

Formation « Python et BioPython pour analyse de données génomiques »

Python est largement utilisé en **bio-informatique** grâce à la disponibilité de **nombreuses bibliothèques et outils** spécialement conçus pour l'analyse et la **manipulation de données biologiques**, la visualisation de résultats, et la modélisation de processus biologiques.

DATES & DURÉE

Les 19 mars et 26 mars et 2 avril 2026
(3 jours en 1j+1j+1j)

LIEU

En distanciel

PUBLIC

Tout public

PRÉ-REQUIS

Avoir des bases de programmation en Python et connaître les bibliothèques de calcul Numpy et Pandas

INTERVENANT

Daniel MOUCHI
Formatech Consulting

METHODES

Apports théoriques suivis d'exercices pratiques

INSCRIPTION

Cliquer sur ce lien pour accéder au [formulaire d'inscription en ligne à compléter](#)
(formulaire unique pour tout agent)

Date limite d'inscription :

Le 31 janvier 2026

CONTACT

Michèle HAYS
Chargée du dvlpmnt des compétences
INSERM DR GO
02 40 35 86 80
michele.hays@inserm.fr

OBJECTIFS :

- Comprendre la syntaxe de programmation du langage Python ;
- Être capable de mettre en œuvre les modules pour l'analyse de données ;
- Apprendre à manipuler et analyser de grands jeux de données génomiques ;
- Savoir modéliser les données et optimiser les modèles ;
- Savoir évaluer les modèles

PROGRAMME :

Chaque journée de formation est dédiée à la présentation des fonctionnalités de certains modules/bibliothèques et à leur manipulation sur des données biologiques. Ces présentations pourront être adaptées en fonction des besoins spécifiques des stagiaires.

Jour 1 - Description et utilisation des modules :

Biopython	offre une gamme complète d'outils pour travailler avec des séquences biologiques, des fichiers de formats courants, des alignements, et bien plus encore
BioPython-Tools	collection d'outils et de scripts basés sur Biopython pour des tâches spécifiques en bioinformatique
DESeq2	bibliothèque pour l'analyse différentielle d'expression génique, utile pour identifier les gènes qui sont régulés de manière significative entre différentes conditions expérimentales
BiGGR	BiGGR (Biclustering Gene Sets) est une bibliothèque pour effectuer des analyses de biclustering sur des données d'expression génique

Jour 2 - Description et utilisation des modules

Bioconda	gestionnaire de packages conçu spécifiquement pour la bioinformatique. Il propose une vaste collection de packages et d'environnements conda pour simplifier l'installation de logiciels bioinformatiques
BioPandas	extension de pandas qui simplifie la manipulation de données moléculaires telles que les structures de protéines et les trajectoires de simulation
Pysam	interface Python pour la bibliothèque Samtools, ce qui permet de manipuler des fichiers de séquençage d'ADN en format SAM et BAM

Jour 3 - Description et utilisation des modules :

Scikit-learn	bibliothèque libre Python destinée à l'apprentissage automatique, elle propose dans son framework de nombreuses bibliothèques d'algorithme à implémenter clé en main, à disposition des data scientists
Scikit-bio	bibliothèque proposant des algorithmes conçus pour l'analyse de données biologiques à grande échelle (génomique, ...) et de données multi-omiques complexes
OpenCV	bibliothèque open source largement utilisée pour le traitement d'images et la vision par ordinateur. Développée à l'origine par Intel, elle est aujourd'hui soutenue par une large communauté de développeurs et chercheurs