

# 農学生命情報科学特論 I(2014 年 6 月 25 日)課題の回答とコメント

受講 ID(5 桁): \_\_\_\_\_

学生証番号(ない人は空欄で可): \_\_\_\_\_

名前: \_\_\_\_\_

課題:アダプター配列除去前後の small RNA-seq データをカイコゲノムにマップし、マップ率を比較する

Q1: マッピング前の総リード数を述べよ

- ・アダプター配列除去前の SRR609266.fastq.gz: 11,928,428 リード
- ・アダプター配列除去後の hoge4.fastq.gz: 11,928,428 リード

502,962,917 と書いてあるヒトが何人かいましたが、これはリファレンス配列の総塩基数です。スライド 53 の対応する箇所(531)と比較すればわかります。ref\_genome.fa は chr1 が 48bp, chr2 が 160bp, chr3 が 100bp, chr4 が 123bp, chr5 が 100bp ですので、 $48+160+100+123+100=531$  という意味です。講義中も何度か述べているように、「疑問に思ったら全貌を完全に把握しているデータで答え合わせをする」です。

Q2: マッピング後の「マップされたリード数」を述べよ

- ・アダプター配列除去前の SRR609266.fastq.gz: 2,257 リード
- ・アダプター配列除去後の hoge4.fastq.gz: 1,308,126 リード

Q3: 結果の考察。

アダプター配列除去は一部を取り除くだけなのでそうリード数は変わらない。アダプター配列を除去するとマップされたリード数が増える。みたいなことが書かれていれば OK です。

課題とは関係ありませんが、k-mer 解析の際に R のバージョン(ver. 3.0.3 と ver. 3.1.0)が違っていると qrac パッケージの k-mer 出現頻度結果が変わるようですね。情報ありがとうございます。

自由記載欄