

Biologie moléculaire et diagnostic de laboratoire :

Les techniques utilisées, leurs limites, leurs mises en œuvre au laboratoire et leurs performances

Les techniques de biologie moléculaire portant sur le génome et le transcriptome ont connu un essor croissant : d'abord réservées aux laboratoires de recherche, elles conquièrent progressivement les laboratoires d'analyse médicale et sont appelées à remplacer les techniques traditionnelles.

La PCR utilisée au départ pour l'établissement de profils génétiques participe maintenant couramment au diagnostic de maladies infectieuses telle que la tuberculose. De nombreuses variantes souvent très sophistiquées de la PCR permettent de réaliser une multitude d'analyses en un temps record.

La technique d'hybridation FISH réalisée sur des coupes de tissus provenant de biopsies ou de pièces opératoires complète les analyses cytogénétiques traditionnelles. Elle contribue au diagnostic des tumeurs et au choix des traitements les plus adaptés pour les patients. La technologie des « puces à ADN », quand à elle, permet d'analyser simultanément des milliers de gènes aboutissant, par exemple, au typage d'une tumeur à partir de sa signature moléculaire.

Les nouvelles méthodes de séquençage à haut débit des acides nucléiques constituent le domaine le plus vaste et de nombreuses technologies concurrentes voient le jour peu à peu. Elles s'inspirent toutes de la technique de Sanger mais les procédés utilisés pour séquencer l'ADN diffèrent que ce soit au niveau de l'appareillage utilisé que du principe même du séquençage: tous les procédés ont en commun leur rapidité d'exécution et leur coût de plus en plus faible . Le séquençage ne concerne pas que le génome humain : par exemple par séquençage on peut identifier par génotypage un microorganisme. Les limites du séquençage concernent maintenant l'utilisation de l'outil informatique avec la quantité colossale de données collectées qui se chiffrent en téraoctets [To] et dont la gestion n'est pas encore bien maîtrisée. Les aspects éthiques de ces nouveaux outils qui font naître une médecine prédictive doivent être envisagés et débattus de manière appropriée.

DIAPOS de 1 à 5

Introduction

La structure en double hélice de l'ADN a été découverte en 1953 par Crick, Watson et Wilkins sans oublier Rosalind Franklin, longtemps ignorée !

Cette découverte a été suivie de la caractérisation, l'isolement et la manipulation des acides nucléiques. Elle a conduit à l'avènement des techniques de biologie moléculaire.

Après du grand public l'étape la plus marquante a été le séquençage du génome humain :

Le projet PGH « projet génome humain » piloté par les « NIH » a été lancé en 1990 avec pour objectif en 2005 le séquençage de 22 chromosomes et des chromosomes X et Y soit 3 milliards de bases (coût prévu de 3 milliards de dollars). Un projet concurrent a été lancé en 1998 dans le privé par la société « CELERA » dirigée par Craig Venter. Les procédés utilisés par le PGH et CELERA sont légèrement différents mais ils aboutiront en 2001 à la publication quasi simultanée de deux séquences « brutes » (incomplètes). En réalité le séquençage expérimental de CELERA n'a été que partiel et Venter a utilisé des séquences publiées au fur et à mesure des résultats par le PGH.

En 2004 a été publiée la séquence complète du génome humain : En réalité le génome publié ne représente pas la séquence exacte d'un génome individuel : il s'agit d'un génome combiné d'une centaine de donneurs anonymes (le séquençage a été réalisé de manière complémentaire dans plusieurs laboratoires européens et américains à partir de différents donneurs).

En 2008 le génome de James Watson (le prix Nobel) est publié un an après celui de Craig Venter (le fondateur de CELERA)...70 génomes individuels seront ainsi obtenus jusqu'en 2012 (dont celui de Steve Jobs pour 100 000 \$)... Depuis de nombreux génomes ont été publiés grâce à la baisse des coûts du séquençage qui est en 2015 de moins de 1000 \$ (750 €). Près de 3000 séquençages ont ainsi été effectués en Islande à l'occasion d'une vaste étude en 2014. En Chine, « Beijing Genomics Institute » prévoit à l'avenir 10 000 séquençages/an.

En parallèle avec les premiers séquençages du génome humain d'autres génomes ont été séquencés dans le cadre du projet PGH: le séquençage complet du génome d'*Escherichia coli* a été finalisé en 1997 et celui de *Saccharomyces cerevisiae* en 1996.

L'intérêt du séquençage est d'identifier les gènes et de les cartographier sur les chromosomes et par la suite d'identifier des mutations génétiques et certaines maladies mais aussi d'évaluer des prédispositions génétiques pour des maladies.

DIAPOS 6 et 7

La biologie moléculaire ne se limite pas au séquençage du génome qu'on abordera à la fin du diaporama. Il existe beaucoup d'autres techniques utilisées en particulier comme outil de diagnostic médical. Toutes les techniques ont schématiquement le même but : identifier et caractériser une des séquences polynucléotidiques au sein d'une molécule d'ADN ou d'ARN.

Les buts recherchés sont ceux évoqués dans les applications du séquençage : par exemple :

- diagnostiquer une maladie génétique en mettant en évidence des mutations dans l'ADN,

- contribuer au diagnostic d'une tumeur,
- évaluer une prédisposition génétique à une maladie au sein d'une famille : hypercholestérolémie familiale (puce ADN),
- pronostiquer l'évolution d'un cancer par analyse du transcriptome des cellules d'une biopsie.

Les techniques de biologie moléculaire sont aussi utilisées pour identifier un agent infectieux dans un produit pathologique ou pour détecter la résistance à un antibiotique d'une bactérie (tuberculose).

Les buts des différentes techniques sont donc souvent similaires et des techniques de principes différents peuvent être concurrentes. Concernant toutes l'étude des acides nucléiques, les techniques de biologie moléculaire utilisent souvent des outils similaires : enzymes, amorces, sondes...

Parmi les techniques de biologie moléculaire d'application courante en recherche médicale et progressivement en analyse médicale les plus importantes sont :

- la PCR «polymerase chain reaction »,
- la technique « FISH » (hybridation in situ en fluorescence),
- les « biochips » ou microarrays (« puces » à ADN)
- le séquençage de l'ADN.

DIAPOS 8 à 10

La PCR

Parmi les premières techniques qui ont révolutionné la biologie moléculaire, la PCR s'est vite implantée dans les laboratoires. La première publication sur la PCR a été faite par Kary Mullis en 1986 (prix Nobel de chimie 1993).

En 1988 a été réalisé la première PCR avec une ADN polymérase thermostable : la Taq polymérase enzyme clé permettant de réaliser rapidement la polymérisation de l'ADN par amplification d'une séquence choisie.

25 ans plus tard la PCR est toujours une technique très utilisée : même si elle a beaucoup évolué et s'est améliorée en particulier par l'apparition de la technique de PCR en temps réel (1992) : qPCR et par des variantes de la PCR tels que la RT-PCR (amplification de l'ARN) ou la PCR Multiplexe (amplification simultanée de plusieurs séquences).

La PCR permet l'amplification d'une séquence d'ADN choisie = obtention de copies multiples de la séquence par des cycles répétés de polymérisation, à l'aide d'amorces (« primers ») spécifiques et d'une ADN polymérase thermorésistante.

Pour justifier les étapes de chaque cycle de PCR il faut rappeler des propriétés caractéristiques de la molécule d'ADN :

- Molécule double brin
- Les deux brins sont anti parallèle : brin « forward » 5'-3' et brin « backward » ou « reverse » 3'-5'
- Les deux brins ont deux séquences complémentaires (complémentarité des bases azotées des désoxyribonucléotides : dNT successifs)

- Par convention un des deux brins est choisi comme « forward » ou « + » : par exemple pour la séquence complète d'un chromosome le brin « forward » débute (extrémité 5') à l'extrémité du bras court p du chromosome .

DIAPOS 11 à 13

Avant de procéder à la PCR on doit réaliser une extraction de l'ADN qui est maintenant faite avec des kits permettant la lyse des cellules et la dénaturation des protéines (système de type « Nucleospin »).

DIAPOS 14 et 15

Chaque cycle de PCR (35 cycles en général) comprend 3 étapes : dénaturation, hybridation et élongation.

- Étape de dénaturation (1)

Pour réaliser l'amplification d'une séquence choisie d'ADN, il faut dénaturer l'ADN de l'échantillon analysé : séparer les deux brins des molécules d'ADN.

Dans chaque molécule d'ADN les deux bras sont liés par des liaisons hydrogène entre les bases azotées A et T d'une part et G et C d'autre part. Pour les dissocier il faut augmenter la température étape de dénaturation de 30 secondes environ à 95°C.

- Polymérisation :

Utilisation d'une ADN polymérase qui permet l'addition des dNT complémentaires de chaque brin: l'ADN polymérase agit dans le sens 5'→3' mais elle ne débute la polymérisation qu'en présence d'une amorce qui fournit une extrémité 3'OH libre à laquelle vont s'ajouter les dNT successifs : pour dupliquer l'ADN il faut une amorce liée à chaque brin :

c'est l'étape d'hybridation (2) qui précède l'étape de polymérisation (élongation) ; l'hybridation nécessite un abaissement de la température à 55 - 65° en fonction de la longueur et de la composition des amorces (interactions G-C plus fortes qui nécessitent un abaissement moins important de la température : G et C s'apparient plus facilement).

Une diminution trop importante de la température risquerait de provoquer des appariements non spécifiques.

Les amorces choisies permettent de borner spécifiquement la séquence que l'on souhaite amplifier : ces amorces « primers » sont constitués d'oligonucléotides (de 20 dNT = 20 « mer » environ : du grec « meros » = parties).

Chaque amorce doit être complémentaire d'une séquence située en amont de celle étudiée ; les deux amorces ne doivent pas s'apparier ensemble : leurs séquences ne doivent pas être complémentaires.

- Étape d'élongation (3)

La Taq-polymérase thermorésistante introduite dès le départ (elle n'est pas dénaturée à 95°) permet de réaliser l'addition successive des dNT à une température de 72°C

Chaque étape d'un cycle de PCR dure entre 30 secondes et une minute.

Les 35 cycles successifs d'une PCR se déroulent dans le même microtube sans addition ultérieure d'autres réactifs : le tube contient l'extrait d'ADN étudié et un « mix » (les 4 dNTP, la Taq-polymerase, les 2 amorces). Les tubes sont introduits dans un thermocycleur qui permet de réaliser des cycles avec des températures et des durées choisies programmées ; la PCR durera au total entre deux et trois heures.

A l'issue de la PCR on obtient un très grand nombre de copies de la séquence initiale que l'on a souhaité amplifier : ces copies sont appelées amplicons.

Les amplicons obtenus à l'issue de la PCR sont analysés par électrophorèse en gel d'agarose : électrophorèse « sous-marine » qui limite évaporation.

DIAPOS 16 à 20

Exemple d'utilisation de la PCR : « Profil ADN » ou « empreintes génétiques »

Identification d'un individu avec des marqueurs discriminants de l'ADN : séquences pour lesquelles il existe de nombreux allèles différents.

Les marqueurs sont choisis dans des régions non codantes de l'ADN pour lesquelles il existe une plus grande variabilité : pour les empreintes génétiques les loci sont choisis au niveau des microsatellites : STR « short tandem repeats » qui sont des séquences caractérisées par la répétition d'un motif de quelques nucléotides (de deux à quatre) ; le nombre de répétitions varie d'un sujet à l'autre : exemple le locus THO1 sur le chromosome 11 : le motifs répété est la séquence TCAT qui peut être répété de trois à 14 fois.

L'amplification des différents allèles conduit à des amplicons de tailles différentes séparés par électrophorèse (la migration se fait du pôle - vers le pôle +). La visualisation des amplicons se fait avec un colorant fluorescent intercalant tel que le BET.

Un seul locus n'est pas assez discriminant pour identifier un individu : en pratique une quinzaine de loci situés sur des chromosomes différents sont étudiés : ils correspondent tous à des STR.

Actuellement la technique d'analyse est accélérée : réalisation d'une PCR multiplexe qui permet d'amplifier simultanément toutes les séquences étudiées.

La séparation des amplicons se fait par ensuite par électrophorèse capillaire ; l'identification des amplicons se fait grâce à leur temps de migration et grâce à un marquage des amorces utilisées lors de l'amplification. Le résultat est obtenu en moins d'une heure (la PCR est plus rapide car elle utilise une polymérase plus active ; les étapes de dénaturation, hybridation, élongation sont raccourcies à moins de 10 secondes : la PCR au total dure 30 minutes)

DIAPOS 21 et 22

Exemple d'utilisation de la PCR : diagnostic de la syphilis par Tp-PCR

Récemment un diagnostic de la syphilis par technique de PCR a été mis au point par le laboratoire du service de dermatologie de l'hôpital Cochin. Le diagnostic direct de la syphilis est délicat car la bactérie (*Treponema pallidum*) n'est pas cultivable : le diagnostic est le plus souvent indirect (sérologie) ce qui ne permet pas toujours de dater l'infection ni le stade de la maladie.

La technique de PCR mise au point est une technique classique de PCR mais réalisée en deux étapes c'est une PCR dite nichée. La séquence amplifiée se situe au niveau du gène tpp 47. Une première amplification est réalisée avec un premier couple d'amorces. Le fragment amplifié à une longueur de 1103 pb. Une deuxième amplification a lieu ensuite à partir des amplicons obtenu : le nouveau fragment amplifié à une taille de 168

pb. L'intérêt de la PCR niché et qu'elle augmente la spécificité du test grâce aux 4 amorces spécifiques utiliser successivement.

Cette méthode a une bonne spécificité mais elle manque de sensibilité. Elle peut remplacer la technique directe de microscopie à fond noir utilisable par des laboratoires spécialisés pour la mise en évidence de *Treponema pallidum* dans un prélèvement à partir d'un chancre ou d'une lésion cutanée mais elle n'est pas applicable à la recherche directe de la bactérie dans le sang (syphilis secondaire).

DIAPOS 23 à 33

Exemple d'utilisation de la PCR : diagnostic de la tuberculose par PCR en temps réel

Technique GeneXpert MTB/RIF (Labo *Cepheid* avec le soutien de NIH « US National Institutes of Health »)

Cette technique permet par PCR de détecter l'ADN du complexe *Mycobacterium tuberculosis* (*M. tuberculosis*, *M. bovis*, *M. africanum*) et la résistance à la rifampicine pour détecter les TB-MR (souches multirésistantes : 150 000 décès/an)

La PCR en temps réel a la même principe général que la PCR classique (en point final). La différence est que l'amplification est mesurée tout au long des cycles de polymérisation (et non détectée après la PCR). La PCR est quantitative. Le thermocycleur utilisé est associé à un fluorimètre qui permet de suivre l'évolution de l'amplification par mesure d'un signal fluorescent.

L'amplification se fait avec les mêmes étapes cyclique que la PCR classique mais des sondes fluorescentes spécifiques sont ajoutées aux réactifs en plus des amorces. À chaque cycle la sonde fluorescente s'hybride en même temps que les amorces avec l'ADN cible ce qui permet l'émission d'un signal de fluorescence. Il existe plusieurs types de sondes ; le système GeneXpert utilise une « balise moléculaire » (« molecular beacon ») : la balise est recyclable: elle possède un groupement suppresseur (« quencher ») et un fluorochrome qui ne fluoresce lors de chaque cycle que lors de l'hybridation avec la séquence cible (l'éloignement du suppresseur provoque la fluorescence).

L'amplification concerne le gène *rpo B* qui code pour la sous-unité β de l'ARN pol qui est la cible de la rifampicine

6 sondes, chacune marquée par un fluorochrome distinct, sont utilisées simultanément : 5 sondes sont spécifiques chacune d'une séquence de la région à amplifier. La sixième sonde permet de contrôler les étapes du processus (elle est spécifique d'un micro-organisme ajouté comme témoin en plus de 2 autres amorces: *Bacillus globigii*). La fluorescence émise augmente d'un cycle à l'autre. On détermine le « CT » « cycle threshold » = seuil à partir duquel la fluorescence mesurée est significative (différente du bruit de fond). Plus le CT est faible (nombre de cycles pour atteindre le seuil faible), plus la séquence était abondante au départ. Le test met ainsi en évidence la charge bactérienne dans l'échantillon analysé.

Dans le test Xpert MTB/RIF, chacune des 5 sondes est spécifique d'une partie distincte du gène *rpoB* : la PCR permet de détecter des mutations du gène responsables de la résistance à la rifampicine.

DIAPOS 34 et 35

En 2010, suite à des recommandations de l'OMS le test Xpert MTB/RIF a été mis en place dans 77 pays dans le cadre du projet EXPAND-TB. En 2015 108 pays sont équipés d'appareils sur 145 éligibles ; plus d'un million de cartouches pour le test ont été utilisées.

Ce test rapide est sensible et spécifique et permet d'adapter le traitement aux malades pour prescrire au plus tôt des médicaments de 2^{ème} intention en cas de résistance : exemples : Amoxicilline, kanamycine ... [4 médicaments associés].

DIAPOS 36 et 37

Variantes techniques de la PCR

- Technique RT- PCR : « *reverse transcription PCR* » utilisée pour amplifier une séquence d'ARN : nécessite une transcription de l'ARN en ADN (transcriptase inverse) avant la PCR : exemple RT-PCR multiplex pour HIV et HCV.

- « LAMP method » (*loop mediated isothermal amplification platform*) : amplification cyclique isotherme (pas de dénaturation nécessaire) mise au point au Japon (utilisation d'une ADN pol à déplacement de brin qui dispense de l'étape de dénaturation de l'ADN)...envisagée pour le diagnostic d'Ebola.

Un réactif fluorescent, la calcéïne permet de révéler l'amplification : sa fluorescence est inhibée par le calcium (= suppresseur) présent dans le milieu.

Le pyrophosphate PPi libéré lors de l'amplification complexe les ions calcium ce qui provoque la fluorescence de la calcéïne : la lecture est visuelle (méthode qualitative)

Testée aussi pour la tuberculose mais pas recommandée par l'OMS (beaucoup de faux positifs). Il s'agit d'une méthode qualitative

DIAPOS 38 à 41

La technique d'hybridation FISH

La technique d'hybridation in situ en fluorescence est une technique de cytogénétique moléculaire.

Elle ne nécessite pas de culture cellulaire (différence par rapport à la caryotypie) et peut être réalisée sur des cellules interphasiques.

Elle utilise des sondes spécifiques de l'ADN chromosomique : oligonucléotides (double brin) marqués par un fluorochrome qui s'hybride avec la séquence étudiée (exemple : FITC)

Réalisable sur des coupes tissulaire (biopsie, pièce opératoire) congelées ou fixées (après déparaffinage et réhydratation).

Protocole

- Perméabilisation: détergent (Triton X 100..) : destruction partielle des membranes.
- Déprotéinisation (ex : pepsine) : élimination partielle des protéines liées à l'ADN des chromosomes pour faciliter l'accès des sondes : 10 minutes à 37°C.
- Dénaturation de l'ADN six et de la tienne de la sonde : cinq minutes à 80°.
- Hybridation : baisse progressive de la température : 15 à 20 heures à 45°C (la température dépend de la sonde : cf Tm).
- Lavage pour éliminer l'excès de sonde et la sonde hybridée non spécifiquement (hétéroduplex) : augmentation de la température et baisse de la concentration saline

(exemple : deux minutes à 70° en solution SSC « saline sodium citrate ») : augmentation de la « stringence » (spécificité).

La lecture se fait au microscope UV équipé pour la fluorescence.

DIAPOS 42 et 43

Application de la technique FISH - Mise en évidence de l'amplification d'un gène

Détection de l'augmentation du nombre de copies d'un gène dans la cellule.

Exemple : cancer du sein : le gène HER-2 « human epidermal growth factor receptor 2 ». Avec deux sondes fluorescentes: la sonde CEN 17 permet le repérage des chromosomes 17 et la sonde HER2 est spécifique de la séquence du gène HER2 : une augmentation du nombre de copies du gène HER2 se traduit par l'apparition d'un excès de taches fluorescentes révélées par la sonde HER2 par rapport à celles révélées par la sonde CEN 17.

Remarque : technique CISH « *chromogenic in situ hybridation* » utilisant la peroxydase.

DIAPOS 44 à 46

Application de la technique FISH - Mise en évidence d'une délétion chromosomique

Exemple : leucémie lymphoïde chronique (LLC) :

- Délétion sur le bras long q du chromosome 11 avec deux sondes fluorescentes: la sonde CEP 11 permet le repérage des chromosomes 11 et la sonde ATM est spécifique de la séquence perdue. Une délétion sur un des chromosomes homologues est visualisée par la disparition d'une des taches révélées par la sonde ATM ;
- Étude similaire de délétion au niveau du bras court p du chromosome 17 : avec la sonde CEP 17 repérant les chromosomes 17 et la sonde p53 repérant la séquence perdue

DIAPOS 47 et 48

Application de la technique FISH - Mise en évidence d'une translocation

Exemple caractérisation du lymphome de Burkitt avec deux sondes de part et d'autres du point de coupure (situé au niveau d'un intron dans le gène MYC) : translocation t (8 ; 14) : ces 2 sondes « de cassure » permettent de mettre en évidence la cassure du gène MYC due à la translocation de l'extrémité du bras long q d'un chromosome 8 sur un autre chromosome (14), la fluorescence des 2 sondes n'étant pas contigüe

DIAPOS 49

Remarque : Technique « SKY » « spectral Karyotype »

Dans cette technique récente, différentes sondes (spécifiques chacune d'un seul chromosome entier, et marquées avec des proportions variables de cinq fluorochromes différents) sont hybridées avec les chromosomes. On obtient ainsi pour chaque paire de chromosomes des couleurs différentes en raison des proportions variables de fluorochrome sur chaque chromosome.

Les chromosomes peuvent être automatiquement identifiés en microscopie de fluorescence couplée avec un interféromètre [imagerie spectrale].

DIAPO 50

Les puces à ADN

Les puces à ADN ou biopuces (« DNA chips, DNA-microarrays, biochips ») sont constituées de fragments d'ADN immobilisés sur un support solide selon une disposition ordonnée.

Leur principe est celui de l'hybridation inverse. Des sondes non marquées sont greffées sur un support miniaturisé, puis les séquences cibles étudiées marquées viennent s'hybrider sur les sondes de la puce.

DIAPOS 51 à 56

Les puces à ADN - Signatures génétiques

L'un des buts majeurs des biopuces est l'étude du transcriptome = ensemble des ARNm de la cellule à un instant donné. L'intérêt des puces à ADN est d'observer l'expression de plusieurs milliers de gènes sur un seul support et en une étape.

Les ARNm de 2 types de cellules sont extraits, purifiés et rétrotranscrits en ADNc qui sont marqués au Cy3-dUTP (vert) pour le premier type et au Cy5-dUTP (rouge) pour le deuxième type

Ces cibles marquées au Cy3 et Cy5 sont cohybridées sur la puce recouverte de séquences des gènes étudiés. Cette hybridation est compétitive. L'intensité de la fluorescence verte ou rouge traduit les proportions des ADNc des 2 types cellulaires étudiés. L'acquisition de l'image se fait grâce à un laser deux couleurs ce qui permet une détection différentielle de la fluorescence des 2 types de cibles.

Les puces à ADN permettent donc de comparer l'expression des gènes de deux types cellulaires différents par exemple de comparer tissus sains contre tissus malades. Cette méthode est utilisée pour donner une signature génétique à différents types de cancer : exemple classification des signatures génétiques des cancers du sein (à comparer à l'étude histochimique des tumeurs) qui permet de prévoir l'efficacité prévisible d'un traitement.

DIAPOS 57 à 61

Les puces à ADN – Polymorphisme génétique

Les SNPs (Single Nucléotide Polymorphism) représentent la variation d'un seul nucléotide dans une séquence d'ADN. Ce sont les variations les plus fréquentes dans l'ADN : elles ont été mises en évidence initialement par la comparaison des trois premiers génomes séquencés : ceux de Watson et de Venter et celui d'un sujet anonyme de nationalité chinoise.

50 millions de SNPs sont été identifiés depuis dans le génome humain. Certains SNP sont liés avec certains allèles pathologiques et peuvent ainsi servir à la cartographie des

loci morbides. De nombreux allèles de risque ont ainsi été localisés caractérisés en tant que gènes de prédisposition à différents cancers

Les SNPs peuvent être également mis à profit dans l'étude de la perte d'hétérozygotie fréquente en oncogenèse : conversion de l'allèle sain en allèle malade chez une personne née hétérozygote pour ce gène. L'individu se retrouve alors homozygote pour la version oncogénique du gène.

DIAPO 62

Le séquençage de l'ADN

DIAPOS 63 et 64

Banque d'ADN génomique

Construction d'une banque d'ADN génomique : l'ADN génomique étudié est fragmenté par des enzymes de restriction et les fragments obtenus sont insérés dans un vecteur tel que les BAC « bacterial artificial chromosomes » ; chaque BAC peut accepter un insert de 150 000 paires de bases. Les BAC sont ensuite intégrés dans des bactéries (*E.coli*) par électroporation. Les bactéries transformées sont isolées puis clonées. La couverture du génome doit être de l'ordre de 5-10 X : 5 à 10 fois la taille du génome étudié [génome humain : $3,2 \cdot 10^9$ pb] est clonée à partir des inserts d'ADN recombinants. Chaque insert cloné sera ensuite à son tour fragmenté et les fragments obtenus séquencés.

DIAPOS 65 à 68

Séquençage par la méthode de Sanger – Électrophorèse capillaire

Séquençage de l'ADN par la méthode de Sanger (1964) : la méthode utilisée est le « Di-deoxy DNA sequencing » ; le brin du fragment à séquencer (matrice) est copié par une ADN polymérase : la polymérisation nécessite une amorce (« primer ») et des 4 dNTPs (désoxyribonucléotides triphosphates) : dATP, dGTP, dCTP et dTTP. La polymérisation est réalisée en parallèle dans 4 tubes : dans chacun des tubes est ajouté un ddNTP (didésoxyribonucléotide triphosphate) : ddATP dans un tube, ddTTP dans le deuxième, ddCTP dans le troisième et ddGTP dans le quatrième : ces ddNTP servent de terminateurs de chaîne : quand l'ADN pol insère un ddNTP la polymérisation s'arrête : on obtient ainsi dans chaque tube des fragments de tailles variables qui se terminent tous par A dans le premier tube, par T dans le deuxième etc... Ces fragments sont séparés en fonction de leur taille par électrophorèse en gel d'agarose. Cette méthode permet ainsi de retrouver la séquence du brin synthétisé.

Depuis 1999 la méthode a été simplifiée en marquant chaque ddNTP par un fluorochrome distinct ce qui permet de réaliser la polymérisation à partir d'un seul tube ; la séparation se fait par électrophorèse capillaire et les différents brins synthétisés sont détectés par un capteur CCD au fur et à mesure de leur migration dans le capillaire. Une centaine de fragments peuvent ainsi être séquencés en parallèle.

DIAPOS 68 et 69

Alignement des séquences

On réalise un assemblage par alignement des séquences des fragments obtenus de tailles différentes (technique du « shotgun »). La comparaison des séquences de différents fragments permet d'aligner les parties qui se chevauchent partiellement. L'assemblage des séquences chevauchantes permet de reconstituer par étapes successives des séquences de plus en plus grandes qui aboutissent au séquençage complet de chaque chromosome : des programmes bioinformatiques permettent la réalisation de cet assemblage.

DIAPOS 70 et 71

Historique des techniques de séquençage

Le séquençage par électrophorèse capillaire a permis d'accélérer et d'automatiser le procédé dans les années 2000. À partir de 2006 sont apparues les méthodes de séquençage dites de « nouvelle génération » NGS qui permettent de séquencer un génome entier beaucoup plus rapidement. Depuis 2014 de nouvelles techniques encore plus rapides sont utilisées : il s'agit des méthodes de séquençage de troisième génération baptisées NNGS « next next generation sequencing ».

L'évolution des techniques a réduit considérablement le coût du séquençage qui est passé pour le génome humain de 3 milliards de \$ (PGH) (le coût actuel de 30 avions Rafale !) à 100 millions de \$ avec l'avènement de la méthode capillaire en 2001, à moins de 10 millions de \$ lors de l'apparition des NGS et ne cesse de baisser depuis. En 2015 un séquençage quasi-complet du génome humain par une méthode de type NNGS est réalisable pour moins de 1000 \$. Parallèlement la durée du séquençage s'est considérablement réduite : de plusieurs années initialement elle est passée à quelques jours seulement ... voire quelques heures à l'avenir.

DIAPOS 72 et 73

Avantages et inconvénients du séquençage par NGS

Le séquençage par NGS (technologies *Illumina* et *SoLID* les plus performantes) permet de séquencer simultanément 100 millions de fragments au lieu de 100 par la méthode de Sanger (technologie ABI) : 1000 milliards de pb sont ainsi lus à chaque run au lieu de 100 000. Mais la qualité des résultats du séquençage a par contre baissé (plus d'erreurs) avec l'avènement des nouvelles techniques de NGS : en effet la taille des fragments séquencés est devenue beaucoup plus courte : de 1000 pb elle est passée à moins de 100 pb : le nombre de séquences à aligner est donc devenu énorme ce qui alourdit considérablement le traitement des données informatiques : de 1 Mo on est passé avec les NGS à plus de 1 To de données.

DIAPPO 74

Clonage de l'ADN *in vitro* et *in vivo*

Quelle que soit la méthode (Sanger ou NGS) le clonage permet d'amplifier les différents fragments du génome (obtenus après l'action de Dnase, nébulisation ou sonication). Dans la méthode de Sanger le clonage est réalisé *in vivo* dans des bactéries grâce à un vecteur (BAC) dans lequel est inséré le fragment qui sera séquencé. Dans les NGS, le clonage et l'amplification des fragments d'ADN est réalisé *in vitro* par PCR.

DIAPO 75

Les 2 technologies les plus utilisées parmi le NGS sont la technologie 454 commercialisée par *Roche* et la technologie *Solexa* commercialisée par *Illumina*.

Dans les deux techniques le clonage est réalisé en utilisant des « adaptateurs » universels qui sont liés (sous l'action de ligases) aux fragments à séquencer. Ces adaptateurs sont des oligonucléotides complémentaires d'amorces fixées sur un support solide: nanobilles dans la technique 454 et nano-carreaux d'une lame dans la technique *Solexa/Illumina*. Le séquençage des fragments ainsi clonés est réalisé ensuite à partir des amorces par polymérisation des dNTP complémentaires du fragment.

DIAPOS 76 à 81

La technologie de séquençage 454 (*Roche*)

La technologie 454 utilise, lors de l'étape de clonage par PCR des fragments, des billes, toutes porteuses de plusieurs milliers d'exemplaires de la même amorce complémentaire de l'adaptateur ajouté à l'extrémité 3' des fragments d'ADN; les proportions de nanobilles et de fragments sont optimisées pour que statistiquement une seule molécule d'un seul fragment se lie à chaque bille de façon à ce que chaque bille puisse donner naissance lors de la PCR à un seul clone.

Chaque bille en émulsion constitue un « microréacteur » : en présence d'ADN polymérase, de dNTP elle est le support de la PCR qui de manière cyclique permet d'obtenir sur chaque bille des milliers de copies du fragment d'ADN à séquencer.

Les billes porteuses chacune d'un clone d'un fragment d'ADN simple brin sont introduites ensuite dans les nanopuits d'une microplaque : la taille des puits est telle que lorsque l'émulsion est introduite dans la microplaque, une seule bille (= un seul clone) sera présent dans chaque puits.

Le séquençage de chaque fragment ainsi cloné est ensuite réalisé en parallèle dans les différents puits par la technique de pyroséquençage.

Le pyroséquençage est une polymérisation réalisée de manière cyclique mettant en jeu des réactifs particulier : à chaque cycle de pyroséquençage, les 4 dNTP non marqués différents sont ajoutés et ainsi testés l'un après l'autre : lorsqu'un dNT est complémentaire de celui du brin matrice il est inséré dans le brin en cours de synthèse par l'ADN polymérase ce qui s'accompagne de la libération de pyrophosphate PPi. Le pyrophosphate réagit avec l'APS (adénosine-5'-phosphosulfate) du milieu sous l'action de l'ATP sulfurylase entraînant la libération d'ATP qui réagit à son tour avec la luciférine du milieu sous l'action d'une troisième enzyme : la luciférase, ce qui conduit à une émission de lumière (bioluminescence). Le signal lumineux est converti en un signal électrique mesuré par une caméra CCD « *charge coupled device* » couplée à un photomultiplicateur : le signal est enregistré sous forme d'un pic sur le tracé de l'enregistrement appelé

« pyrogramme ». La hauteur du pic est doublée si la même base est présente 2 fois de suite : 2 intégrations successives du même dNT donc doublement de la quantité d'ATP produite lors du cycle. A la fin de chaque cycle les dNTP et l'ATP non utilisés sont dégradés par une apyrase.

DIAPOS 82 à 88

La technologie de séquençage *Solexa/Illumina*

La technologie *Illumina* utilise, lors de l'étape de clonage par PCR des fragments, des nano-carreaux d'une lame (cellule de séquençage) , tous porteurs de plusieurs milliers d'exemplaires de 2 amorces distinctes complémentaires des 2 adaptateurs ajoutés à l'extrémité 3' de chaque brin des fragments d'ADN à séquencer. L'amplification (clonage) se fait grâce à la réalisation de ponts « bridges » à la fin de chaque cycle ce qui permet l'extension de la polymérisation de proche en proche autour de chaque fragment initialement fixé : 20 à 30 000 clones différents sont ainsi créés sur chaque carreau.

Le séquençage de chaque fragment cloné est réalisé par la technologie des « terminateurs fluorescents réversibles » : lors de chaque cycle les 4dNTP sont ajoutés simultanément : ils sont chacun marqués par un fluorochrome distinct et leur groupement 3'OH est bloqué (ce qui empêche la fixation d'un autre dNT) : le dNT complémentaire du brin à séquencer est incorporé dans le brin synthétisé et il émet une radiation fluorescente qui est captée par une caméra CCD ceci conditionne à l'apparition d'un spot lumineux localisé au niveau du carreau portant le clone du fragment en cours de séquençage. A la fin du cycle le fluorochrome lié au dNT inséré est éliminé par clivage ainsi que le réactif de blocage du 3'OH : un nouveau cycle peut commencer...et ainsi de suite.

La technologie *Illumina* permet de séquencer simultanément dans une seule cellule de séquençage jusqu'à 100 millions de fragments différents.

DIAPOS 89 et 90

La technologie de séquençage *Ion Torrent PGM (Life Technology)*

Le séquençage se fait après clonage des fragments étudiés sur des nanobilles, comme dans la technologie 454 de *Roche*. Les billes sont introduites dans les puits d'une plaque et le séquençage est réalisé également comme pour la technologie 454 par addition itérative à chaque cycle des 4 dNTP non marqués. L'incorporation d'un dNT dans le brin synthétisé est détectée par potentiométrie : la libération d'un proton à chaque insertion d'un dNT provoque une ddp qui est mesurée.

DIAPOS 91

Les nanotechnologies de séquençage: NNGS

Ces nouvelles technologies miniaturisées apparues il y a quelques années sont appelées à révolutionner les techniques de séquençage en proposant un séquençage « low cost » réalisables en un temps record pour moins de 1000 \$. Leur particularité est qu'elles ne nécessitent pas de clonage préalable puisqu'une molécule unique d'ADN est séquencée entièrement lors du « run ». La fiabilité de ces technologies reste cependant aléatoire avec des erreurs plus nombreuses.

DIAPOS 92 à 94

La technologie des nanopores « *MiniON – Oxford Nanopore* »

Cette technologie de séquençage de troisième génération apparue en 2014 est la plus populaire parmi les NNGS : son apparition a été accompagnée d'une vaste campagne de publicité vantant un séquenceur de la taille d'un smartphone et fiable directement à un ordinateur par une clé USB.

Le séquençage par nanopore utilise une protéine transmembranaire, l'alpha-hémolysine, qui possède un canal interne de 1 nm environ de diamètre. Le nanopore est inséré dans une membrane artificielle à l'intérieur du séquenceur *MiniON* ; un flux d'ions est créé lors du séquençage à travers la membrane ce qui crée un courant électrique transmembranaire. Lorsque un brin de la molécule d'ADN à séquencer pénètre dans le nanopore, le flux d'ions est modifié et l'intensité du courant à travers le nanopore est diminuée différemment selon la nature du dNT du brin qui traverse le canal. Le passage successifs des dNT constituant le brin provoque ainsi des impulsions électriques qui sont enregistrées les unes après les autres et permettent de « lire » la séquence de la molécule.

DIAPO 95

Applications des méthodes de séquençage de 2^{ème} et 3^{ème} génération

Le séquençage ne sert pas qu'à déterminer la séquence d'un génome entier. Il a de nombreuses applications qui nécessitent des séquençages plus restreints : séquençage de l'ADN codant seulement (« ExomeSeq »), séquençage de l'ADN méthylé, séquençage de l'ADN ou de l'ARN immunoprécipités (« ChIPSeq » et « ClipSeq ») pour l'étude des interactions entre ADN ou ARN et protéines, séquençage des ARNm, séquençage des ARN non codants, séquençage des miARN

DIAPOS 96 à 98

Séquençage de génomes bactériens et typage bactérien

Le typage bactérien par séquençage multilocus « *Multilocus Sequence Typing* » proposé au départ vers 1998, est devenue une méthode standardisée: il s'agit d'examiner les séquences nucléotidiques de plusieurs locus codant des gènes de ménage « *housekeeping genes* » codant pour des protéines dont la fonction est essentielle à la viabilité de la cellule. Les données MLST permettent de typer chaque souche bactérienne en lui attribuant un profil allélique issu du séquençage des gènes de ménages choisis.

Le typage bactérien peut aussi être réalisé par séquençage de l'ADNr 16S, codant pour l'ARN 16 S ribosomal : l'ARN 16 S est très conservé dans l'évolution avec des domaines très peu modifiés mais il existe des variations alléliques mineures même entre bactéries très proches qui permettent de les différencier.

Une entérobactérie inconnue jusqu'à présent a ainsi été caractérisée en 2015 comme l'agent pathogène responsable du décès de 3 nourrissons en janvier 2014 à l'hôpital de Chambéry par suite d'une contamination de poches de nutrition.

Cette entérobactérie a été baptisée *Rouxiella chamberiensis*

Les caractéristiques de la souche bactérienne isolée ont été déterminées, en particulier, par l'étude d'un gène codant pour l'ARN 16S et par séquençage multi loci (MLST) ; ces études montrent que cette bactérie est proche sur le plan génotypique des genres *Ewingella*, *Rahnella*, *Yersinia*, *Hafnia* et *Serratia*.

Le nouveau taxon est proposé comme une nouvelle espèce d'un nouveau genre *Rouxiella chamberiensis*.

DIAPOS 99 à 102

Médecine et utilisation de NGS – Rapports de l'Académie de Médecine

« L'évolution technique et technologique dans le cas du séquençage des ADN modifie les concepts et les conditions de l'exercice médical, notamment en cancérologie et en génétique, et soulevant ainsi un grand nombre de questions réglementaires et éthiques.... »

« L'apparition et la diffusion dans les laboratoires hospitaliers des nouveaux appareillages de séquençage à haut débit conduisent à prendre en compte, au cours d'une démarche médicale, l'ensemble du génome ou de ses principaux composants ; de ces moyens analytiques puissants et performants découle le concept de médecine dite « personnalisée » c'est-à-dire qu'un tableau médical peut être conditionné ou modulé par les caractéristiques génétiques individuelles constitutionnelles ou somatiques..... »

Les machines haut de gamme (500 000 euros) permettent de séquencer en 10 jours un génome humain complet mais des appareils de « pailleasse » (50 à 100 000 euros) sont mieux adaptés à une utilisation clinique.

En pratique médicale les analyses peuvent être un séquençage ciblé sur un groupe de gènes (de 50b à 1Mb), un séquençage de l'exome (de 25 à 50Mb) plutôt qu'un séquençage de l'ensemble du génome (6000 Mb).

DIAPOS 103 à 106

Médecine et utilisation de NGS – Applications, limites et problèmes éthiques

Parmi les applications des NGS un dépistage néonatal génotypique de plusieurs centaines de maladies dont les gènes et les mutations pathogènes sont connus est possible pour moins de 300 euros. Les NGS rendent aussi possible le diagnostic préconceptionnel généralisé.

La sensibilité des NGS permet aussi l'analyse de l'ADN fœtal circulant dans le sang maternel.

En oncogénétique constitutionnelle, les NGS réalisent l'analyse simultanée des gènes impliqués dans les formes « héréditaires » de cancers (21 gènes des cancers du sein, 6 gènes des cancers du colon...) Cette analyse peut aussi être envisagée, par exemple, pour le diabète de type 2. Dans les 2 cas, elles peuvent contribuer au choix des traitements les mieux adaptés

Des firmes privées commercialisent des DTC (« Direct Test to Consume ») : «Amby Screen » propose le dépistage de 75 maladies à transmission récessive ou liée à l'X ; «23 and me» (USA) et « deCODEgenetics » (Islande), des tests de génotypage de mutations

responsables de maladies monogéniques, la recherche de polymorphismes associés à des maladies plurifactorielles et un profil pharmacogénétique, pour 99 dollars. En janvier 2015, La FDA a autorisé la mise sur le marché de plusieurs tests de « 23 and me ». « Color Genomics » réalise maintenant un test (effectué sur un prélèvement de salive) permettant le séquençage de 19 gènes (BRCA1 et BRCA2....etc..) liés à une prédisposition aux cancers du sein et de l'ovaire pour 250 \$. (L'actrice Angelina Jolie avait demandé une double mastectomie puis une ovariectomie suite à une mutation détectée du gène BRCA1).

La diapositive 106 illustre les résultats d'un pyroséquençage pour 2 patientes sur le gène BRCA1 qui est un des gènes de prédisposition au cancer du sein : le pyrogramme de gauche montre l'absence d'allèle de prédisposition pour la séquence SNP étudiée: génotype AA/AA ; le pyrogramme de droite montre un allèle de prédisposition : disparition d'un nucléotide [A] par délétion sur un des chromosomes homologues (lors du pyroséquençage la hauteur du pic est diminuée d'un quart : 3 dATP lus au lieu de 4 lors du cycle) : le génotype est AA/A - .

La société Ecoli (laboratoire slovaque) propose un catalogue avec de nombreux kits de dépistage de SNPs de prédisposition à diverses maladies (diapositive 107) Mais cette médecine « prédictive » est souvent à remettre en question car les analyses génétiques ne sont pas forcément toujours dignes de confiance : si les gènes de prédisposition au cancer du sein sont avérés, il n'en est pas de même pour beaucoup d'autres : certains allèles de prédisposition ont été déterminés à partir d'une analyse de populations très réduite : moins de 1000 personnes parfois ! Le gène RUNX3 portant un SNP baptisé rs760805 considéré comme un marqueur de prédisposition au cancer de la vessie (diapositive 61) à partir d'une étude réalisée seulement sur une population de 368 patients chinois

DIAPOS 108 à 110

Il n'y a pas que les techniques de Biologie moléculaire.....

La spectrométrie de masse devient aussi une technique moderne d'identification bactérienne basée sur une analyse des protéines qui se généralise peu à peu dans les laboratoires.

L'excellent article de Jean-François Trucchi paru dans l'Opéron en fait une très bonne synthèse.

<http://www.upbm.org/index.php/publications/revue-l-operon/index-des-numeros-parus/551-operon-numero-72-2014-4?highlight=WyJOcnVjY2hpllo=>

Documentation

1- Découverte de l'ADN

<http://www.ba-education.com/for/science/dnadiscovery.html>

2- Le projet du génome humain

http://en.wikipedia.org/wiki/Human_Genome_Project

<http://www.sciencemag.org/content/291/5507/1304.full>

<http://www.nature.com/nature/journal/v409/n6822/full/409747a0.html>

3- Extraction ADN sur colonnes

<http://www.mn-net.com/tabid/10745/default.aspx>

4- La P.C.R.

<http://www.ens-lyon.fr/RELIE/PCR/principe/anim/presentation.htm>

5- P.C.R. multiplex : Profils ADN

<http://www.sciencedirect.com/science/article/pii/S1687157X12000194>

http://en.wikipedia.org/wiki/Variable_number_tandem_repeat

<http://www.sciencedirect.com/science/article/pii/S1687157X12000194>

6- PCR « nichée »: diagnostic de la syphilis

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3295187/>

7- PCR en temps réel : diagnostic tuberculose et résistance à la rifampicine (test Xpert-MTB/RIF et TB-LAMP)

<http://www.cephheid.com/manageddownloads/xpert-mtb-rif-english-package-insert-301-1404-rev-b-february-2015.pdf>

http://www.molecular-tb.org/gb/pdf/ppt/12_SYMP_CPAudigier_GeneXpert_2802.pdf

http://www.who.int/tb/challenges/hiv/xpert_mtbrif_ds.pdf

http://apps.who.int/iris/bitstream/10665/112469/1/9789241506700_eng.pdf

http://apps.who.int/iris/bitstream/10665/83142/1/WHO_HTM_TB_2013.05_eng.pdf

8- RT-PCR multiplex HIV-HCV

http://en.wikipedia.org/wiki/Reverse_transcription_polymerase_chain_reaction

http://ijbiotech.neoscriber.com/?page=article&article_id=10717

9- Technologie F.I.S.H.

<http://www.afiap.fr/uploads/File/DES%20PARIS/patho%20moleculaire/FISH-CISH-DES-C%20Copie-2011.pdf>

10- Les puces ADN

http://en.wikipedia.org/wiki/DNA_microarray

<http://labmed.ascpjournals.org/content/41/6/364.full.pdf+html>

<http://www.math-info.univ-paris5.fr/~rozen/Analyse-Genome-Tumoral/Exposes%202011/Letourneur.pdf>

http://www.eupedia.com/genetics/cancer_related_snp.shtml

11- Les N.G.S

<http://www.math-info.univ-paris5.fr/~rozen/Analyse-Genome-Tumoral/Exposes%202014/Christine-BOLE-FEYSOT.pdf>

http://users.ugent.be/~avierstr/nextgen/Next_generation_sequencing_web.pdf

<https://s3.amazonaws.com/piazza-resources/i5efj1fiu732d1/i5gx2f4sxtx5g0/Metzker2010NatRevGen.pdf?AWSAccessKeyId=AKIAJKOQYKAYOBKKVTKQ&Expires=1430670481&Signature=zYBbppkgFRPO6noYmm4AeqeM1Xc%3D>

12- Technologie des nanopores

<https://nanoporetech.com/technology/introduction-to-nanopore-sensing/introduction-to-nanopore-sensing>

13- RNA 16 S et MLST

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4371336/>

<http://www.nature.com/nrmicro/journal/v12/n9/full/nrmicro3330.html>

<http://saureus.mlst.net>

<http://ijs.sgmjournals.org/content/early/2015/03/06/ijs.0.000179.abstract>

14- Rapports de l'Académie de médecine : tests génétiques et NGS : perspectives et limites

<http://www.academie-medecine.fr/publication100036098/>

<http://www.academie-medecine.fr/publication100036433/>

<http://www.academie-medecine.fr/publication100100244/>
15- Kits génomiques et DTC
http://www.pcrdiagnostics.eu/files/documents/pyro/brca-screen_la_150811%20en.pdf
<https://www.23andme.com/en-gb/>
<https://getcolor.com/#/>