**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**

***Échantillons***



**Données RNA-seq**

***Détails du séquençage***

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

***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.

***Détails des échantillons liés aux données RNA-seq***

****

**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.**

1. Identifier les gènes qui caractérisent les types de patients : (NTQ, NTA, TNFQ, TNFA) vs CTRL (tous)
2. Identifier les gènes associés à l’activité de la maladie : (NTQ vs NTA)
3. Identifier les gènes associés à l’efficacité du traitement anti-TNF : (NTA vs TNFQ)
4. Identifier les gènes associés à la résistance au traitement anti-TNF : (NTAvsTNFA)
5. Identifier les gènes associés à l’évolution de la maladie

**Pipeline d’analyse**

* **J1- Formatage des données**
* QCs *reads*
* *Mapping* (théorique)
* Quantification (counts) (théorique)
* QC réplicats
* Annotation
* Filtration
* Visualisation : HM, distribution
* **Normalisation**
* **J2 - QC des données**
* **Matrice de corrélations et HC**
* Réduction de dimensions : PCA
* **J3 - Analyses d’expressions différentielles**
  + ***Méthode d’analyse différentielle***
  + *Interactions*
  + *Visualization : MA-plots, Volcano plots, HM*
  + *Diagramme de Venn*

*Feed-back sur la question biologique*

* **J4 - Analyses fonctionnelles** 
  + Analyse d’enrichissement : voie de signalisation, processus biologique, fonction moléculaire
  + *GSEA*
* **J5 - Conclusions** 
  + **Création du rapport. À rendre en fin de journée.**