

Partager



Email destinataire :

Vous devez renseigner un email valide

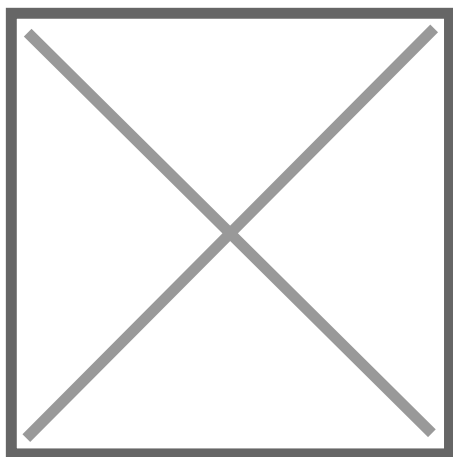
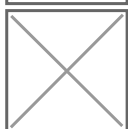
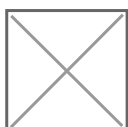
Envoyer

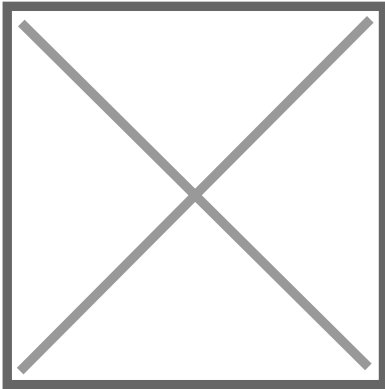
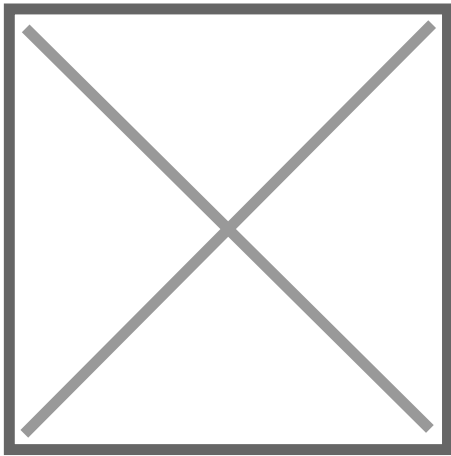
Une erreur est survenue. Merci de réessayer ultérieurement

Le mail de partage a bien été envoyé.

Ajouter au comparateur

# M1 Genomics Informatics Mathematics and Artificial Intelligence for Health and Environment GENIOMHE-AI)





Master

Mention Bio-informatique

Régime(s) d'inscription

Formation initiale Formation continue

Langue(s) d'enseignement

Anglais

Le Master GENIOMHE-AI forme des spécialistes en recherche et développement dans le domaine de la Bioinformatique pour la **génomique** et la **médecine de précision**. Cette formation permet aux étudiants de maîtriser l'**algorithmique**, divers langages et techniques de **programmation** ainsi que des techniques avancées en **science des données**, **intelligence artificielle**, **statistiques** et bases de données pour l'analyse de données massives.

Le M1 GENIOMHE-AI s'organise avec des enseignements de septembre à mars. Un stage obligatoire d'une durée de 2 mois sera réalisé à partir du mois d'avril au sein d'un laboratoire de recherche académique ou d'une entreprise.

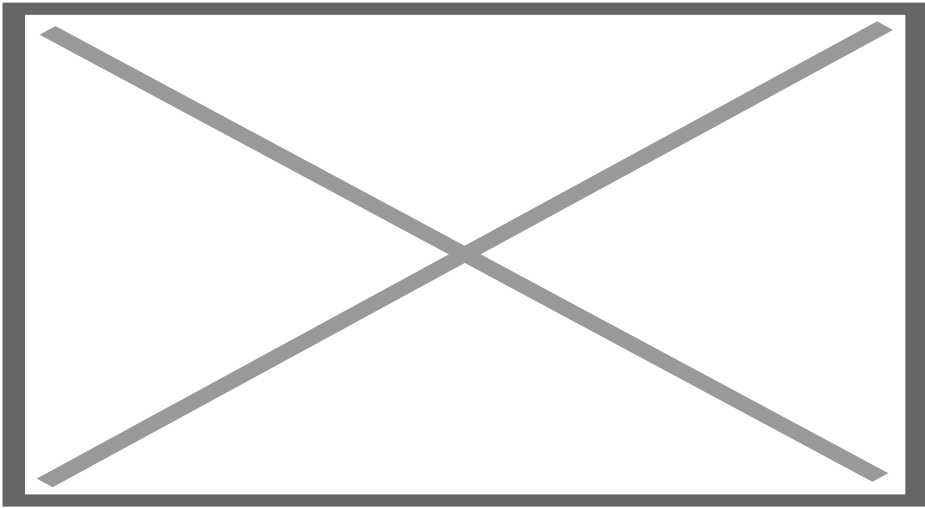
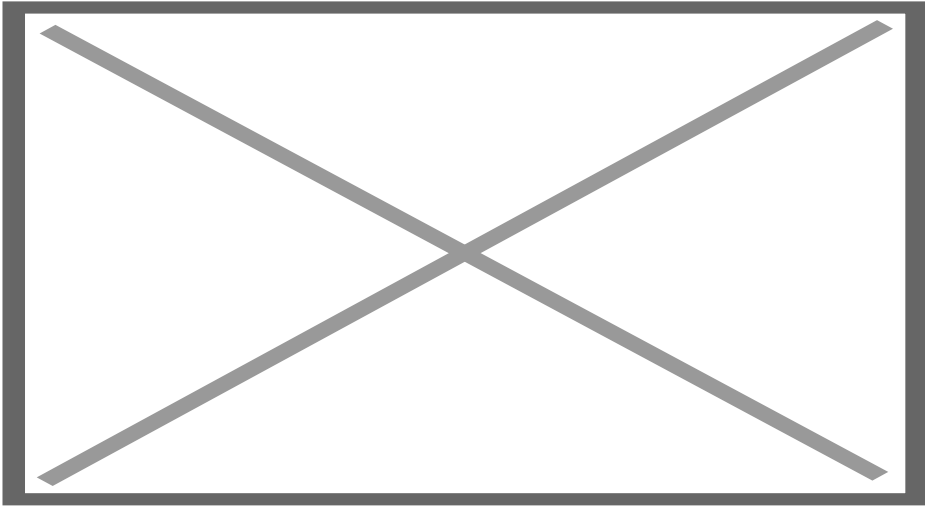
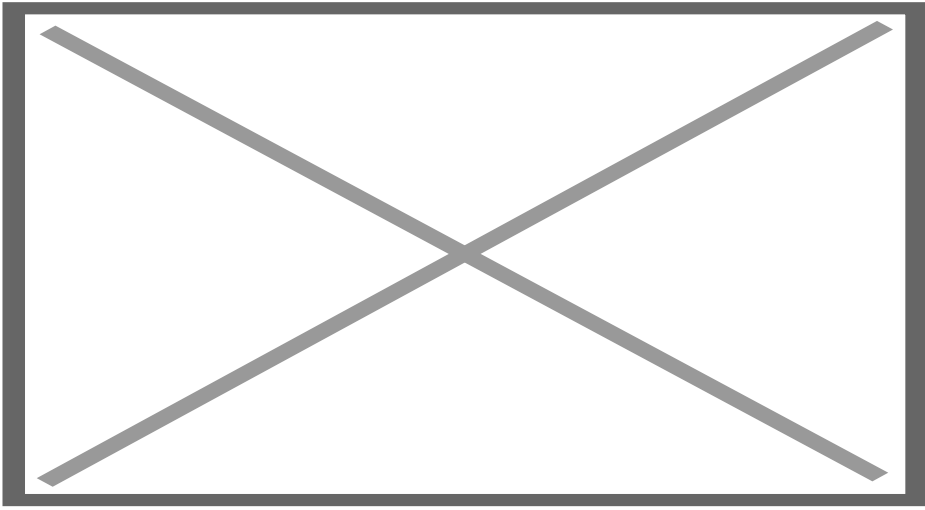
A la fin du M1, les étudiants auront acquis des compétences en génomique et génétique humaine, analyse de données génomiques, outils bioinformatiques et phylogénie, bases de données, apprentissage automatique et exploration de données, apprentissage profond, programmation impérative et orientée objet, modélisation des systèmes biologiques, statistiques multivariées.

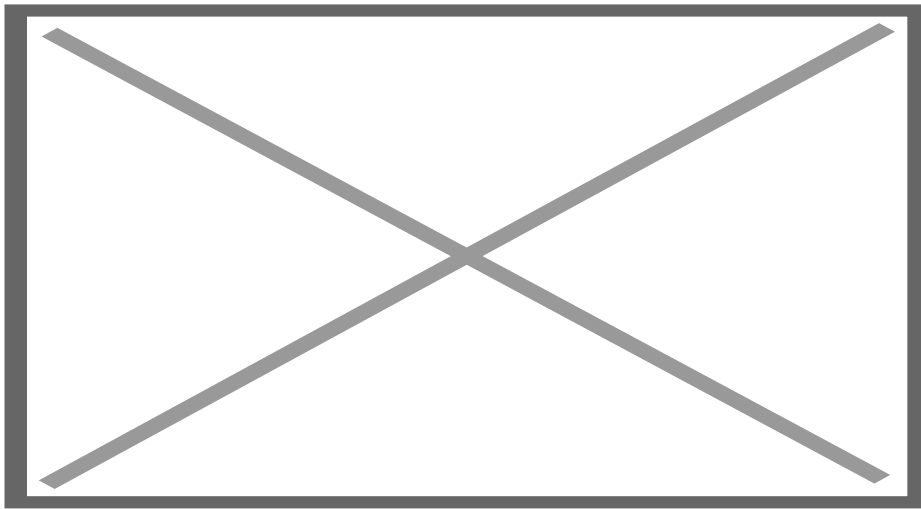
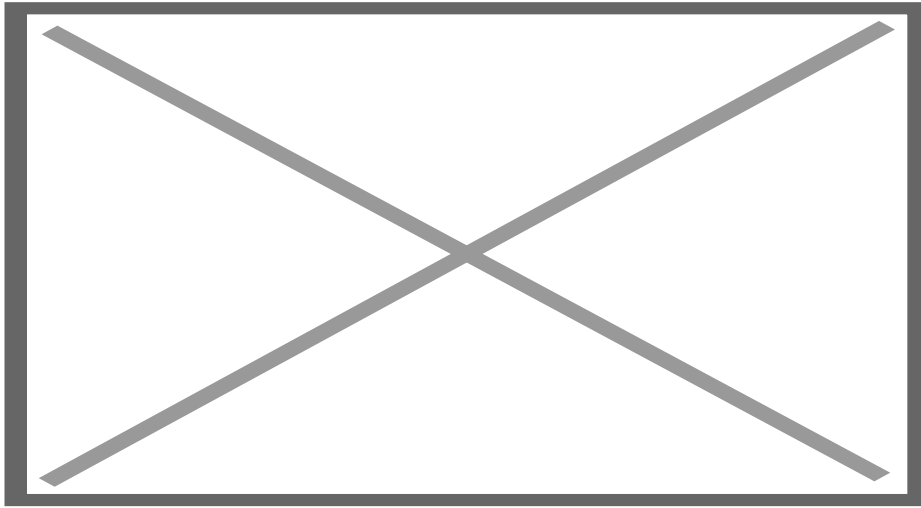
Les étudiants seront également sensibilisés aux principes FAIR et auront des projets à réaliser.

Deux modules d'ouverture seront également obligatoires : un module de développement soutenable et un module à choisir parmi une liste.

[Télécharger la fiche détaillée](#)

[Je candidate](#)





■ ■

## Informations

- [Présentation](#)
- [Admission](#)
- [Programme](#)

Présentation

## Compétences

Maîtriser les concepts fondamentaux en génomique, algorithmique et programmation, intelligence artificielle et statistiques.

## Objectifs pédagogiques de la formation

Ce master est une formation pluridisciplinaire qui permet aux étudiants de : comprendre les enjeux complexes et diversifiés de la biologie afin de répondre aux besoins actuels de recherche et d'innovation dans la recherche académique, l'industrie des biotechnologies, l'industrie pharmaceutique et la santé ; relever les nouveaux défis liés à la médecine de précision rendue possible grâce à la production rapide et massive des données génomiques ; relever les défis de l'utilisation de l'intelligence artificielle dans les domaines de la santé ; maîtriser les techniques d'analyse et de modélisation des données biomédicales ; être autonome dans la gestion de projets de développement d'applications dans divers langages de programmation, proposer des solutions informatiques innovantes.

## **Débouchés**

### **Poursuite d'études**

**Data Scientist, Data Analyst, Ingénieur·e en Machine Learning dans des secteurs innovants (tech, finance, santé, énergie, etc.) ;**

**Master 2**

**Master Bioinformatique**

### **Agenda**

Début de la formation (date indicative)

01/09/2026

Fin de la formation (date indicative)

31/08/2027

Admission

### **Voie d'accès**

Informatique

Sciences de la vie

Sciences de la vie et de la Terre

### **Capacité d'accueil**

#### **Places**

25

### **Public visé et prérequis**

- Etudiants ayant validé des connaissances et des compétences à la fois en Biologie et en Informatique au cours de leur cursus (par exemple : Double licence SDV-Informatique, licence sélective SDV-Info Paris Saclay, licence SDV parcours bioinformatique, licence informatique parcours bioinformatique, Ecole d'ingénieur....) . \* Etudiants ayant un cursus en Biologie mais avec un intérêt particulier et des connaissances et compétences complémentaires validées en informatique.

## Période(s) de candidature

Plateforme Inception

Du 15/01/2026 au 25/04/2026

## Pièces justificatives

### Obligatoires

Lettre de motivation.

Tous les relevés de notes des années/semestres validés depuis le BAC à la date de la candidature.

Attestation de niveau d'anglais.

Curriculum Vitae.

Descriptif détaillé et volume horaire des enseignements suivis depuis le début du cursus universitaire.

### Facultatives

Dossier VAPP (obligatoire pour toutes les personnes demandant une validation des acquis pour accéder à la formation) <https://www.universite-paris-saclay.fr/formation/formation-continue/validation-des-acquis-de-lexperience>.

Document justificatif des candidats exilés ayant un statut de réfugié, protection subsidiaire ou protection temporaire en France ou à l'étranger (facultatif mais recommandé, un seul document à fournir) :

- Carte de séjour mention réfugié du pays du premier asile
- OU récépissé mention réfugié du pays du premier asile
- OU document du Haut Commissariat des Nations unies pour les réfugiés reconnaissant le statut de réfugié
- OU récépissé mention réfugié délivré en France
- OU carte de séjour avec mention réfugié délivré en France
- OU document faisant état du statut de bénéficiaire de la protection subsidiaire en France ou à l'étranger.

Programme

Methods for Genomics 1

| Matières                                | ECTS | Semestre   | Cours | TD   | TP | Cours-<br>TD | Cours-<br>TP | TD-<br>TP | A<br>distance | Projet | Tutorat |
|-----------------------------------------|------|------------|-------|------|----|--------------|--------------|-----------|---------------|--------|---------|
| <a href="#">Bioinformatic<br/>tools</a> | 3    | Semestre 1 | 9     | 13.5 |    |              |              |           |               |        |         |

## Bioinformatic tools

ECTS : 3

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 9

Travaux dirigés : 13.5

Langue d'enseignement

Anglais

Enseignement à distance

non

Prérequis

Theory and practice concerning sequence data, databases and sequence formats, alignments with exact algorithms (dynamic programming) and practice of BLAST on standard databases online.

Basical knowledge on evolution principles and population genetics.

Unix/Linux concepts and utility filters.

Programme / plan / contenus

The course outline:

- Main courses:

- reminder about exact alignment algorithms and alignment scores,
- reminders of basic evolutionary principles,
- construction of scoring matrices (substitution matrices),
- BLAST algorithm, modification of parameters, statistical principles and local usage
- protocols to determine homologies

- Directed work:

- calculation of alignment scores and comparisons, alignment calculations using dynamic programming,
- interpretation of blast results (local and online)
- establishment of a bioinformatics protocol for the determination of protein families and subfamilies

Objectifs d'apprentissage

The objectives of this course are:

- to understand deeply the advantages and limits of sequence comparison through standard bio-informatic tools,
- to be able to set up bioinformatics protocols and to be critical of the established protocols,

| Matières                                | ECTS | Semestre   | Cours | TD   | TP | Cours-<br>TD | Cours-<br>TP | TD-<br>TP | A<br>distance | Projet | Tutorat |
|-----------------------------------------|------|------------|-------|------|----|--------------|--------------|-----------|---------------|--------|---------|
| <a href="#">Structural<br/>Genomics</a> | 3    | Semestre 1 | 9     | 13.5 |    |              |              |           |               |        |         |

## Structural Genomics

ECTS : 3

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 9

Travaux dirigés : 13.5

Langue d'enseignement

Anglais

Enseignement à distance

non

Prérequis

General concepts in genetics and molecular biology usually covered during a Bachelor's degree in Biology or related sciences.

Programme / plan / contenus

- Characteristics of eucaryote genome : limits / avantages regarding sequencing, assembly and annotation
- Sequencing technologies
- Dynamic and evolution of eucaryote genome ; comparative genomics
- Introduction to assembly algorithms for 2nd and 3rd generation sequencing data.
- Key metrics for assessing genome assembly quality
- Overview of scaffolding approaches to obtain chromosome level and T2T quality genome
- Introduction to structural and functional annotation of genomes.
- Practical case study in the form of a project to annotate a genomic region.

Objectifs d'apprentissage

This course provides an in-depth exploration of the structure, organization, and evolution of genomes. Structural genomics focuses on describing the physical and functional organization of genetic material. Students will learn about genome architecture, sequencing technologies, genome assembly, annotation, and comparative genomics.

The course combines theoretical concepts with practical exercises in bioinformatics, using modern databases and tools to analyze genomic structures.

Organisation générale et modalités pédagogiques

This teaching unit is organized in the form of theoretical lecture on the specificities of eucaryote genomes regarding sequencing, assembly and annotation.

| Matières                              | ECTS | Semestre   | Cours | TD   | TP | Cours-<br>TD | Cours-<br>TP | TD-<br>TP | A<br>distance | Projet | Tutorat |
|---------------------------------------|------|------------|-------|------|----|--------------|--------------|-----------|---------------|--------|---------|
| <a href="#">Genome<br/>sequencing</a> | 3    | Semestre 1 | 9     | 13.5 |    |              |              |           |               |        |         |

## Genome sequencing

ECTS : 3

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 9

Travaux dirigés : 13.5

Langue d'enseignement

Anglais

Enseignement à distance

non

Programme / plan / contenus

This course covers genome resequencing studies, from data generation to result interpretation. It explores genomic variation within populations to determine gene functions, decipher pathogenic mechanisms and reconstruct disease history, as well as identify therapeutic opportunities.

Learning objectives:

- Describe the principles of data generation using high-throughput sequencing technologies.
- Explain the steps of data quality control and preprocessing, and their relevance.
- Understand the principles behind the identification of different types of genetic variants.
- Select appropriate variant filtering parameters according to the research question.
- Interpret and critically evaluate the results of publications in functional genomics.
- List the main applications of genome sequencing data in health and precision medicine.
- Understand how to identify altered signaling pathways, biological processes, and biomarkers from gene lists using functional annotation tools.
- Specify the advantages of single-cell approaches for studying tissue complexity or sample polyclonality.

Objectifs d'apprentissage

This course provides knowledge on analyzing genome sequencing data using the most recent deep-sequencing technologies and their numerous applications to specific biological questions.

It will also train students to extract relevant data from scientific articles, identify their limitations, and synthesize them in a journal club format.

Organisation générale et modalités pédagogiques

The teaching is organized in the form of lectures, tutorial sessions involving the analysis of genomic studies based

| Matières                                        | ECTS | Semestre   | Cours | TD   | TP | Cours-<br>TD | Cours-<br>TP | TD-<br>TP | A<br>distance | Projet | Tutorat |
|-------------------------------------------------|------|------------|-------|------|----|--------------|--------------|-----------|---------------|--------|---------|
| <a href="#">Genetics of complex pathologies</a> | 3    | Semestre 1 | 9     | 13.5 |    |              |              |           |               |        |         |

## Genetics of complex pathologies

ECTS : 3

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 9

Travaux dirigés : 13.5

Langue d'enseignement

Anglais

Enseignement à distance

non

Prérequis

Fundamentals of formal genetics.

Programme / plan / contenus

### 1. Introduction

a. Types of human diseases and factors involved

b. Issues raised by multifactorial diseases

### 2. Data production methods

### 3. Data quality control

a. Missing data

b. Hardy-Weinberg model

c. Principal component analysis

d. Other

### 4. Search for causal genetic variants

a. Association analyses on family data (Transmission Disequilibrium Test)

b. Association analyses on population data (case-control studies)

c. Pan-genomic association studies

### 5. Generalizations

a. Association analysis for rare variants

b. Search for interactions

Objectifs d'apprentissage

- Identify the genetic analysis methods to be used for a given problem

- Describe methods for producing genetic data on a large scale

- Be familiar with various examples of complex pathologies

- Select the appropriate statistical tools

- Perform adequate data quality checks

- Conduct a genetic association study

- Know how to interpret the results of a genetic association study

Organisation générale et modalités pédagogiques

The course consists of lectures to learn the theoretical

## Methods for Genomics 2

| Matières                                 | ECTS | Semestre   | Cours | TD   | TP | Cours-<br>TD | Cours-<br>TP | TD-<br>TP | A<br>distance | Projet | Tutorat |
|------------------------------------------|------|------------|-------|------|----|--------------|--------------|-----------|---------------|--------|---------|
| <a href="#">Prokaryotic<br/>genomics</a> | 3    | Semestre 2 | 9     | 13.5 |    |              |              |           |               |        |         |

## Prokaryotic genomics

ECTS : 3

Semestre calendaire : Semestre 2

Détail du volume horaire

Cours magistraux : 9

Travaux dirigés : 13.5

Langue d'enseignement

Anglais

Enseignement à distance

non

Prérequis

? Basic knowledge of molecular biology, microbial genetics, and genomics

? Familiarity with sequence formats (FASTA, GFF, GenBank)

? Awareness of databases such as NCBI, UniProt, and KEGG

Programme / plan / contenus

1. Introduction to Prokaryotic Genomes

2. Genome Annotation

3. Comparative Genomics

Concepts: orthology, synteny, gene conservation, rearrangements

4. Pangenomics and Pangenome Graphs

5. Metabolic Reconstruction and Comparative Metabolism analysis

6. Integrative Project: Comparative Genomics, Pangenome, and/or Metabolic Reconstruction

Objectifs d'apprentissage

By the end of this course, students will be able to:

1. Describe the structure, organization, and variability of prokaryotic genomes
2. Perform genome annotation using reference pipelines and databases.
3. Compare genomes across species to identify orthologous and unique genes.
4. Reconstruct and analyze metabolic pathways from annotated genomes.
5. Build, interpret, and visualize pangenomes using comparative genomics tools.
6. Understand how genome diversity relates to functional and ecological adaptation

Organisation générale et modalités pédagogiques

Teaching methods:

| Matières                                  | ECTS | Semestre   | Cours | TD   | TP | Cours-<br>TD | Cours-<br>TP | TD-<br>TP | A<br>distance | Projet | Tutorat |
|-------------------------------------------|------|------------|-------|------|----|--------------|--------------|-----------|---------------|--------|---------|
| <a href="#">Methods for<br/>Phylogeny</a> | 3    | Semestre 2 | 9     | 13.5 |    |              |              |           |               |        |         |

## Methods for Phylogeny

ECTS : 3

Semestre calendaire : Semestre 2

Détail du volume horaire

Cours magistraux : 9

Travaux dirigés : 13.5

Langue d'enseignement

Anglais

Enseignement à distance

non

Programme / plan / contenus

The following topics will be covered:

- the concept of evolution, homology, orthology and paralogy,
- Phylogenetic representation (networks, trees, roots, dendrograms, topology, branch lengths)
- Phylogenetic inference methods (distances, parsimony, the principle of maximum likelihood, ...)
- Measures of the robustness of phylogenies (bootstrap, topology comparison, multigene corroboration, gene and species trees)

Applications to the phylogeny of some major taxonomic groups and to gene families.

Objectifs d'apprentissage

This course will present the fundamental concepts in molecular evolution and phylogeny.

Practical sessions and tutorials will allow students to become familiar with modern tools for reconstructing phylogenetic trees for species and/or genes.

Organisation générale et modalités pédagogiques

This teaching unit is organized in the form of main courses (9h) and practical work (13,5h) in the computer room. Students will be evaluated through several small exams in class and a final exam.

Compétences

C1.1. Formuler des problématiques et des hypothèses de travail permettant la conception et la réalisation d'un projet dans son domaine d'expertise et /ou à l'interface de plusieurs domaines en biologie.

C2.1. Actualiser ses connaissances théoriques et technologiques par une veille dans son domaine en relation avec l'état de la recherche.

C3.2. Argumenter les choix méthodologiques et les

| Matières                                                | ECTS | Semestre  | Cours | TD | TP | Cours-<br>TD | Cours-<br>TP | TD-<br>TP | A<br>distance | Projet | Tutorat |
|---------------------------------------------------------|------|-----------|-------|----|----|--------------|--------------|-----------|---------------|--------|---------|
| <a href="#">NGS data<br/>processing<br/>methodology</a> | 2    | Annualisé |       |    | 17 |              |              |           |               |        |         |

## NGS data processing methodology

ECTS : 2

Semestre calendaire : Annualisé

Détail du volume horaire

Travaux pratiques : 17

Langue d'enseignement

Anglais

Enseignement à distance

non

Prérequis

Theory concerning NGS data analysis and statistical methods for association analyses.

Programme / plan / contenus

The course outline:

- Perform all bioinformatics steps of sequence data processing (alignment, local realignment, variant identification, variant visualization with IGV)
- Define and apply adequate quality control filters at each step of data processing
- Perform and interpret descriptive statistics at each step of data processing
- Choose and apply filters for the identified variants according to the problem considered (frequency, predicted effect, ...)
- Use the appropriate statistical tools to search for genetic variants associated with a complex disease.
- Interpret the statistical results obtained
- Synthesize the results obtained during the different stages (bioinformatics and statistics).

Objectifs d'apprentissage

The objective of this practical teaching is to manipulate Next Generation Sequencing DNA sequence data in order to identify genetic variants (SNV, indels) associated with a complex disease.

Organisation générale et modalités pédagogiques

This teaching unit is organized in the form of practical work in the computer room. 2/3 will focus on the bioinformatics processing of data, 1/3 on statistical analysis. Students will be evaluated through an online QCM and a report.

Compétences

C1.4. Produire et valider des données scientifiques.

CS1.1. Structurer et organiser les données scientifiques

# IA, Computer Science and Data Sciences for Genomics 1

| Matières                                                                | ECTS | Semestre  | Cours | TD | TP | Cours-<br>TD | Cours-<br>TP | TD-<br>TP | A<br>distance | Projet | Tutorat |
|-------------------------------------------------------------------------|------|-----------|-------|----|----|--------------|--------------|-----------|---------------|--------|---------|
| <a href="#">Refresher course<br/>in Shell, Python,<br/>algorithmics</a> | 1    | Annualisé |       | 40 |    |              |              |           |               |        |         |

## Refresher course in Shell, Python, algorithmics

ECTS : 1

Semestre calendaire : Annualisé

Détail du volume horaire

Travaux dirigés : 40

Langue d'enseignement

Anglais

Enseignement à distance

non

Prérequis

Basic knowledge of scientific computing: computer operation, file hierarchy, and installation of simple software.

Basic knowledge of molecular biology and bioinformatics (sequences, FASTA/FASTQ files, etc.).

Prior experience with a programming language (even superficial, e.g., R, C, or Python).

Motivation to strengthen skills in algorithms and programming.

Programme / plan / contenus

This course is structured in two parts: Algorithmics and Programming.

A. Algorithmics:

This part will be organized as lectures/tutorials where students will write algorithms in pseudocode for searching and updating in various data structures: arrays, matrices, lists, queues, stacks, trees, and graphs.

B. Programming:

1 – Introduction and UNIX/Bash Review (6–7 hours)

Objectives: to master basic navigation and operations under Linux, and to automate small tasks.

Contents:

- Introduction to the UNIX/Linux shell and environment
- Essential commands: navigation (cd, ls, pwd), file management (cp, mv, rm), reading (cat, less, head, tail)
- Redirections, pipes, and filters (>, |, grep, sort, uniq, cut)
- Variables, aliases, and command history
- Practical exercises: exploring a bioinformatics data directory, filtering, and extracting information (e.g., FASTA sequence names)

2 – Bash scripts and automation (6–7 hours)

| Matières                                                                     | ECTS | Semestre   | Cours | TD   | TP | Cours-<br>TD | Cours-<br>TP | TD-<br>TP | A<br>distance | Projet | Tutorat |
|------------------------------------------------------------------------------|------|------------|-------|------|----|--------------|--------------|-----------|---------------|--------|---------|
| <a href="#">Object Oriented<br/>Programming for<br/>Machine<br/>Learning</a> | 3    | Semestre 1 | 9     | 13.5 |    |              |              |           |               |        |         |

# Object Oriented Programming for Machine Learning

ECTS : 3

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 9

Travaux dirigés : 13.5

Langue d'enseignement

Anglais

Enseignement à distance

non

Prérequis

Python

Algorithms

Programme / plan / contenus

- Introduction to software design and development methods
- Machine learning frameworks and the design of libraries
- POO principles
- Classes in UML
- Class, object, methods, attributes, etc.
- Abstraction, Inheritance, Mixins, Interfaces
- Dunder methods
- Class decorators

Objectifs d'apprentissage

This course introduces students to object-oriented programming principles through software modeling activities using UML and extensive Python programming. Throughout the course, students will have the opportunity to study concrete use cases from machine learning frameworks and libraries. This will provide an opportunity to become familiar with the design philosophy of these frameworks. As a final project, students will have the opportunity to work on the modeling, design, and implementation of a machine learning library from scratch, with many features similar to those in the scikit-learn library.

By following this course, the students will be able to:

- Express software specifications in UML
- Implement specifications in Python
- Write clean, readable, and elegant code
- Write extensible and easily maintainable code
- Write code that complies with principles of reproducible research
- Write libraries that can be shared and used by others

| Matières                                              | ECTS | Semestre   | Cours | TD   | TP | Cours-<br>TD | Cours-<br>TP | TD-<br>TP | A<br>distance | Projet | Tutorat |
|-------------------------------------------------------|------|------------|-------|------|----|--------------|--------------|-----------|---------------|--------|---------|
| <a href="#">Multivariate statistics for genomics.</a> | 3    | Semestre 1 | 9     | 13.5 |    |              |              |           |               |        |         |

# Multivariate statistics for genomics.

ECTS : 3

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 9

Travaux dirigés : 13.5

Langue d'enseignement

Anglais

Enseignement à distance

non

Prérequis

- Basic knowledge of R
- Knowledge of univariate and bivariate inferential statistics (estimation, testing, p-values and interpretation of results)

Programme / plan / contenus

- Linear regression
- Logistic regression
- Poisson regression
- Variable selection and validation of models.

Objectifs d'apprentissage

Based on a problem and associated data, be able to recognise a suitable model, write it, implement it with R, and analyse all the results.

Organisation générale et modalités pédagogiques

- 9h lectures and 13.5h practicals
- Evaluations :
- \* Practicals (35% of the final grade)
- \* Final exam (65% of the final grade)

Modalités pédagogiques particulières

Lectures and practicals with computers. Use of R software.

Compétences

C1.2. Concevoir une démarche expérimentale.

C3.2. Argumenter les choix méthodologiques et les interprétations d'une étude.

Bibliographie

\* J.J. Faraway. [https://wangyang-bucket.oss-cn-beijing.aliyuncs.com/cms/image/Linear%20Models%20With%20R\\_1676789263894.pdf](https://wangyang-bucket.oss-cn-beijing.aliyuncs.com/cms/image/Linear%20Models%20With%20R_1676789263894.pdf).

| Matières                                         | ECTS | Semestre  | Cours | TD   | TP | Cours-<br>TD | Cours-<br>TP | TD-<br>TP | A<br>distance | Projet | Tutorat |
|--------------------------------------------------|------|-----------|-------|------|----|--------------|--------------|-----------|---------------|--------|---------|
| <a href="#">Machine learning and data mining</a> | 3    | Annualisé | 9     | 13.5 |    |              |              |           |               |        |         |

# Machine learning and data mining

ECTS : 3

Semestre calendaire : Annualisé

Détail du volume horaire

Cours magistraux : 9

Travaux dirigés : 13.5

Langue d'enseignement

Anglais

Enseignement à distance

non

Programme / plan / contenus

- Introduction to machine learning and data mining
- The clustering problem
- K-means algorithm
- Hierarchical clustering
- Spectral clustering
- Tree-based learning methods (Decision trees)
- Ensemble methods (Boosting, Bagging)

Objectifs d'apprentissage

This course provides an introduction to machine learning and data mining techniques. Students will learn to analyze datasets, discover structure with clustering algorithms, and build powerful and interpretable predictive models with decision trees and ensemble methods. Emphasis is placed on practicals related to multi-omics tabular datasets and applications in precision medicine. Students will have the opportunity to apply their acquired skills in cutting-edge machine learning challenges.

By following this course, students will be able to:

- Understand the underlying mathematical foundations of machine learning and data mining approaches
- Design, train, and evaluate machine learning and data mining pipelines on real-world tabular datasets
- Use popular frameworks like scikit-learn to build and deploy such pipelines
- Engage in cutting-edge real-world machine learning challenges, with a particular focus on applications in precision medicine

Compétences

C1.2. Concevoir une démarche expérimentale.

C1.3. Établir, réaliser et optimiser des protocoles expérimentaux.

C1.4. Produire et valider des données scientifiques.

| Matières                                  | ECTS | Semestre  | Cours | TD   | TP | Cours-<br>TD | Cours-<br>TP | TD-<br>TP | A<br>distance | Projet | Tutorat |
|-------------------------------------------|------|-----------|-------|------|----|--------------|--------------|-----------|---------------|--------|---------|
| <a href="#">Sequence<br/>algorithmics</a> | 3    | Annualisé | 9     | 13.5 |    |              |              |           |               |        |         |

## Sequence algorithmics

ECTS : 3

Semestre calendaire : Annualisé

Détail du volume horaire

Cours magistraux : 9

Travaux dirigés : 13.5

Langue d'enseignement

Anglais

Enseignement à distance

non

Prérequis

Basics of algorithms, data structures

Programme / plan / contenus

Review of algorithmic complexity calculation

Pattern matching

- Naive algorithm for exact pattern matching

- Suffix automata

Repetition matching

- Naive algorithm: matrix usage

- Suffix trees

Dynamic programming

- Longest common subsequence matching

- Pattern matching with k-errors

- Sequence alignment

Objectifs d'apprentissage

This course will cover algorithmic methods for sequence analysis. It will focus on pattern matching, and sequence alignment.

Organisation générale et modalités pédagogiques

EE 100% 1ere et 2eme session

Compétences

Skills:

Master the theoretical and practical foundations of algorithmic methods used for sequence analysis.

Additional skills:

- Formalization of the concept of a biological sequence and its analysis

- Theoretical foundations of algorithms

Nature de l'évaluation

Evaluation Continue non Intégrale

| Matières                                       | ECTS | Semestre   | Cours | TD   | TP | Cours-<br>TD | Cours-<br>TP | TD-<br>TP | A<br>distance | Projet | Tutorat |
|------------------------------------------------|------|------------|-------|------|----|--------------|--------------|-----------|---------------|--------|---------|
| <a href="#">Modeling of biological systems</a> | 3    | Semestre 2 | 9     | 13.5 |    |              |              |           |               |        |         |



## Modeling of biological systems

ECTS : 3

Semestre calendaire : Semestre 2

Détail du volume horaire

Cours magistraux : 9

Travaux dirigés : 13.5

Langue d'enseignement

Anglais

Enseignement à distance

non

Nature de l'évaluation

Evaluation Continue non Intégrale

[Advanced databases](#)

3      Annualisé      9      13.5

## Advanced databases

ECTS : 3

Semestre calendaire : Annualisé

Détail du volume horaire

Cours magistraux : 9

Travaux dirigés : 13.5

Langue d'enseignement

Anglais

Enseignement à distance

non

Prérequis

SQL language and entity-relationship modeling, practical experience with a relational database management system. Unix/Linux concepts and utility filters.

Programme / plan / contenus

The course outline:

\* Main courses:

- UML,
- translation of an object-oriented data representation model into a relational model
- ontologies in biology

\* Directed work:

- conception in UML of different biological problems, mainly in genomics
- translation of the created object-oriented data representation models into a relational model and implementation for PostgreSQL
- practice of Gene Ontology database and data and the underlying data representation formalism

Objectifs d'apprentissage

The objective of this course is to present object-oriented data representation languages with a view to creating models for problems in biology and implementing them under a relational database management system. Ontologies in biology will also be presented, as well as different data representation languages.

Organisation générale et modalités pédagogiques

This teaching unit is organized in the form of main courses (9h) and practical work (13,5h) in the computer room. Students will be evaluated through several small exams in class and a final exam.

Compétences

C1.1. Formuler des problématiques et des hypothèses de

| Matières                                     | ECTS | Semestre   | Cours | TD | TP | Cours-<br>TD | Cours-<br>TP | TD-<br>TP | A<br>distance | Projet | Tutorat |
|----------------------------------------------|------|------------|-------|----|----|--------------|--------------|-----------|---------------|--------|---------|
| <a href="#">Frameworks for Deep Learning</a> | 2    | Semestre 2 |       | 17 |    |              |              |           |               |        |         |

## Frameworks for Deep Learning

ECTS : 2

Semestre calendaire : Semestre 2

Détail du volume horaire

Travaux dirigés : 17

Langue d'enseignement

Anglais

Enseignement à distance

non

Programme / plan / contenus

- Introduction to deep learning frameworks
- Design patterns
- Meta-programming in Python
- Software testing and continuous integration

Objectifs d'apprentissage

This course introduces students to the widely used deep learning frameworks such as PyTorch, Keras, and TensorFlow, from an architectural perspective.

Students will have the opportunity to discover these complex frameworks through a concrete programming project: modeling, designing, and implementing a deep learning framework from scratch, including its fundamental components such as the computation graph and differentiation engine.

This will enable students to understand the underlying design principles of these software libraries and to develop their own libraries that adhere to these standards.

By following this course, the students will be able to:

- become familiar with deep learning frameworks such as PyTorch, Keras, and TensorFlow
- understand the software design principles of these frameworks
- use design patterns to improve software design
- build a deep learning framework from scratch, including its computation graph and differentiation engine
- apply software design principles in future projects

Organisation générale et modalités pédagogiques

Python

Programming

Compétences

C2.3. Se former à de nouveaux outils technologiques.

C3.2. Argumenter les choix méthodologiques et les

| Matières                      | ECTS | Semestre  | Cours | TD | TP | Cours-<br>TD | Cours-<br>TP | TD-<br>TP | A<br>distance | Projet | Tutorat |
|-------------------------------|------|-----------|-------|----|----|--------------|--------------|-----------|---------------|--------|---------|
| <a href="#">Deep learning</a> | 2    | Annualisé | 9     | 9  |    |              |              |           |               |        |         |

## Deep learning

ECTS : 2

Semestre calendaire : Annualisé

Détail du volume horaire

Cours magistraux : 9

Travaux dirigés : 9

Langue d'enseignement

Anglais

Enseignement à distance

non

Programme / plan / contenus

- Introduction to deep learning for precision medicine:

Motivation

- Perceptron and multilayer perceptron

- Gradient descent and the backpropagation algorithm

- Computation graph and automatic differentiation

(autograd) engine

- Optimization (Adam, RMSprop, etc.)

- Regularization techniques

- Convolutional neural networks (CNN) for (medical)

image processing (e.g., CT-scans, etc.)

- Recurrent neural networks (RNN) for (medical) time series (e.g., EEG signals, etc.)

Objectifs d'apprentissage

Deep learning is at the heart of the recent advances in artificial intelligence. Beyond computer vision and natural language processing, deep learning has made significant advances in precision medicine, enabling breakthroughs such as early disease onset prediction, personalized treatment plans, and drug prescription.

This course introduces the foundations of deep learning, from the classic perceptron to modern architectures. It will cover the design, training, and rigorous evaluation of deep learning architectures for various types of data, such as images and time series. Emphasis is placed on practicals related to deep learning for precision medicine applications. Students will participate in cutting-edge AI challenges related to precision medicine using deep learning models.

In parallel, students will follow the course “Frameworks for Deep Learning,” where they will have the opportunity to model, design, and implement a deep learning library from scratch, including a computation graph and an automatic differentiation engine that enables the training of deep learning models.

By following this course, students will be able to:

## Professionalization

| Matières                                    | ECTS | Semestre   | Cours | TD | TP | Cours-<br>TD | Cours-<br>TP | TD-<br>TP | A<br>distance | Projet | Tutorat |
|---------------------------------------------|------|------------|-------|----|----|--------------|--------------|-----------|---------------|--------|---------|
| <a href="#">FAIR project<br/>management</a> | 1    | Semestre 1 | 3     | 6  |    |              |              |           |               |        |         |

## FAIR project management

ECTS : 1

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 3

Travaux dirigés : 6

Langue d'enseignement

Anglais

Enseignement à distance

non

Prérequis

? Basic knowledge of data management and scientific data formats

? Fundamental concepts in bioinformatics and computational biology

? Introductory skills in data analysis and scripting (Python/R/Shell)

? Awareness of ethical and legal issues in research data handling

Programme / plan / contenus

? Introduction to Open Science

The six pillars of openness: Open Access, Open Data, Open Source, Open Peer Review, Open Education, Citizen Science

? FAIR Principles

Overview of FAIR: origins, rationale, and implications

? Tools and Infrastructures for FAIR and Open Bioinformatic : Repositories, Code sharing and reproducibility

? Data privacy and protection : Intellectual property and open licenses, Balancing openness with confidentiality in biomedical research

Objectifs d'apprentissage

By the end of this course, students will be able to:

? Understand the foundations and challenges of Open Science in the context of life sciences and bioinformatics.

? Apply the FAIR principles (Findable, Accessible, Interoperable, Reusable) to biological data and code.

? Identify infrastructures, repositories, and tools supporting data sharing and reproducible research.

? Address ethical, legal, and technical aspects related to open data, including sensitive and personal biological data.

Organisation générale et modalités pédagogiques

| Matières                | ECTS | Semestre   | Cours | TD | TP | Cours-<br>TD | Cours-<br>TP | TD-<br>TP | A<br>distance | Projet | Tutorat |
|-------------------------|------|------------|-------|----|----|--------------|--------------|-----------|---------------|--------|---------|
| <a href="#">Project</a> | 2    | Semestre 2 |       |    | 9  |              |              |           |               |        |         |

| Matières | ECTS | Semestre | Cours | TD | TP | Cours-<br>TD | Cours-<br>TP | TD-<br>TP | A<br>distance | Projet | Tutorat |
|----------|------|----------|-------|----|----|--------------|--------------|-----------|---------------|--------|---------|
| ■        |      |          |       |    |    |              |              |           |               |        |         |

## Project

ECTS : 2

Semestre calendaire : Semestre 2

Détail du volume horaire

Travaux pratiques : 9

Langue d'enseignement

Anglais

Enseignement à distance

non

Prérequis

? Basic knowledge of data management and scientific data formats

? Fundamental concepts in bioinformatics and computational biology

? Introductory skills in data analysis and scripting (Python/R/Shell)

? Awareness of ethical and legal issues in research data handling

Programme / plan / contenus

Each student (or pair) will:

1. Select a dataset or analysis from a previous course module.

2. Describe and document the dataset:

Add structured metadata (sample origin, methods, software, version, author).

Identify suitable file formats.

3. Create a reproducible analysis workflow: Clearly record parameters and software versions.

4. Publish and share:

- Host code and documentation on GitHub or GitLab.

- Deposit dataset or notebook and assign a DOI.

- Apply open licenses.

Objectifs d'apprentissage

By the end of this course, students will be able to:

? Apply the FAIR principles (Findable, Accessible, Interoperable, Reusable) to biological data and code.

? Use of open-source tools for reproducibility and data sharing

? Developed a critical assessment of data openness and reusability

? Improve their documentation and version control for scientific transparency

? Improve scientific reproducibility and transparency

? Learn to use open infrastructures and persistent identifiers



| Matières | ECTS | Semestre | Cours | TD | TP | Cours-<br>TD | Cours-<br>TP | TD-<br>TP | A<br>distance | Projet | Tutorat |
|----------|------|----------|-------|----|----|--------------|--------------|-----------|---------------|--------|---------|
| ■        |      |          |       |    |    |              |              |           |               |        |         |

## Internship

ECTS : 10

Semestre calendaire : Semestre 2

Langue d'enseignement

Anglais

Enseignement à distance

non

Organisation générale et modalités  
pédagogiques

Two-month internship in a research laboratory.

Nature de l'évaluation

Evaluation Terminale

## Lieu(x) d'enseignement

EVRY

## Contact(s)

- **Valérie CHAUDRU**

MCF

<http://www.universite-paris-saclay.fr/formation/master/bio-informatique/m1-genomics-informatics-mathematics-and-artificial-intelligence-health-and-environment-geniomhe-ai>