

Solliciteur : Florent Chuffart

Table des matières

Problématique.....	1
Pistes pour l'analyse des données de lignées cellulaires.....	2
1. Analyse exploratoire	2
2. Analyse en 2 temps	2
3. Modèle hiérarchique ou multiniveaux ou mixte	2
4. Inversion du modèle	2
5. Complément.....	2

PROBLEMATIQUE

<https://github.com/fchuffar/cuter>

Données espèces / races avec des données non observables. Tester si une espèce / race est plus mignonne qu'une autre.

Données de lignées cellulaires mutées (MU) ou non (WT) pour p53, 3 réplicats techniques. Tester si le statut de p53 (WT vs. MU) a un effet sur le niveau d'expression de 20 000 gènes en prenant en compte la variabilité entre les réplicats biologiques au sein de chaque lignée. Pas de résultat attendu.

p53	Lignée
WT	A
WT	A
WT	A
WT	B
WT	B
WT	B
MU	C
MU	C
MU	C
MU	D
MU	D
MU	D
MU	E
MU	E

MU	E
MU	F
MU	F
MU	F
MU	G
MU	G
MU	G
MU	H
MU	H
MU	H

PISTES POUR L'ANALYSE DES DONNEES DE LIGNEES CELLULAIRES

Le petit jeu de données pose problème en modèle linéaire.

Les tests non paramétriques sont adaptés mais s'accompagnent d'une perte de puissance.

1. Analyse exploratoire

Pour dégager une première conclusion : qui du statut de p53 et de la lignée discrimine le mieux ?
Quel critère regroupe les échantillons ?

Analyse factorielle, classification ascendante hiérarchique, arbre de décision.

Cette première analyse sera complétée par des tests statistiques.

2. Analyse en 2 temps

Tests en 2 temps : i) test l'effet de p53 (WT vs. MU) avec Mann-Whitney ; ii) test de l'effet lignée cellulaire au sein des WT avec Mann-Whitney et au sein des MU avec Kruskal-Wallis suivi de tests 2 à 2 si effet global significatif.

3. Modèle hiérarchique ou multiniveaux ou mixte

Avec les librairies R nlme, lme4.

Tutoriel dispo à cette adresse : <https://mate-shs.cnrs.fr/actions/tutomate/tuto28-multiniveau-clementine-cottineau/>

4. Inversion du modèle

Régression logistique : facteur = f(niveau expression d'un gène + autre facteur)

5. Complément

Tracer un histogramme des pvaleurs sous H0 pour vérifier l'uniformité avec la librairie R cp4p (<https://cran.r-project.org/web/packages/cp4p/index.html>).

Recodage des variables.