

Next Generation Sequencing (NGS)

Plan

- **Introduction-Définition**
- **Principes et Méthodes NGS actuelles**
- **Applications**
- **Conclusions**

Introduction-Définition

- Next Generation Sequencing (NGS): révolution biotechnologique de ces dernières années, en permettant de **séquencer de grandes quantités d'ADN en des temps records**
- A titre de Comparaison:
 - Projet Génome Humain: 3 milliards de dollars sur 13 ans entre 1990 et 2003 par la méthode Sanger
 - Séquenceur NGS Illumina Hiseq X: 3 jours 3 génomes humains à 1000\$ chacun
 - La société Illumina a déjà promis le génome à \$100 d'ici deux ans avec le nouveau séquenceur **Illumina Novaseq6000**

Cost per Genome



Diminutions du coût du séquençage par nucléotides au cours des dernières années

Principe et Méthodes NGS actuelles

Préparation de librairie de séquençage

- Une **librairie**, est l'ensemble des fragments d'ADN que l'on veut séquencer.
- Pour créer une librairie, deux méthodes sont à retenir si l'on veut séquencer l'ensemble du génome ou des régions d'intérêts.

Séquençage

Il existe différentes méthodes de séquençage de nouvelle générations

- ✓ Le séquençage par synthèse ([Illumina](#)).
- ✓ Le pyroséquençage ([Roche 454](#)).
- ✓ La ligation ([SOLid](#) [Thermofisher](#)).
- ✓ La détection des ions H⁺ ([Proton](#) [Thermofisher](#)).

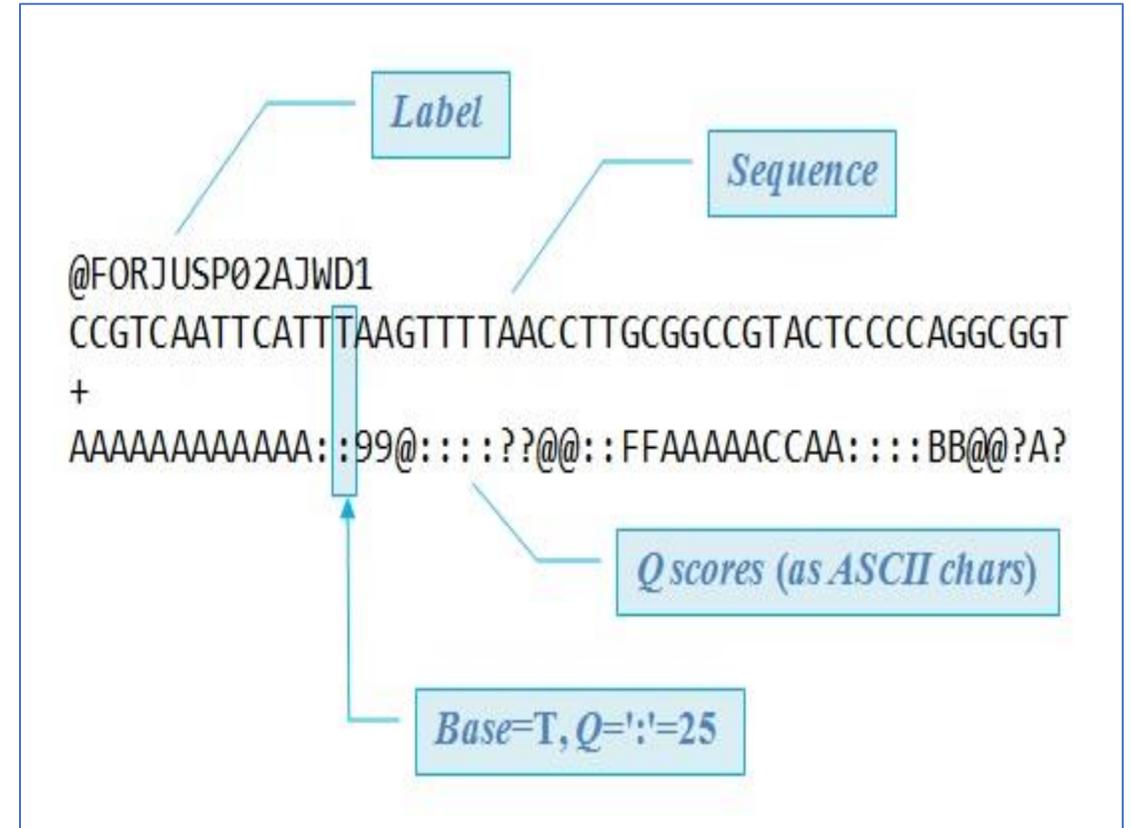
Le principe général reste le même

Dans l'ensemble, le principe général reste le même.

- ✓ Chaque fragment est d'abord cloné plusieurs fois afin d'amplifier le signal.
- ✓ Puis le brin complémentaire de chaque fragment cloné est synthétisé. À chaque incorporation d'un nucléotide, un signal est détecté.
- ✓ De la lumière pour Illumina ou une variation de pH sur du Proton.
- ✓ À la fin du séquençage, chaque fragment est séquencé en parallèle.
L'ensemble des données est enregistré dans un fichier *Fastq*.

Alignement des séquences

À la fin du séquençage, les séquences des fragments, qu'on appelle maintenant des "reads", sont sauvegardées dans un fichier Fastq contenant les séquences et leurs scores de qualité (score Phred). Ce score évalue la confiance du séquençage. Par exemple le séquenceur vous dira que pour tel reads, la probabilité que le quatrième nucléotide soit un 'A' est de 99,9%.

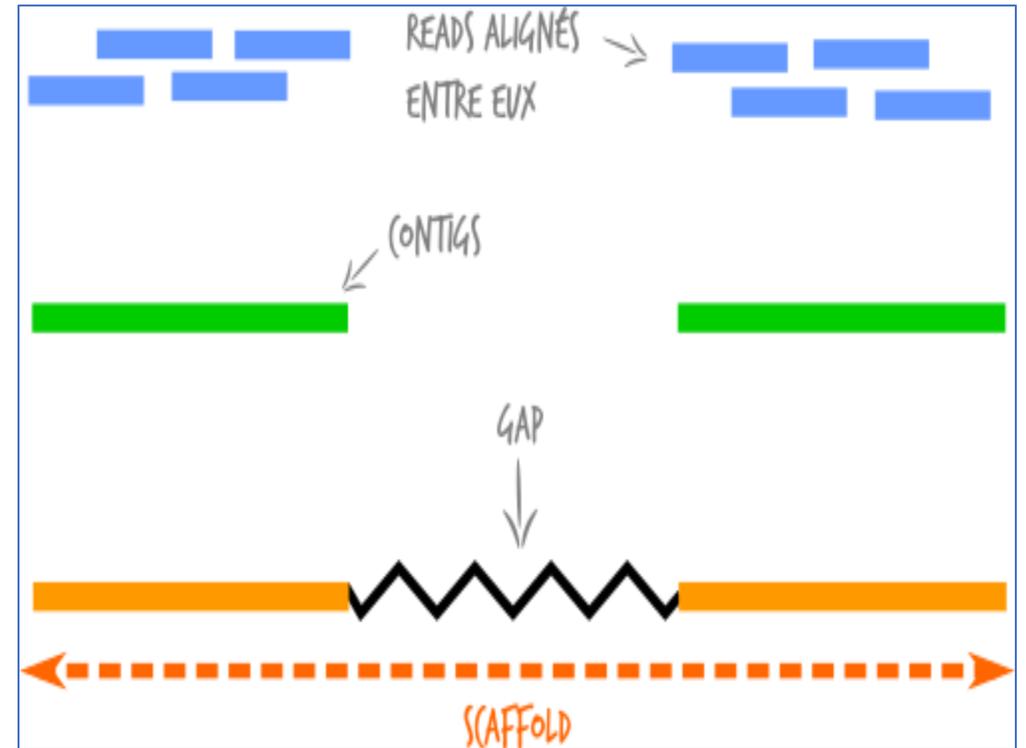


Aperçu d'un read dans un fichier fastq. Le score de qualité associé à chaque nucléotide un caractère ASCII

Alignement

Pour obtenir la séquence complète d'un gène ou d'un génome entier, il faut réaliser un alignement. Deux méthodes existent

- **Assemblage de novo** : Les fragments d'ADN qui sont chevauchants permettent petit à petit de reconstruire ce qu'on appelle un **contig**. L'assemblage des contigs entre eux permet d'obtenir un scaffold. Cette technique est très couteuse en termes de calcul. Des algorithmes bioinformatiques comme les graphes de Bruin, permettent de résoudre ce problème. Cette méthode est principalement employée pour reconstruire des génomes non connus..



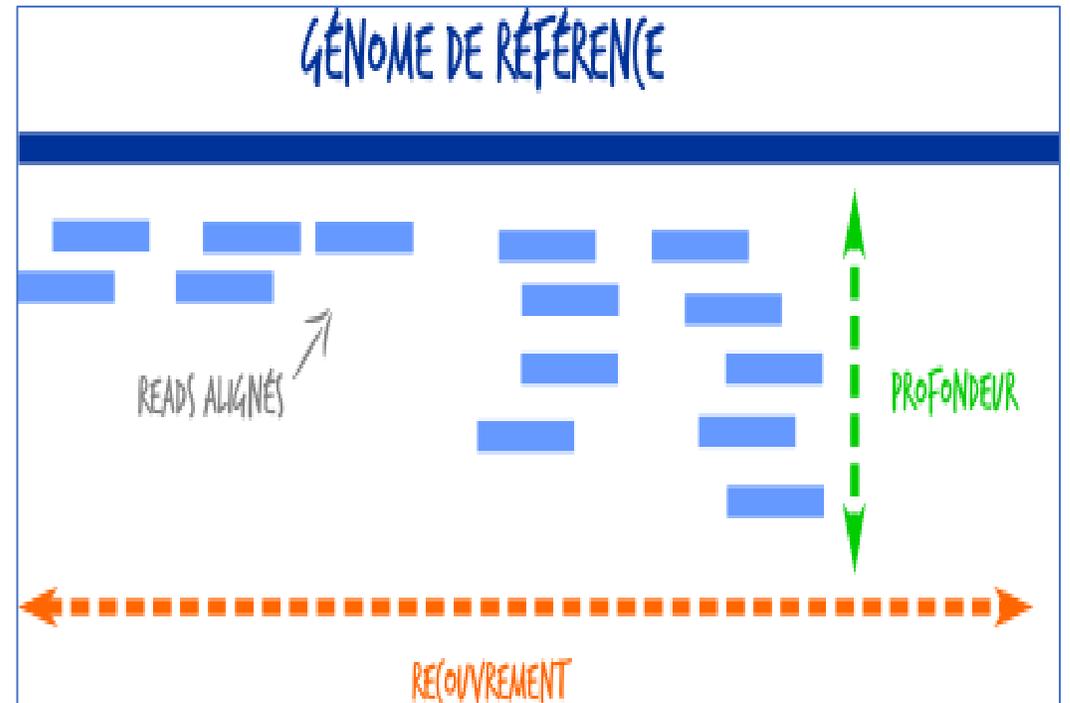
L'alignement de novo consiste à aligner les reads entre eux

Alignement

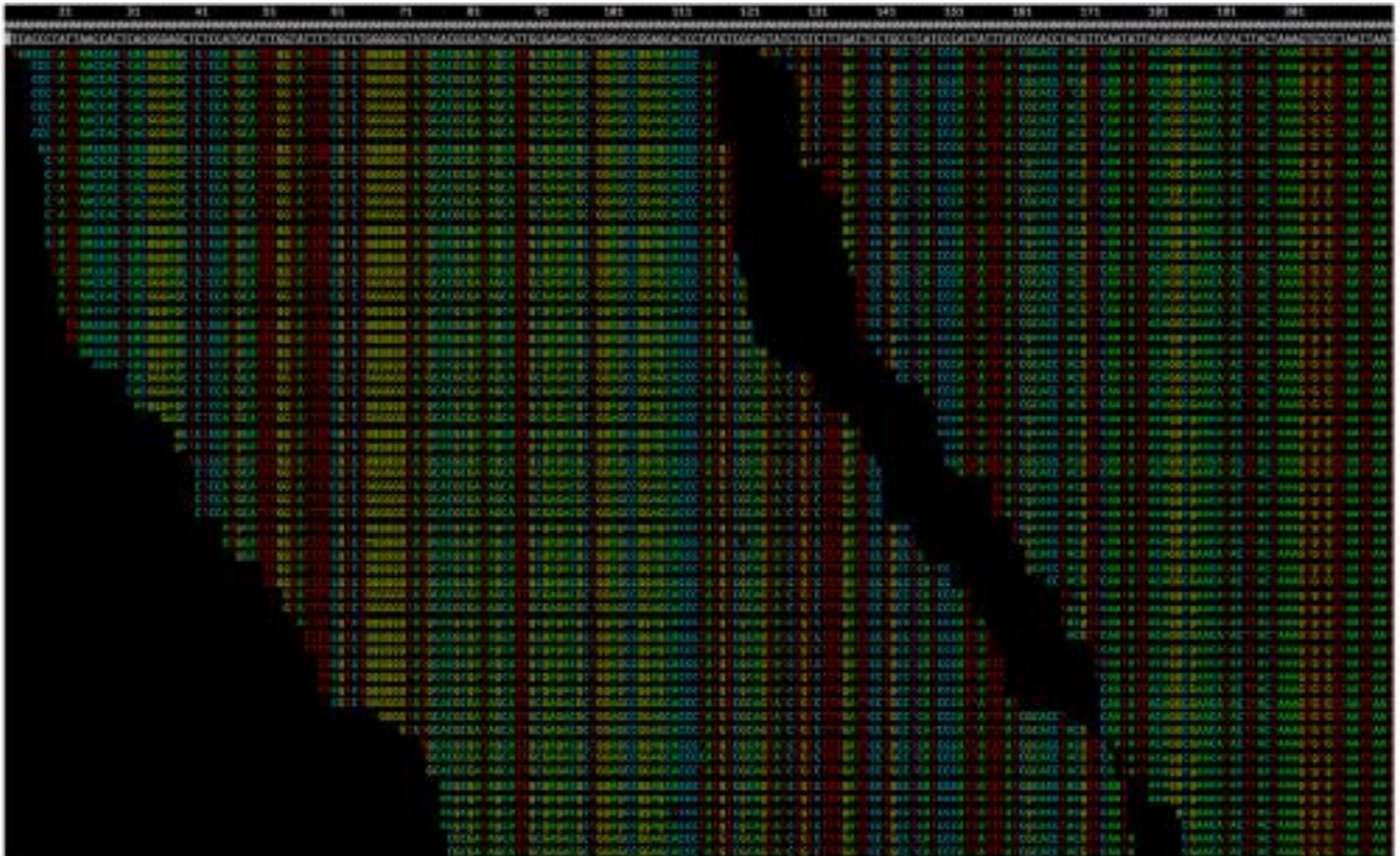
- **Alignement avec référence**

Utilisation d'une version du génome humain (hg19). Chaque read est aligné sur cette référence. La complexité de calcul est plus simple qu'avec l'alignement de novo. On utilise en général un algorithme permettant de rechercher de manière efficace une correspondance entre les reads et la référence. Après cet alignement, on obtient un fichier BAM associant à chaque reads ses coordonnées génomiques. C'est à dire le chromosome et la position.

On appelle la **profondeur**, le nombre moyen de reads qui se superpose et **recouvrement**, l'étalement des reads sur la zone d'intérêt.



L'alignement avec référence consiste à aligner les reads sur une référence



Visualisation d'un alignement réel avec Samtools

Les séquenceurs de 3ème génération

- Capables de générer de très longs reads sans avoir besoin de cloner les fragments pour amplifier le signal. C'est pour cette raison qu'on les appelle aussi "Single molecule sequencing".
- Ces nouvelles techniques produisent encore beaucoup d'erreurs de séquençage.
- Les deux leaders de ce Next Next Generation Sequencing sont Nanopore et PacBio Science. La miniaturisation de ces séquenceurs sera peut être un jour disponible chez tout bon médecin généraliste qui vous diagnostiquera votre prédisposition d'infarctus ou d'Alzheimer en quelques heures. Effrayant ou rassurant, à vous de choisir!



C'est le SmidgION d'Oxford Nanopore, un séquenceur qui se branche sur un iPhone

Applications des NGS

Les applications nécessitant de grandes quantités de données, telles que:

- ✓ le séquençage du génome humain entier (WGS),
- ✓ le séquençage des exomes ultra-profonds
- ✓ le profilage des tumeurs normales, peuvent désormais être réalisées de manière plus rentable.

Applications des NGS

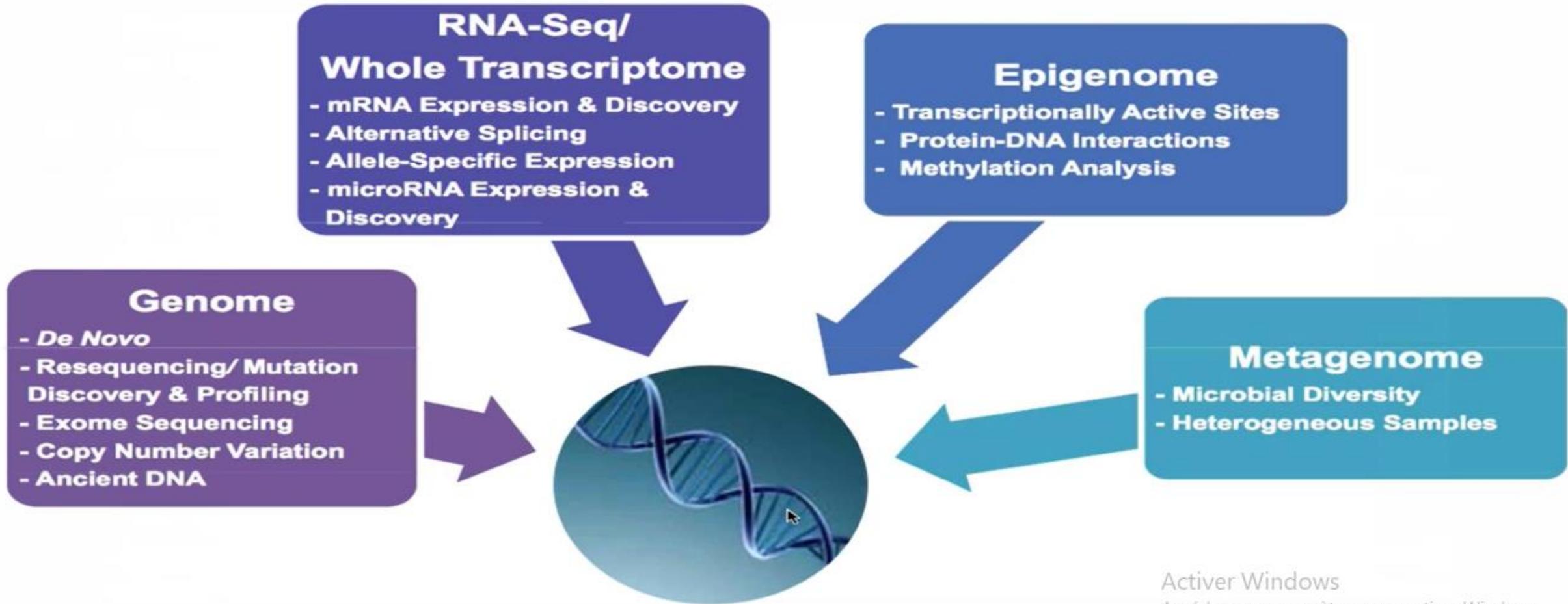
- Médecine personnalisée
 - Conception de traitement taillé sur le profil génétique d'un patient
- Génomique du Cancer
 - Qu'est ce qui retourne une cellule normale en cellule cancéreuse
- Epidémiologie/Médecine légale
 - Epidémie avec flambée de pathogène, criminologie
- Métagénomique
 - Le corps humain est un réservoir d'un nombre important et diversifié de microbes
- Amélioration de la production agricole et animale

Applications: Cas particuliers du SmidgION.

conçu pour être utilisé avec des smartphones ou d'autres appareils mobiles de faible puissance. Il est conçu pour permettre un large éventail d'analyses sur le terrain ;

- ✓ comprendre la surveillance à distance d'agents pathogènes lors d'une épidémie ou d'une maladie infectieuse
- ✓ l'analyse sur site d'échantillons environnementaux tels que des échantillons d'eau ou de métagénomique
- ✓ l'identification en temps réel d'espèces pour l'analyse d'aliments, de bois, d'animaux sauvages ou même d'échantillons inconnus,
- ✓ l'analyse sur le terrain d'environnements agricoles, et bien plus encore.

Applications



Activer Windows
Accédez aux paramètres pour activer Windows.

Évaluation d'un séquenceur

Les capacités d'un séquenceur sont définies par :

- ✓ La longueur des reads produits (L)
- ✓ Le nombre de reads produits (n)
- ✓ Le nombre de nucléotides lu: (L x n)
- ✓ Le temps de séquençage
- ✓ La qualité du séquençage

Les séquenceurs NGS



-Transformer le séquençage en combinant débit, flexibilité et facilité d'utilisation pour pratiquement toutes les méthodes, le génome et l'échelle

Figure 1: The NovaSeq 6000 System

Le système NovaSeq 6000 ouvre une nouvelle ère en matière de séquençage grâce à des innovations révolutionnaires, offrant aux utilisateurs. Le système NovaSeq 6000 représente la plateforme de séquençage Illumina à haut débit la plus puissante, la plus simple, la plus évolutive et la plus fiable à ce jour, produisant des données d'une qualité exceptionnelle.

Avantages technologiques:

- ✓ Le débit, la vitesse et la flexibilité nécessaires pour mener à bien des projets plus rapidement et plus économiquement que jamais auparavant.
- ✓ Une imagerie haute performance à la toute dernière technologie des cellules à flux structuré pour augmenter massivement le débit.
- ✓ Des optiques supérieures offrent un balayage à haute résolution et à grande vitesse, ce qui contribue à faire du système NovaSeq 6000 la plateforme de séquençage Illumina à plus haut débit.

Current NGS Technology

- Illumina HiSeq 2500:
 - 1.5 billion reads per sample
2 X 100 bases = **450 Gbp**
 - or 400 million 2 x 50 bp reads
in 16 hours = **30 Gbp)**
 - Reagent cost ~\$8K per run
- Other machine vendors have fallen behind
- New technologies constantly in development by many companies



Activer Windows
Accédez aux paramètres pour activer Windows.



454 (Roche)

**5500 W Series
Genetic
Analysis
Systems**



**Ion Proton System
for Next-Generation
Sequencing**



Activer Windows
Accédez aux paramètres pour activer Windows

Conclusion et défis bioinformatiques

- ✓ La bio-informatique est essentielle au développement des NGS (Très peu d'investigateurs sont formés)
- ✓ L'analyse des données et la maîtrise des interprétations des données brutes
- ✓ Plus le débit des données augmente, plus le temps consacré l'analyse des données augmentent
- ✓ Rapide changement dans les plateformes d'analyse
- ✓ Evolution rapide des méthodes d'analyse
- ✓ Taille énorme des données (expérimentales et référence)

MERCI