



Université Claude Bernard



Lyon 1

DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **22 octobre 2021**

Nom de famille et prénom de l'auteur : **Madame SESSEGOLO Camille**

Titre de la thèse : « *Développement de méthodes bio-informatiques pour l'étude de l'épissage chez les espèces non modèles : Épissage complexe et apport des technologies de séquençage de 3ème génération.* »

Résumé



Les gènes des organismes eucaryotes sont structurés en exons et en introns. Lors de l'épissage, les introns sont retirés et les exons reliés entre eux. L'utilisation des sites d'épissage par la machinerie cellulaire peut varier d'un transcrit à l'autre pour un même gène. L'épissage alternatif permet alors à un seul gène de produire plusieurs transcrits et parfois plusieurs protéines. L'étude des données issues du séquençage des transcrits (RNAseq) nous permet d'étudier l'épissage. Actuellement deux technologies de séquençage coexistent : les technologies de seconde génération, permettant de produire des lectures courtes (100 à 250pb) avec un taux d'erreur faible et les technologies de 3ème génération permettant de produire des lectures longues (plusieurs kb) avec des taux d'erreur plus élevés. Dans un premier temps, j'ai analysé des jeux de données Nanopore (lectures longues) afin de comprendre comment ces technologies, récentes et en constante évolution, peuvent nous aider à étudier les transcriptomes eucaryotes. Plus particulièrement, je me suis demandée si les quantifications des gènes et des transcrits obtenues étaient fiables. L'utilisation de spikein -transcrits artificiels dont on connaît la quantification- nous a permis de montrer que, parmi les différents protocoles testés, les quantifications obtenues avec le protocole RNA direct sont les plus fiables. De plus, contrairement à ce à quoi l'on s'attendait, les lectures ne couvrent pas systématiquement des transcrits complets. Ensuite, je me suis intéressée à la modélisation des événements d'épissage alternatif complexes chez les espèces non modèles. L'assembleur local de transcriptome, KisSplice, développé dans l'équipe, compare deux à deux les transcrits, même lorsqu'il y a plus de deux transcrits localement. Je propose ici une nouvelle échelle d'étude de l'épissage qui permet de considérer toutes les variations d'épissages observées entre deux exons constitutifs à tous les transcrits d'un gène.