

農学生命情報科学特論 I (NGS ハンズオン講習会第 1 部) 課題

1. ゲノム解析系 (2016 年 7 月 20 日分)

長さ 60 bp、20,000 リードからなる仮想 NGS データ ([kadai_20160720.fasta](#)) から、このゲノムサイズを推定せよ。計算手順や思考回路 (考察) も示せ。

入力ファイルの場所 : http://www.iu.a.u-tokyo.ac.jp/~kadota/kadai_20160720.fasta

考察としては、k-mer の k 値の違いや、どのあたりをシークエンスエラー由来 k-mer とみなしたかなどを想定している。基本的には、講義資料のスライド 128 以降を参考にすればよい。提出物としては、以下のものを想定している。解析環境・解析手順からは結果の再現性を、そして結果と考察からは講義内容の理解度を評価する。(第 2 部の 8/3 参加者は、KmerGenie を実行し、その結果と合わせて考察してもよい。)

- ①解析に用いた R 本体のバージョン :
- ②解析に用いた Biostrings パッケージのバージョン :
- ③解析手順のコード (R コードなど)
- ④k-mer 出現頻度分布 (png ファイル) :
- ⑤ゲノムサイズの推定値 :
- ⑥考察 :

2. トランスクリプトーム解析系 (2016 年 7 月 21-22 日分)

講義で実行しなかった反復あり or なしのデータ を用いて 2 群間比較または 3 群間比較を行い、**結果を考察せよ。**

講義では、ReCount のデータセット (bodymap, gilad, and maqc) や Blekman データセット、あるいはその一部 (サブセット) を用いて 2 群間比較 (反復あり・なし) および 3 群間比較 (反復あり・なし) を行った。例えば、gilad データセットから 4 サンプル分取り出して、(F_SRX014818and9, F_SRX014820and1) vs. (M_SRX014824and5, SRX014828and9) の反復あり 2 群間比較を行えば、(Female の 2 サンプルがクラスターを形成するので) 少しは DEG が得られるのでは?! という思想のもとで解析を行ってみるなどを想定しています。もちろん ReCount データベースの他のデータセットを解析してみるのもよい。課題 1 同様、解析データ、手順、結果と考察を示せ。

課題提出先 : report@iu.a.u-tokyo.ac.jp および kadota@iu.a.u-tokyo.ac.jp (全角の @ を半角の @ に変えてください)

タイトル : 特論 I 課題

提出期限 : 2016 年 8 月 20 日

- ・メール本文に、**所属・氏名を明記**してください。
- ・**課題の添付を忘れない**ようにしてください。
- ・「NGS ハンズオン講習会の受付番号」、「アグリバイオの受講 ID (5 桁)」、東大生は学生証番号などでもできるだけ示してください。
- ・24 working hours 以内に受領メールをお送りします。返事が届かない場合は、'NGS ハンズオン講習会事務局' <NGS@biosciencedbc.jp> 宛てに問い合わせてください。