

SEQUENÇAGE ILLUMINA

I. Objectifs

Le séquençage Illumina est une méthode qui permet le séquençage de plusieurs centaines de millions de fragments d'ADN. L'application de cette « Next generation sequencing » (NGS) permet d'étudier la diversité microbienne dans l'environnement, de séquencer des génomes entiers, ou d'analyser les interactions entre ADN et protéines. Les domaines d'applications sont les suivants : la recherche sur le cancer, l'agrigénomique, les analyses criminalistiques, la FIV (Fécondation In Vitro), le DPI (Diagnostic Préimplantatoire), la découverte de virus ou la caractérisation de micro-organismes non cultivables ³.

II. Principe

Cette technologie se base d'abord sur l'acquisition d'une librairie¹.



Figure 1: Brins d'ADN fragmentés

Les brins d'ADN sont fragmentés en segments de 100 à 500 pb et sont couplés à leurs extrémités à des séquences particulières (une séquence permettant la liaison, des séquences index et une région

complémentaire aux oligonucléotides de la plaque). Ces fragments sont amplifiés en PCR « bridge » en phase solide. Le séquençage est réalisé grâce à un émetteur fluorescent fixé aux nucléotides (un type de nucléotide est associé à un type de longueur d'onde qui lui est propre)³.

La technologie Illumina s'appuie sur différentes techniques qui dont le SBS (*sequencing by synthesis*), la *patterned flow cell technology* et l'acquisition de données de plusieurs téraoctets.

III. Mode opératoire

Les fragments d'ADN sont déposés sur une plaque recouverte de deux types d'oligonucléotides chacun identiques à l'une des extrémités des brins d'ADN. L'ADN complémentaire à ces fragments est synthétisé en se servant des oligonucléotides de la plaque comme support initiateur (Figure 2). Les fragments primaires d'ADN sont dénaturés puis lavés grâce à un tampon spécifique². Le brin se plie et s'hybride au second type d'oligonucléotides présent sur la plaque ³.

Il y a synthèse du brin antisens formant ainsi un pont double brin. Le pont est dénaturé donnant ainsi un brin linéaire sens et un brin linéaire antisens identique au brin d'origine (Figure 3). Cette étape est répétée de nombreuses fois. Après l'amplification, les ponts sont coupés et la partie reverse est lavée grâce à un tampon².

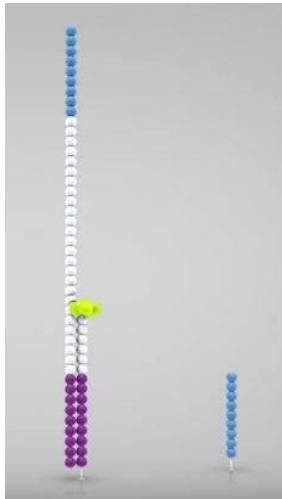


Figure 2: Synthèse du brin complémentaire

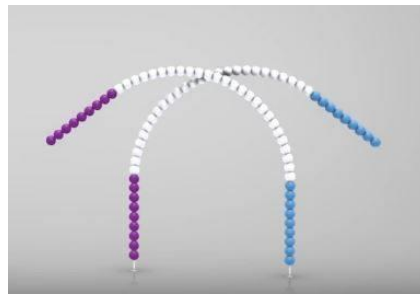


Figure 3: Dénaturation du pont d'ADN

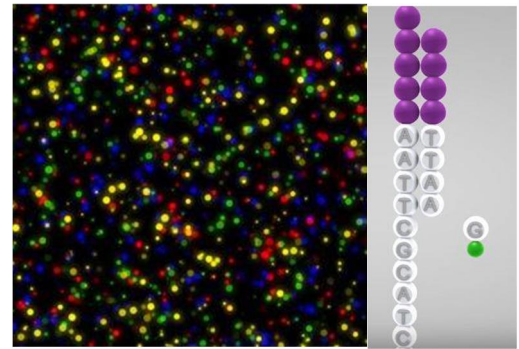


Figure 4: Polymérisation avec des nucléotides émetteurs et excitation de ceux-ci par une longueur d'onde spécifique

Pour assurer la synthèse du brin dans le bon sens, les trois premiers nucléotides attachés aux oligonucléotides du tapis sont « bloqués ».

L'ajout de l'amorce de séquençage permet de guider la polymérisation réalisée avec des nucléotides marqués avec une sonde fluorescente (Figure 4). Les clusters sont excités à une longueur d'onde donnée entraînant ainsi l'émission d'un signal lumineux (ce type de séquençage est appelé « séquençage par synthèse »). Le nombre de cycle détermine la longueur du fragment. La longueur d'onde et l'intensité du signal émis déterminent le nucléotide associé. Tous les brins identiques sont lus en même temps.

Une fois que la première lecture est terminée le brin dernièrement synthétisé est éliminé.

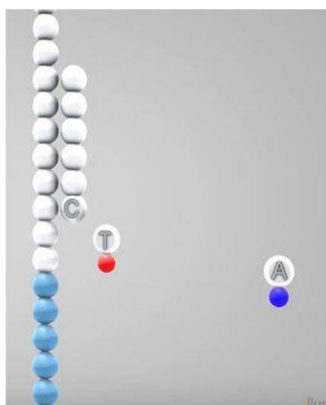


Figure 5: Introduction de l'index 1 et déprotection des 3 derniers nucléotides

L'index 1 est introduit et hybridé au brin puis « complété » par synthèse jusqu'à atteindre les oligonucléotides. Une fois complété, l'index 1 est lavé et les trois derniers nucléotides sont déprotégés¹ (Figure 5).

Le brin se plie alors et se lie au deuxième type d'oligo. L'index 2 est synthétisé de la même manière pour permettre la fixation de l'ADN polymérase qui élongue l'oligonucléotide formant ainsi un pont double-brin qui sera dénaturé, engendrant deux brins linéaires (Figure 6). Les extrémités 3' sont bloquées et le brin sens est éliminé laissant seulement le brin antisens.

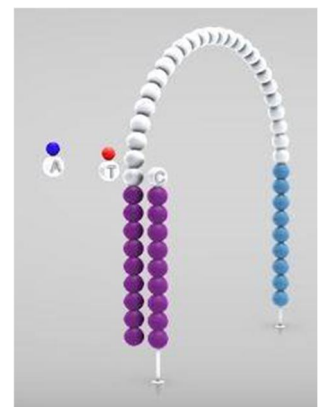


Figure 6: Synthèse de l'index 2

La deuxième lecture est initiée par la fixation de l'amorce de séquence 2 permettant ainsi la synthèse du brin d'intérêt avec les nucléotides excitables. Ce brin nouvellement synthétisé est ensuite éliminé à son tour.

IV. Présentation des résultats

La synthèse des brins complémentaires aux brins fixés sur les oligonucléotides de la plaque entraîne pour chaque nucléotide qui se fixe, l'émission d'une longueur d'onde qui lui est particulière. Les fragments dont la synthèse a été enregistrée sont ensuite triés selon la séquence index qui leur a été attribué lors de la préparation des échantillons (séquence qui a elle aussi été synthétisée, une séquence est propre à un fragment d'ADN). Pour chaque échantillon, les séquences possédant des schémas de nucléotides similaires sont regroupées. Les brins sens et anti sens sont appariés permettant ainsi la construction de séquences continues. Ces séquences continues sont ensuite alignées par rapport au génome de référence⁵. Tout ceci est réalisé grâce à un logiciel interne à l'entreprise : BaseSpace⁶.

V. Interprétation des résultats

L'interprétation des résultats dépend de l'objectif du chercheur. Selon l'absence ou la présence de la séquence identifiée on peut déterminer l'implication d'un gène dans une maladie. Grâce aux transcrits obtenus (dans le cas d'une rétro-transcription d'ARN), on peut aussi identifier des gènes et, par la même occasion, estimer l'abondance des transcrits afin de quantifier la transcription d'un gène.

Les séquences obtenues peuvent ensuite être utilisées pour réaliser les différents tests bioinformatiques dont le chercheur a besoin.

VI. Intérêts et limites

L'avantage de cette méthode est une amélioration dans la vitesse du séquençage en comparaison à une méthode de séquençage « Sanger ». De plus, l'exactitude du séquençage est estimée à 99.9%. L'équipement peut être coûteux (HiSeq1000 : 560k\$), mais en comparaison à d'autres méthodes comme la méthode Sanger, le prix de réalisation du séquençage n'est en lui-même pas cher (10k\$)³.

De plus, contrairement une nouvelle fois à la méthode Sanger, il n'est pas nécessaire de réaliser un clonage bactérien pour assurer l'efficacité du séquençage.

Cependant, les séquences obtenues sont plus courtes du fait de la fragmentation de l'ADN. De plus, un génome de référence doit être établi au préalable, ce qui demande un temps de travail supplémentaire. La quantité de données obtenue (de l'ordre du téraoctet) est non négligeable⁴. Il faut prévoir une grande capacité de stockage et faire un choix dans ce que l'on garde ou pas. La transmission des résultats en sera impactée.

VII. Références bibliographiques

- (1) **Meyer M. and Kircher M.** (2010). Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and sequencing. *Cold spring Harb Protoc.*
- (2) **Hodges E., Rooks M., Xuan Z., Bhattacharjee A., Gordon BG., Brizuela L., McCombie WR., and Hannon GJ.** (2009). *Nat protoc.* 4(6): 960-974.
- (3) <https://www.illumina.com/technology/next-generation-sequencing/sequencing-technology.html>

(4) <https://www.illumina.com/technology/next-generation-sequencing.html>

(5) Illumina Sequencing by Synthesis. Disponible sur:
<https://www.youtube.com/watch?v=fCd6B5HRaZ8>

(6) Site web de la compagnie Illumina. Disponible sur : <https://www.illumina.com/>