

農学生命情報科学特論 I(2014 年 7 月 2 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:シミュレーションデータ(data_hypodata_3vs3.txt)のサンプル間比較解析を行う。

Q1:TCC パッケージを用いた発現変動遺伝子(DEG)検出を行い、FDR 閾値が 0.05、0.20、および 0.40 を満たす遺伝子数を示せ。また、このデータセット中の大まかな DEG 数を示すとともにその根拠を簡単に述べよ。

・FDR 閾値 0.05 を満たす遺伝子数(q -value < 0.05): 1,139 個
1,140 個でもバージョンの違いでこれくらいは起こりうるので OK です。

・FDR 閾値 0.20 を満たす遺伝子数(q -value < 0.20): 1,455 個
1,271 個と書いてあるヒトが何人かいましたが、これは FDR 閾値が 0.1 の結果です。

・FDR 閾値 0.40 を満たす遺伝子数(q -value < 0.40): 1,765 個
1,616 個と書いてあるヒトが何人かいましたが、これは FDR 閾値が 0.3 の結果です。1,766 個でもバージョンの違いでこれくらいは起こりうるので OK です。

・このデータセット中に含まれる推定 DEG 数(偽物を差し引いた本物の DEG 数):
1,131 個でも 1,059 個でも、1,164 個でも、以下の根拠や計算手順が間違っていなければよしとしています。約 1,100 個などでも OK。

・根拠(計算手順と簡単な説明でよい):
1,616 個のうち 70%が本物なので $1,616 \times 0.7 = 1,131.2$ 個程度、みたいなことが書かれていれば OK です。もちろん、1,765 個のうち 60%が本物なので $1,765 \times 0.6 = 1,059$ 個や、1,455 個のうち 80%が本物なので $1,455 \times 0.8 = 1,164$ 個みたいなことが書かれていても OK です。ただし、FDR0.3 の結果である 1,616 個の結果をもとに、 $1,616 \times 0.6 = 969$ 個などというのはやってはいけないミスですので減点です。

Q2:DEG 検出結果の考察。シミュレーションデータ(data_hypodata_3vs3.txt)のサンプル間クラスタリング結果との比較や、実データ(srp017142_count_bowtie.txt)解析結果との比較など自由に述べてよい。

G1 群で高発現の DEG が多い、2,000 個という意図的に作成された DEG 数よりも実際に検出された DEG 数は少ない、あくまで目安程度、などが書かれていれば OK です。基本的には、明らかにおかしい考察をしていなければ OK です。