

Offre de stage SOPHiA GENETICS

VCF format check

Le projet européen REFBIO II vise à amplifier les partenariats et à améliorer la perméabilité et les synergies entre le milieu académique et le monde de l'entreprise dans le domaine des Biotechnologies.

Dans ce cadre, des missions de chercheurs en entreprises sur une durée de 3 mois sont proposées, le financement du doctorant étant pris en charge par REFBIO II.

SOPHiA GENETICS a été fondée en 2011 pour dessiner l'avenir de la génomique clinique. Avec l'adoption des technologies numériques, tel que le séquençage de l'ADN de la prochaine génération (NGS), l'industrie est entrée dans le monde du Big Data, soulevant de nouveaux défis en matière d'analyse et de sécurité des données. Pour résoudre ces défis, une équipe d'experts de renommée mondiale a développé l'Intelligence Artificielle (IA) SOPHiA. Comme les horlogers suisses, la technologie de SOPHiA GENETICS est reconnue pour sa qualité, sa précision et sa fiabilité. Il en résulte des certifications ISO 13485 (Système Management de la Qualité des Dispositifs Médicaux) et ISO 27001 (Système de Management de la Sécurité de l'Information).

SOPHiA GENETICS fournit une technologie avancée et universelle pour l'analyse des données NGS.

SOPHiA, technologie universelle pour la médecine axée sur les données (exclusivement disponible sur SOPHiA DDM) détecte, annoté et pré-classe les variantes génomiques pour aider les cliniciens à les utiliser afin de mieux diagnostiquer et traiter leurs patients.

SOPHiA DDM est une plateforme SaaS (software-as-a-service) qui facilite la visualisation et l'interprétation des variantes, grâce à son interface utilisateur intuitive et à ses fonctionnalités intégrées. De plus, elle assure la protection de toutes les données en les chiffrant selon les normes les plus élevées de l'industrie avant de les stocker de façon redondante dans des centres de données privés et sécurisés.

Elle permet l'analyse et la protection des données cliniques NGS. Nos technologies de base traitent et analysent les données génomiques brutes pour aider les hôpitaux à diagnostiquer avec précision et rapidité les patients dans des domaines tels que l'oncologie, le métabolisme, la pédiatrie, la cardiologie et le cancer héréditaire.

Mission du doctorant dans l'entreprise:

Objet de recherche : Projet 1: VCF format check

Le format de fichier VCF est un format standard pour décrire les différences d'un échantillon par rapport à une référence.

Ces différences peuvent être de petites variantes comme SNP et Indels mais aussi de grandes modifications comme CNVs.

Chez SOPHiA GENETICS, les fichiers VCF sont manipulés à de nombreuses étapes internes du pipeline. Le fichier VCF final est ensuite mis à la disposition du client via notre plateforme SOPHiA DDM.

Puisque plusieurs de nos clients utilisent le fichier VCF pour l'analyse en aval avec l'aide d'autres logiciels tiers, il est essentiel que le fichier VCF soit conforme à la spécification officielle du format VCF.

Par conséquent, nous aimerions implémenter un vérificateur de format VCF qui s'assure que les fichiers VCF sont conformes et qui fournira des messages d'erreur significatifs si ce n'est pas le cas.

A l'avenir, chaque script qui manipule les fichiers VCF devrait exécuter le vérificateur afin de s'assurer que la sortie produite par le script est correcte.

Programme financé par le FEDER

Résultats attendus :

Un outil pour vérifier que le format soit conforme à la spécification du format VCF et qui génère des messages d'erreur significatifs si ce n'est pas le cas. Doit être mis en place efficacement pour qu'un fichier VCF volumineux avec 100 000 variants puisse être vérifié en quelques secondes.

SOPHiA GENETICS pourra ainsi s'assurer que tous les outils développés en interne et produisant des fichiers VCF seront conformes aux spécifications du format VCF. Cela garantira également que le fichier VCF qui peut être téléchargé par le client SOPHiA GENETICS via la plateforme SOPHiA GENETICS DDM est un VCF valide & peut être utilisé dans des logiciels tiers pour d'autres analyses si nécessaire.

Profil et compétences recherchées des doctorants :

Nous recherchons un étudiant avec des compétences en programmation idéalement avec une expérience en Bio-IT, qui soient à l'aise en programmation R, Python ou C++. Le candidat devra savoir gérer un projet depuis le début (recueil des besoins utilisateurs) jusqu'à la finalisation (implémentation qui correspond aux besoins). Il devra également savoir communiquer et documenter son projet correctement de manière à en permettre la maintenance après son départ. La connaissance des technologies d'analyses de données NGS est un atout mais pas un requis pour le stage.

Domaine d'études :

Informatique, bioinformatique ou Biotechnologie

Domaine de recherche :

Développement informatique, analyse de données, data visualisation ; expérience avec des données biologiques serait un plus mais n'est pas obligatoire.

Compétences souhaitées :

Motivé, communication facile, bien organisé, indépendant.

Expérience professionnelle :

Pas de niveau d'expérience requis.

Langues souhaitées :

Anglais obligatoire, langue principalement utilisée chez SOPHiA GENETICS.

Matériel et processus impliqués : Ordinateur fourni.

L'ordinateur sera équipé de toutes les applications et de tous les logiciels nécessaires

Début d'activité : DÈS QUE POSSIBLE

Durée du stage : 3 mois

Lieu du stage :

SOPHiA GENETICS
Technopole IZARBEL
64210 BIDART – France

Contact:

Les candidatures sont à transmettre par mail avant le **31/10/2018** avec CV et le formulaire de candidature complété au Cluster Santé Osasuna

E-mail : clusterosasuna@gmail.com

Téléphone : 05 59 43 54 51

Programme financé par le FEDER