

Méthode d'analyse différentielle par groupe de gènes pour les données *single-cell RNA-seq* appliquées dans la recherche vaccinale

Stage M2/Fin d'étude


Contexte

Le séquençage de l'ARN, grâce à la technologie RNA-seq, s'est imposée comme le nouveau standard pour la mesure de l'expression génique. Les dernières avancées combinant cette technologie aux techniques modernes de micro-fluidique permettent aujourd'hui d'obtenir une image précise de l'expression génique à l'échelle unicellulaire (*single-cell RNA-seq*). La technologie RNA-seq en cellule unique permet de mesurer l'expression génique au sein de centaines, voire de milliers de cellules à partir d'un seul échantillon biologique. Cela permet d'étudier des phénomènes qui ne sont observables qu'à l'échelle uni-cellulaire.

Nous avons récemment développer une nouvelle approche pour l'analyse différentielle de données *single-cell RNA-seq* s'appuyant sur l'estimation et le test de fonctions de répartition conditionnelles¹. Un des avantages de cette approche est sa flexibilité évitant de faire des hypothèses distributionnelles sur les données. L'objet de ce stage est d'étendre cette méthodologie à l'analyse par groupe de gènes², qui permet de prendre en compte des connaissances biologiques *a priori* sur la corrélation et les interactions entre certains groupes de gènes, ce qui augmente la puissance statistique et facilite l'interprétation biologique des résultats.


Ce sujet de stage s'inscrit dans le programme de l'équipe associée Inria [DESTRIER](#) en collaboration avec Denis Agniel (*RAND Corporation, Santa Monica, California USA*). Ce travail de stage sera l'occasion d'échanges avec ces partenaires.

Objectifs

1. Faire une revue de la littérature scientifique recensant les différentes approches par groupe de gènes applicables aux données *single-cell RNA-seq*
2. Implémenter sous  l'extension de `ccdf` à l'analyse par groupe de gènes.
3. Comparer les éventuelles méthodes disponibles à l'extension de `ccdf` sur des données simulées et sur des données réelles (cellules T-CD4 de patients atteint de la COVID-19 à différents stade de sévérité³).

Compétences requises :

- très bonnes connaissances en Biostatistique

- bonnes connaissances de la programmation et du développement sous 
- Master 2 en Biostatistique, école d'ingénieur en statistique ou équivalent
- intérêt pour la recherche médicale, en particulier la recherche vaccinale
- maîtrise de l'anglais scientifique écrit et parlé

Structure d'accueil :**Équipe SISTM**

Inria Bordeaux Sud-Ouest & Inserm U1219 *Bordeaux Population Health*

Lieu de stage :

Centre de Recherche Inserm U1219 *Bordeaux Population Health* – Équipe SISTM

Université de Bordeaux – ISPED

146, rue Léo Saignat

33076 Bordeaux Cedex

Durée :

Stage de 5 à 6 mois

Rémunération :

Gratification selon les grilles (15% du plafond de la sécurité sociale, soit environ 550€/mois)

Contact :

Envoyer une lettre de motivation accompagnée d'un CV détaillé à **Boris Hejblum** :

boris.hejblum@u-bordeaux.fr

Bibliographie

1. Gauthier, M., Agniel, D., Thiébaud, R. & Hejblum, B. P. Distribution-free complex hypothesis testing for single-cell RNA-seq differential expression analysis. *bioRxiv* (2021) doi:[10.1101/2021.05.21.445165](https://doi.org/10.1101/2021.05.21.445165).
2. Ackermann, M. & Strimmer, K. [A general modular framework for gene set enrichment analysis](#). *BMC bioinformatics* **10**, 1–20 (2009).
3. Kusnadi, A. *et al.* [Severely ill patients with COVID-19 display impaired exhaustion features in SARS-CoV-2-reactive CD8+ t cells](#). *Science immunology* **6**, eabe4782 (2021).