

Titre

Intégration de connaissances a priori pour la construction de modèle prédictif dans le cadre de données d'expression de gènes

Présentation de la société

Acteur mondial dans le domaine du diagnostic in vitro depuis 45 ans, bioMérieux est présente dans plus de 150 pays au travers de 39 filiales et d'un large réseau de distributeurs. En 2008, le chiffre d'affaires de bioMérieux s'est élevé à 1,111 milliard d'euros, dont 84% ont été réalisés à l'international.

bioMérieux offre des solutions de diagnostic (réactifs, instruments et logiciels) qui déterminent l'origine d'une maladie ou d'une contamination pour améliorer la santé des patients et assurer la sécurité des consommateurs. Nos produits sont utilisés dans le diagnostic des maladies infectieuses et apportent des résultats à haute valeur médicale pour les urgences cardiovasculaires, le dépistage et le suivi des cancers. Ils sont également utilisés pour la détection de micro-organismes dans les produits agroalimentaires, pharmaceutiques et cosmétiques. bioMérieux est une société cotée sur Euronext Paris.

Contexte du stage

Dans le cadre de la recherche de biomarqueurs pour le développement de tests de diagnostic in vitro, l'étude de l'expression des gènes est considérée comme une approche clé notamment dans le domaine de l'oncologie. La technologie des « puces à ADN » permet aujourd'hui de quantifier simultanément l'expression de l'ensemble des gènes d'un individu dans un tissu donné et constitue ainsi la technologie de référence dans ce domaine. Néanmoins, compte tenu de la taille des échantillons généralement considérés pour une question clinique donnée, les techniques usuelles d'apprentissage supervisé demeurent limitées du fait même de la dimension des données (généralement 100 fois plus de gènes que d'individus). La sélection du sous-ensemble « optimal » de gènes demeure ainsi une question ouverte.

Au sein d'un projet de recherche de biomarqueur pour un test de dépistage du cancer du côlon, le stage vise ainsi à évaluer et développer de nouvelles méthodes de sélection supervisée de gènes par l'intégration de connaissance a priori en lien avec le rôle fonctionnel des gènes. Cette approche dite « intégrative » devrait permettre d'améliorer la robustesse du modèle prédictif final.

Sujet proposé

Le sujet de stage comporte 3 objectifs principaux :

- identifier les sources publiques d'information biologique pertinentes vis à vis de la question clinique (eg. GO, KEGG ...)
- faire une revue de la littérature sur les méthodes d'intégration des différentes sources d'information biologiques et participer à de nouveaux développements méthodologiques.
- mettre en œuvre les outils identifiés/développés sur un jeu de donnée interne

Profil recherché

Niveau Master 2 mathématiques appliquées / biostatistique ou troisième année d'école d'ingénieur (dominante mathématiques appliquées)

Compétences

- Compétences en apprentissage statistique, biostatistique, bioinformatique, mathématiques appliquées
- Maîtrise des outils informatiques : environnement UNIX, langage de programmation (fonctionnel ou objet)

- La connaissance du logiciel R et du langage C serait un plus

Conditions d'accueil

Le stagiaire sera encadré par un ingénieur de l'équipe «biomathématique, recherche de biomarqueurs» dans le département des technologies de l'information. Il/elle sera physiquement basé(e) sur le site de bioMérieux Grenoble.
Stage de six mois rémunéré.

Contact

Laurence.lesueur@eu.biomerieux.com