

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

Stage M2/Fin d'étude


Sujet

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.

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  ventuelles m thodes disponibles   l'extension de **ccdf** sur des donn es simul es et sur des donn es r elles (cellules T-CD4 de patients atteint de la COVID-19   diff rents stade de s v rit ³).

Comp tences requises

- tr s bonnes connaissances en Biostatistique
- bonnes connaissances de la programmation et du d veloppement sous **R**
- Master 2 en Biostatistique,  cole d'ing nieur en statistique ou  quivalent
- int r t pour la recherche m dicale, en particulier la recherche vaccinale
- une ma trise de l'anglais scientifique  crit et parl 

Structure d'accueil

 quipe SISTM

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

Lieu de Stage

Centre de Recherche Inserm U1219 *Bordeaux Population Health* –  quipe SISTM

Universit  de Bordeaux – ISPED

146, rue L o Saignat

33076 Bordeaux Cedex

Dur e

Stage de 5   6 mois

R mun ration

Gratification selon les grilles (15% du plafond de la s curit  sociale, soit environ 550 /mois)

Contact

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).