

Suivi du projet + Ordre du jour

Travail effectué :

- Compte rendu réunion + formalisation des questions.
- Git opérationnel pour tous les membres du groupe.
- Installation sur nos ordinateurs des [données JASPAR](#) (Motif vertébrés, redondants, format MEME, single batch), [des données RNA seq](#) (NCBI Gène Expression Omnibus). + Installation du package GEOquery sur R et téléchargement des données directement avec R. + Installation du [logiciel FIMO](#) (de la suite MEME) pour scanner les séquences.
- Commencer à lire les articles scientifiques.
- Bien définir le sujet. (Ci-après)
- Extraction de premiers résultats avec FIMO.
- Sophie Lèbre (avec Mathilde Robin) ont transmis les séquences ADN associées à chaque gène. + La thèse en français. + Précisions sur les scores.

Travail à faire prochainement :

- Commencer à rédiger/répondre aux questions proposées du 1er rapport intermédiaire à rendre le 15 octobre.
- Faire la liaison des données dans Rstudio, avec les précisions fournies.
- Description de la liste des variables, l'analyse exploratoire du jeu de données et de la matrice des X avec notamment les cours d'analyse multidimensionnelle (ACP, AFC...).

Première définition de sujet :

Le sujet de recherche se concentre sur l'identification des voies de signalisation activées en réponse aux traitements du cancer du sein. Au fil du temps, les cellules cancéreuses ont la capacité de s'adapter et de développer des mécanismes de résistance, y compris face à la chimiothérapie. Comprendre ces mécanismes est essentiel pour améliorer les stratégies de traitement.

Pour ce faire, nous utilisons une analyse approfondie des séquences génétiques afin de déterminer quelles combinaisons de nucléotides obtiennent les scores les plus élevés. Ces scores reflètent la similitude ou la pertinence d'une séquence donnée par rapport à un modèle de référence.

Préalablement, un modèle de régression a été développé pour expliquer l'activité des gènes en utilisant uniquement la séquence d'ADN. Cependant, nous visons maintenant à mettre en place un nouveau modèle de classification, dont les détails seront abordés ultérieurement au cours du projet.

Pour mener à bien notre recherche, nous utilisons la base de données Jaspar, une ressource spécialisée dans les matrices de séquences d'ADN liées aux sites de liaison des facteurs de transcription. Ces matrices, également appelées matrices de poids de position (PWM), offrent une représentation numérique des motifs de liaison des facteurs de transcription. Jaspar fournit des informations détaillées sur la fréquence de chaque base (A, C, G, T) à chaque position au sein de ces motifs de liaison. Cette base de données est

TER Cancer du sein

Version du 15 sept. 2023

Page 1 sur 2

essentielle pour développer des modèles plus précis dans le but de mieux comprendre la régulation génique et les réponses cellulaires aux traitements du cancer du sein.

En complément, nous utilisons pareillement la base de données GEO (Gene Expression Omnibus). GEO stocke une grande variété de données d'expression génique provenant de diverses expériences et technologies, incluant les puces à ADN et les séquençages ARN. Les chercheurs déposent leurs données d'expression génique dans GEO, permettant ainsi à d'autres scientifiques d'y accéder, de les partager et de les analyser. Cette base de données facilite l'exploration et la comparaison des profils d'expression génique dans différents contextes biologiques et expérimentaux.

Jaspar nous aide à analyser les motifs de liaison des facteurs de transcription, tandis que GEO nous permet d'accéder à une vaste gamme de données d'expression génique pour une exploration approfondie des mécanismes biologiques et des réponses aux traitements.

Ce que nous aimerions aborder lors de la réunion :

1. Explications sur l'autre approche du modèle linéaire, les arbres et forêts.
2. Commenter ensemble notre définition de sujet.
3. Fimo délivre également des E-Valeurs, qu'est-ce que c'est ?
4. Rediscuter des scores avec l'appui des formules mathématiques.
5. Comment/à partir de quoi le fichier de séquences des régions promotrices des gènes a été créé ?
6. Parler des deadlines : Est-ce qu'il y en a ? Quels pourraient-être nos objectifs ?
7. Nous avons créé un serveur interne à nous 5 pour le TER. Il est possible de vous inviter. À moins de faire le salon dans le Discord Master Miashs comme convenu.
8. Questions diverses.