

**Unité Mixte de Recherche – Inserm U1283 / CNRS UMR 8199**  
***(Epi)génomique Fonctionnelle et Physiologie Moléculaire***

***Du Diabète et Maladies Associées***

Directeur : Pr. Philippe FROGUEL

**INSERM, CNRS, Université de Lille, Institut Pasteur de Lille, CHU de Lille**

**Ingénieur(e) en Biostatistique et Analyse de Données Génétiques / Data Scientist**  
**Contrat CDD niveau IE de 1 an à temps plein (renouvelable)**

**Qui sommes-nous ?**

L'unité **Inserm U1283 / CNRS UMR 8199** (*(Épi)génomique Fonctionnelle et Physiologie Moléculaire du Diabète et Maladies Associées*) est un laboratoire de recherche regroupant 68 personnes dirigées par le **Professeur Philippe FROGUEL**. Elle fait partie de l'*Institut Européen de Génomique du Diabète (EGID)* et a été lauréate des appels à projets "Laboratoire d'Excellence" (**LABEX**) et "Équipement d'Excellence" (**EQUIPEX LIGAN MP**).

Site web : <http://www.good.cnrs.fr>

Le groupe biostatistique est en charge des analyses statistiques et contribue à l'élaboration du design des études, notamment dans le cadre des demandes de financements. Le groupe biostatistique comporte trois membres et est supervisé par Amélie Bonnefond et Stefan Gaget.

**Que faisons-nous ?**

Nos activités de recherche portent sur la caractérisation de variations génétiques associées à des maladies métaboliques telles le diabète et l'obésité et utilisent les approches modernes de génomique, bioinformatique, biostatistique, biologie moléculaire et modèles animaux.

L'objectif principal du groupe biostatistique consiste à apporter un soutien méthodologique fort aux différentes stratégies proposées par le laboratoire. Cela inclut, des développements méthodologiques et le développement d'outils (programmation), d'analyse et de visualisation, notamment pour les données issues des technologies de puce et de séquençage à haut débit (*p. ex.*, NGS, RNA-seq, Methyl-array, *etc.*).

Site web : <http://www.good.cnrs.fr/bio-informatique-bio-statistique/>

**Nos réalisations**

Notre expertise dans l'utilisation du programme R (<https://www.r-project.org/>) et des extensions qui lui sont associées, nous a permis de développer des applications **shiny** (extension R) pour l'analyse et la visualisation de données de différentes natures (<https://github.com/umr1283>), ainsi que des outils dynamiques (calcul de puissance statistique ou exploration de base de données) dans un rôle de soutien à l'activité de recherche de l'unité.

Le groupe biostatistique a également développé des extensions R (**NACHO "NAostring quality Control dasHbOard"** ; **insane, "INSulin Secretion ANALysEr"** ; **snpEnrichment, "SNPs Enrichment Analysis"**), contribué à de nombreuses publications scientifiques et consortia internationaux ou encore présenté ses réalisations lors de conférences ou colloques.

## Qui êtes-vous ?

### Activités dominantes

Le/la collaborateur/collaboratrice recherché(e) interviendra en amont (design expérimental, calcul de puissance) et en aval des études de type omiques et biologiques réalisées par le laboratoire. Il ou elle participera grâce à ses connaissances en statistiques et en analyse des données à répondre à des questions de recherche variées, et sera capable d'implémenter des idées rapidement et de créer des chaînes de traitement reproductible.

- Réalisation et interprétation des analyses statistiques utilisées dans les études génétiques (*p. ex.*, régressions logistiques, régressions linéaires, analyses de survie, analyses de variance, score génétique).
- Développement et application de chaînes de traitement de contrôle-qualité et d'analyse (de la collecte à l'analyse des données).
- Apporter le conseil aux chercheurs et ingénieurs dans la mise en œuvre des méthodes et des outils statistiques. (Déterminer les méthodes statistiques ainsi que les logiciels/extensions les mieux adaptés pour une étude donnée.)
- Rédaction de rapport d'analyses et des sections des articles scientifiques liées à la méthodologie statistique (*p. ex.*, avec Rmarkdown, shiny ou quarto). Production de figures et de tables sur mesure (Avec des extensions ggplot2, gt, *etc.*).
- Participation à la veille technique et scientifique.

### Profil et compétences recherchés

- Formation supérieure (Master 2, ingénieur ou doctorale) en (bio)statistique, mathématique ou domaine apparenté - Expérience souhaitable, débutants acceptés selon profil.
- Maîtrise théorique et appliquée des statistiques.
- Maîtrise de la collecte et du traitement de données.
- Maîtrise du langage R avec ses outils de visualisation des données. (connaissance de Shiny, ggplot2, targets et renv seraient un plus).
- Connaissance d'autres langages comme Python et/ou bash serait un plus.
- Programmer dans différents environnements informatiques (*p. ex.*, UNIX/Linux, *etc.*).
- Appliquer un dispositif de contrôle de version (*p. ex.*, Git).
- Rédiger de la documentation, des rapports d'analyses et tout ou partie d'article scientifique.
- Connaissance et/ou maîtrise de la génétique et épigénétique serait un atout.
- Expérience ou connaissance dans le développement de modèle (« machine learning », « bootstrapping », approches de régression *etc.*) et dans la réalisation d'analyse prédictive.
- Expérience de développement collaboratif à l'aide des outils git et docker.
- Langue anglaise : B1 à B2 (communication orale régulière et rapport écrit).
- Qualités organisationnelles (autonomie, respect des délais, anticipation des sources de retards, *etc.*), présentation synthétique et didactique des résultats scientifiques.
- Capacité (écrite et orale) de communication, être capable d'expliquer des résultats techniques à une audience non technique ou spécialisé dans le domaine.
- Goût pour le travail en équipe.

### Informations administratives

- Début : Dès que possible.
- Durée : CDD de 1 an à temps plein (renouvelable).
- Rémunération : selon les grilles contractuelles de l'Université de Lille, à partir de 1812€ brut/mois sans expérience.
- Lieu : Au sein du groupe biostatistique de l'unité Inserm U1283 / CNRS UMR 8199.  
EGID - Inserm U1283 / CNRS UMR 8199, Pôle Recherche - 1er étage Aile Ouest  
1, Place de Verdun, 59045 LILLE CEDEX, France

### Contact

Envoi d'un CV et d'une lettre de motivation à :

Stefan Gaget (stefan.gaget@cnrs.fr) et  
Amélie BONNEFOND, Ph.D. (amelie.bonnefond@cnrs.fr)



www.good.cnrs.fr  
UMR1283/8199 - EGID - Faculté de Médecine - Pôle Recherche  
1 Place de Verdun - Aile Ouest - 1er étage - 59045 LILLE CEDEX  
Tél. : 33-(0)3-74-00-81-01 (ou) 81-00 (secrétariat)

