

Contrôle Terminal - Module 3: Analyse statistique avec R

Diplôme Interuniversitaire en Bioinformatique intégrative (DU-Bii 2019)

2019-02-28

Contents

But de ce TP	1
Rendu	1
Jeu de données	1
Chargement des données	1
Exploration des données	1

But de ce TP

Le but de ce travail est de mobiliser les différentes méthodes statistiques enseignées durant les cours et travaux pratiques afin d'analyser un jeu de données, et d'interpréter les résultats.

Rendu

Les énoncés et questions sont fournis sous forme d'un document R markdown, qui servira également de modèle pour répondre aux questions.

En-dessous de chaque question nous vous demandons de fournir d'une part le code qui permet d'y répondre, et une interprétation sommaire des résultats (5 à 15 lignes par question).

Le rapport de travail consistera en un document R markdown permettant aux évaluateurs de reproduire l'ensemble des résultats, et le rapport pdf produit au moyen de ce markdown.

Jeu de données

Pour vous faciliter la tâche, les analyses se baseront sur le jeu de données de transcriptome RNA-seq de l'étude Breast Invasive Cancer du TCGA (**TCGA-BIC**), mais nous sélectionnerons des sous-ensembles de données de tailles raisonnables pour les différentes parties du travail.

Chargement des données

Pour vous éviter de perdre trop de temps avec la localisation des fichiers de données, nous fournissons nous-même les instructions qui permettent de les charger.

Exploration des données

Questions

1. Dessinez un histogramme des comptages bruts pour toutes les valeurs de la table, et commentez la figure en quelques lignes.

```
## Histogram of the raw counts
```

2. Dessinez un histogramme des $\log_2(\text{counts} + \epsilon)$ pour toutes les valeurs de la table, en fixant $\epsilon = 0.1$, et commentez la figure en quelques lignes

```
## histogram of log2-transformed counts
```

3. Sélectionnez aléatoirement un sous-ensemble de 20 échantillons (`sample()`), et dessinez un box plot des comptages bruts et log2-transformés. Commentez la figure en quelques lignes.

```
## Box plot of the raw counts
```

```
## Box plot of the log2-transformed counts
```