

PROJET TER PAR L'ÉQUIPE 5

PROJET TER PAR L'ÉQUIPE 5 RENDU  
INTERMÉDIAIRE 2 : PLANIFICATION ET GESTION  
DE PROJET

Thomas AYRIVIÉ, Mehdi BELKHITER, Jamila CHERKAOUI, Dina EL  
HIJJAWI, Magatte LO.



Département MIASHS, UFR 6 Informatique, Mathématique et Statistique  
Université Paul Valéry, Montpellier 3

Novembre 2023

SOU MIS COMME CONTRIBUTION PARTIELLE  
POUR LE COURS TER TV15MI - TV25MI

---

## Déclaration de non plagiat

---

Nous déclarons que ce rapport est le fruit de notre seul travail, à part lorsque cela est indiqué explicitement.

Nous acceptons que la personne évaluant ce rapport puisse, pour les besoins de cette évaluation:

- la reproduire et en fournir une copie à un autre membre de l'université; et/ou,
- en communiquer une copie à un service en ligne de détection de plagiat (qui pourra en retenir une copie pour les besoins d'évaluation future).

Nous certifions que nous avons lu et compris les règles ci-dessus.

En signant cette déclaration, nous acceptons ce qui précède.

Signature : *Thomas AYRIVIÉ*, n°22000580, Date : 10/11/2023.

Signature : *Mehdi BELKHITER*, n°21813356, Date : 10/11/2023.

Signature : *Jamila CHERKAOU*, n°22309204, Date : 10/11/2023.

Signature : *Dina EL HIJJAWI*, n°22310171, Date : 10/11/2023.

Signature : *Magatte LO*, n°22311161, Date : 10/11/2023.

---

## Préface

---

Dans le cadre de notre master 1 Miashs, les étudiants réalisent un **projet de Travaux d'Études et de Recherche** (TER) en lien avec un commanditaire. Ce présent document est un rapport intermédiaire dans lequel nous présenterons la définition de notre projet ainsi que sa planification et gestion.

**Sujet 6 :** « Identification de voies de signalisation activées par des traitements de cancer du sein ».

### Encadrement pédagogique :

Mme Sophie Lèbre, [sophie.lebre@univ-montp3.fr](mailto:sophie.lebre@univ-montp3.fr) (Institut Montpelliérain Alexander Grothendieck) avec Mathilde Robin (Institut de Recherche en Cancérologie de Montpellier, LIRMM Laboratoire d'informatique, de robotique et de microélectronique de Montpellier), Charles Lecellier (LIRMM) et Laurent Bréhélin (LIRMM).

### Composition du groupe :

Dina EL HIJJAWI, n°22310171, [dina.el-hijjawi@etu.univ-montp3.fr](mailto:dina.el-hijjawi@etu.univ-montp3.fr). Coordinatrice.

Thomas AYRIVIE, n°22000580, [thomas.ayrivie@etu.univ-montp3.fr](mailto:thomas.ayrivie@etu.univ-montp3.fr).

Mehdi BELKHITER, n°21813356, [mehidi.belkhiter@etu.univ-montp3.fr](mailto:mehidi.belkhiter@etu.univ-montp3.fr).

Jamila CHERKAOUI, n°22309204, [jamila.cherkaoui@etu.univ-montp3.fr](mailto:jamila.cherkaoui@etu.univ-montp3.fr).

Magatte LO, n°22311161, [magatte.lo@etu.univ-montp3.fr](mailto:magatte.lo@etu.univ-montp3.fr).

---

## Table des matières

---

Chapitre 1	Coordination du projet	1
1.1	Tâches à effectuer dans ce semestre . . . . .	1
Chapitre 2	Attribution des tâches	3
Chapitre 3	Suivi du projet et rencontres	4
Chapitre 4	Echéances universitaires	6

---

# CHAPITRE 1

## Coordination du projet

---

Le team leader est **Dina El Hijjawi**. Elle a réparti les tâches et le travail de la manière ci-après décrite. En premier temps, et pour mieux comprendre le projet sur lequel nous travaillons, tous les membres du groupe ont fait des analyses exploratoires sur les données, la lecture des articles mis à disposition par notre coordinatrice **Mme Sophie Lebre**.

Étant à l'aise en Python, **Jamila Cherkaoui**, **Magatte Lo** et **Mehdi Belkhiter** travaillent sur la construction du jeu de données. Ces données sont ensuite regroupées dans une matrice qui croise les gènes de chaque séquence et leurs motifs avec des p-valeurs associées pour chaque croisement. Une fois cette matrice créée, ils ont fait une jointure avec les données GEOquery NCBI (Patientes suivies pour le traitement de la maladie cancer du sein).

À partir de là, nous serons dans un problème de classification, abordé par **Dina El Hijjawi** et **Thomas Ayrivié**. Nous devons rechercher des groupes de variables corrélées pour trouver un modèle explicatif qui utilise la structure de corrélation des variables avec l'utilisation de R.

Comme plusieurs tâches dépendent de la réalisation complète d'autres tâches et de leur validation par notre coordinateur, lorsque Jamila, Magatte et Mehdi travaillaient sur la création du jeu de données, Thomas et Dina complétaient les rendus et les agendas. Nous inverserons les tâches par la suite. De plus, les groupes de personnes se sont entraïdés chaque fois que quelqu'un en avait besoin.

### 1.1 Tâches à effectuer dans ce semestre

- Créer et comprendre le jeu de données. **Fait**
- Effectuer des analyses exploratoires du jeu de données. **En cours**
- Comprendre le projet et le vocabulaire avec une première approche de la recherche. **En cours**
- Répartir en classe les résultats. up, down et neutre en considérant un seuil à 2 pour le up (resp. 1/2 pour les down) pour le ratio des comptages de read ARNm ("fold change") **En cours**
- Proposer un premier modèle de classification. 800 modèles car 800 variables, on garde le meilleur modèle. **En cours**
- Remplacer ces valeurs nulles par  $10^{-3}$  multiplié par 2 <- Plus quelque chose. **Mehdi**

- Matrice X : Passer à  $-\log(10)$  pour étaler les valeurs. Cela créera un score de présence de la valeur. Selon les conseils de Mathilde Robin. **Mehdi**
- Matrice Y : Répartir en classe les résultats: up, down et neutre en considérant un seuil à 2 pour le up (resp. 1/2 pour les down) pour le ratio des comptages de read ARNm (“fold change”). **Thomas**
- Regarder les liens entre les variables, à l’intérieur de la matrice X (régression deux à deux, conditionnement de X transpose X). **Jamila**
- Faire une Classification hiérarchique. **Magatte**
- Faire ACP + Clustering avec la matrice X. **Dina**
- Découvrir GLMnet, tester sur les données. **Thomas** et **Magatte**
- PRÉVOIR DES CRITÈRES D’ÉVALUATION DU MODÈLE. **Tous**
- Calculer la précision et le rappel. **Magatte**
- Représentations générales pour l’analyse des données. **Dina** et **Jamila**
- Représentation sous forme de heatmap, cases colorées. **Thomas**

---

## CHAPITRE 2

### Attribution des tâches

---

Les tâches confiées et divisées aux membres de l'équipe sont attribuées dans les fichiers minutes ou sur notre tableau de bord et de suivi Trello. Nous avons choisi cet outil pour la facilité d'utilisation, la personnalisation, et son accès facile en collaboration.

Lien de notre tableau de bord :

<https://trello.com/invite/b/7t00yIVU/ATTI59bddd29acb96bcddeab5e2ed89e0cd8B21571projetter>

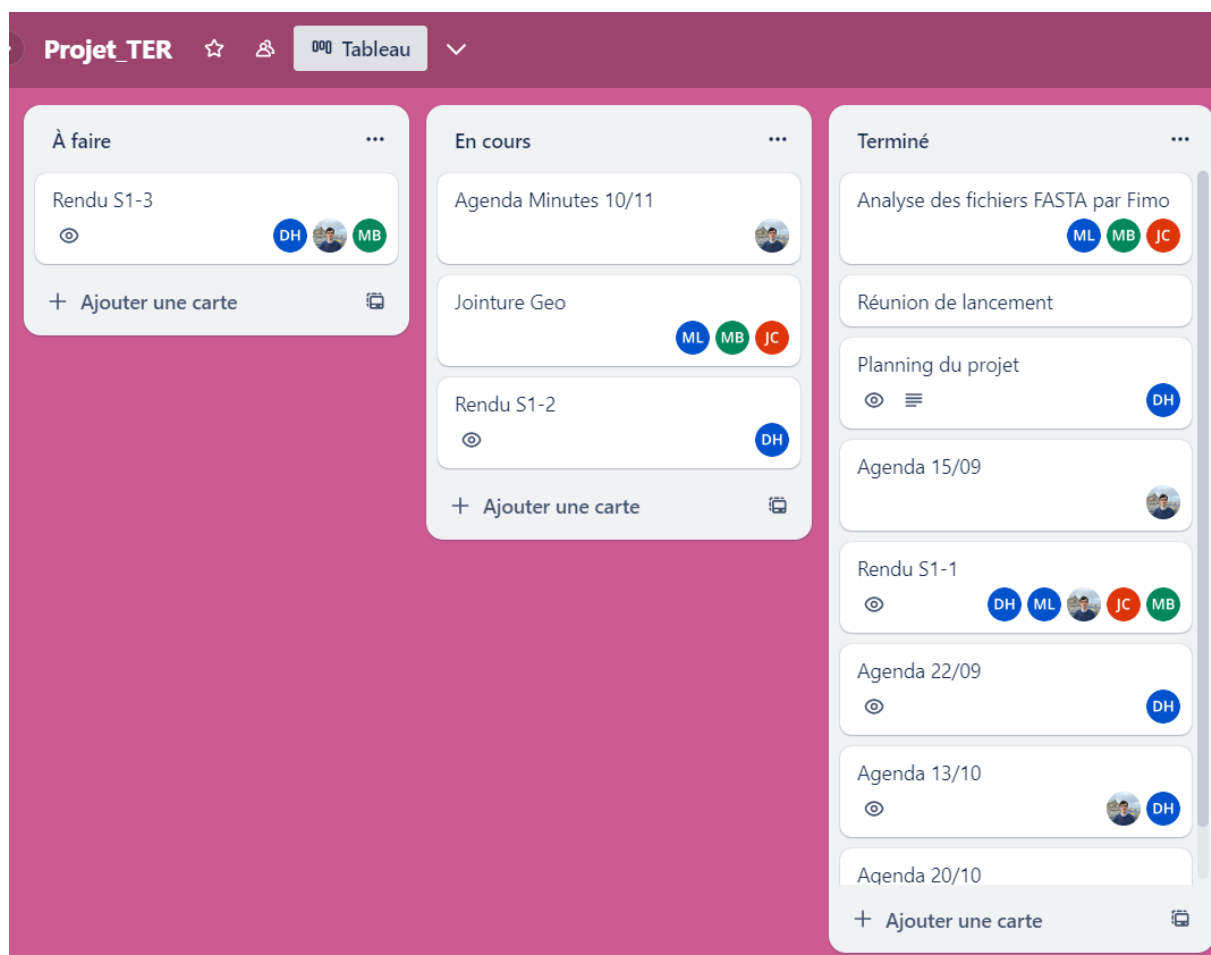


Figure 2.1: Aperçu du Trello

Les dates de rencontres sont décrites dans notre Google Agenda :

<https://calendar.google.com/calendar/u/0?cid=NmE4M2NlOWZkYmMyYjRlODljODg4MzFiZW>

Pour faciliter notre communication et l'encourager, on a créé un serveur sur Discord. Cela permet à tous les membres du groupe de se tenir au courant de ce que font les autres. Nous nous posons également des questions et y répondons en permanence:



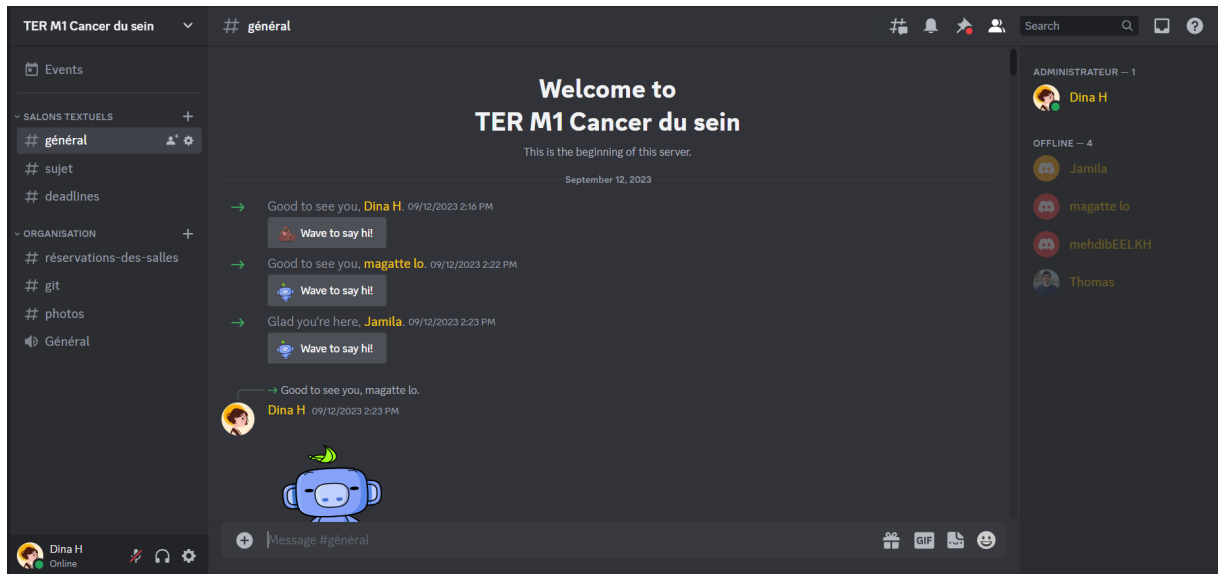


Figure 3.2: Aperçu du discord

---

## CHAPITRE 4

### Echéances universitaires

---

- Comprendre et explorer les données avant le 17/11/2023
- Revue de la littérature avant le 06/12/2023
- Soutenance le 15/12/2023
- Description des choix de modélisation et d'outils avant le 10/02/2024
- Description des choix de visualisation avant le 10/03/2024
- Interprétation des résultats avant le 14/04/2024
- Rendu final avant le 09/05/2024
- Soutenance finale le 13/05/2024