

| | | |
|--------------|----------------|----|
| 5.00 crédits | 30.0 h + 9.5 h | Q2 |
|--------------|----------------|----|

| | |
|---|--|
| Enseignants | Symul Laura ; |
| Langue d'enseignement | Français > English-friendly |
| Lieu du cours | Louvain-la-Neuve |
| Préalables | Concepts et outils équivalents à ceux enseignés dans les UEs LSTAT2020 Logiciels et programmation statistique de base LSTAT2110 Analyse des données |
| Thèmes abordés | <ul style="list-style-type: none"> • Introduction aux données omiques (définitions, motivations pour leur génération, exemples) • Caractéristiques statistiques des données omiques (type et nature des données, distributions typiques, etc.) • Méthodes de correction pour tests multiples • Détails des protocoles expérimentaux et des méthodes de prétraitement et d'analyses de données omiques les plus courantes (transcriptomique, métabolomique, protéomique, metagenomique, cytométrie de flux, et single-cell transcriptomics) • Revues des méthodes d'analyses multivariées supervisées (classification et régression, PLS(-DA), O-PLS, Lasso & ridge regression, SVM) et non-supervisée (PCA, MDS, clustering) et des modèles à composantes de variances (ASCA, APCA). • Les méthodes d'intégration de données (analyse de données multitableaux) • Les méthodes mathématiques et statistiques pour le prétraitement de données spectrales (ex : modèles semi-paramétrique de lissage pour correction de ligne de base, alignement de pics) • Méthodes de correction d'effets de batch et planification expérimentale pour les éviter. • Revue et utilisation des packages R pour l'analyse des données omiques (typiquement les packages BioConductor) • Application sur des données réelles. |
| Acquis d'apprentissage | <p>A la fin de cette unité d'enseignement, l'étudiant est capable de :</p> <p>Eu égard au référentiel AA du programme « Le master en statistique, orientation biostatistique », cette activité permet aux étudiants de maîtriser</p> <ul style="list-style-type: none"> • De manière prioritaire les AA suivants: 1.4, 1.6, 2.2, 2.4, 2.5, 2.6, 3.1, 3.2, 3.3, 3.4, 3.5, 4.3, 4.5, 5.4, 5.6 • De manière secondaire les AA suivants: 5.2, 5.7 <p>1</p> <p>Eu égard au référentiel AA du programme de « Le master en statistique, orientation générale », cette activité permet aux étudiants de maîtriser</p> <ul style="list-style-type: none"> • De manière prioritaire les AA suivants : 1.4, 1.6, 2.2, 2.4, 2.5, 2.6, 3.1, 3.2, 3.3, 3.4, 4.3, 5.3 |
| Modes d'évaluation des acquis des étudiants | <p>Dans le cadre de ce cours, les étudiant-es sont évalué-es de deux manières :</p> <ul style="list-style-type: none"> • l'évaluation continue certificative incluant: <ul style="list-style-type: none"> • des travaux obligatoires à remettre en cours de quadrimestre selon un calendrier fixé en début de quadrimestre (40% de la note finale) • et un projet final à présenter oralement lors du dernier cours (40% de la note finale) • un examen oral à livre ouvert (20% de la note finale) <p>Une note globale de 10/20 ou plus ne sera acquise que si l'étudiant a obtenu pour chaque partie de l'évaluation (travaux obligatoires, projet final, et examen) une note individuelle de minimum 8/20. Dans le cas contraire, la note minimale prévaut.</p> |
| Méthodes d'enseignement | <p>Le cours est composé d'une série d'activité qui amènent l'étudiant à se plonger activement dans le monde des données -omiques. Il propose:</p> <ul style="list-style-type: none"> • des exposées par des spécialistes actifs dans le domaine, • des mini-projets de traitement de données à réaliser chaque semaine, • un projet final sur des données proposées par les différents intervenants du cours ou de data repositories. |
| Contenu | <p>La plupart des chapitres sont construits autour d'une méthode omique d'intérêt et introduisent parallèlement les concepts statiques nécessaires à l'analyse des données générées. Une ou deux études de cas seront présentées pour chaque chapitre et la plupart sont suivis d'un mini-projet (lui-même consistant en une petite analyse de cas) à réaliser seul ou en binome.</p> |

Partim A (3C, 20h + 8h)

Introduction

- Révision des bases de biologie moléculaire
- Introduction aux données omiques (Pourquoi? Comment? Quoi?)
- Reproductibilité scientifique et technique. Outils et bonnes pratiques pour améliorer la reproductibilité.

Genomique et transcriptomique, méthode de correction pour tests multiples, régression robuste

- Méthodes de quantifications d'ADN et d'ARN (qPCR, micro-array, séquençage)
- Méthodes de correction pour tests multiples
- Calibration, contrôles positifs et négatifs, régression robuste
- Modélisation statistique pour l'analyse d'expression différentielle entre 2 ou plusieurs conditions (glm, negative binomial glm, etc).

Métagenomique, analyses statistiques de données "écologiques",

- Introduction aux microbiotes et aux méthodes de génération de données pour la quantification de la composition des microbiotes
- Distances adaptées pour les données écologiques
- Dimension de réduction et projections de matrices de distance (MDS)

Identification de biomarqueurs ou de caractéristiques spectrales associés à un diagnostic ou état clinique

- Méthodes de classifications multivariées
- Méthodes de régression multivariées

Proteomique et gestion des données manquantes

- Méthodes d'acquisition de données protéomiques
- Méthodes de prétraitement et d'analyse de données protéomiques
- Mécanismes de génération des données manquantes et méthodes d'imputation

Partim B (2C, 10h + 2h)

Metabolomique et méthodes d'analyses de données spectrale, méthodes d'analyse de variance multivariée

- Introduction à la métabolomique et méthodes de quantification des métabolites
- Focus sur les méthodes 1H-RMN et les méthodes de pré-traitement associées
- Méthodes d'analyse de variance multivariée (ASCA, APCA)

Effets de groupe et leur correction

- Définition des effets de groupe/lot (batch effects).
- Méthodes de correction des effets de groupe
- Prévenir la confusion entre effets techniques (e.g., batch effects) et les effets biologiques

Méthodes "single-cells" et méthodes de réduction de dimension non-linéaires

- Cytométrie de flux
 - Techniques de génération de données
 - Contrôle de qualité
 - Pré-traitement et analyse statistique des données
- Single-cell transcriptomics
 - Techniques de génération de données
 - Pré-traitement et analyse statistique des données
 - Méthode de réduction de dimension non-linéaires

Méthodes d'intégration multi-omiques

- Buts et motivations
- Méthodes d'intégration supervisées
- Méthodes d'intégration non-supervisées

Simulations de données et analyse de puissance par simulation








- Buts et motivations
- Évaluer la qualité des données simulées
- Exemples d'analyse de puissance par simulation.

Ressources en ligne

Site Moodle: <https://moodle.uclouvain.be/course/view.php?id=2964>

Faculté ou entité en charge:

LSBA

| Programmes / formations proposant cette unité d'enseignement (UE) | | | | |
|--|---------|---------|-----------|--|
| Intitulé du programme | Sigle | Crédits | Prérequis | Acquis d'apprentissage |
| Master [120] en science des données, orientation statistique | DATS2M | 5 | |  |
| Master [120] en statistique, orientation biostatistiques | BSTA2M | 5 | |  |
| Master [120] en sciences mathématiques | MATH2M | 5 | |  |
| Master [120] en statistique, orientation générale | STAT2M | 5 | |  |
| Master [120] : bioingénieur en chimie et bioindustries | BIRC2M | 5 | |  |
| Certificat d'université : Statistique et science des données (15/30 crédits) | STAT2FC | 5 | |  |
| Master [120] : bioingénieur en sciences agronomiques | BIRA2M | 5 | |  |