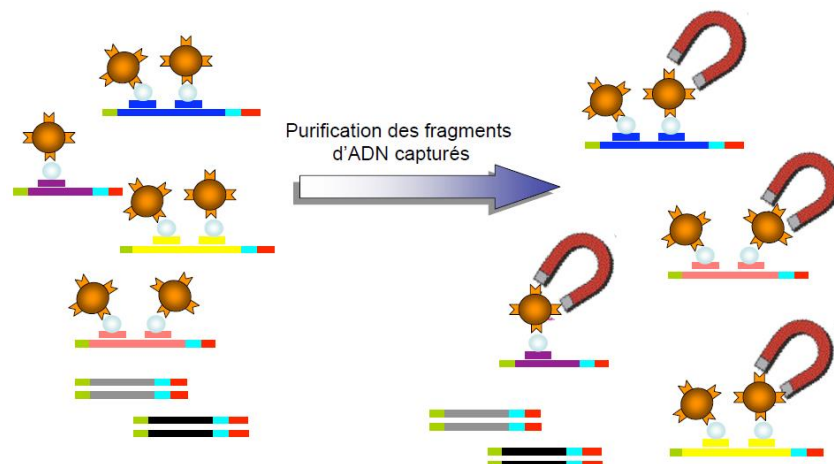
Une fois que cette hybridation est faite, on veut **capturer/récupérer seulement les régions d'intérêt**.

On va donc rajouter à notre mélange des **billes** (des **nanobilles** plus précisément, c'est-à-dire des billes de l'ordre du **nanomètre**) qui ont la particularité d'être **magnétiques** et recouvertes de **streptavidine**.

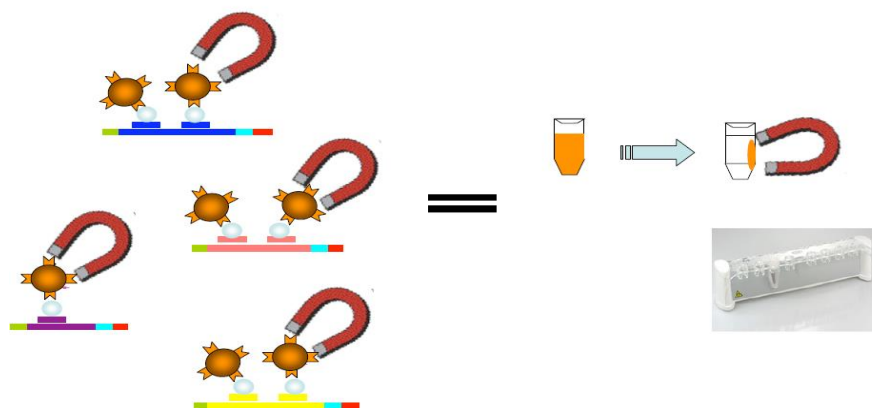
**La biotine et la streptavidine forment un complexe de très grande affinité !**



Ces billes **magnétiques** recouvertes de **streptavidine** vont donc **accrocher** l'ADN où l'ARN **biotinylé** s'est **hybridé** ! Autrement dit ces billes magnétiques ne vont se fixer qu'au niveau des fragments correspondant à nos **régions exoniques d'intérêt**.

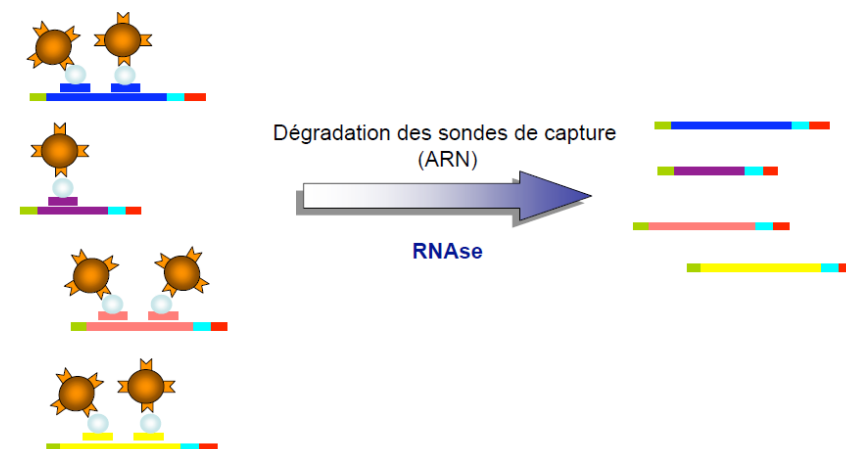


Pour **capturer** et **isoler** ces billes, on place les tubes à essai dans un **aimant**. L'aimant contre la paroi permet d'**attirer les billes magnétiques** dont la streptavidine est liée à la biotine des sondes de capture, elles-mêmes appariées à nos régions exoniques d'intérêt. Les autres fragments sans billes sont **éliminés** après plusieurs **lavages** et **élutions**.



## E. Dégradation des sondes de capture

On dégrade les sondes de capture ARN grâce à une RNase pour récupérer les fragments d'ADN simple brin correspondant à nos régions exoniques d'intérêt que nous avons réussi à isoler.



Les séquences d'intérêt ont été **isolées**, il faut maintenant les **séparer** et les **amplifier** par PCR avant de débuter leur **séquençage** !

A ce stade nous avons donc **terminé la préparation de nos échantillons** : nous avons extrait le matériel génétique du patient, on a fragmenté son ADN en utilisant des endonucléases non spécifiques, puis nous avons ajouté à ces fragments des adaptateurs pour n'utiliser que deux primers lors de la PCR pour tous nos fragments, ainsi que des BC pour attribuer chaque fragment à un patient ce qui nous permet de les mélanger.

Nous avons par la suite **identifié** les séquences exoniques d'intérêt et nous les avons **isolées** pour ne travailler maintenant qu'avec celles-ci. Nous avons donc **uniquement les fragments qui nous intéressent** et qui sont presque prêts pour le séquençage, nous devons juste **multiplier le matériel génétique** dont nous disposons avec une **technique d'amplification** un peu particulière, propre au NGS : **l'amplification clonale par PCR en émulsion** !

## V. Amplification clonale par PCR en émulsion

### A. Définitions

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

**Instant culinaire :** Une **émulsion** est un mélange de **deux substances liquides non miscibles**, l'une étant dispersée sous forme de **petites gouttelettes** dans l'autre. Ce sont toujours deux liquides qui ne se mélangent pas spontanément, comme **l'eau et l'huile**, mais qui vont grâce à des opérations spécifiques adopter un aspect macroscopiquement homogène, mais microscopiquement hétérogène : ces fameuses gouttelettes !

La **mayonnaise** est l'exemple le plus typique d'émulsion : on forme des gouttelettes d'huile dans de l'eau !

### Chaque gouttelette d'huile forme un microréacteur

Lors d'une **PCR en émulsion** on reproduit cette configuration de gouttelettes d'huile éparpillées dans de l'eau (c'est-à-dire une **émulsion**). Chacune de ces gouttelettes constitue ce que l'on appelle un **microréacteur**.

On s'arrange pour que chaque gouttelette, c'est-à-dire chaque microréacteur, ne contienne qu'**un seul fragment d'ADN** !

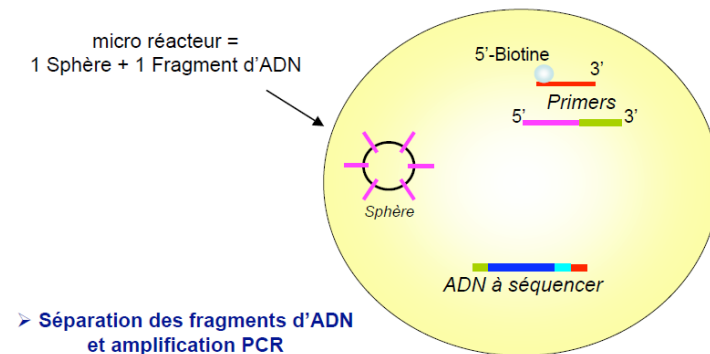
#### ○ Un microréacteur est toujours composé de :

- **1 sphère** sur laquelle sont fixés des primers
- **2 primers** dont un est **biotinylé** complémentaires des adaptateurs
- **ADN polymérase**
- Des **dNTP**
- Un **tampon**

### Il y a un seul fragment d'ADN donné pour une sphère !

Formation de « micro réacteurs » (Emulsion)

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



**Remarque :** On retrouve les **mêmes étapes** que pour la **PCR classique**, à savoir : dénaturation, hybridation, élongation, avec les mêmes variations cycliques de température, n fois.

**Remarque 2 :** Le **tout premier cycle** de la PCR en émulsion ne nécessite **pas d'étape de dénaturation** car l'ADN est déjà simple brin suite à la **dégradation des sondes de capture biotinylées**. Cependant cette information est anecdotique, retenez vraiment que les étapes de la PCR en émulsion sont **les mêmes** que celles d'une PCR classique, **dénaturation y compris** !

**Rappel :** Les **adaptateurs** sont tous **les mêmes** pour **TOUS** les fragments, ce qui fait que peu importe quel fragment est placé dans le réacteur, les mêmes primers seront utilisés partout il n'y aura aucune conséquence ! Ces adaptateurs sont l'équivalent de la borne d'amont et de la borne d'aval d'une PCR classique

## B. Amplification de l'ADN

**1. Dénaturation par la chaleur**  
(rend l'ADN à séquencer simple brin)



**2. Hybridation de l'amorce P1**  
sur l'extrémité 3' du fragment d'intérêt



**3. Elongation** de 5' en 3' par la polymérase



**4. Dénaturation** (rend l'ADN nouvellement synthétisé simple brin)



**5. Hybridation de l'amorce A biotinylée**  
sur l'extrémité 3' du brin nouvellement synthétisé



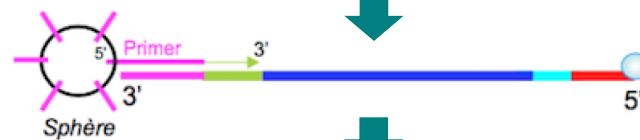
**6. Elongation** de 5' en 3' par la polymérase



**7. Dénaturation** (rend l'ADN nouvellement synthétisé simple brin)



**8. Hybridation** du brin nouvellement synthétisé sur le **primer de la sphère** par complémentarité



**9. Elongation**



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

Contrairement au fragment d'ADN d'intérêt de base où on avait que la **séquence de l'adaptateur en 3'** (en vert) les **brins fils** sont composés d'une **séquence supplémentaire** (partie rose/violette) qui a une **séquence complémentaire** de celle du **primer** qu'on a utilisé ! Cette extrémité nouvellement synthétisée sur nos brins d'ADN est complémentaire aux primers fixés sur notre sphère.

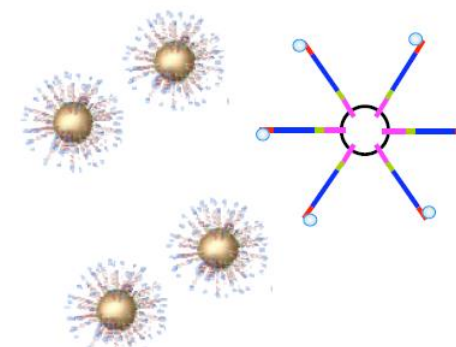
**Remarque :** on en déduit que la **séquence du primer utilisé est identique** à la séquence des **primers fixés sur la sphère**, puisque la séquence **complémentaire du primer** utilisé est également **complémentaire de la séquence des primers de la sphère** !

Donc après la **dénaturation** de l'ADN nouvellement synthétisé, on peut **hybrider** ce dernier au **primer de la sphère** par complémentarité. Après ça, la polymérase synthétise le brin complémentaire à partir de la sphère dans le **sens 5' 3'** !

Au bout de **n cycles d'amplification** on obtient une sphère qui correspond à un **fragment d'ADN amplifié n fois**. Il est important de comprendre que l'**extrémité 5'** de tous nos fragments **double brin** rattachés à la sphère est **biotinylée** ! Nos sphères sont donc ainsi recouvertes de **produits PCR avec une extrémité 5' biotinylée**.

**Une sphère = 1 fragment d'ADN amplifié**

**Remarque :** certains de nos **microréacteurs** contiennent des ADN **sans sphère** et d'autres des **sphères avec plusieurs fragments** différents mais le traitement informatique **détecte ces erreurs** pour ne pas fausser nos résultats.



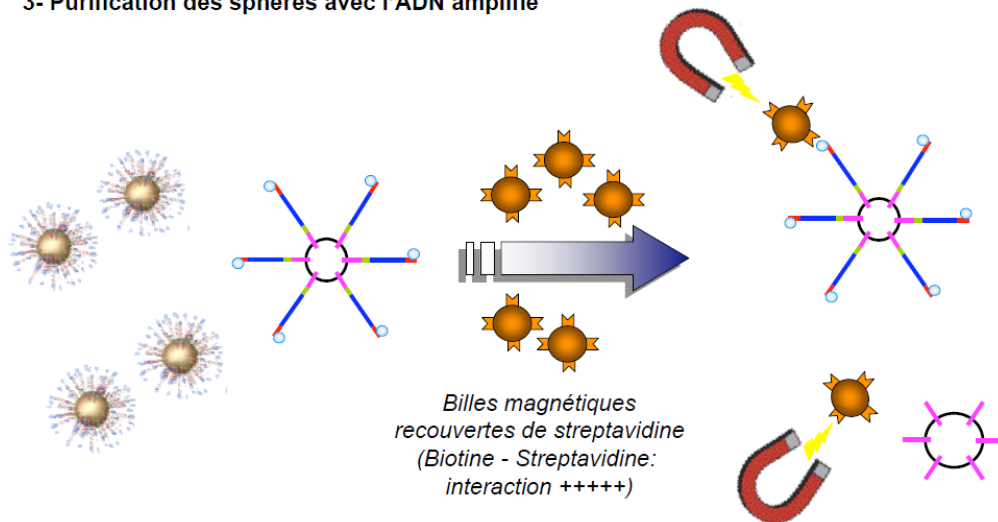


## IV. Purification des sphères avec l'ADN amplifié

Après **amplification de nos fragments d'ADN**, il faut **purifier les sphères**. On ne veut en effet récupérer que les **sphères** qui ont à leur **surface** des **fragments d'ADN amplifiés** et se débarrasser des sphères qui n'ont fixé aucun fragment d'ADN (= sphères sans extrémités biotinylées).

Nous allons donc réutiliser une seconde fois l'**interaction biotine/streptavidine**. 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 à une sphère.

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



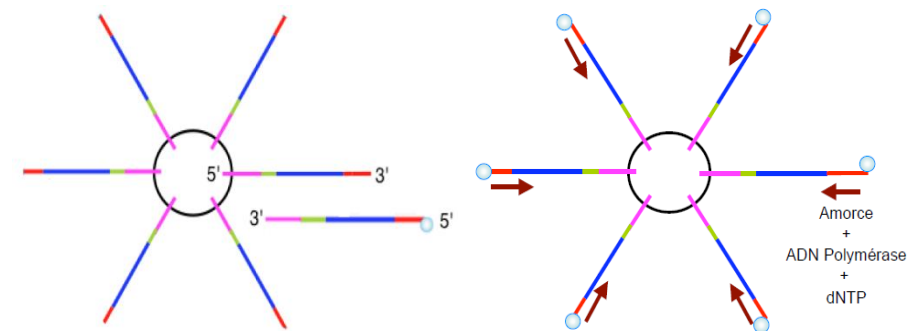
**On récupère les sphères avec de l'ADN amplifié et on élimine les sphères sans ADN amplifié (sans extrémité biotinylée)**

## V. Séquençage individuel des sphères

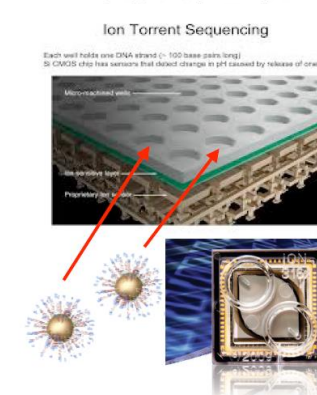
### A. Séquençage

#### 1. Dénaturation/Hybridation des produits d'amplification rattachés aux sphères juste avant de les charger sur une puce :

- **Dénaturation** à 95°C de l'ADN double brin fixé sur la sphère afin de **libérer le brin biotinylé en 5'** (ne servant pas pour le séquençage)
- **Hybridation d'une amorce** (complémentaire du primer A) sur l'extrémité 3' du brin non biotinylé.



On utilise une **amorce**, qui est la même pour toutes les sphères grâce aux adaptateurs, une **ADN polymérase** et des **dNTP**.



Pour le séquençage, on utilise une **puce** avec un **support métallique composé d'un grand nombre de puits** (1 à 11 millions) dans lesquels on va placer les fragments d'ADN à séquençer.

**Remarque :** Le nombre de puits étant **variable en fonction de la puce utilisée**, le nombre de réactions de séquençage possible varie aussi !

## Chaque puits de la puce ne peut contenir qu'une seule sphère !

Maintenant que les fragments que l'on veut séquencer sont sur une **sphère** dans une **puce** avec une **amorce**, une **polymérase** et des **dNTP**, il ne nous reste plus qu'à trouver comment **lire la séquence nucléotidique**.

Les sociétés Roche et Illumina utilisent dans ces puits contenant chacun une sphère des **dNTP fluorescents**.

Mais dans le cadre de ce cours nous avons déjà dit que nous ne nous intéressons qu'à la méthode développée par Life Technologies.

Nous avons déjà vu la technique particulière de **PCR en émulsion** utilisant des sphères développée par cette société, nous allons donc terminer avec leur méthode pour lire la séquence nucléotidique : **interpréter des variations de pH !**

### 2. Injection des dNTP les uns après les autres par la machine :

- **S'il y a complémentarité** : synthèse par la polymérase d'une **liaison de type phosphoester** (le H du groupement 3'-OH est libéré pour permettre la liaison avec l'extrémité 5'-P) + **libération d'un ion H<sup>+</sup>**

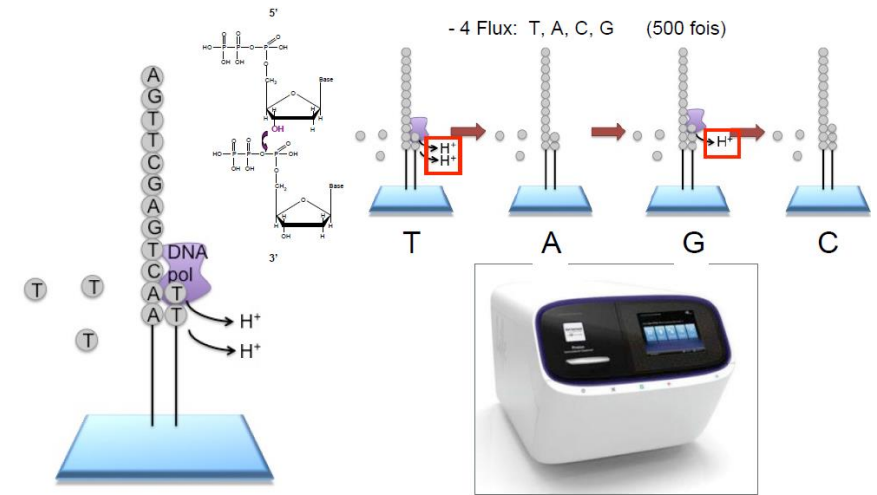
On obtient la création d'une **variation du pH** (détectée par la machine) !

- **S'il n'y a pas complémentarité** : on n'observe **pas de variation du pH**.

L'appareil **mesure le pH en continu** et **envoie les dNTP un par un**, donc 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 !

Par exemple si on envoie une guanine et qu'on n'observe aucune acidification **du milieu réactionnel**, cela signifie qu'**aucun proton n'a été libéré**, la polymérase n'a donc **pas formé de liaison** de type phosphoester et ce nucléotide n'est donc **pas le bon dans la séquence**.

Si maintenant on envoie par exemple une cytosine et qu'on observe une **diminution du pH**, cela signifie au contraire que la polymérase a réussi à **former une liaison** de type phosphoester et qu'elle a ainsi **libéré un proton** ! La cytosine est donc le **nucléotide suivant** de la séquence. Bien sûr, tout s'interprète par complémentarité des bases, la séquence du patient qui nous intéresse étant complémentaire du résultat que l'on obtient après séquençage.



**Remarque** : La réaction de séquençage dure de 2 à 3 h.

**Attention on n'utilise PAS la technique de séquençage automatisée avec des ddNTP fluorescents !**

### B. Analyse informatique

Lors de l'étape précédente de **séquençage**, nous avons réussi à déterminer en utilisant des **variations de pH** la **séquence nucléotidique** des fragments que nous avons amplifiés par PCR en émulsion.

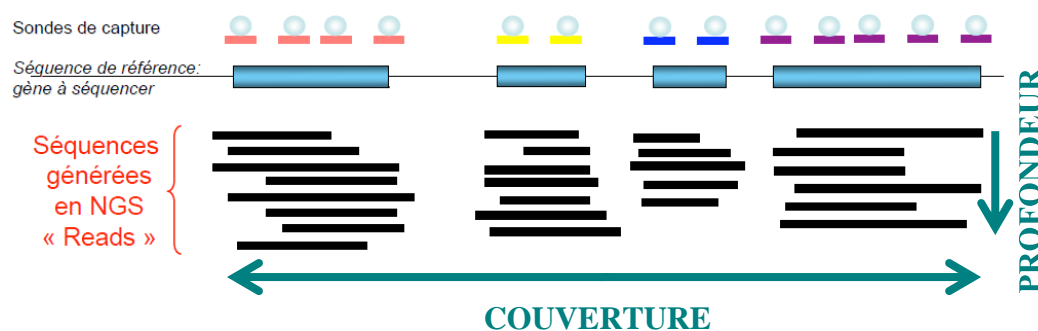
Il ne nous reste plus qu'à **comparer ces séquences obtenues à une séquence de référence** pour mettre en évidence de potentielles **mutations** : on va pouvoir détecter les changements nucléotidiques.

Une fois que ces **multiples mutations** auront été **répertoriées/cartographiées** sur le génome, nous pourrions potentiellement **poser un diagnostic de pathologie multigénique**, comme le syndrome de Charcot-Marie-Tooth !

La seule méthode dont nous disposons pour comparer nos séquences amplifiées avec les séquences originelles de référence est de procéder à une **analyse informatique**.

L'analyse informatique est très complexe : on a un **tri** et un **réalignement** sur la **séquence de référence** de toutes les séquences lues par la machine. Pour faire simple, la machine va **calquer la séquence** de chaque région d'intérêt exonique que l'on a **amplifiée** sur la séquence de l'**exon original d'un gène type**.

Je vous rappelle que l'on **connaît** la séquence caractéristique des exons du gène qui nous intéresse, car on a réussi à **séquencer l'intégralité du génome** et a fortiori **des exons** (on parle d'exome).



**Couverture :** Pourcentage des régions **séquencées** de notre région d'intérêt. Ainsi, une couverture de 100 % signifie que l'on a réussi à séquencer intégralement notre région d'intérêt !

**Profondeur :** Nombre de séquences générées, autrement dit c'est le nombre de « **reads** » que l'on a pour **une seule et même séquence**, ou encore le **nombre de fois qu'une même position nucléotidique a été lue** : de 1000 à 3000 fois, le NGS est une technique extrêmement puissante !

**Read :** Une séquence générée (chaque bâtonnet noir sur le schéma).

La vraie difficulté n'est pas de séquencer massivement plusieurs gènes. Ce n'est pas la technique de séquençage en elle-même qui pose problème, mais plutôt l'**interprétation** qui en découle.

En effet, on ne sait pas encore comment interpréter de façon satisfaisante et précise les résultats du NGS dans la pratique, car c'est une technique beaucoup **trop puissante** qui génère énormément de séquences, et on est dépassés par la quantité d'informations obtenues. Il nous faudra donc développer des techniques pour **identifier les variants** et déterminer leur **pathogénicité**.

## VI. Applications du NGS

### A. Généralités

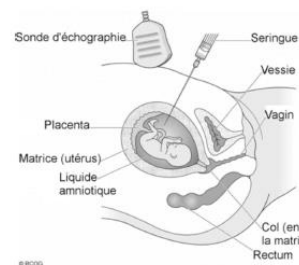
Le séquençage NGS est de plus en plus utilisé : c'est devenu une **technique de base** dans certains laboratoires (immunologie, virologie, etc...).

- Séquençage de **plusieurs gènes** (panel de plusieurs centaines de gènes)
- Séquençage d'**exome** (séquences codantes d'un génome)
- Séquençage du **génome entier** : « whole genome sequencing » (peu utilisé en diagnostic)
- **Diagnostic prénatal non invasif** ou **DPNI** (recherche des trisomies 13, 18, 21)

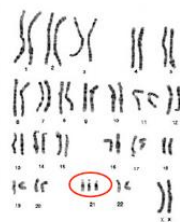
Pour terminer ce cours, nous allons nous intéresser d'un peu plus prêt à **une application précise du NGS**, le diagnostic prénatal non invasif (DPNI)

### B. Diagnostic prénatal non invasif (DPNI)

Diagnostic Prénatal	Diagnostic Prénatal Non Invasif
Prélèvement <b>très invasif</b>	Prélèvement <b>non invasif</b>
<b>Amniocentèse</b> (récupération du liquide amniotique par ponction)	<b>Prise de sang</b> (récupération de l'ADN fœtal circulant par prise de sang)
Réalisation d'un <b>caryotype</b>	<b>NGS</b>
<b>Risque important de fausse couche</b>	<b>Aucun risque pour le fœtus</b>

Diagnostic  
PrénatalDiagnostic Prénatal  
Non Invasif

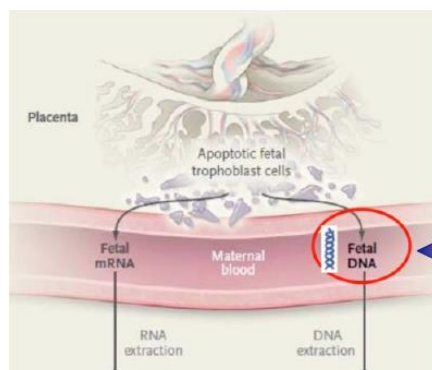
Amniocentèse

Caryotype  
Trisomie 21Prise de sang  
(ADN fœtal circulant)

NGS

Dans le **sang maternel** de la femme enceinte circule une **toute petite quantité d'ADN en provenance du fœtus** que l'on appelle **ADN fœtal circulant**. On va pouvoir le séquencer grâce au NGS.

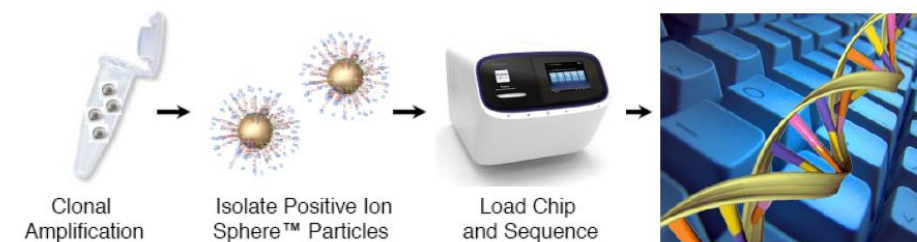
Cet ADN fœtal circulant vient des **cellules trophoblastiques** et est d'**apparition précoce** dans la circulation maternelle avec **une augmentation de sa quantité au cours de la grossesse**.

Prélèvement sanguin  
(tubes Streck)

Le séquençage de l'ADN fœtal circulant reprend **les mêmes principes que le NGS classique**, à la différence près que cet ADN fœtal circulant est **déjà fragmenté** ! Il n'y a donc **pas besoin d'utiliser des endonucléases**. On extrait notre ADN fœtal circulant à partir du **plasma**, puis on ajoute nos **barre-codes** et nos **adaptateurs**. On séquence et on procède à l'**analyse informatique**.

○ **Etapes du DPNI :**

1. Extraction de l'ADN fœtal circulant à partir du plasma
2. Préparation des échantillons (adaptateurs, barre code, etc...)
3. Amplification clonale (PCR en émulsion)
4. Purification des sphères
5. Séquençage
6. Analyse bio-informatique



**On n'utilise pas d'endonucléases car l'ADN fœtal circulant est déjà fragmenté !**

On n'a cependant **pas suffisamment de matériel génétique** pour **séquencer plusieurs fois** l'ADN fœtal, mais on a suffisamment de données pour **quantifier le nombre de régions** de chacun de nos chromosomes et chercher une **surreprésentation chromosomique**.

Il est important de comprendre que le NGS n'a **pas toujours pour but de déterminer une séquence nucléotidique**, comme c'est le cas ici ! Le NGS est utilisé dans un but **quantitatif** et **non qualitatif**.

On ne **détermine pas la séquence** de notre ADN fœtal circulant, il n'y en a pas assez pour ça, on veut simplement déterminer s'il y a **trop de fragments en provenance d'un chromosome 13, 18 ou 21** parmi tous les ADN fœtaux que l'on a récupérés pour diagnostiquer une trisomie !

Les fragments d'ADN sont **amplifiés**, on obtient ainsi de **nombreux reads** que l'on va séquencer et aligner sur le **génom de référence** via des **algorithmes de bio-informatique**.



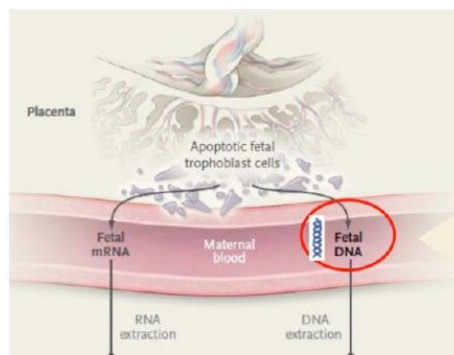
On obtient alors un fichier contenant la **liste des reads** associés à leur position sur le génome. Autrement dit, pour **chaque fragment séquencé, son chromosome lui est associé !**

Une **trisomie 21** se manifestera donc par un **excès de reads** s'alignant sur le **chromosome 21**. Pour quantifier précisément cet excès il nous faut des **valeurs de référence** obtenues chez des femmes enceintes témoins dont le fœtus est sain. Avec un nombre suffisant de témoins, la moyenne et l'écart-type du nombre de reads par chromosome sont calculés, ce qui nous permet d'obtenir un point de comparaison !

Le NGS permet juste de savoir dans ce cas à **quelle portion de quel chromosome correspondent nos reads** mais on ne les **séquence pas intégralement !**

**Le DPNI est une analyse QUANTITATIVE car on recherche une surreprésentation chromosomique (chromosome surnuméraire) et non pas une variation nucléotidique**

**Le DPNI ne s'applique qu'au diagnostic des trisomies 13, 18 et 21**



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

**Remarque :** si on détecte la présence d'un **chromosome surnuméraire**, on est obligés de faire une **amniocentèse** pour **confirmer le diagnostic par séquençage Sanger**, l'amniocentèse nous permettant de récupérer **suffisamment de matériel génétique** pour faire un séquençage complet.

## VII. Conclusion

On utilise en théorie le séquençage haut débit pour tous types de recherches de mutations dans plusieurs gènes. Ainsi, si dans un QCM on vous propose d'utiliser le NGS pour rechercher une mutation ciblée (*achondroplasie, cours 1*) ou une mutation dans un gène donné (*syndrome de Wolfram, cours 2*) vous devez compter l'item faux ! Il faut vraiment garder le schéma bilan ci-dessous en tête !

**Le NGS s'applique à la recherche de plusieurs mutations dans plusieurs gènes**

On utilise par exemple le NGS dans le diagnostic de la maladie de **Charcot-Marie-Tooth**, pathologie **neurodégénérative** dans laquelle **plusieurs gènes sont impliqués**.

