

Proposition de stage de M2

Durée : 6 mois

Lieu : Hospices Civils de Lyon, Groupement Hospitalier Est, Centre de Biologie et Pathologie Est

Contexte

Les variants structuraux (SV pour « *Structural Variants* ») désignent des remaniements dans l'organisation du génome pouvant concerner plusieurs millions de paires de nucléotides. Deux catégories d'événements peuvent être distingués : (i) les variants équilibrés tels que les translocations ou les inversions ; (ii) les variants affectant le nombre de copies, que sont les délétions et duplications (CNV, pour *Copy Number Variants*).

Le projet de Recherche « ANI », mené par le Dr Caroline Schluth-Bolard, s'est intéressé à l'efficacité du séquençage de génomes complets quant à la détection de ces événements. En effet, cette stratégie permet théoriquement de séquencer l'ensemble des points de cassure associés aux SV. Dans le cadre de ce projet, un pipeline de détection des SV basé sur le croisement des résultats de plusieurs logiciels a été mis en place. Ce pipeline s'est montré efficace, permettant le diagnostic de plus de 50% des patients inclus dans le projet ANI. Le pipeline a ainsi été transféré en routine diagnostique dans le Laboratoire de Cytogénétique des Hospices Civils de Lyon (HCL).

Ce pourcentage de diagnostics pourrait encore être amélioré grâce à l'ajout de nouvelles fonctionnalités. Notamment, en intégrant dans le pipeline un ou plusieurs logiciels de détection des CNV utilisant la profondeur de séquençage, et en incluant des annotations à partir de bases de données publiques (base de données gnomAD-SV par exemple), qui sont un support efficace pour filtrer les variants polymorphiques.

Missions et profil

Le candidat sélectionné aura pour mission principale d'améliorer le pipeline de détection des SV des HCL. Il devra tout d'abord identifier dans la bibliographie un lot de logiciels de détection des CNV pertinents pour des données « génome complet ». À l'aide d'un jeu de CNV de référence, il effectuera la comparaison de ces logiciels, et implémentera dans le pipeline les plus efficaces d'entre eux. Il validera enfin son outil (sensibilité/spécificité...) sur un second jeu de CNV. D'un autre côté, il améliorera l'annotation des SV, en utilisant notamment l'information contenue dans des bases de données d'intérêt pour les biologistes. Ces derniers valideront ces nouvelles annotations grâce à des tests en aveugle sur des patients déjà diagnostiqués.

Le candidat sélectionné devra être familier avec la lecture d'articles en anglais. Il devra être capable de travailler en autonomie au sein d'une petite cellule de bioinformatique, où la communication est toutefois primordiale. Il devra être à l'aise dans un environnement *linux*, connaître un langage de script (*perl*, *python*, ...) et savoir utiliser les outils de développement collaboratif (*git*). Il devra être sensible aux questions d'ergonomie, et s'assurer que les résultats fournis sont faciles à manipuler par les biologistes.

Contacts ;

Pierre-Antoine Rollat-Farnier : pierre-antoine.rollat-farnier@chu-lyon.fr