

CHAPITRE IV

TECHNOLOGIES DE SÉQUENÇAGE D'ADN

Le séquençage de l'ADN constitue une méthode dont le but est de déterminer la succession linéaire des bases A, C, G et T prenant part à la structure de l'ADN. La lecture de cette séquence permet d'étudier l'information biologique contenue par celle-ci. Étant donné l'unicité et la spécificité de la structure de l'ADN chez chaque individu, la séquence de l'ADN permet de nombreuses applications dans le domaine de la médecine, comme, par exemple, le diagnostic, les études génétiques, l'étude de paternité, la criminologie, la compréhension de mécanismes physiopathologiques, la synthèse de médicaments, les enquêtes épidémiologiques.

1. Première génération de séquençage

1.1. Méthode de Sanger

La méthode de séquençage de Sanger (dite par terminaison de chaîne) utilise des nucléotides appelés didésoxyribonucléotides (ddNTP) qui ont un atome d'hydrogène à la place du groupement OH sur le carbone 3' du ribose.

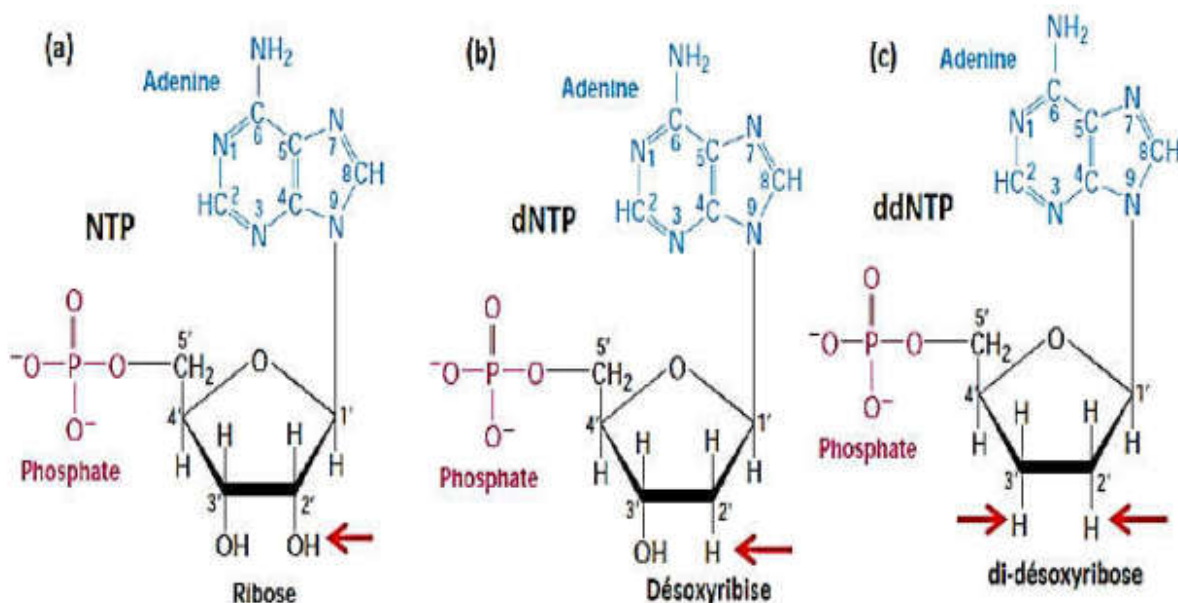


Figure 1 : Différentes formes des nucléotides.

Dans un premier temps, il est nécessaire d'amplifier l'ADN cible par PCR, puis de le dénaturer afin d'obtenir un ADN simple brin. À l'aide d'une amorce spécifique et

complémentaire du brin étudié (sens ou antisens), identique ou différente de celle utilisée pour la PCR, une ADN polymérase effectue alors la synthèse de l'ADN complémentaire à partir de cette amorce. De l'extrémité 5' vers l'extrémité 3', cette enzyme ajoute les désoxyribonucléotides-triphosphates (dNTP) complémentaires et de manière aléatoire et inconstante des didéoxyribonucléotides triphosphates (ddNTP), par exemple un ddGTP sera parfois ajouté à la place d'un dGTP. La réaction se faisant dans un seul tube, les ddNTP (ddATP, ddGTP, ddCTP et ddTTP) sont marqués à l'aide de fluorophores différents pour chaque ddNTP (fluorophores "quatre couleurs").

Lorsqu'un ddNTP est incorporé à la place d'un dNTP, l'ADN polymérase ne peut plus continuer sa polymérisation. Il se forme donc une multitude de brins complémentaires inachevés, chacun terminé par un didéoxyribonucléotide fluorescent caractéristique de la base de ce dernier nucléotide. En séparant par électrophorèse ces fragments complémentaires on sépare dans le gel chacun des fragments, du plus petit au plus grand et on les détecte au passage par un faisceau laser qui excite la fluorescence et une cellule photoélectrique qui lit la lumière émise à chacune des longueurs d'onde des fluorescences caractéristiques des quatre bases (Figure 2). L'ordinateur reçoit donc une série de mesure d'intensité lumineuse en forme de pics correspondant au passage de chacun des fragments : en interprétant la couleur de la fluorescence de chaque pic l'ordinateur écrit la séquence d'ADN.

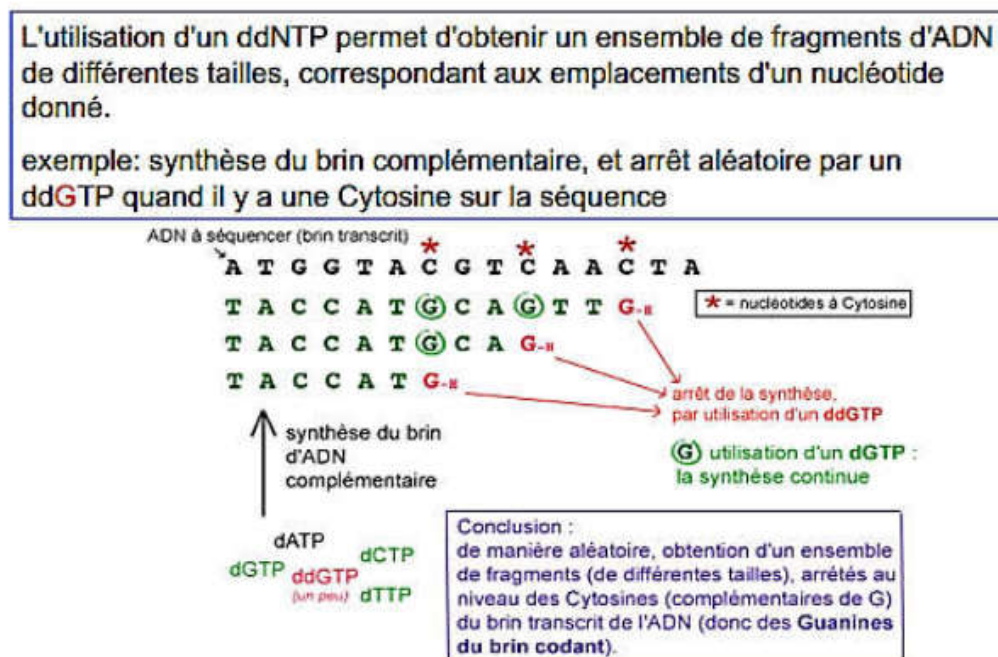


Figure 2 : Principe de Sanger.

4 mélanges sont préparés:

- le fragment qui doit être séquencé cloné dans un vecteur
- un petit morceau d'ADN dont la séquence est complémentaire à l'extrémité 3' du fragment à séquencer = **amorce**
- les 4 dNTP's (dCTP, dATP, dGTP, dTTP)
(Un des nucléotides est marqué au ^{35}S)
- l'ADN polymérase

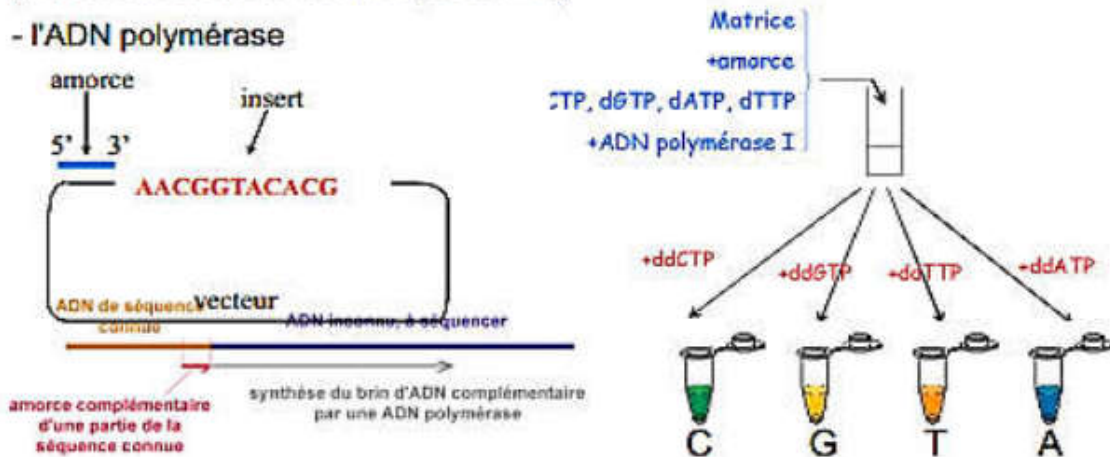


Figure 3: Mélange réactionnel de Sanger.

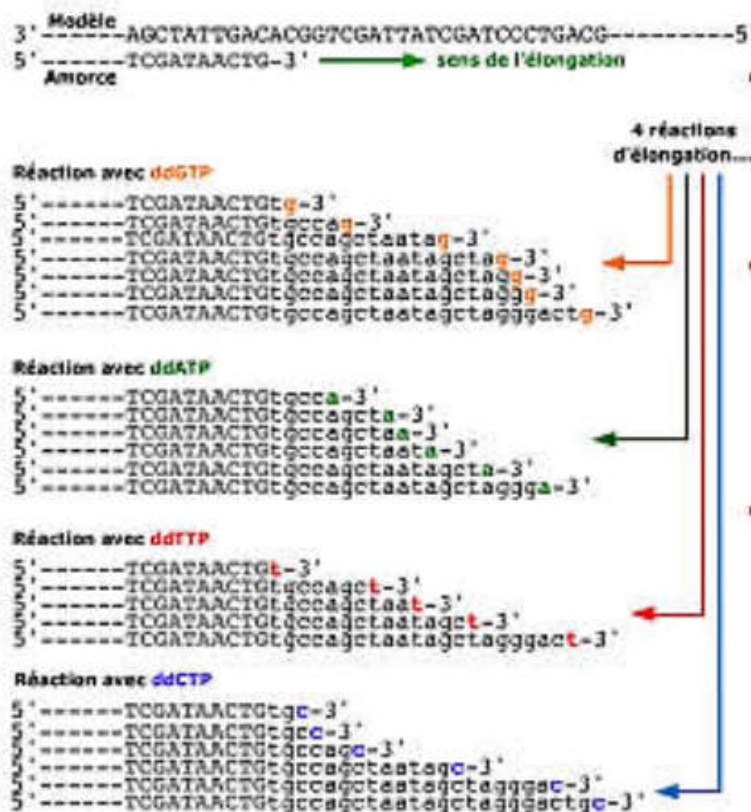


Figure 4 : Ensemble des fragments obtenus en fin de séquençage.

- Électrophorèse sur gel d'acrylamide
- Un des dNTP est marqué au ^{35}S , rendant le fragment radioactif
- Détection des fragments d'ADN, en exposant un film photographique au gel.
- Lecture de la séquence

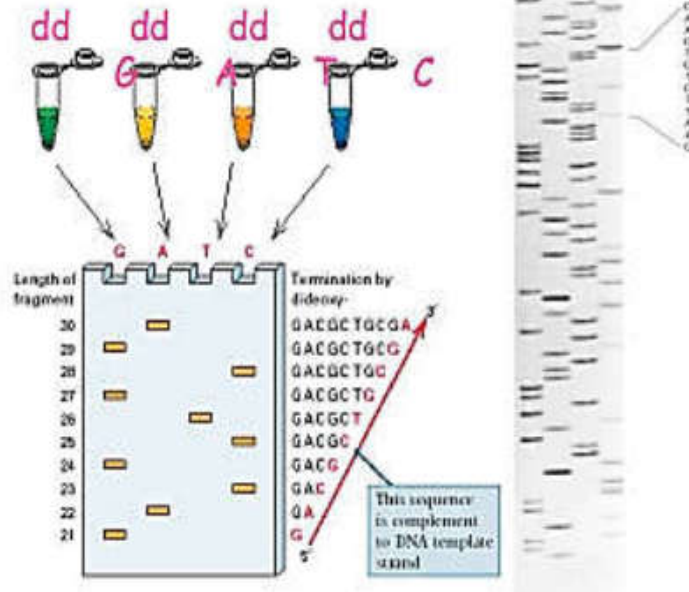


Figure 5 : Migration et visualisation des produits du séquençage.

L'Evolution de la Technique de Sanger est affectée par Deux changements:

- 1- Didésoxynucléotides marqués avec des fluorophores
- 2- Automatisation

1. 2 . Technique chimique de Maxam et de Gilbert

La méthode de Maxam et Gilbert a été développée en 1977. Elle est également appelée méthode de clivage chimique ; elle est connue sous le nom de méthode de dégradation chimique. Elle repose sur le clivage des nucléotides par des produits chimiques et est plus efficace avec les petits polymères de nucléotides.

Le traitement chimique génère des cassures au niveau d'une petite proportion d'une ou deux des quatre bases nucléotidiques dans chacune des quatre réactions (C, T+C, G, A+G). Cette réaction donne des séries de fragments marqués qui vont se séparer selon leur taille par électrophorèse.

Cette méthode est réalisée sans clonage d'ADN. Elle est également considérée comme dangereuse car elle utilise des produits chimiques toxiques et radioactifs. L'autobiographie est utilisée pour visualiser la séparation des fragments d'ADN. Grâce à l'extrémité ^{32}P radio-marquée de l'ADN, les bandes d'ADN sont visualisées par autoradiographie, comme le montre la figure ci-dessous (Mehdi Kchouk1, 2017) (G. Dorado, 2019).

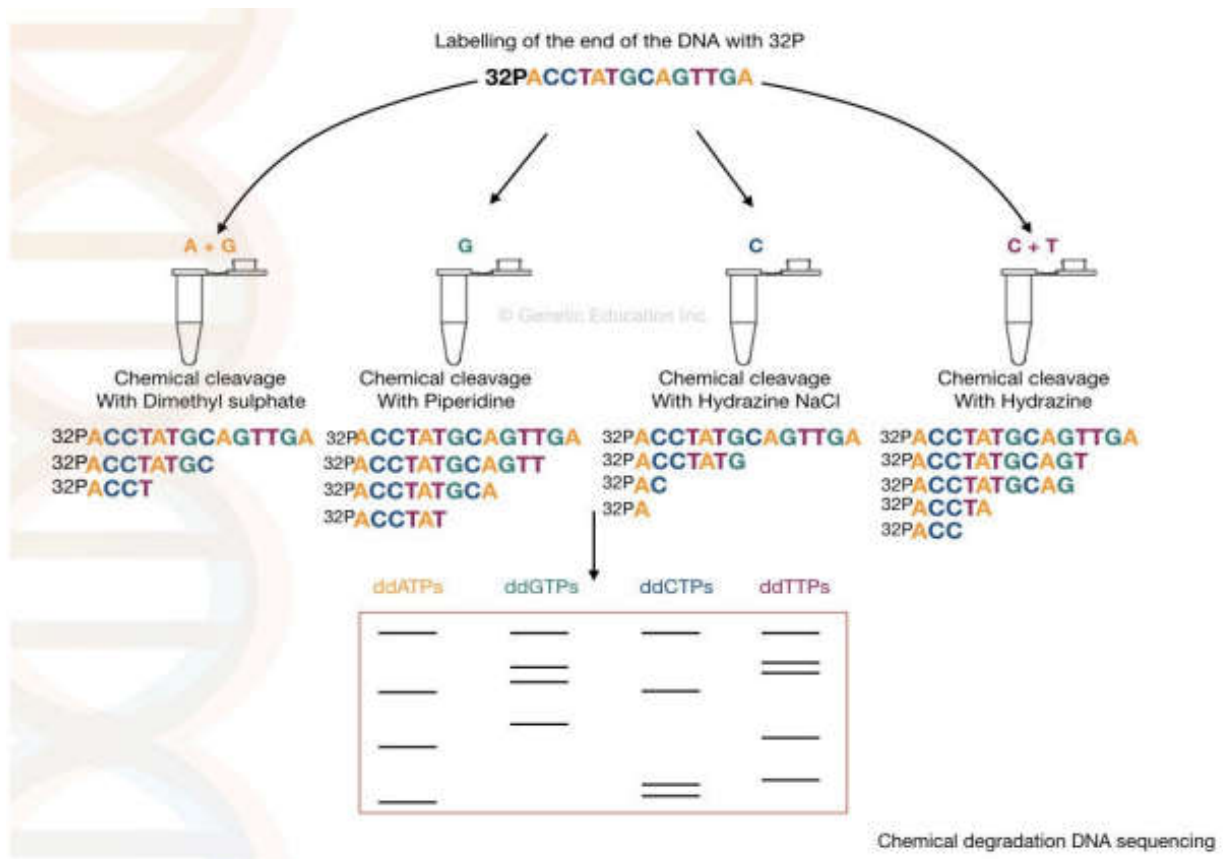


Figure 6 : Méthode de clivage chimique Maxam- Gilbert pour le séquençage de l'ADN (DNA Sequencing: History, Steps, Methods, Applications and Limitations, 2019).

2. Deuxième génération de séquençage

2.1. Séquençage Roche/454

Le séquençage Roche/454 est une technique qui utilise le **pyroséquençage**, c'est-à-dire la détection du pyrophosphate libéré après chaque incorporation de nucléotide dans le nouveau brin d'ADN synthétique.

La technique de **pyroséquençage** est une approche de séquençage par synthèse. Les échantillons d'ADN sont fragmentés de façon aléatoire et chacun d'entre eux est attaché à une bille qui porte des amorces ayant des oligonucléotides complémentaires aux fragments d'ADN, de sorte que chaque bille est associée à un seul fragment.

Ensuite, chaque bille est isolée et amplifiée à l'aide d'une émulsion PCR qui produit environ un million de copies de chaque fragment d'ADN à la surface de la bille.

Les billes sont ensuite transférées sur une plaque contenant de nombreux puits appelée plaque picotitre (PTP) et on applique la technique de **pyroséquençage** qui consiste à activer une série de réactions en aval produisant de la lumière à chaque incorporation de nucléotide.

En détectant l'émission de lumière chaque incorporation de nucléotide, la séquence du fragment d'ADN est déduite. L'utilisation de la plaque picotitre permet à des centaines de milliers de réactions de se produire en parallèle, augmentant considérablement le débit de séquençage.

Le Roche/454 est capable de générer des lectures relativement longues qui sont plus faciles à mapper à un génome de référence. Les principales erreurs détectées lors du séquençage sont des insertions et des délétions dues à la présence de régions homopolymères (Lin Liu, 2012) (Erwin van Dijk, 2021).

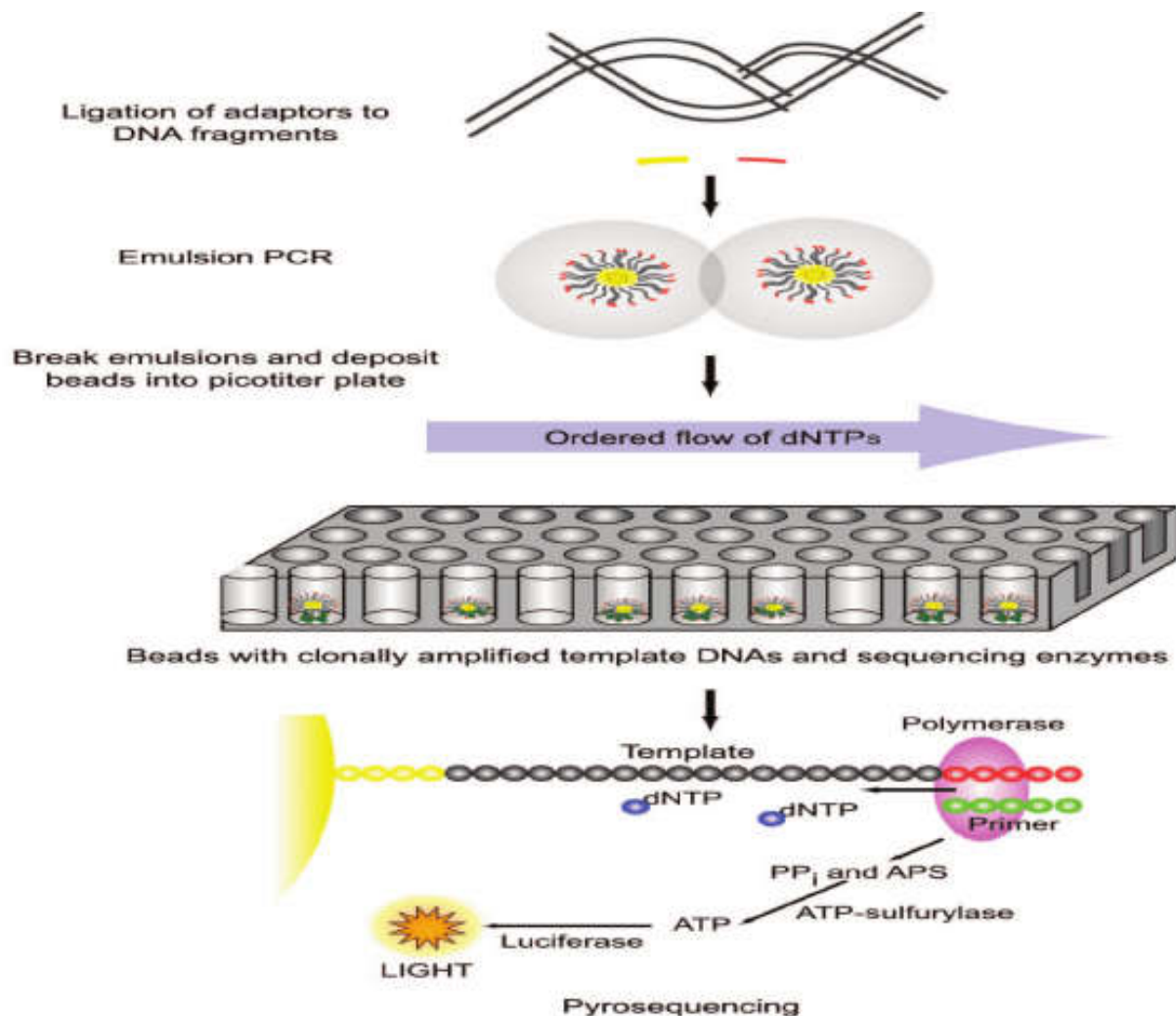


Figure 7: Séquençage Roche 454 GS FLX (Karl V Voelkerding, 2009)

2.2. Séquençage Ion Torrent

En 2010, Life Technologies a dévoilé la technologie de séquençage à semi-conducteurs Ion Torrent. Contrairement aux autres technologies de deuxième génération, elle n'utilise pas de nucléotides marqués par fluorescence.

Elle est basée sur la détection de l'ion hydrogène libéré lors du processus de séquençage. Concrètement, Ion Torrent utilise une puce qui contient un ensemble de micro-puits et dont chacun comporte une bille contenant plusieurs fragments identiques. L'incorporation de chaque nucléotide avec un fragment dans la perle, un ion hydrogène est libéré qui change le pH de la solution.

Ce changement est détecté par un capteur fixé au fond du micro-puits et converti en un signal de tension qui est proportionnel au nombre de nucléotides incorporés. Les séquenceurs Ion Torrent produisent des longueurs de lecture de 200 bp, 400 bp et 600 bp avec un débit qui peut atteindre 10 Gb pour le séquenceur ion proton.

Les principaux avantages de cette technologie de séquençage sont axés sur les longueurs de lecture, qui sont plus longues que celles des autres séquenceurs SGS, et sur la rapidité du séquençage (entre 2 et 8 heures). L'inconvénient majeur est la difficulté d'interpréter les séquences homopolymères (plus de 6 pb) qui provoquent des erreurs d'insertion et de délétion (indel) avec un taux d'environ ~1% (Barton E. Slatko, 2018).

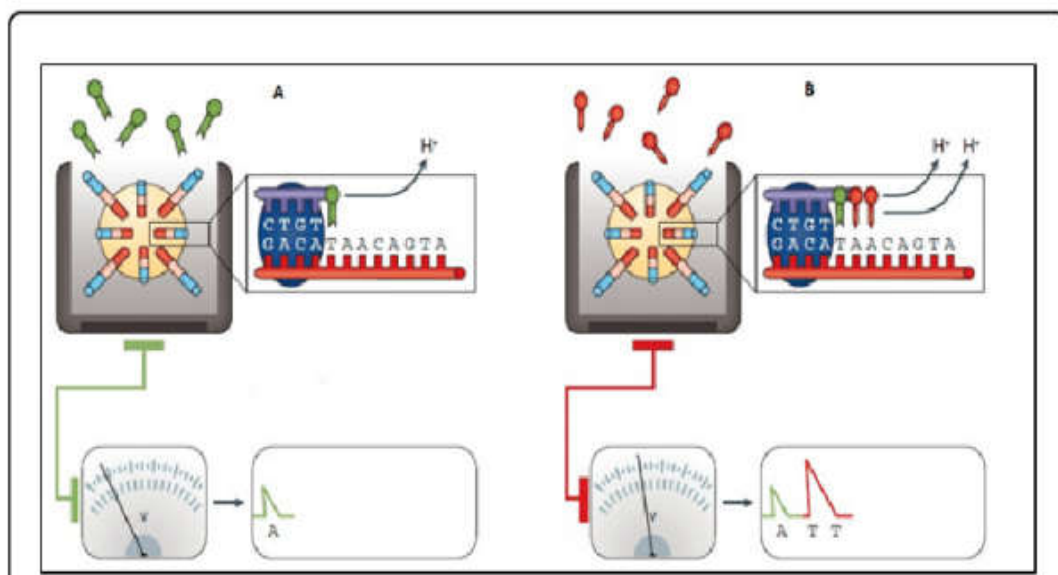


Figure 8: Technologie de séquençage Ion torrent (Mehdi Kchouk1, 2017)

2.3. Séquençage Illumina/Solexa

Actuellement, la technologie la plus utilisée sur le marché du NGS est le séquenceur Illumina/Solexa Genome Analyzer (GA). Le séquenceur adopte la technologie du séquençage par synthèse (SBS).

Au cours de la première étape, les échantillons d'ADN sont fragmentés de manière aléatoire, puis les molécules d'ADN avec des adaptateurs appropriés ligaturés à chaque extrémité sont utilisées comme substrats pour des réactions répétées de synthèse par amplification sur un support solide (lame de verre) qui contient des séquences d'oligonucléotides complémentaires à un adaptateur ligaturé.

Au cours de la deuxième étape, chaque séquence fixée sur le support solide est amplifiée par "amplification par pontage PCR" qui crée plusieurs copies identiques de chaque séquence ; un ensemble de séquences fabriquées à partir de la même séquence d'origine est appelé un cluster. Chaque cluster contient environ un million de copies de la même séquence originale.

La dernière étape consiste à déterminer chaque nucléotide des séquences, Illumina utilise l'approche de séquençage par synthèse qui emploie des terminateurs réversibles dans lesquels les quatre nucléotides modifiés, les amorces de séquençage et les ADN polymérases sont ajoutés sous forme de mélange, et les amorces sont hybridées aux séquences.

Ensuite, des polymérases sont utilisées pour étendre les amorces en utilisant les nucléotides modifiés. Chaque type de nucléotide est marqué par un spécifique fluorescent afin que chaque type soit unique.

Les nucléotides ont un groupe 3'-hydroxyle inactif qui garantit qu'un seul nucléotide est incorporé. Les clusters sont excités par un laser pour émettre un signal lumineux spécifique à chaque nucléotide, qui sera détecté par une caméra à dispositif à charge couplée (CCD) et des programmes informatiques traduiront ces signaux en une séquence de nucléotides.

Le processus se poursuit par l'élimination du terminateur avec le marqueur fluorescent et le démarrage d'un nouveau cycle avec une nouvelle incorporation.

Les premiers séquenceurs Illumina/Solexa GA ont été capables de produire des lectures très courtes et ils avaient l'avantage de pouvoir produire des lectures courtes en paires (PE), dans lesquelles la séquence aux deux extrémités de chaque cluster d'ADN est enregistrée.

Les données de sortie des derniers séquenceurs Illumina sont actuellement supérieures à 600 pb et les longueurs des lectures courtes sont d'environ 125 bp.

L'un des principaux inconvénients de la plateforme Illumina/Solexa est le besoin élevé de contrôle de la charge de l'échantillon, car une surcharge peut entraîner un chevauchement des clusters et une mauvaise qualité de séquençage.

Le taux d'erreur global de cette technologie de séquençage est d'environ 1 %. Les substitutions de nucléotides sont le type d'erreurs le plus courant dans cette technologie, la principale source d'erreur étant due à une mauvaise identification du nucléotide incorporé (Christophe Audebert¹, 2014) (Weinbrecht², NextGeneration Sequencing: Methodology and Application, 2013) (Lin Liu, 2012).

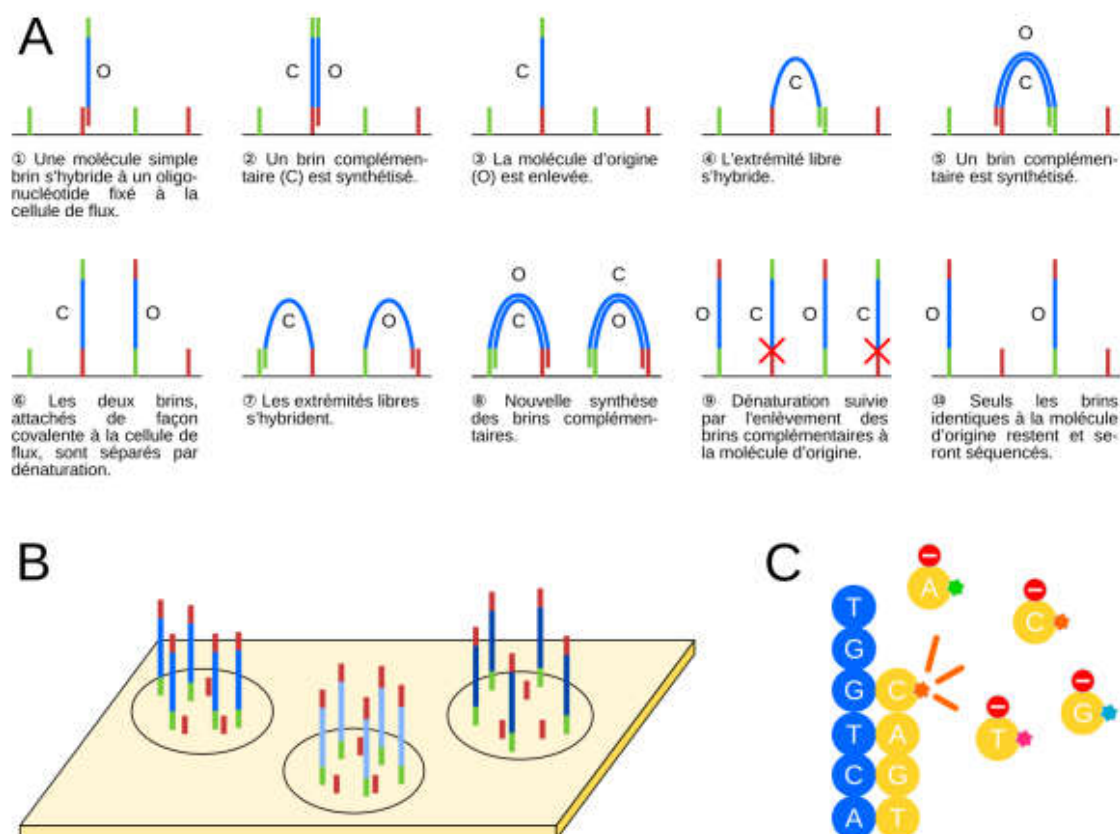


Figure 9 : Le principe de la technologie Illumina (Erwin van Dijk, 2021).

3. Troisième génération de séquençage

Alors que le séquençage de nouvelle génération est de plus en plus utilisé et modifié, le séquençage de troisième génération apporte un nouvel éclairage sur le séquençage.

Le séquençage de troisième génération présente deux caractéristiques principales. Premièrement, la PCR n'est pas nécessaire avant le séquençage, ce qui raccourcit le temps de préparation de l'ADN pour le séquençage.

Deuxièmement, le signal est capturé en temps réel, ce qui signifie que le signal, qu'il soit fluorescent (Pacbio) ou électrique (Nanopore), est contrôlé pendant la réaction enzymatique d'ajout de nucléotides dans le brin complémentaire (James M Heather 1, 2016) (Mehdi Kchouk1, 2017).

3.1. Molécule unique en temps réel (SMRT)

Il s'agit d'une méthode développée par Pacific Bioscience, qui utilise une enzyme modifiée et l'observation directe de la réaction enzymatique en temps réel. La cellule SMRT est constituée de millions de guides d'ondes en mode zéro (ZMW), dans lesquels sont intégrés un seul ensemble d'enzymes et une matrice d'ADN qui peut être détectée tout au long du processus.

Au cours de la réaction, l'enzyme va incorporer le nucléotide dans le brin complémentaire et cliver le colorant fluorescent précédemment lié au nucléotide. Ensuite, la caméra à l'intérieur de l'appareil va capturer le signal sous forme de film pour une observation en temps réel. Cela donnera non seulement le signal fluorescent mais aussi la différence de signal à long terme, ce qui peut être utile pour la prédiction de la variance structurelle dans la séquence, particulièrement utile dans les études épigénétiques telles que la méthylation de l'ADN (Barton E. Slatko, 2018) (Weinbrecht2, Next-Generation Sequencing: Methodology and Application, 2013).

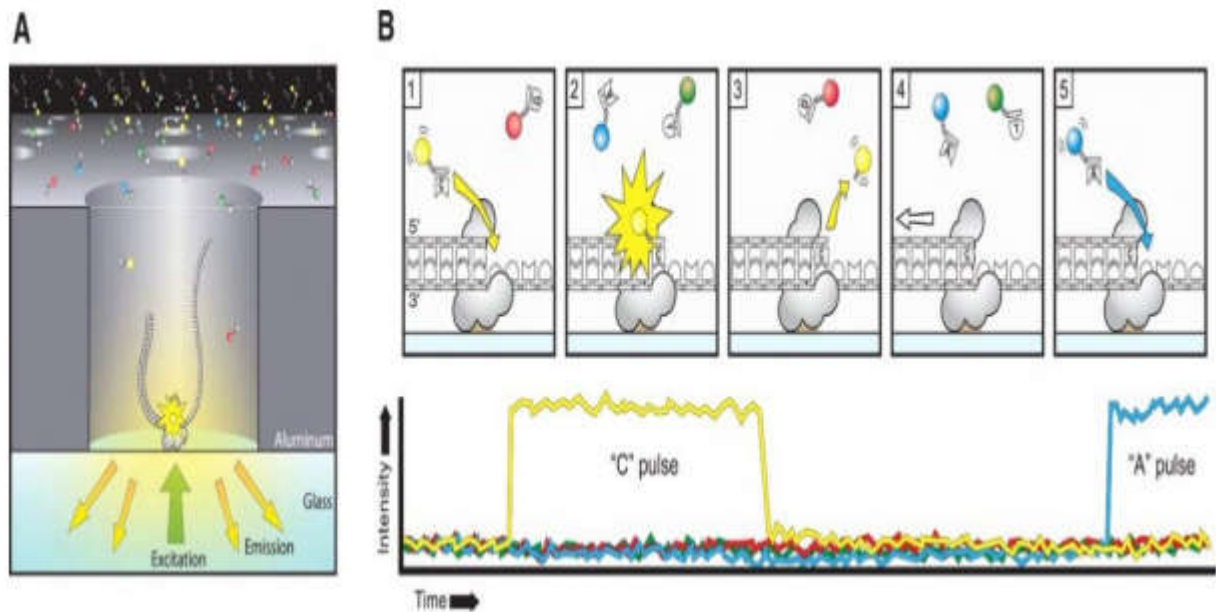


Figure10 : Principe du séquençage "SMRT" (John Eid*, 2009).

3.2. Séquençage Oxford Nanopore

Le séquençage Oxford Nanopore (ONT) a été développé comme une technique permettant de déterminer l'ordre des nucléotides dans une séquence d'ADN.

En 2014, Oxford Nanopore Technologies a mis sur le marché le dispositif MinION qui promet de générer des lectures plus longues qui assureront une meilleure résolution structurale du contenu des variantes et répétitions génomiques.

Ces séquenceurs utilisent des nanopores protéiques dans une membrane polymère résistante à l'électricité, à travers laquelle des changements de courant caractéristiques se produisent lorsque chaque nucléotide passe dans le détecteur.

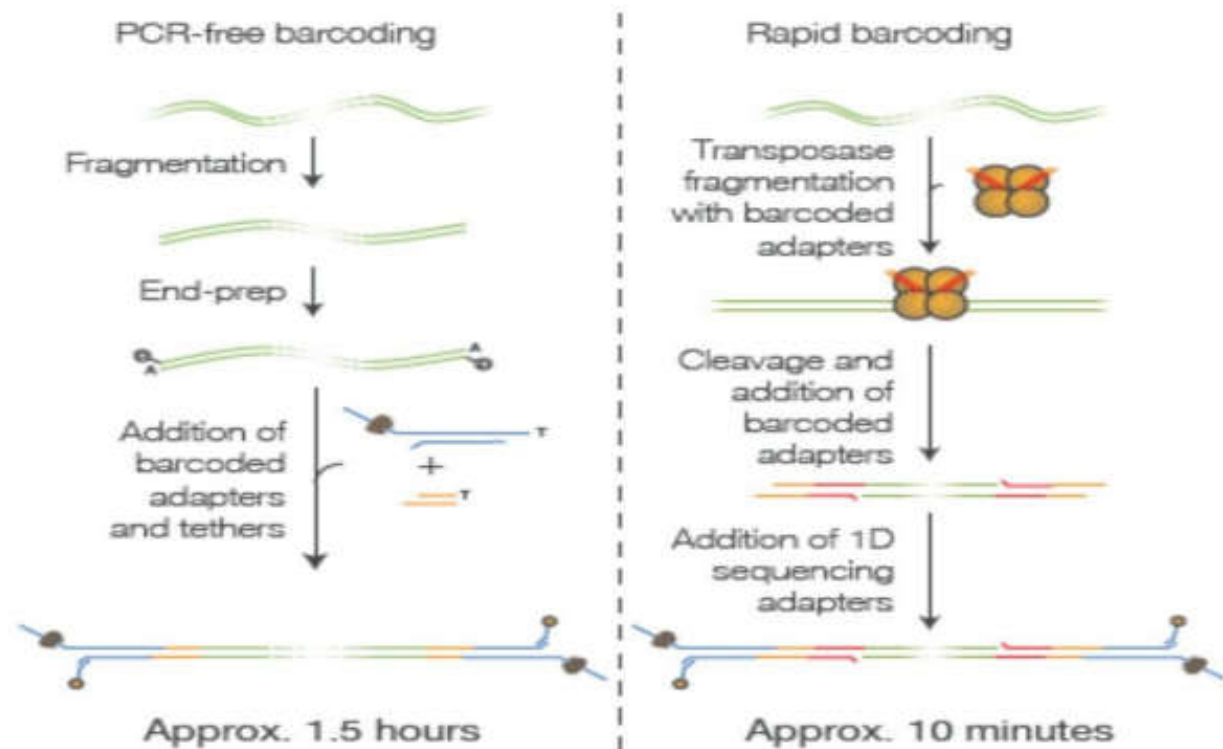


Figure 11 : Préparation des banques (ONT)

(Séquençage à « longues lectures ONT », 2021).