

UE 1B :  
Biomolécules, génome, bioénergétique,  
métabolisme

**ACTUALISATION**  
Fiche de cours n°25

**Approches pan-génomiques :  
Bio-informatique**

- ☆☆ Notion tombée 1 fois au concours
- ☆☆ Notion tombée 2 fois au concours
- ☆☆☆ Notion tombée 3 fois ou plus au concours

LES APPLICATIONS PRINCIPALES DU NGS	
	<ul style="list-style-type: none"> <li>▪ A visée diagnostique</li> <li>▪ Oncogénétique ou génétique constitutionnelle</li> <li>○ Diagnostic moléculaire des syndromes génétiques</li> <li>▪ Génétique tumorale, maladies métaboliques :               <ul style="list-style-type: none"> <li>○ Cibles : mutations causales</li> <li>○ Panel de gènes : un même syndrome peut être lié à des mutations causales de gènes différents</li> </ul> </li> </ul>
	<ul style="list-style-type: none"> <li>▪ A visée plutôt de recherche</li> <li>▪ Description de paysages mutationnels ou mutational landscape               <ul style="list-style-type: none"> <li>○ Grâce à la profondeur du séquençage, on a accès au % de chaque mutation dans une tumeur</li> <li>○ Permet de reconstruire l'arbre phylogénétique du développement tumoral</li> </ul> </li> <li>▪ Séquençage de toutes les séquences codant des protéines ou exome</li> <li>▪ Séquençage du génome entier</li> <li>▪ Exemples en oncologie :               <ul style="list-style-type: none"> <li>○ Cancers du sein et des ovaires héréditaire : portrait moléculaire de l'évolution tumorale</li> <li>○ On focalise sur 25 gènes grâce à des amorces spécifiques</li> </ul> </li> </ul>
	<ul style="list-style-type: none"> <li>▪ Etude de la variabilité génétique : génotypage, SNP, haplotypage</li> <li>▪ A visée plutôt de recherche</li> <li>▪ Dans une population « tout-venant »</li> <li>▪ Découverte fortuite de mutations causales</li> </ul>
	<ul style="list-style-type: none"> <li>▪ A visée diagnostic</li> <li>▪ Le NGS est capable de détecter des gains ou des pertes d'ADN comme en CGH-array grâce à sa profondeur de séquençage</li> <li>▪ CGH-array non indiqué dans les remaniements saufs les translocations</li> </ul>
	<ul style="list-style-type: none"> <li>▪ Séquençage de l'ensemble des transcrits d'une cellule</li> </ul>

INFORMATIQUE ET BIOLOGIE MOLÉCULAIRE	
	<ul style="list-style-type: none"> <li>▪ Avec le développement des techniques pan-génomiques, le biologiste n'est plus capable d'interpréter seul le résultat de ses expériences : micro-arrays, SNG...</li> </ul>
	<ul style="list-style-type: none"> <li>▪ Utilisation par le biologiste de l'outil informatique</li> <li>▪ Les bio-informaticiens :               <ul style="list-style-type: none"> <li>○ Développent des outils informatiques : science de l'information</li> <li>○ Ces outils permettent d'extraire une signification biologique de résultats expérimentaux : science du vivant</li> </ul> </li> </ul>
	<ul style="list-style-type: none"> <li>▪ EBI : European Bioinformatics Institute, à Cambridge en Angleterre</li> <li>▪ NCBI : National Center for Biotechnology Information, à Bethesda aux USA</li> </ul>

TRIÉ L'INFORMATION		
	Informations en ligne	<ul style="list-style-type: none"> <li>Retrouver des informations stockées en banque de données               <ul style="list-style-type: none"> <li>Interrogations par des requêtes</li> <li>Soumission</li> <li>Maintenance</li> </ul> </li> </ul>
	Dry lab	<ul style="list-style-type: none"> <li>Biologie sèche</li> </ul>
	Compilateurs de bases de données	<ul style="list-style-type: none"> <li>Méta-moteurs ou méta-banques de données</li> <li>Ils répertorient l'ensemble des informations au niveau génétique               <ul style="list-style-type: none"> <li>Entrez NCBI</li> <li>Genecards</li> <li>Genome Browser</li> </ul> </li> <li>Le journal NAR répertorie l'ensemble des bases de données disponibles sur la génomique et la génétique moléculaire</li> </ul>
	Recherche bibliographique	<ul style="list-style-type: none"> <li>PubMed</li> </ul>
	Sites données spécialisées	<ul style="list-style-type: none"> <li>Séquences des gènes :               <ul style="list-style-type: none"> <li>Refseq</li> <li>GenBank</li> </ul> </li> <li>Variations polymorphes de l'ADN :               <ul style="list-style-type: none"> <li>dbSNP</li> </ul> </li> <li>Séquences et fonctions des protéines :               <ul style="list-style-type: none"> <li>UniProt</li> </ul> </li> </ul>

BIO-INFORMATIQUE LIÉE AU NGS : FICHIERS DE SORTIE	
	<ul style="list-style-type: none"> <li>Fichiers de sortie de run : lecture des données des machines de séquençage</li> <li>Fichiers de séquence : Fastq               <ul style="list-style-type: none"> <li>À chaque nucléotide est associé une donnée de qualité</li> </ul> </li> </ul>
	<ul style="list-style-type: none"> <li>Alignement sur le génome de référence : fichiers Bam</li> <li>Détection des variants par rapport au génome de référence : fichiers Vcf</li> </ul>
	<ul style="list-style-type: none"> <li>Annotation des variants</li> <li>Interprétation : permet d'extraire la signification biologique du résultat du séquençage</li> </ul>
	<ul style="list-style-type: none"> <li>Exemples               <ul style="list-style-type: none"> <li>IGV</li> <li>Alamut</li> </ul> </li> </ul>

STOCKER L'INFORMATION : MÉTA-BANQUES DE DONNÉES	
	<ul style="list-style-type: none"> <li>▪ Localisation : NCBI – USA</li> <li>▪ Compilateur de base de données : renvoie à des sites spécifiques</li> <li>▪ <b>Banques de données universelles</b> ⚡ :</li> <li>○ Génome <ul style="list-style-type: none"> <li>– Gène</li> <li>– Localisation chromosomique</li> <li>– Séquence génomique</li> </ul> </li> <li>○ Transcrits dont épissages alternatifs</li> <li>○ Protéines <ul style="list-style-type: none"> <li>– Description</li> <li>– Localisation tissulaire et cellulaire</li> <li>– Anticorps...</li> </ul> </li> <li>○ Fonctions, voies fonctionnelles, classification « ontologique »</li> <li>○ Dysfonction, pathologies</li> <li>○ Références bibliographiques</li> </ul>
	<ul style="list-style-type: none"> <li>▪ <b>Photographie d'un gène</b> ⚡ : toutes les informations connues ⚡ et catégorisées</li> <li>▪ Compilateur de bases de données</li> </ul>
	<ul style="list-style-type: none"> <li>▪ Localisation : Université de Californie, Stanford, USA</li> <li>▪ Outil de cartographie du génome ⚡⚡⚡ : <ul style="list-style-type: none"> <li>○ Représentation graphique du gène, à l'échelle</li> <li>○ <b>Structure du gène</b> ⚡ : exons et introns</li> </ul> </li> </ul>

BIBLIOGRAPHIE	
	<ul style="list-style-type: none"> <li>▪ Affiche les articles scientifiques ⚡ mentionnant le mot clef recherché</li> <li>▪ Classe les publications par ordre décroissant de date de sortie : la plus récente en premier ⚡</li> </ul>

SITES SPÉCIALISÉS NUCLÉOTIDES	
	<ul style="list-style-type: none"> <li>▪ DDBJ au Japon, la moins utilisée</li> <li>▪ European Nucleotide Archive ENA en Europe</li> <li>▪ GenBank aux USA, par le NCBI, la plus utilisée</li> <li>▪ <b>Sites qui rassemblent toutes les séquences connues des gènes de toutes les espèces</b></li> </ul>
	<ul style="list-style-type: none"> <li>▪ RefSeq : <ul style="list-style-type: none"> <li>○ NCBI</li> <li>○ Fournit la <b>séquence de référence</b> ⚡⚡ : un identifiant unique à une position nucléotidique couverte par plusieurs séquences primaires</li> </ul> </li> <li>▪ Genatlas : <ul style="list-style-type: none"> <li>○ France ⚡</li> <li>○ <b>Structure intron-exon d'un gène</b> ⚡⚡⚡</li> </ul> </li> </ul>

SITES SPÉCIALISÉS POLYMORPHISMES	
	<ul style="list-style-type: none"> <li>▪ Correspond à une variation d'une à quelques bases de l'ADN sans retentissement sur la structure, la fonction ou l'expression de la protéine</li> <li>▪ Définit un allèle : version alternative d'un gène</li> <li>▪ Le pourcentage de détection du polymorphisme dans une population correspond à une fréquence allélique</li> <li>▪ Polymorphisme simple nucléotidique (SNP) : seul un nucléotide varie à une position donnée</li> </ul>
	<ul style="list-style-type: none"> <li>▪ NCBI – USA</li> <li>▪ Répertorie l'ensemble des variations des gènes à une position donnée 🌐</li> </ul>

SITES SPÉCIALISÉS PROTÉINES	
	<ul style="list-style-type: none"> <li>▪ Localisation : Europe</li> <li>▪ <b>Trouver la protéine de référence</b></li> <li>▪ Liste de protéines de toutes les espèces ou possédant le mot clé dans son nom</li> <li>▪ <b>Genecards permet de choisir la bonne référence protéique</b></li> </ul>
	<ul style="list-style-type: none"> <li>▪ Séquence</li> <li>▪ Structure 🌐</li> <li>▪ Fonction 🌐🌐</li> <li>▪ Localisation 🌐</li> <li>▪ Spécificité d'expression</li> <li>▪ Localisation tissulaire</li> <li>▪ Modification post-traductionnelle</li> </ul>

BANQUES DE DÉPÔT	
	<ul style="list-style-type: none"> <li>▪ <b>Déposer le résultat d'expériences</b> dans le cadre académique</li> <li>▪ <b>Partage</b> : utilisation par d'autres équipes</li> <li>▪ <b>Validation <i>in silico</i></b> : partage de données et comparaison</li> </ul>
	<ul style="list-style-type: none"> <li>▪ <b>GEO</b> : Gene Expression Omnibus, NCBI <ul style="list-style-type: none"> <li>○ Résultats CGH-array et biopuces d'expression</li> </ul> </li> <li>▪ <b>Array-express</b></li> <li>▪ Donnent des informations sur : <ul style="list-style-type: none"> <li>○ La plateforme où a été faite l'expérimentation par biopuce</li> <li>○ Le code de la série expérimentale pour le sujet donné</li> <li>○ Les résultats de chaque expérience d'hybridation</li> </ul> </li> </ul>

APPLICATIONS PRATIQUES ADN	
	<ul style="list-style-type: none"> <li>▪ <b>RefSeq</b> pour la séquence de référence               <ul style="list-style-type: none"> <li>○ Choisir "Nucleotide"</li> <li>○ Donner le nom du gène</li> </ul> </li> <li>▪ <b>Trier par des filtres :</b> <ul style="list-style-type: none"> <li>○ Espèce</li> <li>○ Demander une séquence de référence</li> </ul> </li> <li>▪ On obtient des séquences de référence en ADN génomique, en ARN...</li> <li>▪ Les contigs sont également référencés               <ul style="list-style-type: none"> <li>○ Ensemble de séquences assemblées qui ont servies à déduire la séquence de référence</li> </ul> </li> <li>▪ <b>Tri de l'information nécessaire :</b> <ul style="list-style-type: none"> <li>○ D'autres gènes possédant le mot clef dans leur titre peuvent également apparaître : gènes dont le produit interagit avec la requête par exemple</li> </ul> </li> <li>▪ Une fois choisie la bonne référence, on arrive sur <b>Genbank</b> qui donne la séquence de référence</li> </ul>
	<ul style="list-style-type: none"> <li>▪ <b>Genatlas :</b> <ul style="list-style-type: none"> <li>○ Aller dans Gene Database et donner le nom du gène</li> <li>○ On obtient une représentation graphique du gène</li> <li>○ Choisir de voir les exons</li> <li>○ On obtient le codage des RefSeq et la longueur des fragments des RefSeq</li> </ul> </li> <li>▪ Codage de couleur des séquences proposées :               <ul style="list-style-type: none"> <li>○ Partie d'introns encadrant les exons : en bleu minuscule</li> <li>○ Exons : noir majuscule</li> <li>○ Séquences codantes : en gras</li> <li>○ Codon ATG initiateur : rouge</li> </ul> </li> </ul>

APPLICATIONS PRATIQUES TROUVER DES AMORCES DE PCR	
	<ul style="list-style-type: none"> <li>▪ Cribler et/ou séquencer un gène ou un exon à la recherche de mutations ponctuelles délétères</li> <li>▪ Pour un diagnostic mutationnel et/ou biologique</li> <li>▪ Dans les produits de la PCR, la taille des amorces est incluse</li> <li>▪ Les amorces doivent être situées 30 à 50 nucléotides en amont pour l'amorce sens et 30 à 50 nucléotides en aval pour l'amorce antisens</li> </ul>
	<ul style="list-style-type: none"> <li>▪ <b>Pubmed</b> avec le « nom du gène » et « PCR » comme mots d'entrée</li> </ul>
	<ul style="list-style-type: none"> <li>▪ <b>Genatlas</b> : pour obtenir la séquence de l'exon à amplifier et de la séquence environnante</li> <li>▪ <b>Primer 3</b> : pour déterminer la séquence des amorces :               <ul style="list-style-type: none"> <li>○ On copie la séquence de l'exon avec les séquences introniques adjacentes</li> <li>○ Des symboles « &lt; &gt; » sont placés de part et d'autre de la séquence exonique pour interdire que les amorces soient dans cette zone</li> <li>○ Fabrication d'un primer ou amorce forward ou sens et d'un primer reverse ou antisens</li> <li>○ <b>Les amorces générées doivent avoir un Tm proche</b></li> </ul> </li> <li>▪ <b>Alamut</b> : visualiser la représentation du gène et la position des amorces créées               <ul style="list-style-type: none"> <li>○ Permet de vérifier les possibles polymorphismes et les mutations délétères, notamment au niveau de la position des amorces</li> </ul> </li> </ul>
	<ul style="list-style-type: none"> <li>▪ <b>Blast</b> : Comparaison de séquences               <ul style="list-style-type: none"> <li>○ Aligner la séquence de chacune des amorces afin de voir si la bonne séquence nucléotidique est retrouvée et voir si elle est unique</li> <li>○ Par alignement de séquences et code couleur des scores d'alignement</li> <li>○ Le classement se fait par ordre décroissant de similarité</li> </ul> </li> <li>▪ <b>Genbank</b> : on peut vérifier l'emplacement des 2 amorces sur la séquence génomique</li> </ul>

APPLICATIONS PRATIQUES PATHOLOGIES	
	<ul style="list-style-type: none"> <li>▪ <b>COSMIC</b> : Catalogue of Somatic Mutations in Cancer</li> <li>▪ Représentation des mutations le long de la séquence codante du gène.</li> <li>▪ Limitation possible à un seul exon</li> <li>▪ Répartition des mutations dans les cancers humains par tissus</li> </ul>
	<ul style="list-style-type: none"> <li>▪ OMIM : Online Mendelian Inheritance in Men</li> <li>▪ Classification à partir du gène ou des pathologies</li> <li>▪ Donne une importante somme d'informations sur le gène et la protéine <ul style="list-style-type: none"> <li>○ Historique des clonages</li> <li>○ Structure du gène</li> <li>○ Localisation</li> <li>○ Fonction</li> <li>○ Génétique moléculaire : données de pathologies</li> <li>○ Variants alléliques</li> <li>○ Références</li> <li>○ Contributeurs</li> <li>○ Date de création et historique.</li> </ul> </li> </ul>
	<ul style="list-style-type: none"> <li>▪ <b>Orphanet</b> 🌐🌐, France</li> <li>▪ Permet de trouver une consultation en Europe</li> <li>▪ Pas d'information sur le gène en 1<sup>ère</sup> intention 🌐 mais des informations liées à la pathologie : <ul style="list-style-type: none"> <li>○ Résumé apportant toutes les informations nécessaires et fiables, écrit par un expert</li> <li>○ Description du gène</li> <li>○ Classification</li> <li>○ Publications</li> <li>○ Centres experts : Consultations et conseils en Europe</li> <li>○ Diagnostic</li> <li>○ Associations de malades</li> <li>○ Thérapies ciblées, médicaments</li> </ul> </li> </ul>