**Feuille de route de l’atelier bioinformatique - CMI-SCM6U03L-2020**

" Analyse de données RNA-seq provenant de maladies inflammatoires chroniques de l’intestin "

**Contexte biologique**

Il s’agit de RNA-seq à partir de biopsies de colon de patients atteints d’une maladie inflammatoire chronique de l’intestin (MICI), soit la maladie de Crohn ou bien la rectocolite hémorragique (RCH/Ulcerative colitis). Ces patients sont sous traitement ou pas. Il y a aussi une cohorte témoin de colons sains. Comme ces maladies évoluent par poussées inflammatoires, il y a chez les patients non-traités des états quiescents ou actifs de la maladie. Chez les patients traités il y a des non-répondeurs aux traitements, ainsi on retrouve des états quiescents quand le traitement est efficace ou actifs chez qui le traitement n’a pas d’effet curatif.

Cette étude cherche à identifier des marqueurs potentiellement utilisables en clinique et/ou utiles pour la compréhension des mécanismes physiopathologiques de ces maladies. Ainsi les questions posées sont les suivantes**\***:

→ Est-il possible d’identifier des marqueurs associés à la sévérité de ces maladies ?

→ Est-il possible d’identifier des marqueurs de résistance aux traitements ?

**Dans cet atelier nous nous centrerons sur la maladie de Crohn.**

**Échantillons**

***Organisation des échantillons***

Prélèvements humains à partir de biopsie des différentes maladies indiquées, ou de patients sains (contrôles)



**Figure 1. Répartition des échantillons biologique utilisés pour les analyses RNA-seq**

***Détails des échantillons (formats, réplicats, batch ,etc)***

Row cells HS292, HS297, HS301 et HS302 en 50-length Single-Read.



***Détails du séquençage***

***Mapping***

Reads mapés sur le genome *hg38* Homo sapiens avec Tophat 2.0.14 [1,2] and bowtie version 2-2.1.0[3]. Seul les reads mapé une seul fois ont été retenus pour les étapes suivantes d’analyse.

***Quantification***

L’expression des gènes ont été quantifié avec HTSeq-0.6.1 [4] et les annotations de *Ensembl 87*. Seuls les reads assignés de façon non-ambigue ont été retenus pour les étapes suivantes d’analyse.

**Objectifs spécifiques aux données RNA-seq des échantillons Crohn**

**Faire réfléchir les étudiants sur les objectifs spécifiques, une question par groupe pour les DEG. Montrer les questions dans un 2ème temps.**

**Options (à valider)**



**Pipeline d’analyse**

**T= Théorie ; P=Pratique**

**AM : Matin ; PM : Après-midi**

* ***(Jour 1, mardi 2 Juin) Pre-processing (T : AM/PM, P : PM) faisable ? Sinon reaprtir sur J2)***

1. *QCs reads* ***(T)***
2. *Mapping* ***(T)***
3. *Quantification* ***(T)***
4. QC réplicats **(T+P)**
5. *Normalisation* **(T+P)**
6. QC réplicats **(T+P)**

* ***(Jour 2, mercredi 3 juin) Analyse exploratoire ou non-supervisé (avant DEG ) (T : AM ; P PM)***

1. Reduction de dimensions : PCA **(T+P)**
2. Clustering hierarchique (HC) **(T/P)**

* ***(Jour 3, jeudi 4 juin) Analyses différentielles par questions biologiques (T : AM ; P PM)***
  1. *Vulcano plots* **(T+P)**
  2. *HM et liste de gènes* **(T+P)**
  3. *Diagramme de Venn (****T+P****)*
* ***(Jour 4, vendredi 5 juin ) Analyses fonctionnelles (T : AM ; P PM)***
  1. *GSEA***(T+P)**
  2. *Pathway analysis* **(T+P)**
  3. *Comparaison avec signatures d’intérêt ? Franck/Philippe en avez-vous en tête pour quelles soient dispo pour le cours et si le temps le permet*
* ***Conclusions (Jour 5, lundi 8 juin ) et feed-back question biologique***

**Scriptes R**

Implémenter sur .rmd par intervenants bioinfo