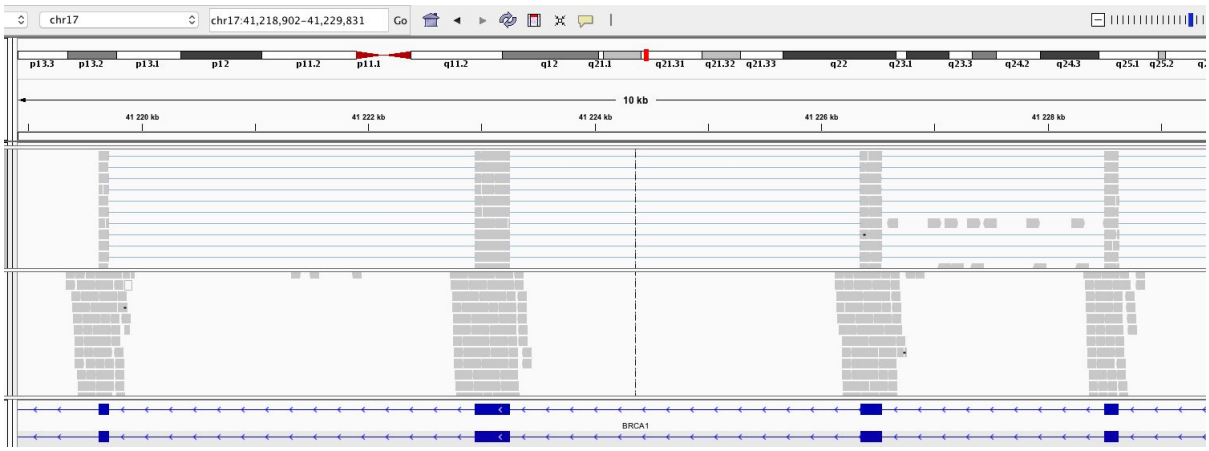
Examen écrit

Durée : 1 h 30

<u>Consignes</u> : Les documents et ordinateurs sont autorisés. Internet non autorisé. Les questions sont indépendantes et seront corrigées par chaque enseignant. Chaque question compte pour ⅓ de la note finale.

● Question 1 (D. Gautheret)

DG1: Dans l'image ci-dessous, deux fichiers BAM sont visualisés, alignés sur un fragment de gène humain (visible en bas avec 4 exons). D'après la structure des *reads* du BAM, que pouvez-vous dire de la nature des banques NGS les ayant produits? Justifiez votre réponse.



DG2: Soit un *dataframe* R possédant les colonnes "Gene" et "Somatic.P.value", comment afficher les gènes ayant une valeur "Somatic.P.value" inférieure à 10e-4?

● Question 2 (M. Gallopin)

- Pourquoi est-il préférable de supprimer les gènes peu exprimés avant d'effectuer l'analyse différentielle sur l'ensemble des gènes?

- Quelle est l'utilité d'une analyse en composante principale sur les échantillons biologiques avant d'effectuer l'analyse différentielle?

- **Question 3 (G. Lelandais)**

Pourquoi la statistique utilisée dans la méthode LIMMA est-elle plus performante que celle utilisée dans un test de Student ?

N'oubliez pas de rendre votre liste de gènes, avec l'explication et la justification des choix que vous avez réalisé.