

Données RNA-seq semi-synthétiques : pourquoi normaliser alors qu'on peut générer des faux positifs ?

vendredi 17 novembre 2023 10:55 (25 minutes)

En revisitant les analyses de données RNA-seq semi-synthétiques simulées par Li et al. (*Genome Biology* 2022), nous soulignons l'importance de la prise en compte de la taille de librairie et de l'utilisation d'une normalisation adéquate avant d'envisager toute manipulation des observations supposant leur échangeabilité. Après avoir proposé une nouvelle stratégie de simulation corrigée, nous montrons que seules certaines méthodes d'analyse différentielle souffrent réellement d'une inflation exagérée du taux de faux positifs dans les études de grandes tailles avec une forte variabilité biologique (telles que les études chez l'humain) —en particulier edgeR et DESeq2. De plus, nous montrons que le comportement du test des rangs de Wilcoxon n'est en rien supérieur à d'autres tests spécialisés pour l'analyse différentielle de données RNA-seq tels que dearseq. Ces derniers sont de surcroît souvent plus versatiles et peuvent prendre en compte des designs expérimentaux relativement complexes, tandis que le test des rangs ne peut s'appliquer que pour la comparaison non-ajustée entre deux groupes.

Auteurs principaux: HEJBLUM, Boris (Inserm U1219); BA, Kalidou; THIÉBAUT, Rodolphe; AGNIEL, Denis

Orateur: HEJBLUM, Boris (Inserm U1219)

Classification de Session: Apprentissage statistique et grande dimension