

Nouveaux outils en biologie

Le séquençage haut-débit

DES d'hématologie

16 janvier 2015

Paris

Alice Marceau-Renaut
Laboratoire d'hématologie
CHRU Lille

NGS = Next-Generation Sequencing

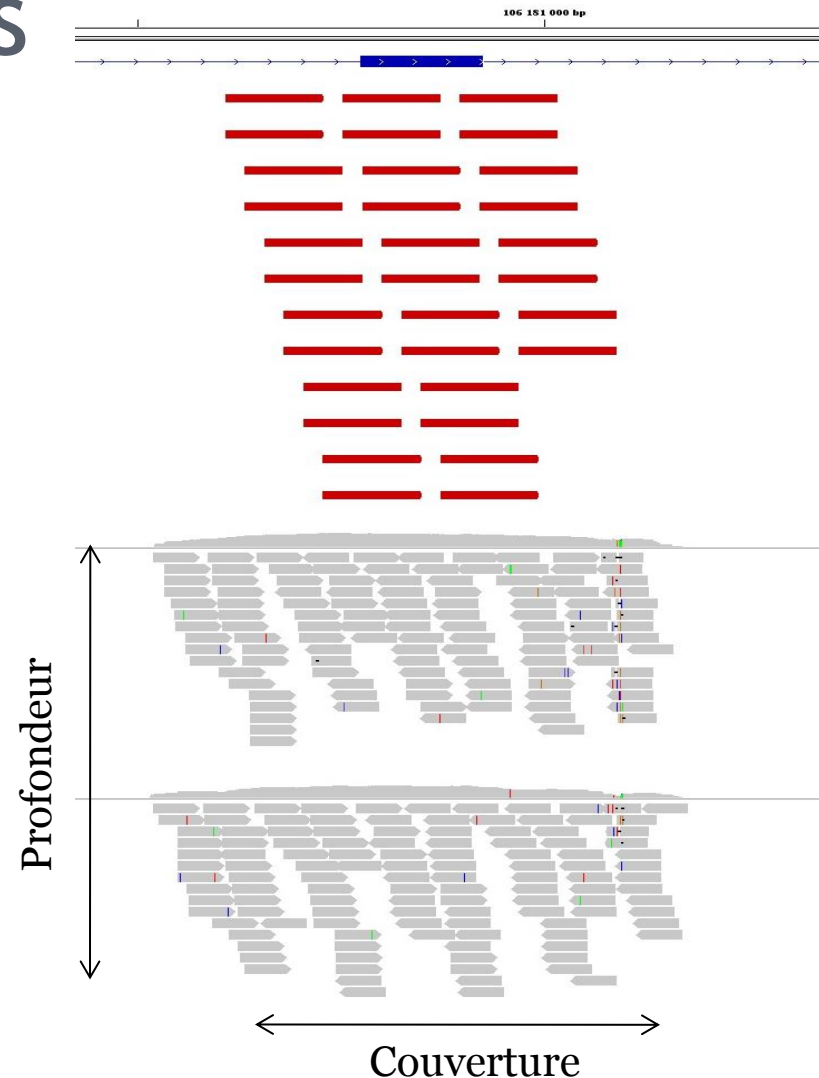
- Whole-genome
- Whole-exome
- RNA-seq (transcriptome)
- Méthylome
- Reséquençage ciblé

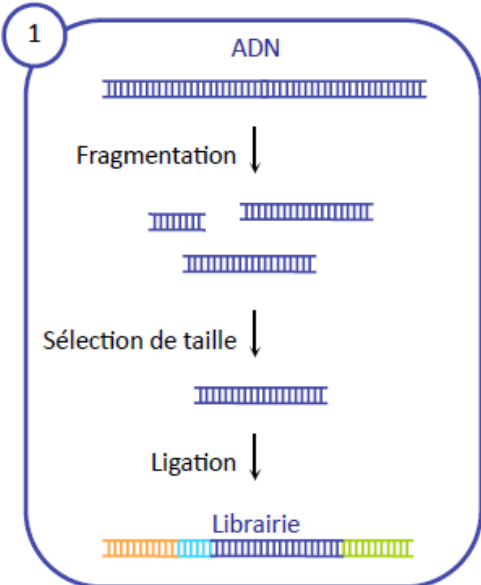
Quelques définitions

- Reads = séquences
- Amplicon = fragment d'ADN amplifié par PCR
- Librairie = collection de fragments d'ADN représentative de l'ADN à séquencer
- Index ou barcode = séquence de 8 à 12 bases permettant l'identification de l'échantillon
- Flow cell ou chip = cellule de séquençage

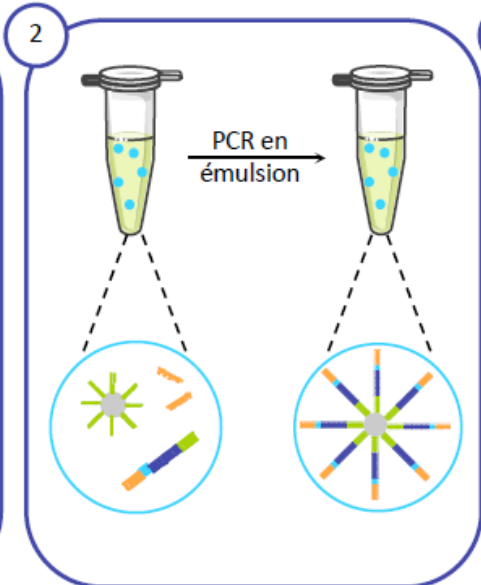
Quelques définitions

- Couverture
= zone du génome couverte
par un nombre suffisant de
lecture
- Profondeur
= nombre de lecture de
chaque base

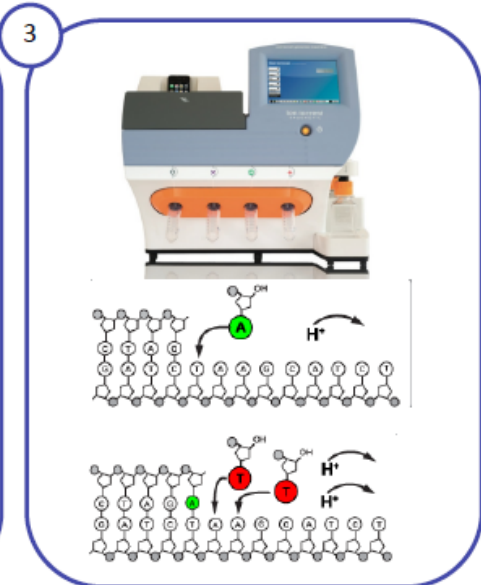




Construction des bibliothèques

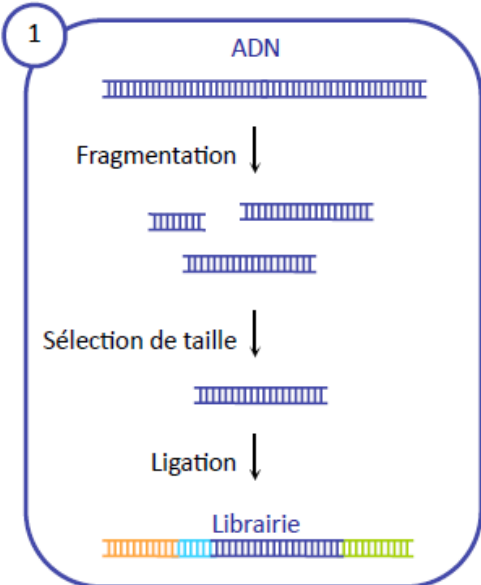


Préparation de la matrice

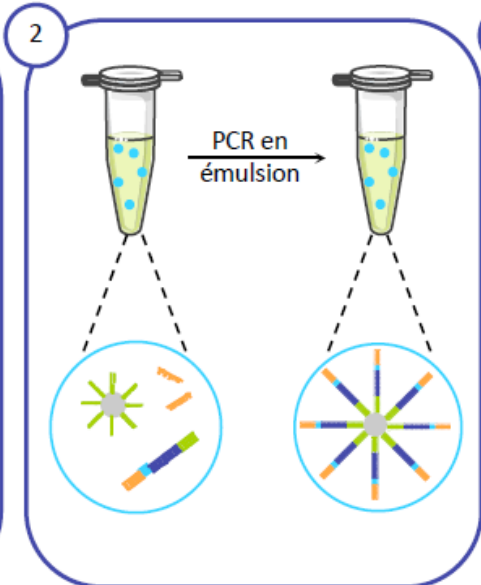


Analyse des données

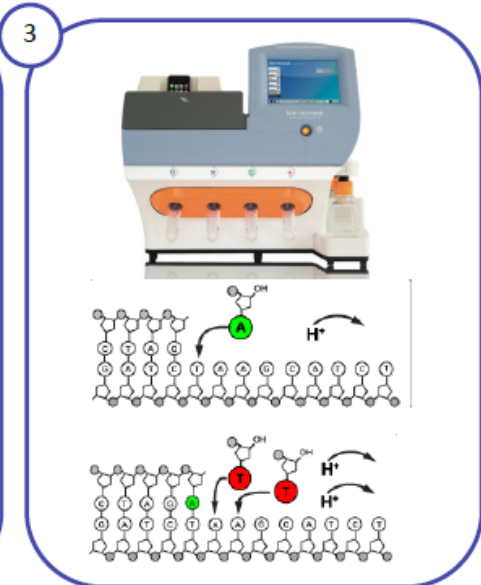




Construction des bibliothèques



Préparation de la matrice



Ampliseq
(PCR multiplex)



Life Tech
(PCR en émulsion)



Life Tech
(PGM/Proton)

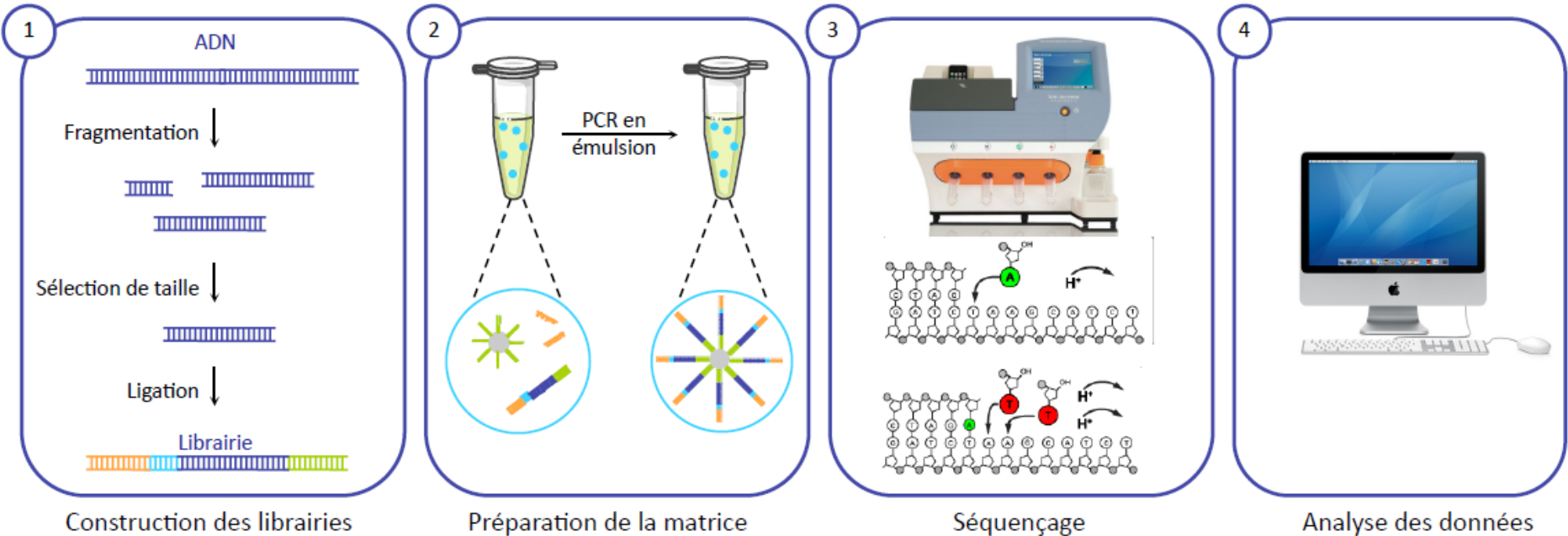
Haloplex
(circularisation)



Illumina
(Bridge PCR)



Illumina
(MiSeq/NextSeq)

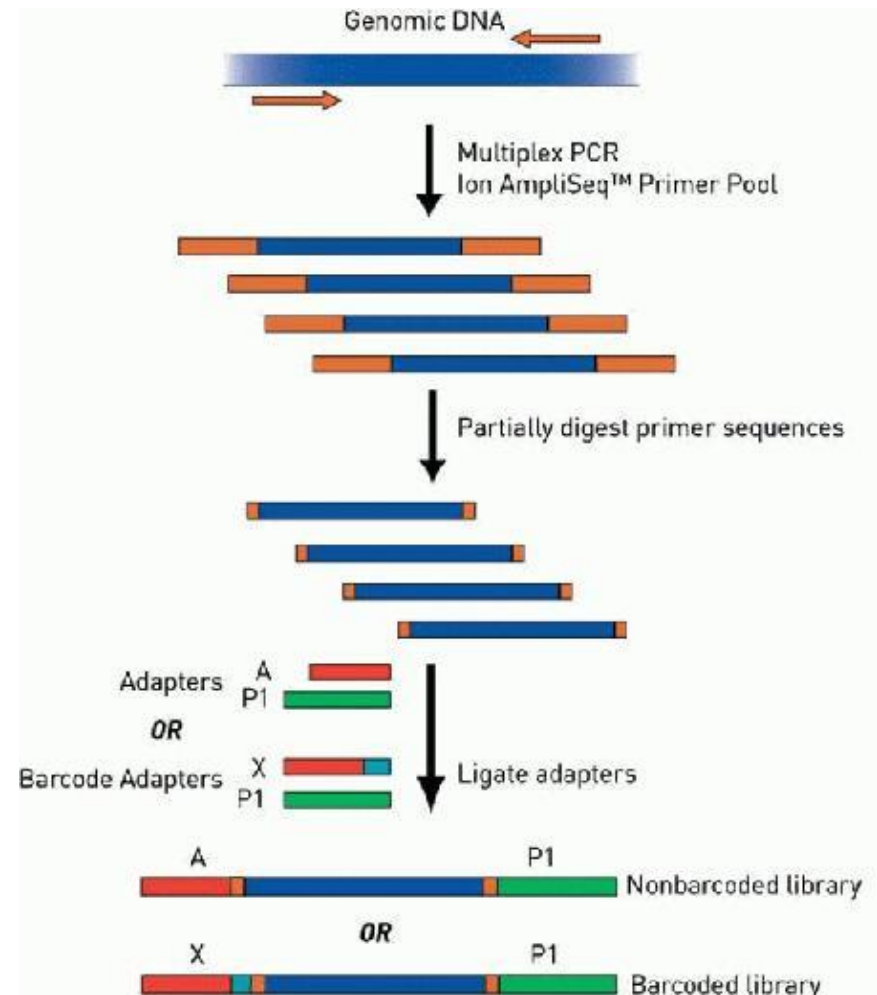


Ampliseq (PCR multiplex) → Life Tech (PCR en émulsion) → Life Tech (PGM/Proton)

Haloplex (circularisation) → Illumina (Bridge PCR) → Illumina (MiSeq/NextSeq)

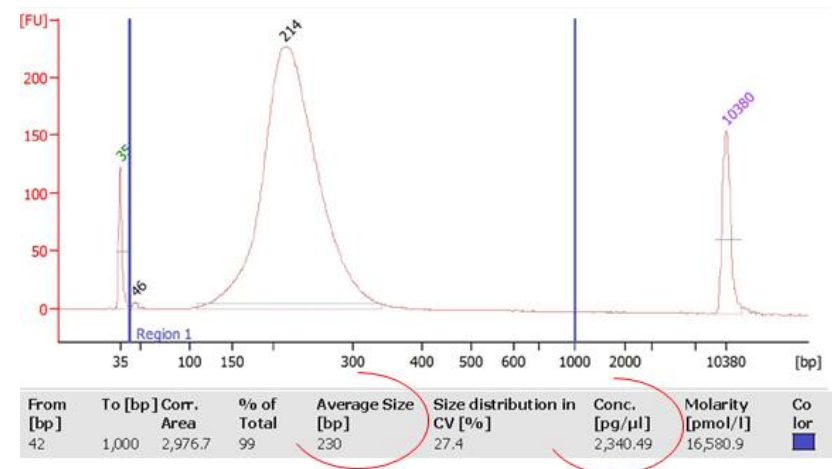
Préparation de la librairie (Ampliseq)

- 20ng d'ADN
- PCR multiplex (primers en 2 à 4 pools)
- Ligation des adaptateurs et barcodes
- Double purification sur robot



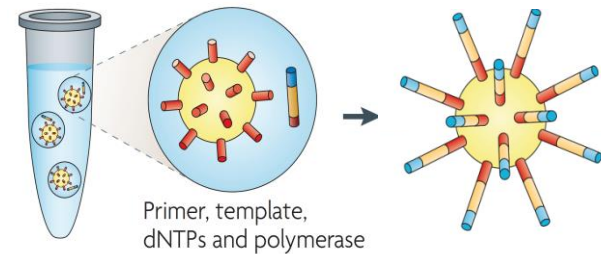
Préparation de la librairie (Ampliseq)

- Quantification de chaque PCR multiplex sur TapeStation
- Normalisation sur robot
- Composition de la librairie finale (X amplicons pour Y échantillons)
- Quantification de la librairie sur bioanalyseur et dilution

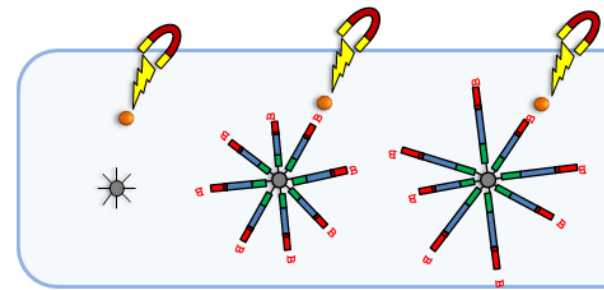


Préparation de la matrice de séquençage

- Amplification clonale
PCR en émulsion sur One Touch 2

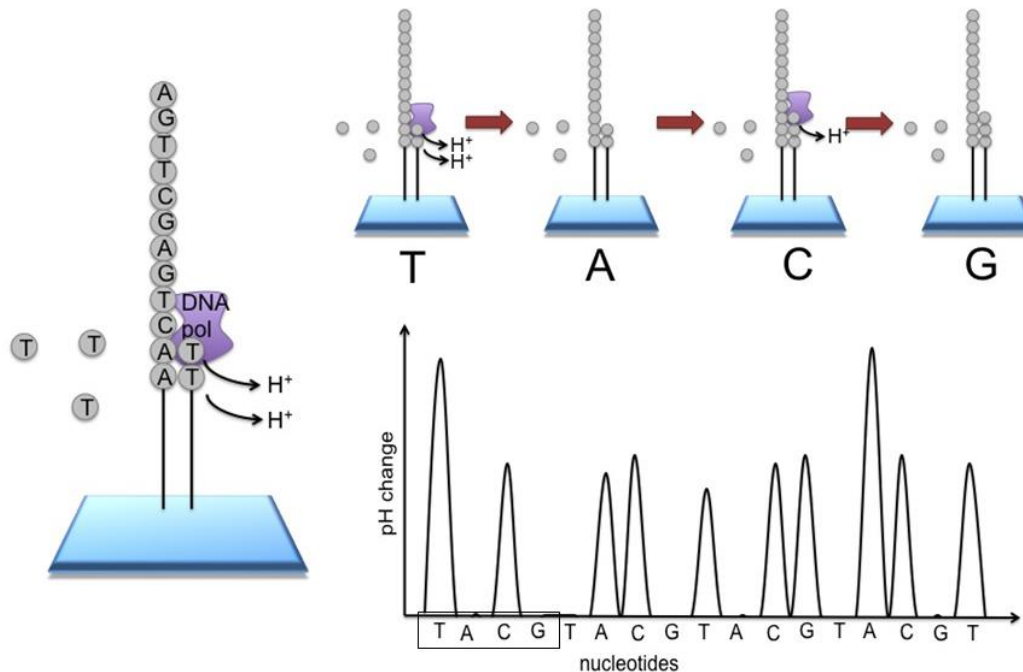


- Enrichissement des ISP
par des billes magnétiques sur
OneTouch ES



Séquençage (Life Tech)

- Séquençage par synthèse
- Détection du signal par variation de pH



Les séquenceurs de Life Tech

PGM

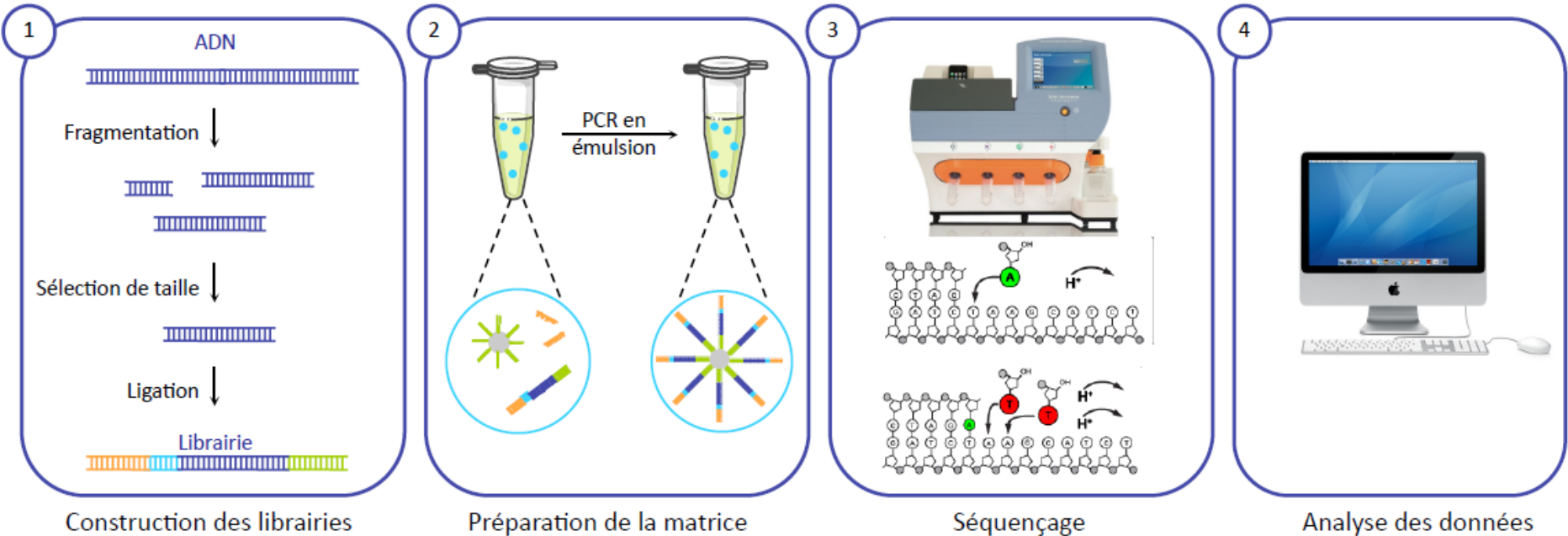


3 puces de 20Mb à 1Gb

Proton



1 puce de 10Gb



Ampliseq
(PCR multiplex)



Life Tech
(PCR en émulsion)



Life Tech
(PGM/Proton)

Haloplex
(circularisation)

→

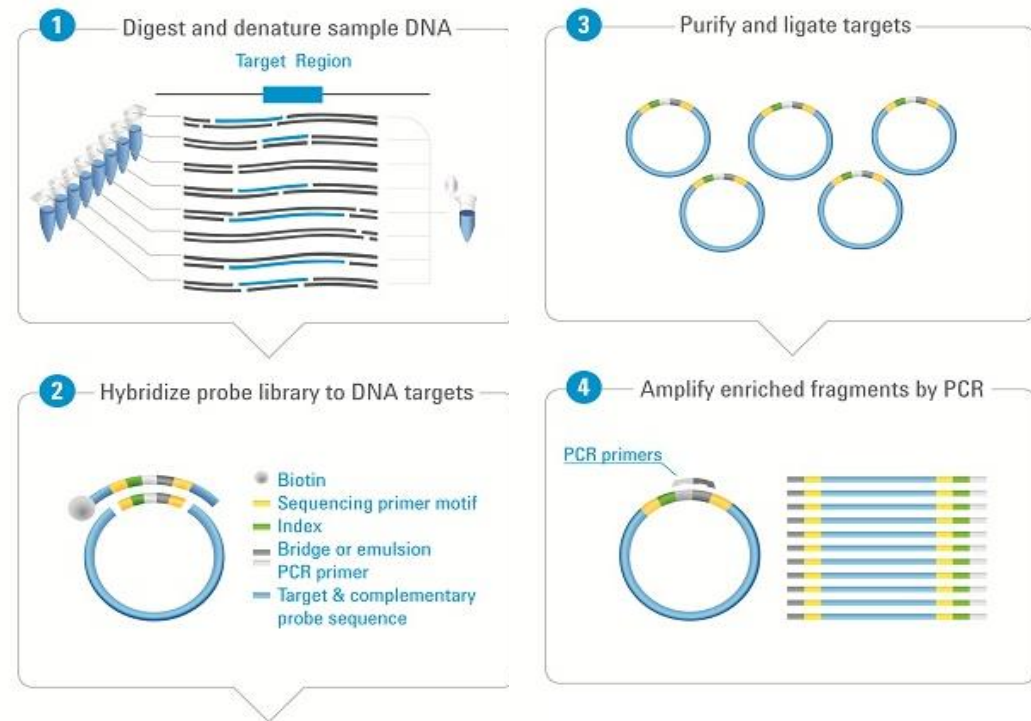
illumina
(Bridge PCR)

→

illumina
(MiSeq/NextSeq)

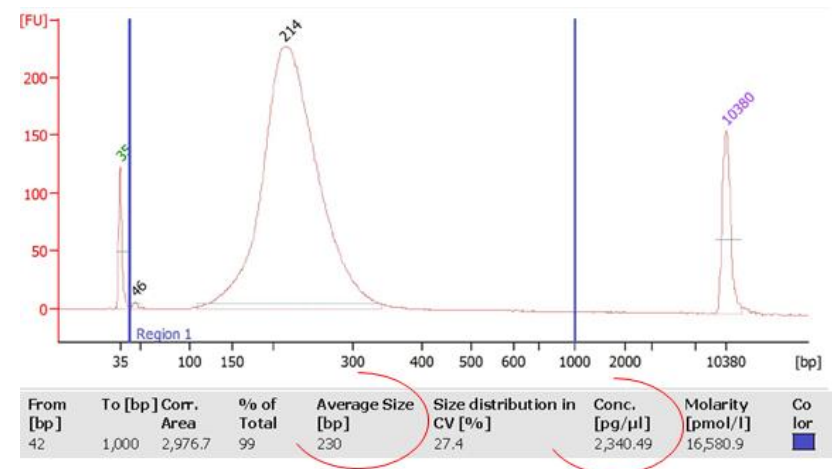
Préparation de la librairie (Haloplex)

- 225ng d'ADN
- Digestion enzymatique
- Hybridation avec les sondes Haloplex
- Circularisation de l'ADN
- Capture des hybrides par billes magnétiques
- Amplification par PCR



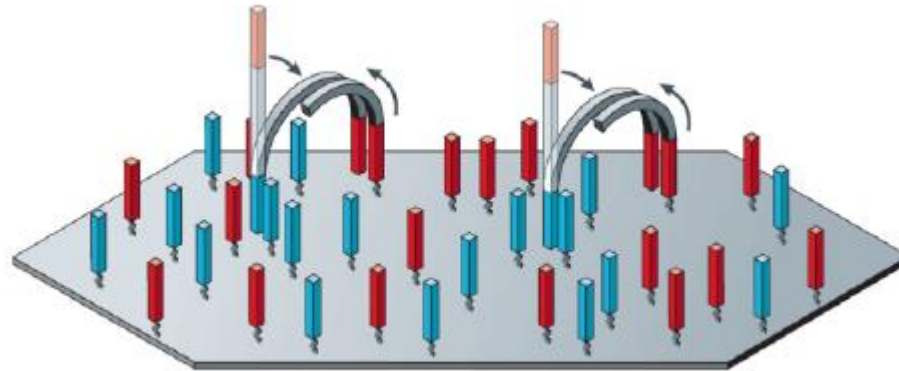
Préparation de la librairie (Haloplex)

- Quantification de chaque PCR sur TapeStation
- Normalisation sur robot
- Composition de la librairie finale (X amplicons pour Y échantillons)
- Quantification de la librairie sur bioanalyseur et dilution



Préparation de la matrice de séquençage (Illumina)

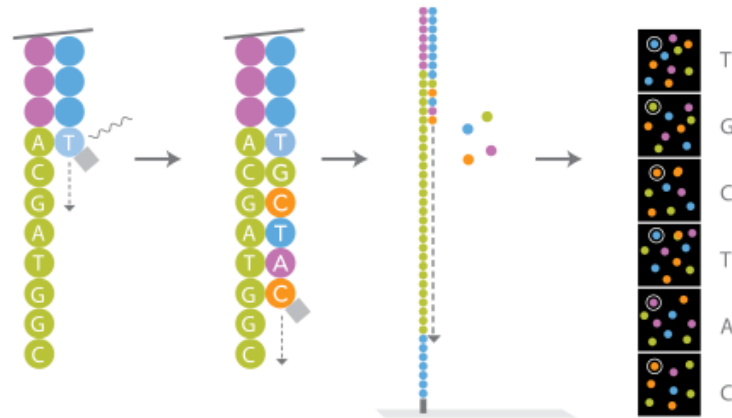
- La librairie est mise directement dans le séquenceur
- Amplification par bridge PCR (génération des clusters)



Bridge amplification

Séquençage (Illumina)

- Séquençage par synthèse
- Détection optique du signal par nucléotides marqués



Les séquenceurs d'Illumina

MiSeq



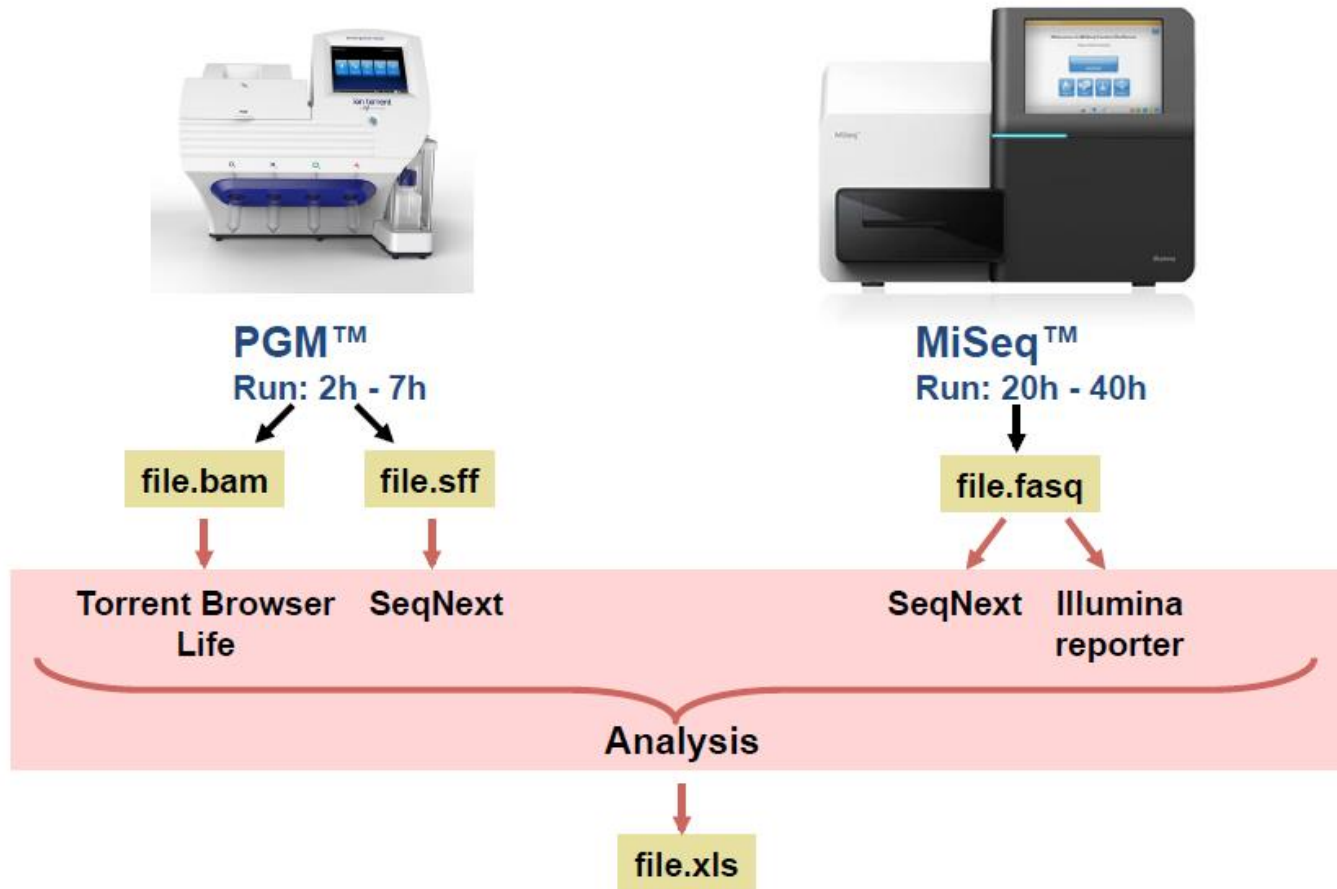
8Gb

NextSeq500



16 à 120Gb

Analyse des données



Analyse bio-informatique des données

- **Torrent Browser**

- Recalibration des bases
- Filtrage et découpage des séquences
- Alignement des séquences sur la référence
- Recherche des variants à l'aide de 3 algorithmes (SNP, ins/del, longues indels)
- Annotation des variants par logiciel Annovar
- Filtre à 1% d'allèle muté

- **Logiciel SeqNext**

- Alignement, recherche et annotation des variants
- Visualisation des séquences

Validation et classification des variants identifiés

- Base de données de SNP



dbSNP
Short Genetic Variations

- Base de données de mutations



- Intérêt de disposer de matériel germinale pour identifier formellement les SNP rares et les mutations

Tiers of Variants

Somatic variants that are identified on whole-genome sequencing and other large-scale sequencing analyses are often categorized according to their likely effect on biologic function. In this study, the somatic variants were divided into four tiers.

Tier 1: Changes in the amino acid coding regions of annotated exons, consensus splice-site regions, and RNA genes (including microRNAs).

Tier 2: Changes in highly conserved regions of the genome or regions with regulatory potential.

Tier 3: Changes in the nonrepetitive part of the genome that do not meet the criteria for tier 2.

Tier 4: Changes in the remainder of the genome.

Avantages du NGS en pratique clinique

- **Rapidité** : plus de patients, plus de gènes étudiés en simultané
- **Coût** : inférieur au séquençage de Sanger
- **Sensibilité** : supérieure au Sanger, détection de sous-clones, suivi de MRD
- **Exhaustivité** : plus d'exons, gènes entiers
- **Économie du matériel biologique** : qq ng d'ADN pour un grand nombre de gènes

Précautions

- **Choix du séquenceur** : capacité différente, profondeur de séquençage souhaitée
- **Choix de la préparation de la librairie** : capture, circularisation, PCR multiplex
- **Robotisation** : étapes de purification
- **Validation des résultats**
 - vérification par Sanger
 - sous-clones : nécessité d'une autre technique de NGS
- **Analyse des données** : bio-informaticien indispensable
 - Évaluation de la qualité du run
 - Filtres : éliminer les artefacts mais pas les variants
- NGS peut être **inadapté** pour les grandes indels, homopolymères, régions riches en GC

Conclusion

- Progrès technologique spectaculaire
- Bien connaître les gènes étudiés
- Robotisation indispensable
- Bio-informatique +++
- Diagnostic vs suivi : objectifs différents

Remerciements

- Nicolas Duployez
- Aline Renneville
- Olivier Nibourel
- Sandrine Geffroy
- Pauline Grave
- Maxime Bucci
- Christophe Demay
- Martin Figeac
- Claude Preudhomme



**Centre Hospitalier Régional
Universitaire de Lille**



