

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

Biostatisticien-ne / Data Scientist

Contrat post-doctoral ou Contrat CDD de 1 an à temps plein (renouvelable)

L'unité Inserm U1283 / CNRS UMR 8199

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 60 personnes dirigées par le Professeur Philippe FROGUEL. Elle fait partie de l'Institut Européen de Génétique du Diabète (EGID) et a été lauréate en 2011 (renouvelé jusqu'en 2025) des appels "Laboratoire d'Excellence" (LABEX) et "Équipement d'Excellence" (EQUIPEX LIGAN MP).

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.

Le groupe biostatistique

Qui sommes-nous ?

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 supervisée par Mickaël CANOUIL, Ph.D.

Que faisons-nous ?

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 de visualisation et d'analyse, notamment pour les données issues des technologies de puce et de séquençage à haut débit (*p. ex.*, Methyl-seq, RNA-seq, *etc.*).

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 la visualisation et l'analyse de données de différentes natures et en particulier les données de transcriptomiques (*p. ex.*, NanoString, qPCR, *etc.*), ainsi que des outils dynamiques de calcul de puissance statistique, dans un rôle de soutien à l'activité de recherche de l'unité.

Le groupe biostatistique a également développé des extensions R :

- **NACHO** "*NA*nostring quality Control *dasHbOard*";
- **insane**, "*INS*ulin Secretion *AN*alys*Er*";
- **snpEnrichment**, "*SNPs* Enrichment Analysis";
- **clere**, "*Simultaneous Variables Clustering and Regression*".

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.
- Production de figures et de tables sur mesure pour des articles scientifiques.
- Rédaction des sections des articles scientifiques liées à la méthodologie statistique.
- Participation à la veille technique et scientifique.

Profil recherché

- Formation supérieure (Master 2 / ingénieur) ou doctorale en biostatistique en (bio)statistique, mathématique ou domaine apparenté - Expérience souhaitable, débutants acceptés selon profil.
- Très bonnes connaissances théoriques et appliquées en statistique.
- Expérience ou connaissance dans le développement de modèle (« machine learning », « bootstrap », *etc.* et/ou approches de régression) et dans la réalisation d'analyse prédictive.
- Maîtrise du langage R avec ses outils de visualisation des données (*p. ex.*, Shiny et ggplot2).
- Connaissance et/ou maîtrise des extensions : **targets** et **renv**.
- La maîtrise d'autres langages comme Python ou C/C++ serait un plus.
- Bonne connaissance de l'environnement Unix/Linux (en particulier Debian).
- Expérience de développement collaboratif à l'aide des outils **git** et **docker**.
- Connaissances dans le domaine de la génétique et de l'épigénétique serait un atout.
- Maîtrise écrite et orale de l'anglais.
- Qualités organisationnelles (autonomie, respect des délais, anticipation des sources de retards, *etc.*), présentation synthétique et didactique des résultats scientifiques.
- Grande capacité (écrite et orale) de communication, être capable d'expliquer des résultats techniques à une audience non technique ou spécialisé dans le domaine
- Une expérience en management serait appréciée.

Informations administratives

- *Début* : Dès que possible.
- *Durée* : CDD de 1 an à temps plein (renouvelable).
- *Lieu* :
Au sein du groupe biostatistique de l'unité Inserm U1283 / CNRS UMR 8199.
EGID - Inserm U1283 / CNRS UMR 8199
Pôle Recherche - 1^{er} étage Aile Ouest
1, Place de Verdun
59045 LILLE CEDEX, France

Contact

Envoi d'un CV et d'une lettre de motivation à Mickaël CANOUIL, Ph.D. et Amélie BONNEFOND, Ph.D. :

- mickael.canouil@cnrs.fr
- amelie.bonnefond@cnrs.fr



N° ordre (Philippe FROGUEL) : 590784922
UMR1283/8199 - EGID – Faculté de Médecine – Pôle Recherche
1 Place de Verdun – Aile Ouest – 1^{er} étage – 59045 LILLE CEDEX
Tél. : 33-(0)3-74-00-81-01 (ou) 81-00 (secrétariat)

