Cours n°4 – 06/12/12

1e PARTIE : MÉTHODES DE BIOLOGIE MOLÉCULAIRE ADAPTÉES À LA GÉNOMIQUE DES TUMEURS SOLIDES

Par Pr Laurence Bianchini

(La précision de diapos entre parenthèse concerne des diapos plutôt inutiles, les diapos importantes sont dans la fiche ;)).

Méthodes à connaitre car primordiales pour le diagnostic (dg) au niveau du laboratoire, en particulier le dg des tumeurs solides (sarcomes +++).

I. TECHNIQUES BASÉES SUR LA PCR

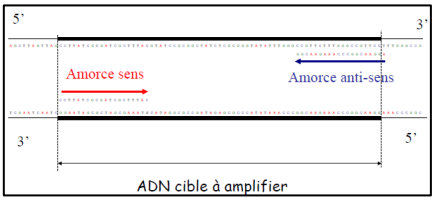
PCR = Polymerase Chain Reaction = amplification en chaine par polymérase

1. Principe de la PCR

- C'est une procédé permettant de **synthétiser in vitro une partie spécifique d'un acide nucléique donné** (matrice d'ADN ou d'ARN mais il faut transformer l'ARN en cDNA pour que la réaction puisse se faire) afin d'en obtenir une **quantité suffisante** pour le détecter (on doit connaitre la séquence à amplifier).
- Technique mise au point par Kary B. Mulis entre 1983 et 1985 puis prix nobel en 1993 ; elle a révolutionné la bio mol, essor très rapide et utilisée dans tous les laboratoires.

2. PCR conventionnelle

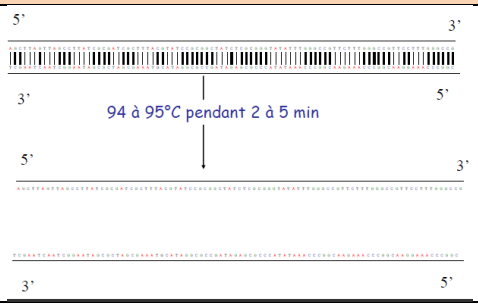
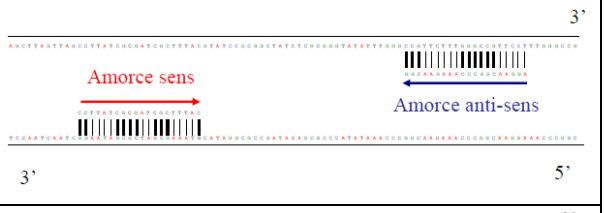
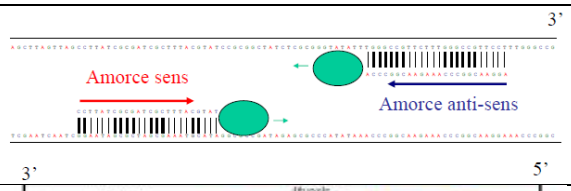
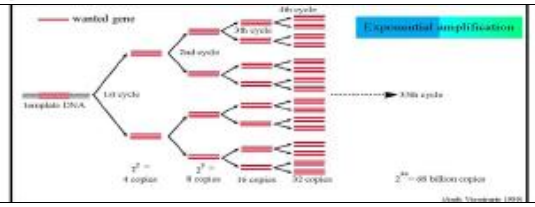
- Elle permet d'obtenir, à partir d'un échantillon complexe et peu abondant, **d'importantes quantités d'un fragment d'ADN spécifique et de longueur définie** (million de copies en qlq heures → c'est ça qui est vraiment révolutionnaire !).
- C'est une succession de réactions de **réplication d'une matrice double brin d'ADN grâce à des amorces ou "primers"**.
- Important : utilisation des produits de chaque étape de synthèse comme matrices pour les étapes suivantes **→amplification exponentielle.**



- Au niveau expérimental, on a besoin :
  * d'un ADN cible ou matrice, double brin évidemment.
  * de dNTP : dATP, dGTP, dCTP, dTTP.
  * de 2 oligonucléotides ou amorces : sens et antisens (primers forward et reverse) →
  * d'une enzyme : la Taq DNA polymérase (thermus aquaticus).
  * d'un tampon de la Taq DNA polymérase contenant du magnésium en particulier ($MgCl_2$).

→ On ajoute tout ça dans un tube à PCR, et la réaction va se faire dans un thermocycleur.

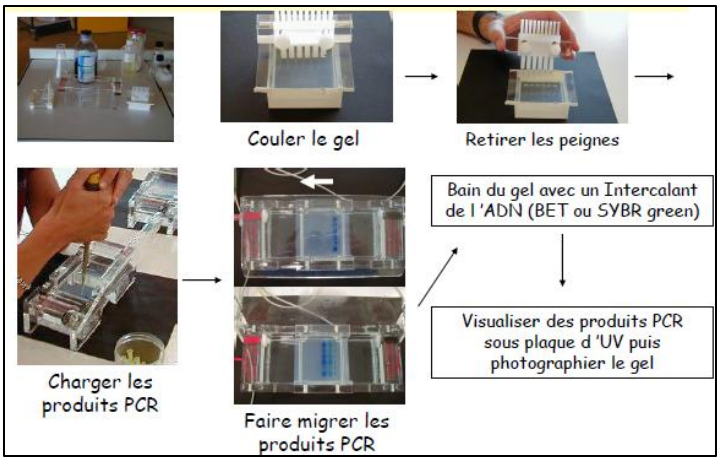
***Internet** (pour ceux que ça intéresse !) : le brin sens est celui qui a la même séquence nucléotidique que l'ARN messager transcrit (mis à part T/U). Le brin antisens sert de matrice à la polymérase.*
L'ARN est polymérisé du 5' au 3'. La matrice (brin "antisens") peut donc être représenté du 3' au 5'.
Le brin complémentaire (brin "sens") est représenté du 5' au 3'.
Sens : 5'-ATTTGCTGCCTT...TGC-3' Antisens : 3'-TAAACGACGGAA..ACG-5' Messager :5'-AUUUGCUGCCUU..UGC-3'
Pour ce qui est des amorces lors de la PCR :
-l'amorce "sens" s'hybride en 3' du brin antisens : *Brin antisens :.....3' TTAGGCATTGCAGTACCCG 5'*
Amorce sens :.....5' AATCCG 3' ====>
-l'amorce "antisens" s'hybride en 3' du brin sens : *Brin sens :.....5' AATCCGTAACGTCATGGGC 3'*
Amorce antisens :.....<==== 3' TACCCG 5'
Si vous regardez bien, l'amorce antisens est identique au début (5') de la séquence du brin antisens, donc elle est COMPLÉMENTAIRE de la fin (3') du brin sens, où elle s'hybride. Même chose avec l'amorce sens.

DANS LE THERMOCYCLEUR, 3 ÉTAPES ESSENTIELLES À L'AMPLIFICATION		
1e étape	**Dénaturation** des 2 brins de l'ADN matrice. → 94-95°C pendant 2 à 5min.	
2e étape	**Hybridation** des 2 amorces (rappel : elles ont été mises dans le tube dès le début). → 45 à 55°C, pendant 30 à 45sec (ça dépend du Tm=temps de fusion des amorces).	
3e étape	**Élongation** des 2 amorces grâce à la Taq polymérase (la réaction progresse dans les 2 sens). → 72°C pdt Xs selon la taille de la cible à amplifier 1000pb/min.	
Précision ***+ diapo 44 pour illustrer la PCR conventionnelle (mais pas nécessaire au fond !)***	On a les 2 produits de la 1e réaction qui servent de matrice pour la 2e réaction et dc de façon exponentielle, les réactions continuent : technique extrêmement puissante.	

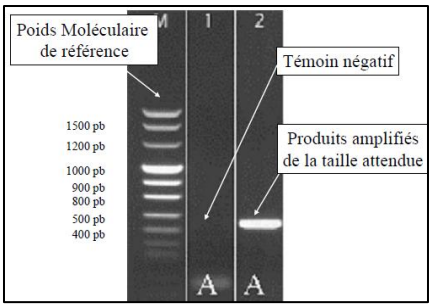
NB : Tm est la température de fusion des amorces. Elle dépend de la séquence, et d'autre chose. Donc suivant la séquence des oligo, on a des formules qui nous permettent de dire "on choisit ce Tm", et donc d'après le Tm de ces 2 amorces on va choisir un temps d'hybridation située entre 45 et 55°C. + la température est haute et + la température est stringente, cad spécifique.

❖ *Expérimentation en laboratoire*

- On remplit les tubes et on les met dans le thermocycleur, puis réaction (30-40 cycles).
- Une fois la réaction terminée, on coule un gel d'agarose, on remplit les petits puits des produits de PCR.
- Ensuite, on a besoin d'un agent permettant de visualiser l'ADN (bromure d'éthidium souvent) que l'on met soit dans les produits PCR soit dans le bain.
- Puis visualisation sous éclairage UV avec prise de photos pour voir les bandes.



Cours n°4 – 06/12/12



❖ *Image d'un gel d'électrophorèse*

- PM de référence très important.
- On obtient une seule bande pour les produits amplifiés (à savoir : si ça correspond à la taille attendue, on a amplifié comme il faut).
- Le témoin négatif est là pour vérifier qu'il n'y ait pas de contamination.

❖ *Optimisation du choix des oligonucléotides et des conditions de PCR*

- Nombre de cycles : en général on fait 30 cycles, mais + on fait de cycles, + la réaction est stringente (entre 25 et 40).
- Optimisation des conditions de PCR :

* concentration en MgCl2 : Mg^{2+} en quantité trop faible → rendement pas assez bon (on n'a aucune bande) mais Mg^{2+} en excès → altère la spécificité de la réaction (gel totalement saturé).

* quantité d'ADN matrice.

* concentration en amorces.

* spécificité des amorces : 18 à 24 nucléotides (en général 20), 40 à 60% de GC, température d'accrochage des amorces (phase d'annealing) calculée en fonction du Tm (temps de fusion) des 2 amorces cad de 45 à 55°C (on peut vraiment faire varier les conditions pour optimiser et obtenir la bande recherchée, correspondant au fragment qu'on veut amplifier).

3. RT-PCR = Reverse Transcription-PCR

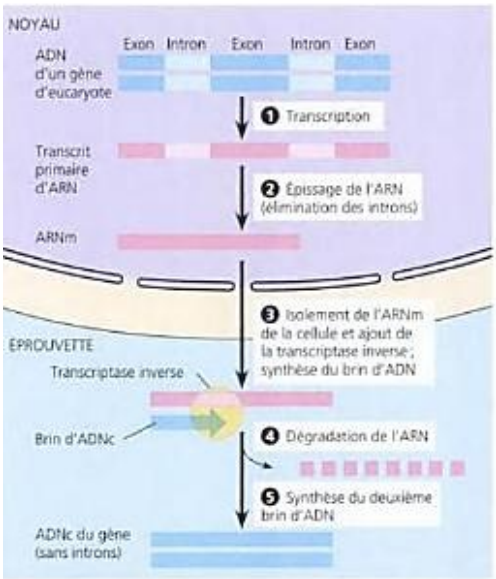
- Elle est réalisée après **transcription inverse d'un ARN en ADN complémentaire** (ADNc) car **on ne peut pas faire de PCR directement sur de l'ARN.**
- Elle est mise au point pour utiliser les **ARN** comme matrice d'amplification de la PCR après la reverse transcription.
- Méthode la + sensible pour détecter les **ARNm** au niveau d'un organe, d'un tissu ou d'une cellule.

Ex d'application : on veut voir la quantité, l'expression d'ARN messagers dans une tumeur par rapport au tissu sain.

Résumé : *** PCR : se fait sur de l'ADN qui est une matrice double brin.**
*** RT-PCR : analyse de l'ARN mais transcription inverse obligatoire en 1er lieu pour faire de la PCR.**

- Dans l'ADN d'un gène eucaryote il y a alternance d'exons (codants) et d'introns (souvent + longs que les exons).
- <u>In vivo</u> → transcription de l'ADN, puis épissage du transcrit primaire pour éliminer les introns et former l'ARNm.

- <u>In vitro</u> → quand on veut aboutir à la synthèse d'ADNc, on part d'une cellule qu'on broie, on isole les ARNm et on ajoute une transcriptase inverse qui a pour rôle, sur la matrice de l'ARN, de synthétiser le brin d'ADNc. Puis il y a synthèse du 2e brin d'ADN par la Taq polymérase → On a une matrice double brin pour faire la réaction de PCR classique.



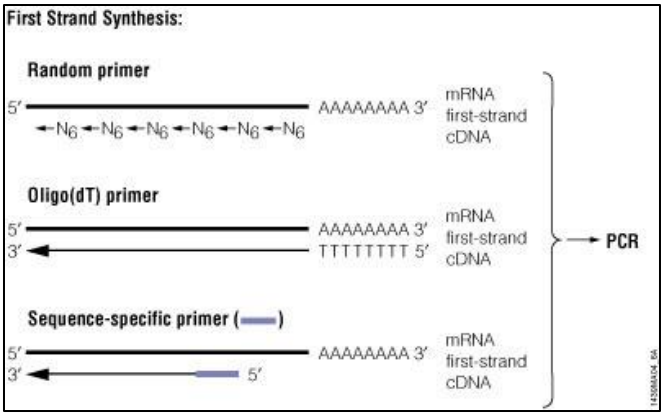
Cours n°4 – 06/12/12

Question : Comment peut-on être sûr d'avoir que des ARN épissés et pas des transcrits primaires?
*Pcqu'on a des techniques qui nous permettent de récupérer vraiment que les ARNm, et on utilisera ensuite des oligonucléotides qui se mettent à la jonction **exon-exon** dc si on a un transcrit primaire, on ne pourra pas faire d'amplification. De plus, quand on fait de la RT-PCR, on fait aussi un traitement à la DNase pour être sûr d'éliminer tout ce qui est ADN, donc on n'a plus que de l'ARNm.*

❖ Amorçage de la transcription inverse

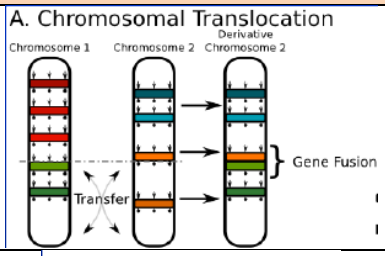
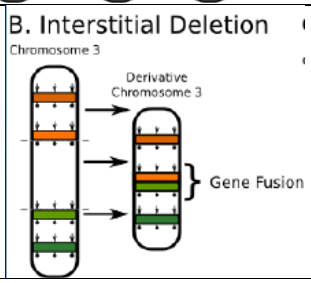
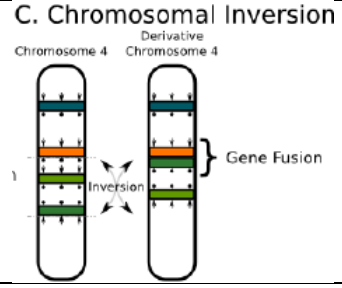
→ 3 techniques.

- Random primer : utilisation d'hexamères N_6 qui vont s'hybrider de façon aléatoire tout le long de la séquence de l'ARNm et chacun synthétise un bout du brin.
- Oligo(dT) primer : complémentaire de la queue polyA, puis la reverse transcriptase synthétise le **tout** brin.
- Sequence-specific primer : se fixe sur une séquence précise mais on perd la séquence avant le primer (voir le 3e dessin : pas de synthèse à droite du primer). On l'utilise quand on ne veut pas synthétiser l'ADNc tout entier.



❖ Application de la RT-PCR : recherche d'un transcrit de fusion spécifique d'un type de tumeur

- Les translocations entre différents K vont donner des gènes de fusion spécifiques d'un certain type de tumeurs, suivant évidemment l'endroit où les cassures s'effectuent au niveau du K.

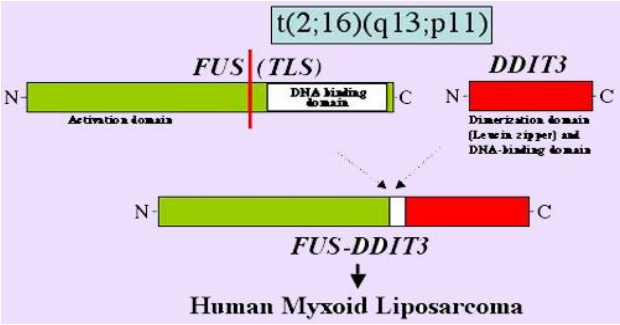
3 GRANDS MÉCANISMES DE FORMATION DES GÈNES DE FUSION	
Translocation chromosomique entre un K1 et un K2 : si on se concentre sur le K2 (et sur la région transloquée) dans un caryotype, on aura un K2 normal avec son allèle normal et un der2 (K dérivé 2) avec une juxtaposition de 2 segments de K qui ne sont normalement absolument pas juxtaposés, cad un der2 avec un gène de fusion.	
Délétion interstitielle : il ne s'agit que d'**un seul K**, le K 3 par ex, il y a délétion d'un morceau du K et donc rapprochement de 2 bandes chromosomiques, ce qui peut former un gène de fusion. Dans le caryotype, on aura alors un K3 normal avec son allèle normal et un der3 (K dérivé 3) avec son gène de fusion.	
Inversion au niveau d'un K : il ne s'agit que d'**un seul K** encore une fois, mais il y a inversion de 2 parties de ce K cette fois. Si on prend le K4 par ex, dans le caryotype on aura alors un K4 normal avec son allèle normal et un der4 avec son gène de fusion.	

Cours n°4 – 06/12/12



❖ *Exemple du liposarcome myxoïde & transcrit FUS-DDIT3*

- Les sarcomes sont des tumeurs rares **(1% des tumeurs malignes**). Ce sont des tumeurs d'origine mésenchymateuses (de tissus extra squelettiques non épithéliaux). Elles peuvent être très agressives.
- Liposarcome : tumeur du tissu adipeux.
- Liposarcome myxoïde : caractérisé par une translocation entre K12 (q13) et K16 (p11) → génération d'un gène de fusion entre FUS et DDIT3 : cassure dans le gène FUS et cassure sur le K16 au niveau de DDIT3 ce qui donne un gène de fusion FUS-DDIT3, qui donne naissance à un **transcrit FUS-DDIT3**, puis à une protéine chimérique. **C'est ce transcrit qui sera détecté par RT-PCR**.
- Mécanisme : on sait où ça casse dans les 2 gènes, donc on va prendre un primer dans le gène FUS, un primer dans le gène DDIT3, et par RT-PCR on va pouvoir faire un dg de liposarcome myxoïde.
- La protéine de fusion FUS-DDIT3 va jouer le rôle de facteur de transcription et induire l'expression de différents gènes intervenant dans la tumorigénèse des liposarcomes myxoïdes.



❖ *Translocation et gènes de fusion dans les sarcomes*

(Diapo 26, mais pas trop d'intérêt je pense…)

- C'est dans les sarcomes et les hémopathies que les gènes de fusion ont été d'abord identifiés car leurs caryotypes sont relativement simples, avec des anomalies faciles à repérer.
- Sarcome d'Ewing : on a toujours le partenaire EWS dans le gène de fusion, cad que le K22 est toujours impliqué. En grande majorité détecté : EWS-FLI1 donc on le recherche en 1e intention mais s'il n'est pas détecté, ça ne veut pas pour autant dire qu'on a rien, c'est peut être juste pcq le partenaire de EWS est autre chose que FLI1.

Donc, même si ça ne représente qu'un faible pourcentage des sarcomes d'Ewing, c'est au niveau du dg que c'est très important pcq souvent (au niveau de l'anapath) on n'est pas sûr à 100% que c'est un sarcome d'Ewing, et grâce à des techniques de bio mol, avec les RT-PCR, et avec la détection de ces gènes (transcrits) de fusion, on a vraiment la preuve qu'il s'agit de ce type de tumeur.

- Synovialosarcome : translocation entre KX et K18.
- Dermatofibrosarcome protuberans (DP) : translocation entre K17 et K22.

Question : pq ne fait-on pas une PCR conventionnelle sur la zone du gène de fusion mais une RT-PCR?
Pcq'il peut y avoir le gène de fusion sans le transcrit de fusion (rappel : c'est le transcrit qui donne naissance à la protéine).
En effet, le gène de fusion peut être détecté avec la FISH, mais après il peut y avoir un gène de fusion qui ne donne pas un transcrit de fusion donc, avec la RT-PCR, on sait qu'il y a le transcrit de fusion (« et après, est ce que le transcrit de fusion va donner une protéine de fusion ? Ce n'est pas sûr non plus »).
Question : et pq ne pas chercher directement la protéine de fusion dans ce cas?
Car c'est beaucoup + compliqué ! Alors qu'un transcrit de fusion, une fois qu'on sait où ça casse, c'est + "simple".

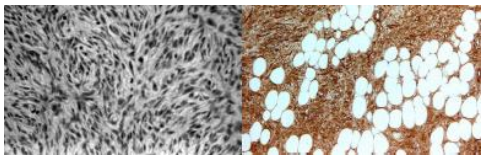
❖ *Translocation et gènes de fusion dans les carcinomes*

- Il y a qlq années, on pensait que les gènes de fusion étaient propres aux sarcomes et aux hémopathies, et pas forcément présents dans les carcinomes. Et depuis qlq années, on s'est rendu compte que c'est pcqu'on ne savait pas les rechercher.
- Dans les carcinomes thyroïdien, prostatique, pulmonaire, mammaire sécrétoire, et l'adénocarcinome pléiomorphique (*elle n'a cité qu'eux*), on a des anomalies chromosomiques qui donnent des gènes de fusion. Les partenaires de fusion impliqués sont souvent soit des facteurs de transcription, soit des Rc à activité tyrosine kinase. Ceci est très important pour avoir un dg sûr de la tumeur étudiée.

Cours n°4 – 06/12/12

❖ *Dermatofibrosarcoma Protuberans (DP)*

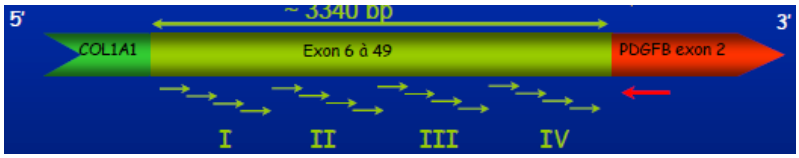
- Sarcome rare : 0,1% des tumeurs.
- Rarement métastatique (dans environ 11% des cas).
- Évolution lente.
- Classiquement chez le jeune adulte.
- Tumeur infiltrante, récidivante.
- **Traitement chirurgical par exérèse large** (marges de 5cm pour éviter tout risque de récidives mais marge énorme et problématique si au niveau du visage !!).
- Anapath : prolifération fuso-cellulaire dermo-hypodermique dense de disposition "storiforme" avec marquage CD34+.



- Caryotype (*diapo 29, mais sans grand intérêt*) :
  * chez l'adulte : présence d'un K en anneau.
  * cas pédiatrique : translocation entre K17 et K22.

Mais même si l'anomalie du caryotype est différente, le DP est caractérisé par le même **gène de fusion COL1A1-PDGFB**, car on observe un rapprochement des signaux (normalement sur 2 K différents).

- Dg moléculaire par RT-PCR :
  * le COL1A1 a un nombre d'exons très important contrairement au PDGFB.
  * on sait que la cassure dans PDGFB se fait toujours exactement au niveau de l'exon 2, tandis que la cassure dans le gène COL1A1 se fait dans les exons 6 à 49 donc gros problème dg.

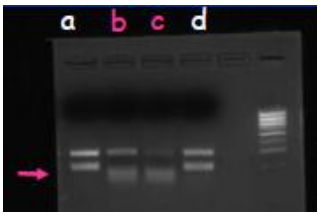


→ Donc, quand on fait de la RT-PCR, on a un primer reverse dans PDGFB, et on fait 4 PCR différentes avec 4 primers différents pour COL1A1 (=PCR multiplexe, donc manipulation compliquée pour COL1A1).
Stratégie : 1. Multiplexes / 2. Simplexes / 3. Clonage / 4. Séquençage.
Multiplexe = on met plusieurs oligo sens et antisens dans la même réaction de PCR.

❖ *Exemple de dg par RT-PCR : le synovialosarcome*

- Caractérisé par translocation entre KX et K18, et un transcrit de fusion SYT-SSX.





1e étape : détection du gène(transcrit) de fusion SYT-SSX chez les patients b et d.
→ Bande obtenue par RT-PCR en prenant un primer pour SYT et un primer pour SSX.

2e étape : distinction entre les 2 isoformes SSX1 et SSX2 (ils ont des séquences très similaires, et il existe aussi SSX3 et SSX4) comme partenaires de sites pour le synovialosarcome grâce à une coupure enzymatique (Taq1). Mais rien ne prouve que le pronostic soit différent pour un des 2 pour le moment (b et c ont un transcrit SYT-SSX1).

On ne peut pas faire une RT-PCR sur un gène de fusion, ce sont des fragments beaucoup trop importants : il faut qu'il y ait le transcrit de fusion.

Cours n°4 – 06/12/12

4. Techniques de PCR non conventionnelle

→ Utilisées en recherche, mais pas en routine pour les dg car techniques compliquées à mettre en place.

❖ *DOP PCR = Degenerated Oligonucleotide-Primed PCR*

- Méthode universelle pour l'amplification non spécifique et uniforme de l'ADN.
- Peut être utilisée comme étape de pré-amplification.
- But : passer de 5-10 copies d'ADN à un nombre beaucoup + important de façon à ensuite incorporer des nucléotides fluo et faire du FISH.
- Donc, c'est une PCR avec :

  * des amorces nucléotidiques "dégénérées" qui vont permettre d'amplifier un maximum de séquences cible d'ADN (car on ne connait pas la séquence à amplifier).

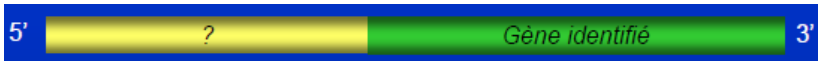
  * une amorce de Télénius faite pour reconnaitre le + grand nombre de sites dans le génome humain, constituée de 22 bases (10 bases en 5' spécifiques, les 6 suivantes dégénérées, 6 bases en 3' spécifiques).

➢ A retenir : on peut faire de la PCR et amplifier comme ça avec des amorces qui sont presque des amorces aléatoires, avec un design fait pour amplifier un maximum de séquences.

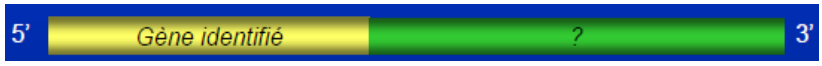
❖ *RACE PCR=Rapid Amplification of cDNA ends*

Rappel : la PCR se fait sur un matériel de départ en très petite quantité, la RT-PCR se fait sur ADNc obtenu à partir d'ARNm avec séquence connu et des amorces spécifiques.

- 5' RACE PCR : PCR sur ADNc obtenu à partir d'ARNm, **séquence connue en 3'** avec utilisation d'une amorce spécifique, **séquence inconnue en 5'**.



- 3' RACE PCR : PCR sur ADNc obtenu à partir d'ARNm, et **c'est l'inverse** (séquence connue en 5' avec utilisation d'une amorce spécifique, séquence inconnue en 3').

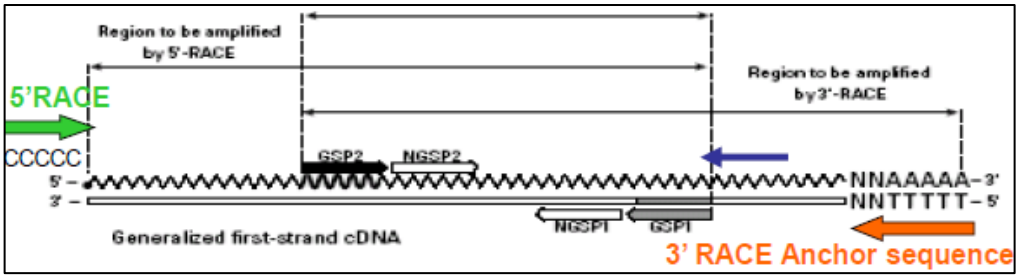


- Utilisation de cette technique **lorsqu'on veut identifier un partenaire de fusion d'un gène qu'on connait** (ex du sarcome d'Ewing : EWS fusionné avec un autre gène que l'on connait ou pas).
- Évidemment, + l'anomalie est sous représentée, + elle sera difficile à détecter.

- Technique :

  * en 3' RACE on utilise un primer anchor qui s'hybride avec la queue polyA de l'ARNm.

  * en 5' RACE on utilise une enzyme qui va greffer un stretch de C et on va avoir un oligo avec des G qui va s'hybrider dessus.



➢ A retenir : il y a des techniques qui permettent **d'identifier un partenaire de fusion** d'un gène qui a déjà été identifié.

❖ *Ex de l'utilisation de 3' RACE PCR : détection du gène de fusion HMGA2-MSRB3-NFIB dans un lipome*

- Gène HMGA2 : gène composé de 5 exons remodelant l'architecture de la chromatine, il est très important dans les tumeurs adipeuses (et dans d'autres tumeurs). La cassure se fait souvent au même endroit ; plusieurs partenaires de fusion sont identifiés pour HMGA2 : **la partie 5' de HMGA2 va fusionner avec la partie 3' d'un gène sur un autre K**.
- Ici, MSRB3 se retrouve au milieu, mais le lipome ne dépend « que » de la fusion entre HMGA2 et NFIB.
- Au niveau FISH : la sonde rouge reconnait les exons 1-3 de HMGA2 et la sonde verte reconnait les exons 4-9 de NFIB : signal fusionné jaune montrant qu'il y a bien rapprochement entre les 2 *(diapo 39 mais il n'y a rien de fou niveau signal !).*

- Au niveau du transcrit : utilisation de la 3' RACE PCR. On a identifié le gène NFIB, et utilisation d'un primer anchor pour le HMGA2 puis gel d'électrophorèse → bandes → séquençage de ces bandes, et lorsqu'on les confronte avec des banques de données, on s'aperçoit qu'on a la séquence de HMGA2 juxtaposée ac la séquence de MSRB3 et ac la séquence de NFIB.
- On se retrouve avec un gène hybride, entre une partie de HMGA2 des exons 1-3 et une partie de NFIB (en l'occurrence ici, c'est l'exon 9).
- De plus, comme il y a des éléments de régulation négative dans la région 3' de HMGA2, quand il y a cassure puis fusion on perd ces éléments, et on a une surexpression de HMGA2 (**normalement pas exprimé** dans les tissus adultes).

5. Risques de contamination

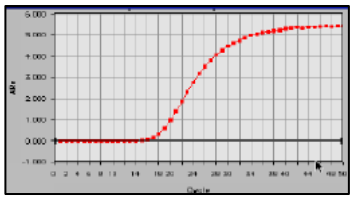
- La PCR est une méthode très très efficace car elle permet une amplification considérable, mais les risques de contamination sont très élevés.
- Travail sous une hotte.
- Circuit monodirectionnel (obligatoire en PCR dg) :
  * extraction et purification.
  * préparation des échantillons pour la PCR.
  * analyse des échantillons sur gel d'agarose.

6. Inconvénients de la PCR classique

- Etape de révélation sur gel d'électrophorèse.
- Risque de contamination quand ouverture du tube.
- N'est pas quantitative.

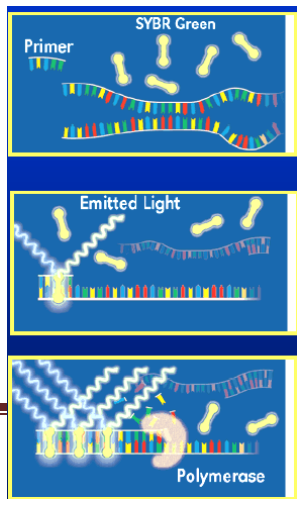
7. PCR quantitative (q-PCR) ou en temps réel (Real Time PCR)

- Technique qui a été un très grand progrès pour accéder à, et évaluer l'expression des gènes dans nos échantillons (très important dans le dg car si surexprimés, ça participe au mécanisme de tumorigénèse).
- La PCR en temps réel permet la **détection et** la **quantification** d'un reporter fluorescent dont l'émission est directement proportionnelle à la quantité d'amplicons générés pendant la réaction de PCR (en temps réel).



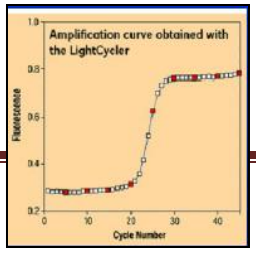
- Courbe obtenue avec la PCR en temps réel : au fur et à mesure des cycles de PCR, la fluorescence augmente. Au début de la courbe, on a du non spécifique. La fin de la courbe est la phase plateau, elle dépend du nombre de cycles. Il n'y a pas de gel en fin de réaction : des logiciels extrapolent, à partir de ces courbes, l'expression des gènes d'intérêt.

- Il existe 3 méthodes de détection :
  * sondes d'hydrolyse (TaqMan, Beacons, Scorpions).
  * sondes d'hybridation (Light Cycler).
  * composés liant l'ADN (SYBR Green).



❖ ***SYBR Green (composés liant l'ADN)***

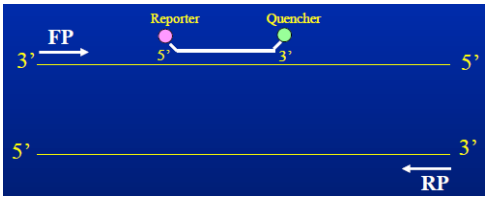
- C'est un intercalant de l'ADN qui a l'avantage d'émettre une fluorescence très importante **une fois qu'il est lié** à l'ADN double brin.
- On enregistre la fluorescence émise et ceci forme une courbe (mesure de fluorescence en fin d'élongation).
- Dans cette méthode, on a des primers comme dans la PCR normale.



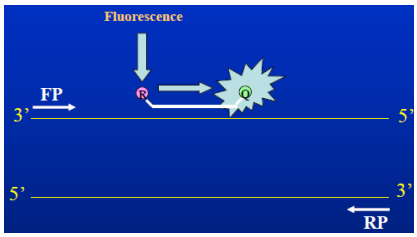
❖ *Sonde TaqMan (sonde d'hydrolyse)*

- Méthode + spécifique : la sonde va être composée d'un primer sens et d'un primer anti sens (jusque là, comme dans la PCR normale) et on a, en +, la **présence d'une sonde fluorescente** qui va augmenter la spécificité. Ces 3 choses vont s'hybrider.
- Cette méthode repose sur 2 principes :
  * la présence de 2 fluorochromes sur la même sonde.
  * l'activité 5'-exonucléase de la Taq Pol.

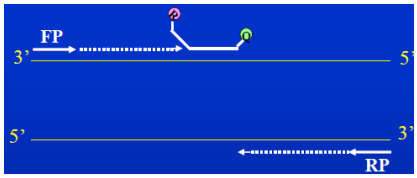
→ **Spécificité de l'hybridation de la sonde TaqMan** (alors qu'il y a spécificité de l'amorce pour la PCR).



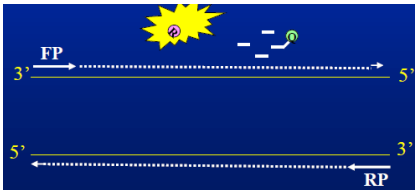
- Excitation des fluorochromes : le fluorochrome suppresseur (**quencher**) **inhibe l'émission du fluorochrome** émetteur (**reporter**) lorsqu'ils sont à proximité → **pas de signal fluorescent émis par le reporter quand la sonde est intacte**.
(le quencher n'émet pas de fluo, il absorbe juste).

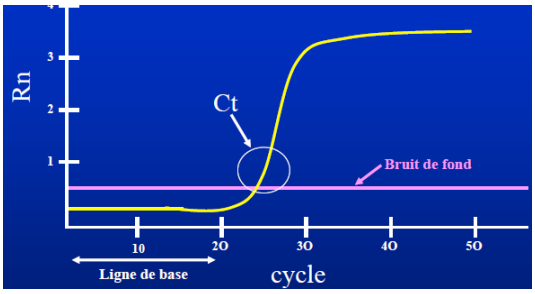


- Au cours de l'élongation (même principe que la PCR normale) catalysée par la Taq Pol, l'activité 5' nucléase déplace et hydrolyse la sonde.



- Quand la sonde va être hydrolysée, le quencher ne va plus absorber la fluorescence émise par le reporter → émission de fluorescence du reporter. Lorsque la polymérisation est complétée, **pour chaque molécule d'ADN synthétisée**, un **reporter émet** de la fluorescence.





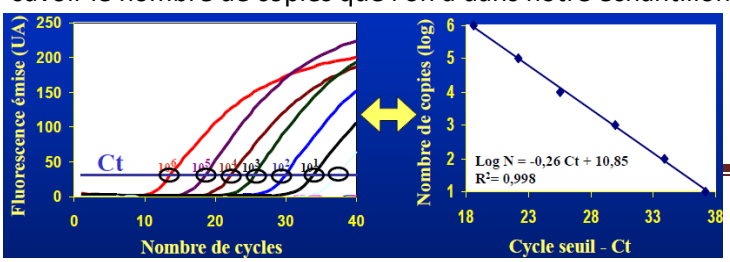
- Voici la courbe obtenue en fonction du nb de cycles. Nous avons une ligne de base, et un bruit de fond cad la fluorescence non spécifique (valable dans les cycles précoces).
Ct : cycle seuil, cycle de PCR pour lequel le signal fluorescent atteint la valeur du seuil (phase exponentielle). C'est ceci qu'on va enregistrer et qui va être important pour ensuite extrapoler le niveau d'expression de notre gène d'intérêt.

Rn : signal fluorescent normalisé par une référence fluorescente passive.
Seuil : calculé sur le signal normalisé durant les cycles précoces de la PCR (ligne de base).

- Courbes de calibration : comparer les valeurs de Ct

Le nombre de cycles nécessaires pour atteindre une fluorescence donnée (phase log-linéaire) est fonction du nombre d'ADN cibles initialement présents : + il y a un nombre de cycles important, - on a de molécules présentes initialement dans notre solution. Autrement dit, - le gène sera présent dans notre solution, + le Ct sera élevé.

On va de 10 en 10 (Ct) et avec ces valeurs là, on établit une droite. Grâce à la droite formée, on peut extrapoler et savoir le nombre de copies que l'on a dans notre échantillon.



Ceci permet donc de faire de la quantification absolue d'un gène : **on peut savoir combien il y a de copies de ce gène dans la tumeur étudiée**.

Cours n°4 – 06/12/12

❖ Application de la PCR quantitative : quantification relative des ARNm par RT-PCR quantitative

- On fait beaucoup de RT-PCR quantitative : on veut quantifier l'ARNm dans l'échantillon, donc à partir d'un tissu → extraction de l'ARN → reverse transcriptase (transcriptase inverse de l'ARN en ADNc) → PCR quantitative.
- Ce n'est plus de la quantification absolue mais relative, on compare le pathologique au normal, appelé échantillon calibrateur.
- Donc il faut déterminer :

* **un échantillon calibrateur :** par rapport auquel sera déterminé le niveau d'expression du gène cible dans les échantillons inconnus. Il **représente l'état normal** en terme physiologique (normalisation des résultats)
(ex : culture cellulaire non traité par une drogue, organe non atteint dans une pathologie, lignée sauvage).

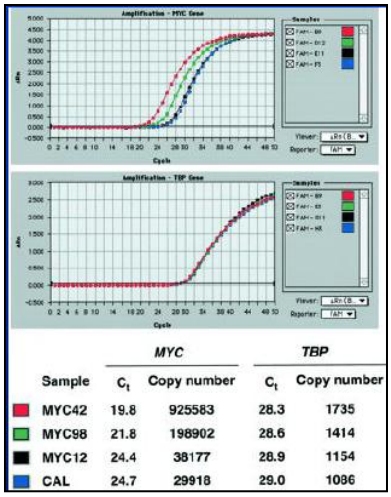
* **un gène de référence :** normalisation de la quantité d'ARN sur laquelle on effectue la RT-PCR. C'est un gène dont l'expression doit être la + stable possible. L'utilisation d'un **gène domestique** (housekeeping gene) comme référence permet de normaliser le processus d'extraction, la qualité de l'ARN, l'efficacité de l'étape de transcription inverse.
→ Importance du choix du gène de référence le + pertinent.
(ex : glyceraldehyde-3-phosphate deshydrogenase, beta-actin, RPLP0 (ribosomal protein large P0)).

- Quand on va faire les calculs, **on va tenir compte de l'échantillon tumoral et de l'échantillon calibrateur**.

➢ A retenir : lorsqu'on fait une manipulation de RT-PCR quantitative et qu'on veut mesurer par ex HMGA2 dans un liposarcome, il nous faut le liposarcome **et du tissu adipeux sain.** Et si on veut faire une PCR quantitative sur HMGA2, on doit aussi la faire, et dans les mêmes conditions, **sur le gène de ménage** (=gène domestique).
➢ **A retenir +++ : gène de référence et échantillon calibrateur dans cette manipulation.**

❖ Quantification de l'expression du gène Myc dans des tumeurs du sein par RT-PCR en temps réel

- **Gène d'intérêt : Myc**.
- **Gène domestique : TBP**.
- Echantillons différents : rouge (MYC42) vert (MYC98) noir (MYC12).
- Echantillon calibrateur : bleu.



	MYC		TBP	
Sample	C_t	Copy number	C_t	Copy number
MYC42	19.8	925583	28.3	1735
MYC98	21.8	198902	28.6	1414
MYC12	24.4	38177	28.9	1154
CAL	24.7	29918	29.0	1086

- 1er graphique : les courbes de Myc sont décalées, la courbe rouge est en avant de la courbe verte qui est en avant de la courbe bleue (la courbe noire est superposée sur la courbe bleue) donc il y a + d'expression de Myc dans les courbes rouge et verte : surexpression de Myc dans les tumeurs MYC42 (19x) et MYC98 (5x).
- 2e graphique : courbes de TBP confondues, donc à peu près la même quantité de TBP.
- Avec ce genre de technique on accède à la **quantification de l'expression** du gène d'intérêt dans notre tissu d'intérêt. Ceci nous permet de savoir si notre gène d'intérêt est surexprimé ou au contraire inhibé dans une tumeur, **comparé à un tissu sain**.

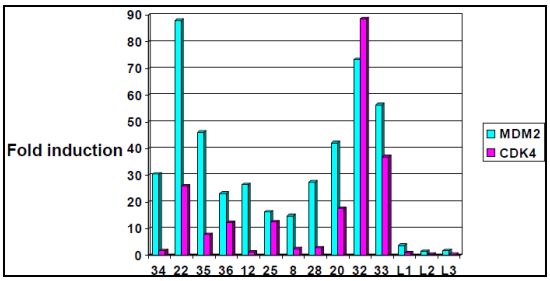
Question : comment sait-on que cette surexpression est pathologique?
On fait toujours l'expérience en regardant **par rapport à un tissu sain**! → Donc il y a toujours comparaison.
Et il faut aussi regarder au niveau de la protéine comment ça se traduit.

❖ Liposarcomes bien différenciés

- Ils sont caractérisés par un K surnuméraire : soit un K géant soit un K en anneau
- Il y a qlq années, grâce au FISH, on s'est rendu compte qu'il y avait un grand nombre de copies de MDM2 sur le K géant. Aujourd'hui, on sait qu'il y a un amplicon du K12 (= séquences du K12 amplifiées de façon très importantes sur ces K surnuméraires, et notamment MDM2).

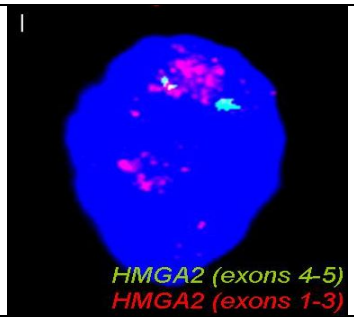
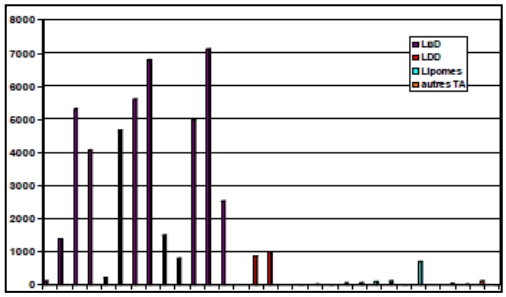
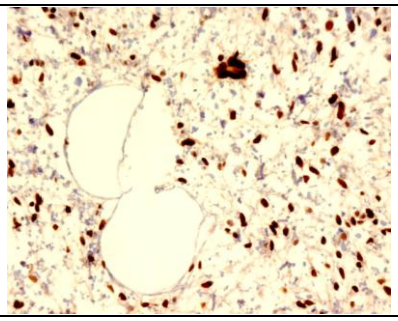
Cours n°4 – 06/12/12

- Analyse par RT-PCR en temps réel des gènes MDM2 et CDK4 dans des liposarcomes bien différenciés et des lipomes :
  * ces 2 gènes sont sur le K12.
  * fold induction correspond à la comparaison avec le tissu sain.
  * dans les **lipomes** L1 L2 L3 (tumeurs adipeuses bénignes), il n'y a **pas de surexpression de ces 2 gènes**.
  * **MDM2 est surexprimé dans presque 100% des liposarcomes** et ceci se traduit au niveau de la protéine → **marqueur des liposarcomes bien différenciés.**
  * **CDK4 surexprimé et amplifié dans 80-85% des liposarcomes**.



→ C'est grâce à cette technique que l'on peut distinguer un lipome (tumeur bénigne pratiquement la + répandue) d'un liposarcome, si au niveau de l'anapath ce n'est pas évident à voir → **marqueur diagnostic**.

❖ *HMGA2 : un gène clé dans la physiopathologie des tumeurs adipeuses*

		
FISH : multiples copies de HMGA2 exons 1-3.	RT-PCR en temps réel : surexpression	Immuno histochimie : surexpression traduite au niveau de la protéine.

→HMGA2 est amplifié et surexprimé dans les LBD/LDD.

Diapos 68-69-70 (*qui ne servent à rien*) : thermocycleur et autres machines, 96 puits traités en environ 2h.

❖ *Avantages de la PCR quantitative*

- L'amplification peut être suivie en temps réel.
- Rapidité et possibilité de traiter un grand nombre d'échantillons en même temps.
- Pas de traitement post-PCR, meilleur contrôle de la contamination.
- Gamme dynamique de concentrations très large (5 à 8 log).
- Nécessite 1000 fois moins d'ARN qu'une méthode conventionnelle.
- Détecte des variations d'expression jusqu'au seuil minimal d'un facteur 2.
- Méthode la + sensible, la + précise, la + reproductible (et c'est la seule permettant de quantifier).
- Pas + chère que la PCR conventionnelle (à l'exception du coût de l'équipement).

Résumé des techniques de PCR :

*** Matériel de départ en très petite quantité.**
*** PCR conventionnelle : sur ADN, séquence du fragment à amplifier connue, amorces spécifiques.**
*** RT-PCR : sur ADNc obtenu à partir d'ARNm, séquence connue, amorces spécifiques.**
*** DOP PCR : sur ADN, séquence inconnue, amorces dégénérées.**
*** 5' RACE PCR : sur ADNc obtenu à partir d'ARNm (=RT PCR), séquence connue en 3' avec amorce spécifique, séquence inconnue en 5'.**
*** PCR temps réel : sur ADN, séquence connue, amorces spécifiques.**
*** RT-PCR temps réel : sur ADNc obtenu à partir d'ARNm, séquence connue, amorces spécifiques.**

Cours n°4 – 06/12/12

I. TECHNIQUES NON BASÉES SUR LA PCR : Southern Blot & Northern Blot

- Techniques peu utilisées de nos jours, et plus du tout utilisées en dg.
- Elles nécessitent un matériel de départ en **grande quantité** (de l'ordre du microgramme, et pas de l'ordre du nanogramme comme dans les autres techniques) **:**
  * ADN pour le Southern.
  * ARN pour le Northern . *(protéine pour le Western)*

1. Southern Blot

- Se fait sur de l'**ADN**.
- Les étapes : *(Diapos 77 & 78 pour les dessins, mais non nécessaires je pense [et ils auraient pris bcp trop de place !])*

1- **Clivage enzymatique de restriction** : on va utiliser des enzymes de restriction (on sait où elles coupent) pr cliver l'ADN et avoir des fragments de restriction.
2- **Electrophorèse sur gel d'agarose** : séparation des fragments en fonction de leur taille (ac du bromure d'éthidium=BET).
3- **Dénaturation dans le gel des fragments de restriction**.
4- **Transfert sur support solide par capillarité** (buvardage).
5- **Fixation de l'ADN sur la membrane** (réplication exacte de ce qu'il y avait sur le gel) par cuisson (membrane de nitrocellulose) ou exposition UV (membrane de nylon).
6- **Hybridation de la membrane avec une sonde marquée complémentaire** de ce que l'on recherche.
7- **Lavages** : élimination de tout ce qui est non spécifique.
8- **Autoradiographie** : on a les bandes correspondant au signal.

Problème : technique très longue et matériel de départ en quantité beaucoup + importante que tout ce qui est PCR.

2. Northern Blot

- Principe identique à celui du Southern mais étude sur des **ARNm** (pas de digestion par enzymes de restriction).
- La visualisation d'un ARN par une sonde (fluorescente ou radioactive) permet de :
  * apprécier sa distribution dans les tissus, étudier son abondance relative.
  * déterminer sa taille : gros avantage de cette technique.
  * détecter les intermédiaires de maturation et les différentes formes d'épissage de l'ARN (migration à des tailles différentes) : autre avantage de cette technique.
- Les étapes : *(Diapos 81 & 82…)*

1- **Extraction des ARNm et purification**.
2- **Electrophorèse sur gel d'agarose en conditions dénaturantes**.
3- **Visualisation des ARNm** par coloration au BET (intercalant de l'ADN) sous UV.
4- **Transfert des ARNm du gel sur une membrane** (nitrocellulose ou nylon).

→ **2 ARNm différents** A et B sont liés à la membrane → incubation de la membrane avec une **sonde spécifique** (ADN simple brin, ARN ou oligonucléotide) **homologue à l'un des 2 ARNm** présents sur la membrane. La sonde porte un indicateur (fluo ou radioactif) permettant la détection ultérieur du signal → **lavages** éliminant la sonde qui n'est pas liée de façon spécifique à l'ARNm sur la membrane → l'ARNm A reste lié à la membrane : c'est l'ARNm spécifique complémentaire à la sonde.

(Diapos 85-86-87-88 non traitées)

Cours n°4 – 06/12/12

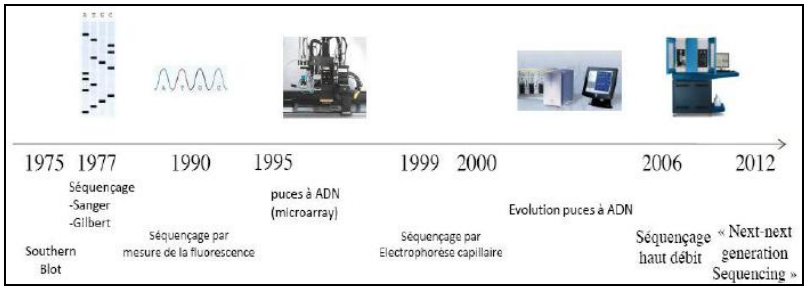
2e PARTIE : SÉQUENÇAGE À HAUT DEBIT ET APPLICATIONS

Par Pr Valérie Duranton-Tanneur

La prof a lu ses diapos durant pratiquement toute cette 2e partie…

I. HISTORIQUE DES TECHNOLOGIES

- Années 70 : apparition du séquençage Sanger et Gilbert (techniques très fastidieuses).
- Années 2000 : automatisation du séquençage.
- Au cours des années 2000 : apparition des puces à ADN. Elles ont permis d'obtenir un très grand nombre de données.
- En 2005 : techniques de séquençage à haut débit (on en est actuellement à la 3e génération).



1. Méthode Sanger « historique »

- Technique qui reposait sur **4 réactions** de séquences (mises en parallèle ensuite) contenant :
  * ADN matrice.
  * l'amorce (ou primer).
  * l'ADN polymérase I.
  * les nucléotides dATP, dCTP, dGTP, dTTP dont un étant un di-déoxynucléotide marqué au 35S (arrêt de synthèse de la PCR à l'incorporation de ce di-déoxynucleotide).

→ La méthode qualitative repose sur l'utilisation de ddNTP qui va terminer la séquence de manière aléatoire.

- Les réactions étaient migrées sur gel en fonction de la taille et on en déduisait la séquence : tout ça était manuel, très fastidieux *(« et autant vous dire qu'on ne séquençait pas de très longues séquences d'ADN »*).

- Son évolution : 20ans + tard, automatisation du séquençage avec l'utilisation de di-désoxynucleéotides marqués avec des fluoroflores différents ce qui permettait de faire **une seule réaction PCR dans le mm tube**.
- Cette réaction de PCR est ensuite passée sur un séquenceur capillaire ou sur gel et, grâce à un algorithme, le séquenceur et l'ordinateur sortaient la séquence.
- Toujours utilisée à l'heure actuelle dans les techniques de gènes candidats, et cette évolution de la technique a permis un gain de temps considérable.

2. Puces à ADN

- Analyse quantitative et qualitative des séquences.
- Elles permettent des études d'expression de gènes.
- La technique repose sur une hybridation compétitive d'ADN d'un témoin et d'ADN expérimentale sur une puce, ce qui permet de voir des anomalies K (gain/perte de K, ou d'une partie).

Cours n°4 – 06/12/12

II. SÉQUENCAGE HAUT DÉBIT

Aussi appelé NGS (Next Generation Sequencing) / HTS

- Les grands principes :
  * intégration (système combinant les avantages de la PCR et des puces).
  * parallélisation (il s'agit d'une vaste PCR multiplex).
  * miniaturisation (on va vers des systèmes de + en + petits, de moins en moins couteux et de + en + performant).
- Le NGS a permis le séquençage entier du génome humain.

1. Principe *(Diapo 8 si besoin…)*

- Matériel biologique : ADN ou ARN.
- Etapes communes aux différentes technologies :
  * fragmentation enzymatique de l'ADN.
  * préparation d'une banque d'ADN (library) par ligation d'adaptateurs à chaque extrémité des fragments d'ADN.
  * amplification clonale (équivalent d'une PCR multiplex).
  * séquençage générant des signaux (luminescent ou fluorescent), les signaux varient en fonction des technologies utilisées.
  * détection des signaux émis et conversion en séquence.

2. Technologies *(Diapos 9-10 si besoin…)*

- Actuellement, 3 compagnies majeures se partagent le marché : ILLUMINA, ROCHE et LIFE TECHNOLOGIES avec des séquenceurs qui sont de + en + petits pour faire une application + adaptée au dg (=séquenceurs de paillasse).
- Comparaison des technologies : il existe différents appareils, différents types d'amplifications, les performances sont évaluées en Gb ou Mb, et le temps pour sortir un run peut varier (à l'heure actuelle, en moins d'une journée, on peut séquencer des millions de bases).

NB : un run (réalisation d'un processus complet par la machine) produit un grand nombre de lectures (reads) correspondant à des séquences d'ADN ou d'ARN. La capacité de la machine se mesure en nombre total de bases séquencées.

3. Préparation des banques (« library »)

- Fragmentation de l'ADN ou du cDNA et/ou sélection des molécules à séquencer (ARN, ADN immuno-précipités).
- Ligation à bout franc d'adaptateurs spécifiques à chaque technologie pour permettre une amplification par PCR.

4. Amplification clonale *(Diapo 12 si besoin…)*

- 2 technologies se partagent le marché :
  * PCR en émulsion : ligation des fragments d'ADN sur une petite bille qui est émulsionnée dans un système huile/eau et qui va amplifier à l'intérieur de ce système.
  * technique d'immobilisation sur phase solide (développée par illumina) : ligation des adaptateurs, puis bridge amplification (amplification clonale), et ensuite vous avez le système qui va se copier de multiples fois pour amplifier le signal.

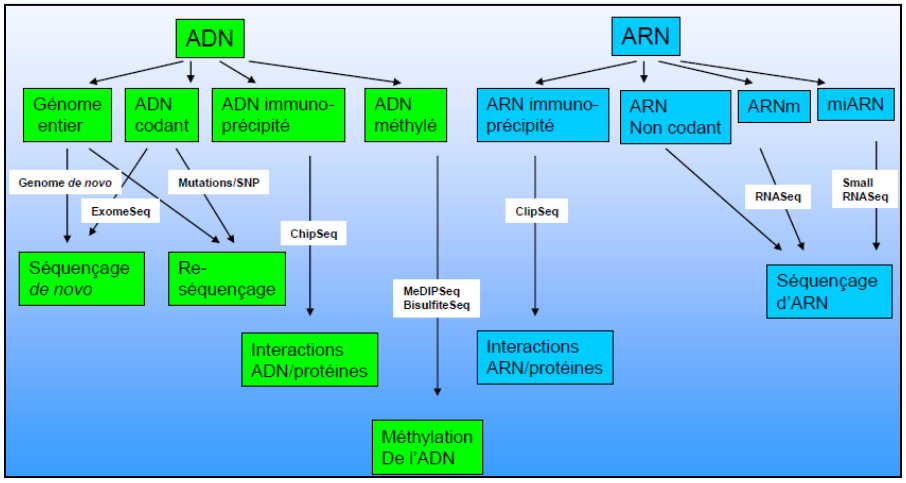
→ Obtention jusqu'à 1000 copies identiques de chaque molécule ("clusters").
→ Jusqu'à 10 millions de clusters par cm².

Cours n°4 – 06/12/12

5. Séquençage et analyse

- Chaque plateforme possède sa propre méthode de séquençage :
  * pyroséquençage - ROCHE.
  * Reverse Dye Terminator (RDT) - ILLUMINA.
  * séquençage par ligation - 5500 series SELiD sequencers (LIFE TECH).
  * séquençage par mesure ionique - Ion Torrent (LIFE TECH)

6. Applications du NGS



- Les principales applications :
  * <u>études génomiques :</u>
    - **DNA-Seq :** à partir de l'ADN, séquençage de fragments d'ADN, comparaison de souches bactériennes ou virales, comparaison de régions ciblées au sein d'une espèce (applications dans le domaine de la génétique constitutionnelle).
    - **DNA Exome-Seq** *(ceci nous intéresse particulièrement, rappel : exons=1% du génome complet mais comprennent 80% des mutations pathogènes chez l'homme)* : séquençage de fragments d'ADN de régions exoniques choisies → détection et identification de SNP (=signal nucléotide polymorphisme, donc mutations ponctuelles de l'ADN).
  * <u>études transcriptomiques :</u>
    - **RNA-Seq :** à partir de l'ARN, séquençage de fragments transcrits (transcriptome), comparaison de l'expression des gènes entre conditions biologiques/échantillons, découverte de nouveaux transcrits, identification de sites d'initiation de la transcription.
  * <u>études épigénétiques :</u>
    - **MeDIP-Seq** (Methylated DNA immunoprecipitation) : séquençage de fragments d'ADN isolés par immunoprécipitation avec un Ac dirigé contre la 5-méthylcytosine permettant d'étudier la méthylation de l'ADN.
    - **CHIP-Seq** (Chromatin Immunoprecipitation) : séquençage de fragments d'ADN en interaction avec une protéine d'intérêt.

7. Avantages du NGS

- Très grande sensibilité de détection (1 copie par cellule).
- Lecture de + en + longue en un temps de + en + court.

ex : en 2006 : 20Mb/run - 100bp/read
en 2012 : 700Mb/run - 700bp/read

- Coût de - en - élevé avec l'apparition de NGS de paillasse.

→ Critères rendant envisageable une application en dg.

8. Limites du NGS

- Quantité de données recueillies (traitement et stockage des données, recours à un bio-informaticien, et le matériel informatique doit aussi évoluer).
- Problèmes au niveau insertions/délétions, ne sont pas forcément détectées par le système.
- Ethique (que faut-il faire de toutes ces données? car il y a bcp de données "à coté" de ce qui nous intéresse).

9. NGS dans le cadre des thérapies ciblées du cancer

- 2006 : création des plateformes de génétiques hospitalières des cancers.
- Mise au point de 14 tests déterminants pour l'accès aux thérapies ciblées.
- Actuellement 1 test=1 technique (c'est assez laborieux). L'objectif serait de remplacer ces 14 tests par un seul test.

10. Exemples *(Diapos 19-20 si besoin…)*

- Adénocarcinome pulmonaire : détection de la mutation EGFR grâce au système Ion Torrent + qlq autres mutations.
- Autre cas d'adénocarcinome pulmonaire : avec mutation EGFR + autres mutations mais qui n'ont pas d'implications au niveau clinique → Donc que faire de ces mutations? Faut-il rendre une carte complète? Que faut-il faire de toutes ces données? → Attente des directives de l'INCA à propos de ces résultats (pbs éthiques).

Question : pas entendu dsl…
Mais réponse : il n'y a pas de thérapeutique pour ttes les mutations malheureusement (thérapeutique existante pour EGFR, KRAS, ...). Ce qui ne remet pas en cause le fait que cette technique ait beaucoup apporté dans le domaine de la génétique constitutionnelle, pour la recherche de facteurs de prédisposition à l'origine de pathologies.

Question : quelle est la différence entre PCR et séquençage haut débit?
Le NGS repose sur une PCR en gros.
PCR : un seul gène candidat encadré par un couple de primers.
NGS : là on a des millions de primers qui amplifient tous en même tps tous les fragments que l'on a, donc on est à un niveau bien supérieur!!

Question : ça représente combien de primers pour le NGS?
Je ne sais pas exactement, mais tout ce fait dans la même réaction donc séquençage en une journée (il faut compter une autre journée pour la préparation des "library").