

Intitulé de l'UE	Langage R et bioinformatique
Section(s)	- (4 ECTS) Master en Sciences de l'Ingénieur industriel orientation Life data technologies / Cycle 2 Bloc 1

Responsable(s)	Heures	Période
Vincent BRANDERS	50	Quad 1

Activités d'apprentissage	Heures	Enseignant(s)
Application du langage R à la transcriptomique	30h	Axel THIEFFRY
Introduction à la programmation R	20h	Vincent BRANDERS

Prérequis	Corequis

Répartition des heures
Application du langage R à la transcriptomique : 10h de théorie, 20h d'exercices/laboratoires
Introduction à la programmation R : 5h de théorie, 15h d'exercices/laboratoires

Langue d'enseignement
Application du langage R à la transcriptomique : Français, Anglais
Introduction à la programmation R : Français, Anglais

Connaissances et compétences préalables
- Principes de bases en biologie moléculaire et technologies de séquençage haut-débit (Illumina) - Utiliser la ligne de commande (bash) dans un système LINUX/UNIX

Objectifs par rapport au référentiel de compétences ARES
Cette UE contribue au développement des compétences suivantes
<ul style="list-style-type: none"> • Identifier, conceptualiser et résoudre des problèmes complexes • Concevoir et gérer des projets de recherche appliquée • Communiquer face à un public de spécialistes ou de non-spécialistes, dans des contextes nationaux et internationaux • S'engager dans une démarche de développement professionnel

Objectifs de développement durable

Aucun

Acquis d'apprentissage spécifiques

- Produire des figures scientifiques pertinentes et attractives en terme de visualisation mais également en terme de communication effective de résultats
- Familiarisation au langage de programmation R, le Tidyverse, et Bioconductor, et leurs utilisation dans le cadre d'analyses de données de séquençage
- Développer une capacité critique vis-à-vis de résultats de l'analyse de données CAGE et de leur distinction par rapport aux méthodes classiques (mRNA-seq)

Contenu de l'AA Application du langage R à la transcriptomique

Au travers d'une analyse complète d'un jeu de données de type CAGE (Cap Analysis of Gene Expression):

1. Nettoyage de données et contrôle qualité (LINUX/Unix, command line, bash/zsh)
2. Analyses bioinformatiques des données nettoyées: mapping, quantification, annotation
3. Quantification et expression différentielle (R, Tidyverse, Bioconductor)
4. Présentation des résultats sous la forme de figures scientifiques

Contenu de l'AA Introduction à la programmation R

Au terme de l'AA Programmation R, l'étudiant sera capable :

- d'expliquer la syntaxe de base de R et d'utiliser des concepts tels que les boucles, fonctions et apply(s)
- d'identifier et de manipuler les types de données courants en R
- d'importer, lire et écrire des données dans différents formats (CSV, TSV, Excel, ...)
- d'appliquer des opérations de base sur les données, telles que la sélection, le filtrage et le tri
- de créer des graphiques de base pour explorer et représenter des données et d'utiliser des packages tels que ggplot2 pour créer des visualisations plus avancées
- de réaliser de la manipulation avancée de données
- de respecter les conventions de codage R et de maintenir un code propre et bien documenté
- de collaborer efficacement en utilisant des outils de gestion de versions comme Git

Méthodes d'enseignement

Application du langage R à la transcriptomique : travaux de groupes, approche par projets, approche interactive, approche par situation problème, approche déductive, étude de cas, utilisation de logiciels

Introduction à la programmation R : travaux de groupes, approche par projets, approche interactive, approche par situation problème

Supports

Application du langage R à la transcriptomique : copies des présentations, notes d'exercices, Ressources et scripts disponible via un répertoire GitHub public

Introduction à la programmation R : copies des présentations, notes d'exercices

Ressources bibliographiques de l'AA Application du langage R à la transcriptomique

- Collection de "cheat sheets" pour la programmation R, la manipulation de données, et la production de figures scientifiques: <https://www.rstudio.com/resources/cheatsheets/>
- Guide pour la création et validation de figures scientifiques: <https://www.data-to-viz.com>
- Répertoire GitHub public créé spécifiquement pour le cours

Ressources bibliographiques de l'AA Introduction à la programmation R

- Collection de "cheat sheets" pour la programmation R, la manipulation de données, et la production de figures scientifiques: <https://www.rstudio.com/resources/cheatsheets/>
- Guide pour la création et validation de figures scientifiques: <https://www.data-to-viz.com>
- Cours en ligne de datacamp
- Golemund, Garrett, and Hadley Wickham. 2017. R for Data Science. Sebastopol, CA: O'Reilly Media.

Évaluations et pondérations	
Évaluation	Note globale à l'UE
Langue(s) d'évaluation	Français, Anglais
Méthode d'évaluation	<p>La note finale est définie comme suit :</p> <ul style="list-style-type: none"> • 60 % pour l'AA <i>Compléments de techniques bioinformatiques 2</i> <ul style="list-style-type: none"> ◦ Une évaluation de la participation tout au long du cours (présence, attention, motivation, questions, réponses, etc.) ◦ Complétion d'un examen dont le format (individuel/groupe ; interne ou <i>via</i> service tiers) sera défini en fonction de l'avancement par rapport au programme prévu • 40 % pour l'AA <i>Programmation R</i> : <ul style="list-style-type: none"> ◦ Travail continu (certifications DataCamp) ◦ Réalisation d'un projet
Report de note d'une année à l'autre pour l'AA réussie en cas d'échec à l'UE	
Application du langage R à la transcriptomique : oui Introduction à la programmation R : oui	

Année académique : **2024 - 2025**