

Analyse des séquences répétées dans les données omics par l'approche des kmers

Encadrants : Anthony Boureux et Nicolas Gilbert (IRMB, inserm U1183)

Contexte :

Il est clair que les éléments mobiles ont, et continuent de contribuer à la structure, la variabilité et la stabilité des génomes. Les éléments transposables, et en particulier les rétrotransposons, sont abondants dans le génome humain. LINE-1 (L1) est le seul élément actif autonome connu et plusieurs copies sont encore capables de mobilité à travers un intermédiaire ARN. Dans les cancers, la perturbation des mécanismes cellulaires qui répriment normalement l'activité de L1 peut faciliter son effet mutagène. En effet, des travaux récents ont montré l'activité de L1 dans plusieurs cancers (Kong et al. 2019; Rodriguez-Martin et al. 2020).

Projet :

- Outils informatiques de gestion de très grands jeux de données (Milliard de k-mers) qui permettent la combinaison des localisations génomiques et des métadonnées.

Les objectifs étant de réduire les rétrotransposons à une combinaison de k-mers uniques tout en conservant les informations importantes liées à chaque élément.

- Application dans l'analyse de transcriptomes

L'objectif est d'identifier et de caractériser l'activité transcriptionnelle de chaque rétrotransposons à l'échelle individuelle (LINE – SINE – hERV) dans les cancers à l'aide de données de séquençage à haut débit (milliers de données RNAseq).

- Développement d'outils informatiques pour établir des « clusters d'expression »
- Application à d'autres données « omics » notamment pour l'analyse de ChipSeq (protéines de liaison à l'ADN)

L'objectif est de mesurer un enrichissement ou appauvrissement de chaque rétrotransposons à l'échelle individuelle.

Logiciels utilisés et développés au laboratoire (<https://bio2m.fr>) :

- Kmerator, génèse des kmers
- Reindeer, outils d'indexation et comptage des kmers

Méthode : développement R, python et C++

Projet développé en vue de réaliser une thèse par la suite.

Contacts : anthony.boureux@umontpellier.fr et nicolas.gilbert@inserm.fr