

Histoire évolutive d'une tumeur

Projet

Au cours d'un cancer, différentes mutations s'accumulent sur les cellules cancéreuses, générant plusieurs lignées cellulaires qui co-existent dans une tumeur donnée. L'objectif de ce projet est d'étudier l'histoire évolutive d'une tumeur à partir de données de séquençage haut débit dites "bulk", c'est-à-dire mélangeant les différentes cellules de la tumeur.

Ces données sont complexes à la fois pour des raisons biologiques et techniques. L'évolution d'un cancer dépend en effet de nombreux processus biologiques, qui induisent notamment des mutations, des altérations structurelles de certaines régions du génome dans certaines cellules, ainsi que des variations de tailles de la tumeur. D'un point de vue technique, le séquençage haut-débit ne fournit pas des séquences entières de génomes, mais renvoie un très grand nombre de petits fragments, appelés "reads", que l'on place sur une séquence de référence pour pouvoir les exploiter. Dans le cas de données *bulk* où l'on séquence plusieurs cellules, il n'est de plus pas possible de déterminer directement à quelle cellule appartient tel ou tel *read*.

L'objet général du stage est de développer un modèle stochastique de données de séquençage *bulk* d'une tumeur. Un tel modèle se décompose naturellement deux parties. Une première partie modélise l'évolution biologique de la tumeur, qui détermine sa composition, c'est-à-dire le nombre de cellules de chaque lignée présente, à un temps donné. Le nombre de cellules de chaque type peut en particulier être vu comme le résultat d'un processus de naissance et mort, les mutations créant les nouvelles lignées suivant un processus de Poisson donnant les différentes lignées. La seconde partie modélise le processus de séquençage haut débit, soit l'ensemble des *reads* observés sachant l'effectif de chaque lignée cellulaire.

Ce modèle pourra en premier lieu être utilisé pour simuler des données de séquençage sous diverses hypothèses biologiques, afin de tester la robustesse et la précision des méthodes de reconstructions déjà existantes, telles que *Pairtree* (Wintersinger et al., 2022). Un second objectif, plus ambitieux, sera de calculer la vraisemblance des données sous ce modèle, notamment en se basant des résultats obtenus dans Didier and Laurin (2020). L'étude statistique de ce modèle pourra être approfondie lors d'une éventuelle thèse suivant le stage.

Environnement et compétences attendues

Le stage se déroulera à l'**Institut Montpelliérain Alexander Grothendieck (IMAG)**. Il sera encadré par Gilles Didier et Paul Bastide, en collaboration avec Alice Cleynen et Sophie Lèbre. Le projet s'inscrit dans l'**ANR IdenTHiC** (*Identification of Tumor HHistory at the Clone level*), qui porte plus généralement sur l'étude de données cliniques de patients atteints de cancers pour l'aide au diagnostic. Là où le stagiaire pourra interagir avec les autres membres de l'ANR, et une bourse de thèse sur le sujet pourra être disponible. Le stage, d'une durée de 4 à 6 mois, pourra commencer dès février ou mars 2024. Il sera rémunéré en fonction du taux légal d'indemnité de stage en vigueur. Le développement de l'outil de simulation nécessitent un goût pour la programmation, notamment en R ou python. L'étude du modèle met en oeuvre des compétences en probabilité et statistiques et bénéficierait d'un intérêt pour les applications biologiques.

Contact

CV et lettre de motivation sont à adresser à : Gilles Didier, Paul Bastide, Alice Cleynen et Sophie Lèbre (prenom.nom@umontpellier.fr).

Références

Didier, Laurin. 2020. *Systematic Biology*. 69:1068–1087.

Wintersinger, Dobson, Kulman, et al. 2022. *Blood Cancer Discovery*. 3:208–219.