



Fiche de poste

Ingénieur(e) de Recherche bioinformaticien(ne)

Description des équipes :

L'équipe 1 de l'UMRS 1166 (<https://recherche-cardiovasculaire-metabolique.fr/>) est impliquée dans la recherche des causes génétiques et cellulaires des pathologies cardiovasculaires tant sur des modèles animaux et humains (tissus cardiaques, cellules pluripotentes induites (iPSC) ou bien modèle murin) que sur des données génomiques chez l'homme, notamment par le biais d'analyses bioinformatiques.

Ces études portent principalement sur les cardiomyopathies qui constituent l'une des causes principales d'insuffisance cardiaque et de mort subite du sujet jeune.

Jusqu'à présent ces approches bio-informatiques ont principalement consisté en des analyses d'association génome entier (GWAS) sur des cohortes de cas sporadiques et des témoins, ou bien dans le séquençage d'exome à la recherche de variants pathogéniques impliqués dans les formes familiales. Ces approches ne suffisent cependant pas à identifier tous les facteurs génétiques impliqués.

Récemment, l'équipe a obtenu un nouveau financement *via* l'appel à projet « maladie rare_résoudre les impasses diagnostiques » financé par l'ITMO Génétique, Génomique et Bioinformatique (ITMO GGB) d'Aviesan dans le cadre du Programme Prioritaire de Recherche sur les maladies rares. Cet appel à projet a été demandé en collaboration avec d'autres équipes, dont celle d'Antonio Rausell ([Laboratoire de bioinformatique clinique, Institut Imagine, https://www.institutimagine.org/en/RausellLab](https://www.institutimagine.org/en/RausellLab)), pour développer de nouvelles approches afin de résoudre ces impasses diagnostiques.

Objectif et missions :

Le poste s'adresse à un(e) candidat(e) motivé(e) par une activité de recherche qui aura la charge de mener ces analyses bioinformatiques innovantes, sur les données génomiques déjà constituées, grâce à des outils de détection récemment développés mais aussi *via* des approches innovantes basées sur l'intelligence artificielle (AI), afin d'identifier de nouveaux facteurs génétiques dans les régions codantes et non codantes du génome tels des variations du nombre de copies (CNV) ou des mutations pertes de fonction (LoF).

Le poste aura une double localisation à la faculté de médecine Sorbonne Université site de la Pitié Salpêtrière, au sein de l'UMR_S 1166 (91 boulevard de l'Hôpital, 75013 Paris) et à l'Institut Imagine (24 Boulevard du Montparnasse, Hôpital Necker AP-HP, 75015 Paris).

Activités principales :

- Réaliser les contrôles qualité sur les données génomiques et cliniques afférentes aux différents projets des équipes
- Analyser les données génomiques de séquençage haut-débit, de puces à ADN, des outils de détection des CNV et de LoF produites dans les deux équipes
- Participer au développement et à la réalisation des approches d'IA développées par le laboratoire de bioinformatique clinique de l'Institut Imagine
- Concevoir ou développer de nouveaux outils bioinformatiques pour répondre aux besoins rencontrés par les chercheurs de l'unité
- Interagir étroitement avec les biologistes de l'unité responsable des divers projets génomiques menés afin d'optimiser le processus d'interprétation de l'impact moléculaire et cellulaire des variants

identifiés, sa cohérence avec les signes clinique des patients, ainsi que son rapport avec la pathophysiologie des maladies étudiées

Connaissances

- Analyses bioinformatiques et statistiques de données de puces à ADN
- Analyses bioinformatiques des données de séquençage (notamment exome/génome et ARN messager) dans le cadre de la recherche clinique
- Connaissances dans le domaine du machine learning
- Expérience dans l'analyse statistique de données
- Connaissances générales en biologie moléculaire (génétique) afin de comprendre et analyser le problème scientifique posé et de pouvoir interagir au mieux avec les chercheurs des unités
- Langue anglaise : B2 à C1

Compétences opérationnelles

- Maîtriser des techniques de programmation sous environnement Unix/Linux ainsi que des langages de programmation Python et R
- Maîtriser des logiciels d'analyse des données de puces d'ADN (Plink, etc.) et des analyses de type GWAS ou d'imputation
- Maîtriser l'analyse d'exomes / génomes y compris l'annotation des effets de variants et la priorisation à l'aide de scores de pathogénicité
- Maîtriser l'utilisation des clusters/grilles de calculs pour le calcul scientifique haute-performance et le calcul en parallèle.
- Maîtriser les bases de données publiques en génomique (NCBI, GenCode, Biomart, etc.)
- Savoir gérer le stockage et la manipulation des données de grande dimension
- Rédiger des rapports de synthèse et les présenter oralement (anglais et français) lors de réunion de travail.

Compétences comportementales :

- Rigueur scientifique et expérimentale
- Capacité à travailler en équipe

Expérience souhaitée :

- Débutant(e) accepté(e)

Formation souhaitée :

- Diplôme d'ingénieur(e) ou thèse

Niveau de recrutement selon les grilles de rémunération Sorbonne Université

CDD de 2 ans avec une double affectation dans l'équipe 1 de l'UMR_S 1166 ICAN et au Laboratoire de bioinformatique clinique de l'Institut Imagine (UMR 1163)

Date d'ouverture du poste : dès septembre 2021

Contact : Envoyer CV et lettre de motivation par email à Sophie Garnier, sophie.garnier@sorbonne-universite.fr

Job Description

Research Engineer in Bioinformatics

Teams' description

Team 1 of UMRS 1166 (<https://recherche-cardiovasculaire-metabolique.fr/>) is involved in the research of the genetic and cellular causes of cardiovascular diseases both in animal and human models (cardiac tissues, induced pluripotent cells (iPSC) or mouse model) and in human genomic data, notably through bioinformatic analyses.

These studies focus on cardiomyopathies, one of the leading causes of heart failure and sudden death of the young person.

Until now, the bioinformatic approaches have mainly consisted on whole genome association analyses (GWAS) on sporadic cases cohorts and controls, or on exome sequencing to search for pathogenic variants involved in familial forms. However, these approaches are not sufficient to identify all the genetic factors involved.

Recently, the team obtained a new funding « maladie rare _ résoudre les impasses diagnostiques » (“rare disease _ solving diagnostic impasses”) through the Aviesan ITMO Genetics, Genomics and Bioinformatics (ITMO GGB) call for Rare Disease Research Projects funded under the Priority Programme for Rare Disease Research. This call for funds was requested in collaboration with other teams, including Antonio Rausell’s (Clinical Bioinformatics Laboratory, Imagine Institute, <https://www.institutimagine.org/en/RausellLab>), to develop new approaches to address these diagnostic impasses.

Objectives and missions

The position is destined to a candidate motivated for a research activity. He/She will be responsible for conducting the innovative bioinformatics analyses, based on the genomic data already compiled, thanks to recently developed detection tools but also through innovative approaches based on artificial intelligence (AI), in order to identify new genetic factors in coding and non-coding regions of the genome such as variations in the number of copies (CNV) or loss of function mutations (LoF). The position have a double location : Sorbonne University School of Medicine, within the UMR_S 1166 (91 boulevard de l'Hôpital, 75013 Paris) and Imagine Institute (24 Boulevard du Montparnasse, Necker Hospital, 75015 Paris).

Key Activities

- Perform quality controls on genomic and clinical data related to the different projects of the teams
- Analyze genomic data from high-throughput sequencing, DNA chips, CNV and LoF detection tools produced in both teams
- Participate in the development and implementation of AI approaches developed by the Clinical Bioinformatics Laboratory of the Imagine Institute
- Design or develop new bioinformatics tools to meet the needs of units' researchers
- Close interaction with the biologists of the units responsible for the various genomic projects carried out, in order to optimize the process of interpretation of the molecular and cellular impact of the identified variants, its coherence with the patients' clinical signs, and its relationship with the pathophysiology of the studied diseases

Knowledge

- Bioinformatic analyses and statistics of DNA chip

- Bioinformatics analysis of sequencing data (including exome/genome and messenger RNA) in clinical research
- Machine learning knowledge
- Experience in statistical data analyses
- General knowledge in molecular biology (genetics) in order to understand and analyze the scientific problem in question and to be able to interact with the researchers of the units
- English language: B2 to C1

Skills

- Master Unix/Linux programming techniques as well as Python and R programming languages
- Proficiency in DNA chip data analysis software (Plink, etc.) and GWAS or imputation type analyses
- Mastering analysis of exomes / genomes including annotation of effects of variants and prioritization using pathogenicity scores
- Mastering the use of clusters/grids for high-performance scientific computing and parallel computing
- Mastering public databases in genomics (NCBI, GenCode, Biomart, etc.)
- Ability to manage large data storage and handling
- Write synthesis reports and present them orally (English and French) during lab meeting

Behavioral skills

- Scientific and experimental rigor
- Ability to work among a team environment

Desired experience

- Beginner accepted

Training desired

- Engineering degree or thesis

Recruitment level according to Sorbonne University remuneration grid

2-year fixed-term contract with a double assignment in Team 1 of UMR_S 1166 ICAN and in the Laboratory of Clinical Bioinformatics of the Imagine Institute (UMR 1163)

Position opening date: September 2021

Contact: send CV and cover letter by email to Sophie Garnier, sophie.garnier@sorbonne-universite.fr