



الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي و البحث العلمي



République Algérienne démocratique et populaire

Ministère de l'enseignement supérieur et de la recherche scientifique

Université Mohamed Sedik Benyahia, Jijel

Faculté des Sciences de la Nature et de la Vie

Département de biologie moléculaire et cellulaire

TP : Logiciels libres et Open source.

Présenté par :

-Boussafi Khadîdja

-Lakhlef fatima

prof : Dr. Bensalem

Année universitaire : 2025/2026

General introduction:

Programs and applications, also known as logiciel, are used to manage and process information on computers and digital devices, and they represent a fundamental part of modern technology relied upon in scientific, educational, and administrative fields. These tools aim to simplify complex tasks and transform them into precise, executable steps efficiently and quickly. They also enable the analysis and processing of large amounts of data that would be difficult for humans to handle manually. Additionally, they support education and learning by providing interactive simulation environments, digital exercises, and tools to illustrate scientific concepts in a practical and visual way, enhancing understanding for students and researchers. They also play a major role in scientific research, allowing scientists to organize and manage experimental results, extract accurate conclusions, and support decision-making based on reliable data.

In the field of biology, the use of such programs has become essential and multifaceted. For example, BLAST is used to analyze gene sequences and compare genetic codes, PyMOL for designing and modeling 3D protein structures, ImageJ to analyze microscopic images of cells and biological structures, and programming languages like R and Python to process large biological datasets such as gene expression and proteomics studies. These tools also help simulate biological processes like cell growth, cellular interactions, and disease spread, allowing researchers to study these processes in a virtual environment before conducting practical experiments. In this way, applications and programs have become vital tools in supporting modern biological research, combining the ability to handle large amounts of information with the facilitation of experiments and a better understanding of living processes.

Objectives:

- Study the biopython library and its main functionalities.
- Analyze the strengths and limitations of biopython in biological data analysis.
- Explore the Zenodo platform and its role in open science.
- Create a GitHub repository and upload the final report.



PARTIE 01:

Etude théoriques de Biopython

1. Définition de Biopython :

Biopython est une bibliothèque open source en langage python destinée à la bioinformatique. Elle offre des outils pour manipuler et analyser les séquences biologiques telles que l'ADN, l'ARN et les protéines, d'accéder à des bases de données biologique comme NCBI ou uniProt, et réaliser des alignements de séquences ou des analyses phylogénétiques.

Biopython est utilisé dans des domaines variés notamment la recherche génomique et protéomique, l'analyse des gènes et protéines, ainsi que l'étude de l'évaluation des espèces. Il a été développé afin de faciliter le traitement des données biologique, d'uniformiser les outils informatiques en python et rendre la bioinformatique accessible aux chercheurs et étudiants.

2. Les fonctionnalités principales de Biopython :

- Manipulation et analyse des séquences biologiques (ADN, ARN, protéinés).
- Conversion entre différents formats de fichiers biologiques (FASTA, GenBank...).
- Alignement de séquences pour l'étude de la similarité génétique et protéique.
- Accès aux bases de données biologique telles que NCBI, uni Prot et PDB.
- Lecture et étude des structure moléculaires 3D (protéines, macromolécules).
- Outil des calcul bioinformatique pour l'analyse des séquences, traduction protéines, recherche des motifs conservés.

3. Aspect technique de Biopython :

- ❖ Langage de développement : Biopython est développé en python, ce qui assure la simplicité, la lisibilité et la flexibilité pour le traitement des données biologiques.
- ❖ Compatibilité multiplateforme : fonctionne sur Windows, Linux et MacOs, facilitant son utilisation dans divers environnements académique et professionnels.

- ❖ Installation et environnement : installation standardisée via le gestionnaire de paquets python (pip), utilisation dans Anaconda, Jupyter Notebook et tout environnement de développement compatible python.
- ❖ Intégration avec d'autres outils : compatible avec la bibliothèque python scientifique (Numpy ; Pandas, Matplotlib)
Pour l'analyse statistique et la visualisation avancée.
- ❖ Gestion des formats et des bases des données : prise en charge des principaux formats biologiques et accès direct aux bases des données publique (NCBI, uni Prot), permettant l'extraction et des données biologique.

4.Points forts de biopython :

- ❖ Bibliothèque open source et gratuite, largement utilisé en recherche scientifique
- ❖ Développée en langage python, offrant simplicité, lisibilité et flexibilité.
- ❖ Compatible avec les principaux systèmes d'exploitation (Windows, Linux, macOS)
- ❖ Pris en charge de nombreux formats biologique standards (FASTA, GenBank ...)

5.Points faibles de biopython :

- ❖ Nécessité de compétences de base en programmation python pour une exploitation optimale.
- ❖ La documentation bien que riche, peut être complexe pour les débutants, en particulier pour les utilisateurs n'ayant pas de bases solides en programmation python.
- ❖ Certaines fonctionnalités avancées en bioinformatique (analyse des données, modélisation structurale...) sont moins développées ou nécessitent l'utilisation de bibliothèques externes.
- ❖ Certaines parties de la bibliothèque sont mise à jour de manière irrégulière, pouvant entraîner des difficultés de compatibilité avec les nouvelles versions de python .

6. Conclusion :

Biopython est un ensemble de bibliothèques open source en Python dédié à la bioinformatique, facilitant l'analyse et la gestion des données biologiques. Grâce à sa simplicité, sa flexibilité et son intégration avec l'écosystème Python, il constitue un outil efficace pour la recherche scientifique. Malgré certaines limites liées aux performances pour de grands volumes de données, Biopython reste largement utilisé en génomique, protéomique et pharmacologie.



PARTIE 02 :

Étude pratique – Exploration de la plateforme Zenodo

1.Présentation de la plateforme Zenodo :

Zenodo est une plateforme en libre accès qui permet de stocker et de partager les travaux scientifiques. Elle a été créée par le CERN pour encourager la science ouverte.

Elle permet aux chercheurs de déposer différents types de productions comme des articles, des données, des présentations ou des rapports, tout en assurant leur conservation et leur accessibilité.

Zenodo est un dépôt généraliste, c'est-à-dire qu'il accepte tous les domaines de recherche. Il est surtout utile pour les travaux qui n'ont pas de dépôt spécialisé.

➤ Il permet de déposer plusieurs types de contenus, comme :

- des datasets (jeux de données),
- des articles scientifiques,
- des codes sources,
- des thèses,
- des posters et présentations,
- des rapports techniques.

L'intérêt de Zenodo pour la recherche en sciences de la nature et de la vie réside dans le fait qu'elle facilite :

- le partage des données expérimentales,
- la réutilisation des résultats scientifiques,
- la traçabilité grâce à l'attribution d'un DOI,
- la transparence scientifique,
- la promotion de la science ouverte.

2. Description des étapes réalisées :

2.1 Recherche du dataset :

Nous avons effectué une recherche sur la plateforme Zenodo en utilisant le mot suivant : **cell cancer**. Cette requête nous a permis d'obtenir plusieurs données en relation avec la biologie cellulaire et la recherche biomédicale.

The screenshot shows the Zenodo search interface. The search bar at the top contains the query 'CELL CANCER'. Below the search bar, it displays '136,552 result(s) found'. The results are sorted by 'Best match'. The first result is a dataset titled 'A Single-cell Perturbation Landscape of Colonic Stem Cell Polarisation' uploaded on July 24, 2023. It has 129,579 versions and 1199 citations. The second result is a dataset titled 'Processed counts data of cfMeDIP-seq profiles of small cell lung cancer patients' uploaded on December 22, 2022. It has 101,084 versions and 509 citations. The interface includes filters for 'Versions', 'Access status' (Open, Restricted, Embargoed), and 'Resource types' (Publication, Dataset). A cookie consent banner at the bottom asks if the site uses cookies and provides options to accept all or essential cookies.

Figure 1 : Capture d'écran montrant la requête de recherche sur Zenodo.

2.2 Sélection du dataset :

Nous avons sélectionné le dataset intitulé : « **CSF1R+ myeloid-monocytic cells drive CAR-T cell resistance in aggressive B cell lymphoma** », grâce à ça relationne au domaine des sciences biologiques .

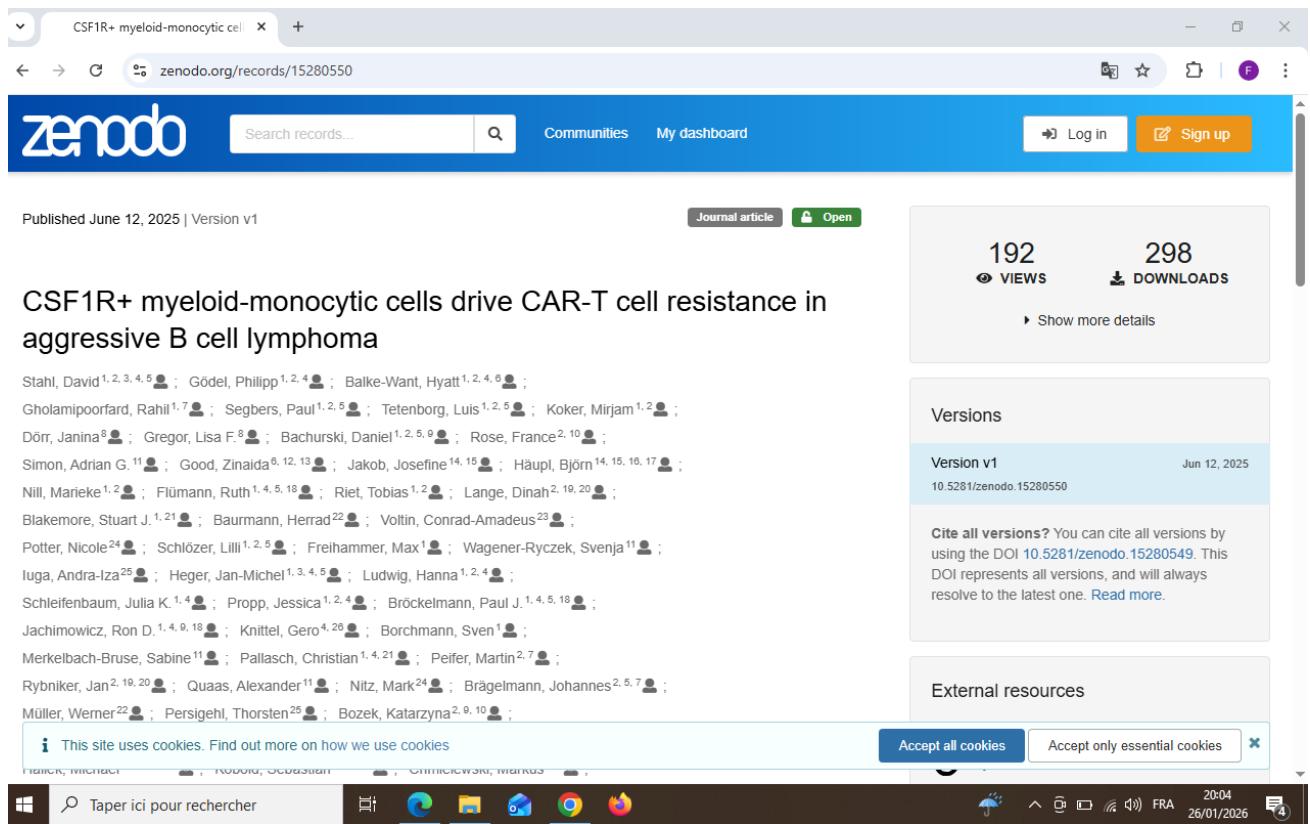


Figure 2 : Capture d'écran montrant la page du dataset sélectionnée sur Zenodo .

2.3 Téléchargement du dataset

Nous avons utilisée la section « **Files** » qui comprend trois fichiers :

- `adata_scRNASeq-human.h5ad`` (272 MB) : fichier principal contenant les données de single-cell RNA sequencing.
- `Human_B-NHL_Bulk_RNA_count.csv`` (5.6 MB) : fichier CSV contenant des données d'expression génique.
- `'Human_Ex-vivo_DIA_Protein.csv`` (1.2 MB) : fichier CSV contenant des données protéomiques.

CSF1R+ myeloid-monocytic cell x Devoir Étude Outil x +

zenodo.org/records/15280550

ENSG000000000419	DPM1	protein_coding	113	63	53	55	119	160
ENSG000000000457	SCYL3	protein_coding	17	33	46	58	87	244
ENSG000000000460	C1orf112	protein_coding	45	32	0	30	52	97
ENSG000000000938	FGR	protein_coding	19	2	172	53	150	53
ENSG000000000971	CFH	protein_coding	12	31	63	111	49	156
ENSG000000001036	FUCA2	protein_coding	35	23	32	49	60	62

Files (278.7 MB)

Name	Size	Actions
adata_scRNAseq-human.h5ad md5:a4de50ea043f9dfc54f5889c39a10729	272.0 MB	Download Preview
Human_B-NHL_Bulk_RNA_count.csv md5:fd9e38772f0314322590291622177f1	5.6 MB	Download Preview
Human_Ex-vivo_DIA_Protein.csv md5:37fc24244bbf4c33f58957364bab2304	1.2 MB	Download Preview

Additional details: i This site uses cookies. Find out more on how we use cookies

Accept all cookies Accept only essential cookies

Style APA

Export

Windows taskbar: Taper ici pour rechercher, Task View, File Explorer, Edge, Google Chrome, Firefox, Internet Explorer, 20:15 FRA 26/01/2026

Figure 3 : Capture d'écran montrant la section « Files ».

Nous avons après téléchargé le dataset en utilisant l'option « Download all » :

CSF1R+ myeloid-monocytic cell x zenodo.org/records/15280550 x +

ENSG000000000419	DPM1	protein_coding	113	63	53	55	119	160
ENSG000000000457	SCYL3	protein_coding	17	33	46	58	87	244
ENSG000000000460	C1orf112	protein_coding	45	32	0	30	52	97
ENSG000000000938	FGR	protein_coding	19	2	172	53	150	53
ENSG000000000971	CFH	protein_coding	12	31	63	111	49	156
ENSG000000001036	FUCA2	protein_coding	35	23	32	49	60	62

Files (278.7 MB)

Name	Size	Actions
adata_scRNAseq-human.h5ad md5:a4de50ea043f9dfc54f5889c39a10729	272.0 MB	Download Preview
Human_B-NHL_Bulk_RNA_count.csv md5:fd9e38772f0314322590291622177f1	5.6 MB	Download Preview
Human_Ex-vivo_DIA_Protein.csv md5:37fc24244bbf4c33f58957364bab2304	1.2 MB	Download Preview

Additional details: i This site uses cookies. Find out more on how we use cookies

Accept all cookies Accept only essential cookies

Style APA

Export

Windows taskbar: Taper ici pour rechercher, Task View, File Explorer, Edge, Google Chrome, Firefox, Internet Explorer, 20:09 FRA 26/01/2026

Figure 4 : Capture d'écran montrant l'option de téléchargement .

3. Métadonnées du dataset (Norme Dublin Core) :

Nous avons extrait les métadonnées directement depuis la page Zenodo du dataset selon la norme Dublin Core.

Champ (Dublin Core)	Valeur
Title	CSF1R+ myeloid-monocytic cells drive CAR-T cell resistance in aggressive B cell lymphoma
Creator	Stahl D., Gödel P., Balke-Want H., et al.
Publisher	Zenodo
Date	2025
Identifier	https://doi.org/10.5281/zenodo.15280550
Description	Dataset contenant des données transcriptomiques et protéomiques permettant d'étudier le rôle des cellules myéloïdes CSF1R+ dans la résistance aux cellules CAR-T.
Language	English
Type	Dataset scientifique
Format	H5AD, CSV
Rights	Creative Commons Attribution 4.0 International (CC BY 4.0)
Subject	cell, cancer, RNA-seq, proteomics, CAR-T

This screenshot shows a Zenodo record page for a dataset titled "CSF1R+ myeloid-monocytic cells drive CAR-T cell resistance in aggressive B cell lymphoma".

Files:

- Human_B-NHL_Bulk_RNA_count.csv

Details:

- DOI:** DOI 10.5281/zenodo.15280550
- Resource type:** Journal article
- Publisher:** Cancer Cell
- Languages:** English

Rights:

- Creative Commons Attribution 4.0 International

License:

- Accept all cookies | Accept only essential cookies

Figure 5 : Capture d'écran montrant le DOI et les informations de citation .

This screenshot shows a Zenodo record page for the same dataset.

Files (278.7 MB):

Name	Size	Action
adata_scRNAseq-human.h5ad	272.0 MB	Download
Human_B-NHL_Bulk_RNA_count.csv	5.6 MB	Preview Download
Human_Ex-vivo_DIA_Protein.csv	1.2 MB	Preview Download

Additional details:

Dates	Accepted
	2025-05-22

License:

- Creative Commons Attribution 4.0 International

Citation:

Stahl, D., Gödel, P., Balke-Want, H., Gholamipoorfar, R., Segbers, P., Tetenborg, L., Koker, M., Dörr, J., Gregor, L. F., Bachurkis, D., Rose, F., Simon, A. G., Good, Z., Jakob, J., Häupl, B., Nill, M., Flümann, R., Riet, T., Lange, D., ... Ullrich, R. T. (2025). CSF1R+ myeloid-monocytic cells drive CAR-T cell resistance in aggressive B cell lymphoma.
<https://doi.org/10.5281/zenodo.15280550>

Style: APA

Export:

- JSON
- Export

Accept all cookies | Accept only essential cookies

Figure 6 : Capture d'écran montrant la section « Rights » et la licence CC BY 4.0

4. Conclusion de la partie pratique :

Cette étude pratique a permis de découvrir le fonctionnement de la plateforme Zenodo et son importance dans le cadre de la science ouverte. L'exploration du dataset sélectionné a montré l'utilité des métadonnées pour identifier, décrire et réutiliser correctement un jeu de données scientifique. L'utilisation de la norme Dublin Core facilite l'organisation et la compréhension des informations essentielles associées au dataset.

Reference:

- 1.** Biopython Tutorial and Cookbook. Retrieved from <http://biopython.org/DIST/docs/tutorial/Tutorial.html>
- 2.** Zenodo. (2026). *Zenodo: Research. Shared.* <https://zenodo.org>
- 3.** Dublin Core Metadata Initiative. (2024). *Dublin Core Metadata Element Set, Version 1.1.* <https://www.dublincore.org/specifications/dublin-core/dces/>
- 4.** CERN. (2023). *Zenodo – Research data repository.* <https://home.cern/science/computing/zenodo>
- 5.** OpenAIRE. (2022). *Open Science and Open Research Data.* <https://www.openaire.eu>