OFFRE DE STAGE : Développement d'un package R pour l'annotation de données de cytométrie

Laboratoire : CIPHE - Centre d'immunophénomique

Lieu : Marseille

Début du stage : Janvier 2024

Durée : 6 mois

Profil recherché : Étudiant(e) en Master 2

**Description** :

De nombreux algorithmes d’analyse automatique non-biaisée par l’opérateur ont été développés pour analyser des jeux de données complexes de cytométrie par regroupement de cellules en sous-populations au sein de clusters. L'identification de ces clusters non-supervisés afin de leur attribuer une annotation porteuse de sens pour un biologiste exprimant le lignage ou l’état de maturation d’un type cellulaire requiert l’inspection manuelle de ces clusters par le biologiste en fonction des marqueurs exprimés. Il en résulte une annotation chronophage, subjective donc potentiellement biaisée par les connaissances de l’analyste et source d’imprécision ou d’erreur. Nous disposons d’un jeu de donnée répertoriant la dynamique d’expression de 428 molécules à la surface de 16 populations principales constituant le système immunitaire chez la souris en conditions basales ou inflammatoires. Ce jeu de donnée peut servir de base à une annotation supervisée de clusters de population du système immunitaire murin. C’est pourquoi l’équipe Immunomics du CIPHE est à la recherche d'un(e) stagiaire passionné(e) par l'analyse de données et la programmation pour rejoindre notre équipe et relever ce challenge. Le but principal de ce stage sera de développer un package R destiné à l'annotation supervisée de données de cytométrie.

Missions :

1. Conception et développement d'un package R pour l'annotation de données de cytométrie.
2. Benchmarking de différentes approches pour l'annotation des cellules. Vous explorerez, entre autres, les méthodes suivantes:
3. Approche basée sur le clustering (ex. K-means, DBSCAN)
4. Approche basée sur l'apprentissage supervisé (ex. SVM, forêts aléatoires)
5. Approches basées sur l'apprentissage profond (ex. réseaux de neurones)
6. Approche basée sur la réduction de dimensionnalité (ex. PCA, t-SNE)
7. Comparaison et évaluation des performances de ces méthodes par rapport à une annotation de référence.
8. Intégration des meilleures méthodes dans le package R.
9. Contribution au développement d'une application R Shiny destinée aux biologistes de l'équipe, facilitant l'exploitation et la visualisation des données.
10. Documentation et formation pour les utilisateurs finaux.

Compétences requises :

* Bonne maîtrise du langage de programmation R
* La connaissance de R Shiny est un plus, mais n'est pas obligatoire
* Connaissances en analyse de données et en algorithmes d'apprentissage automatique
* Capacité à travailler en équipe et à communiquer efficacement

Avantages :

* Travail au sein d'une équipe dynamique et passionnée
* Possibilité de contribuer à des projets ayant un impact réel dans le domaine de la recherche en cytométrie
* Apprentissage de la vulgarisation scientifique pour présenter efficacement les résultats bioinformatiques à une équipe de biologistes
* Acquérir de l'expérience dans le développement d'applications web front-end

Pour postuler :

Merci d'envoyer votre CV et lettre de motivation à [george-alehandro.saad@inserm.fr](mailto:george-alehandro.saad@inserm.fr) et herve.luche@inserm.fr avant la date limite du 31/10/2023