

Stage M1 Nanopore transcriptomics pipeline in HBV

Stage · Stage M1 · 4 mois · Bac+5 / Master · INSERM U1052 · Lyon (France)

Date de prise de poste : 1 Avril 2023

Mots-Clés

RNA-seq, Nanopore, long-reads, Hepatitis B virus, Nextflow, nf-core, splicing

Description

Titre : Optimisation, adaptation et déploiement d'un workflow d'analyse du transcriptome de HBV

Encadrant : Xavier Grand (xavier.grand@ens-lyon.fr)

Objectif :

L'objectif du stage est de poursuivre l'implémentation d'un workflow d'analyse du transcriptome du Virus de l'Hépatite B. Les étapes du workflow sont identifiées et testées. Les outils sont containerisés (docker, singularity et charliecloud). Une première trame en Nextflow a été produite. L'objectif du stage est une mise à jour et une adaptation au standard nf-core du workflow pour être diffusée à la communauté scientifique HBV. Nous cherchons un candidat ayant une formation initiale en informatique/programmation se spécialisant en bioinformatique.

Résumé :

Le virus de l'Hépatite B est la principale cause de maladies hépatiques aiguës et chroniques. Il est un des responsables majeurs du carcinome hépatocellulaire, deuxième mortalité par cancer dans le monde. En 2016, on recense 240 millions de personnes infectées et 650000 décès, dus à une insuffisance hépatique liée à cette maladie, malgré la disponibilité d'un vaccin.

Le VHB possède un génome circulaire de 3,2 kb de long, portant 4 ORF différents sur le brin négatif partageant un unique site de terminaison. Il code 8 ARN majeurs et 22 variants d'épissage connus à ce jour. Les ARN les plus longs, pg et preC, sont plus longs que le génome lui-même. Par conséquent, il existe une redondance de séquence aux extrémités 5' et 3' de ces ARN. De plus, les ARN les plus petits sont, dans de diverses proportions, des sous-ensembles les uns des autres en termes de séquence.

Ainsi, les technologies conventionnelles de séquençage short-reads ne permettent ni l'identification ni la quantification des ARN du VHB, en raison de la structure du génome. Les technologies récentes de séquençage long-reads, comme Nanopore, ont permis de capturer et séquencer les ARN complets du VHB. C'est dans ce contexte qu'il est aujourd'hui nécessaire développer un workflow d'analyse bioinformatique de ces données de séquençage.

Ce stage requiert de l'aisance en programmation et en informatique, la partie biologie sera apportée par l'encadrant. Le candidat sera formé au développement de pipelines sous Nextflow et leur utilisation sur un ordinateur haute performance (CPU et GPU).