

Paris, le 12 juin 2026

Présentation des TP

Evaluation de l'apport de l'IA pour le développement logiciel et l'analyse des données biologiques

Bertrand Cosson, Jacques van Helden et Thomas Denecker



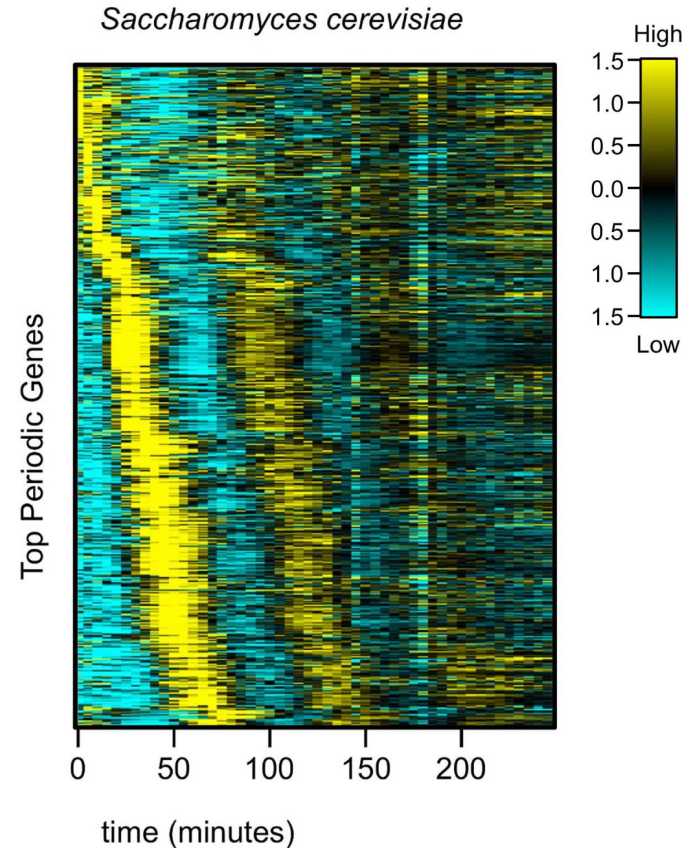
Source des données:

- Kelliher, C. M., Leman, A. R., Sierra, C. S. & Haase, S. B. Investigating Conservation of the Cell-Cycle-Regulated Transcriptional Program in the Fungal Pathogen, *Cryptococcus neoformans*. *PLoS Genet* 12, e1006453 (2016). DOI [10.1371/journal.pgen.1006453](https://doi.org/10.1371/journal.pgen.1006453)

Pré-traitement

- Effectué par Gaëlle Lelandais et les étudiants du Diplôme Universitaire Omique (DUO)
- La table de profils transcriptomiques contient les comptages normalisés avec DeSEQ2 (en unités [RPKM](#)) de **1705 gènes** présentant des oscillations d'expression durant le cycle cellulaire, sur **50 points temporels**

But de l'exercice : interagir avec une IA générative pour reconstruire le pipeline d'analyse et pour produire un code de bonne qualité permettant de reproduire l'analyse, et réutilisable pour analyser d'autres jeux de données.



Reproductibilité et FAIRisation

- **Configuration** : on fournit à l'IA une info exhaustive (article, matériel supplémentaire, données brutes et normalisées)
- **But** : s'appuyer sur une IA générative pour
 - **extraire le protocole** détaillé utilisé par les auteurs de l'article
 - **implémenter** un code de qualité permettant de **reproduire** les analyses et la figure
 - publier le code sous une forme respectant les **principes FAIR**
- **Langages** : python ou R (classique ou tidyverse)
- **Encadrement**:
 - Thomas Denecker
 - Imane Messak
 - Baptiste Rousseau



Ingénierie inverse

- **Configuration** : on fournit à l'IA une **info minimale** (tableau de données + figure)
- **But** : s'appuyer sur une IA générative pour
 - proposer un plan d'analyse qui permette de générer une figure similaire à partir des données
 - implémenter un code de qualité qui met en oeuvre ce plan d'analyse
- **Langage** : R (classique ou tidyverse)
- **Encadrement**
 - Jacques van Helden
 - Bertrand Cosson,
 - Lilia Younsi
 - Benjamin Saintpierre
 - Olivier Kirsh



Objectif général

- Évaluer dans quelle mesure les modèles d'IA générative peuvent assister des biologistes expérimentaux et des bioinformaticiens dans :
 - la conception d'une stratégie d'analyse
 - l'implémentation d'un code robuste et réutilisable

Objectifs spécifiques

- Impact du modèle de langage
- Stratégies de formulation des requêtes (prompting)

Méthodologie

- Réalisation d'exercices de **bioscripting assisté par IA** en groupes de 2 à 3 participants.
- Recueil immédiat des **retours d'expérience** via un questionnaire individuel.
- **Évaluation indépendante des productions** par deux experts externes, selon une grille de critères prédéfinie.
- **Analyse comparative des résultats**
- **Rédaction d'un article de synthèse.**

Revues pressenties

- PLOS Computational Biology (rubrique Education)
- Briefings in Bioinformatics
- Autres revues en bioinformatique ou pédagogie scientifique

Contribution des participants

- **Auteurs individuels** : contributeurs impliqués dans l'analyse des résultats et la rédaction du manuscrit.
- **Consortium « AI & Bioscripting 2026 »** : ensemble des participants à l'atelier pratique. Les membres du consortium seront listés selon les modalités prévues par la revue (généralement dans un fichier supplémentaire).

Principe

- Déclaration préalable des questions de recherche, hypothèses, méthodologie et plan d'analyse avant la collecte et l'analyse des données
- Embargo optionnel : le protocole est horodaté immédiatement mais peut rester non public jusqu'à la réalisation de l'étude

Remplissage d'un formulaire "OSF Preregistration"

Description explicite de

- Hypothèses scientifiques
- Questions de recherche
- Distinction explicite entre analyses planifiées (confirmatoires) et analyses exploratoires
- Design expérimental
- Variables et critères d'évaluation
- Plan d'analyse
- Règles d'inclusion / exclusion
- Gestion des données manquantes
- Un tas d'autres questions assez fondamentales pour la pré-conception d'un projet de recherche

→ 19 pages au total

Bénéfice secondaire

- Aide à structurer et clarifier un projet de recherche avant sa mise en œuvre

Astuces

- Le dialogue avec l'IA a été utilisé comme assistant à la rédaction de la préinscription, mais les décisions méthodologiques ont été validées par les chercheurs..
- Bon cadrage initial (explication de notre démarche) et recadrage au fil des questions à l'IA.
- Une approche itérative (question par question) permet d'obtenir des réponses plus précises et mieux adaptées au projet qu'une génération complète du formulaire en une seule requête.

En pratique

- Titre : Assessing Generative AI for Reverse Engineering Bioinformatics Workflows and Producing Reusable Code from Published Scientific Results
- Les hypothèses, questions de recherche et critères d'évaluation ne sont pas communiqués aux participants afin de limiter les biais comportementaux et les effets d'attente.
- Si vous désirez participer à l'expérience, signature d'un formulaire de consentement



*Les fautes d'orthographe, c'est la seule chose
que les IA nous ont laissée.*

Hervé Ménager, le 8 juin 2026