

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

INFORMATIQUE ET BIOLOGIE MOLÉCULAIRE

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

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

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

STOCKER L'INFORMATION : MÉTA-BANQUES DE DONNÉES

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

BIBLIOGRAPHIE

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

SITES SPÉCIALISÉS NUCLÉOTIDES

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

**SITES SPÉCIALISÉS
POLYMORPHISMES**

	<ul style="list-style-type: none">■ 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■ Définit un allèle : version alternative d'un gène■ Le pourcentage de détection du polymorphisme dans une population correspond à une fréquence allélique■ Polymorphisme simple nucléotidique (SNP) : seul un nucléotide varie à une position donnée
	<ul style="list-style-type: none">■ NCBI – USA■ Répertorie l'ensemble des variations des gènes à une position donnée 

**SITES SPÉCIALISÉS
PROTÉINES**

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

BANQUES DE DÉPÔT

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

APPLICATIONS PRATIQUES ADN

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

APPLICATIONS PRATIQUES TROUVER DES AMORCES DE PCR ⓘ

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

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