



Université Claude Bernard  Lyon 1

DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **11 juillet 2023**

Nom de famille et prénom de l'auteur : **Madame ZHAI Yue**

Titre de la thèse : « *Évaluation et amélioration des performances du séquençage de nouvelle génération en absence de gold standard* »



Résumé

Les travaux de cette thèse ont porté sur la place du modèle de concordance-discordance dans l'évaluation de la performance du séquençage à haut débit et sur des comparaisons de performance entre modèles de classification dans la reconstitution de résultats de séquençage haut débit à partir de réplicats techniques.

Le premier travail a étudié le problème de l'évaluation du séquençage haut débit en l'absence de 'gold' standard et, dans ce cadre, la pertinence des critères de concordance-discordance. Il a examiné les relations entre le taux de discordance et le taux d'erreur dans diverses situations théoriques. Il a ensuite analysé les effets des covariables sur ces deux taux en utilisant un modèle additif généralisé avec des données réelles issues de réplicats de séquençage du génome NA12878.

Le second travail a étudié le problème de la fusion de résultats de séquençage haut débit de réplicats techniques en vue d'obtenir un nouveau jeu de données susceptible de comporter moins d'erreurs. Il a évalué et comparé les aptitudes des principaux modèles de partitionnement à améliorer la performance finale du séquençage à partir des résultats de trois séquençages du génome NA12878. L'étude fournit des arguments pour choisir le modèle le plus convenable et utiliser ces résultats en matière de diagnostic ou de médecine de précision.

