

機能ゲノム学 課題の解答

解答部分は赤字で示してあります。

— 2018年5月8日実施分 —

1.

	chromosome	plasmid1	plasmid2	③Ensembl
①Gene transcripts (transcript_nrow)	2,262	99	51	2,412
②Coding genes (cds_nrow)	2,194	99	51	2,344

2.

転写物数：2262、最大値：11196、最小値：71、平均値：888.x (xはなんでもよい)

感想やコメントについて

- ・最大値と最小値の差が大きすぎるとか、最大値が大きすぎるとか様々な感想がありました。最大値をもつ転写物を調べてみるといいでしょう。
- ・shRNAやCRISPRなどで用いるデータ解析については、残念ですが私にはできません。
- ・「礼賛(らいさん)」が正解ですね。ご指摘ありがとうございますm(_ _)m

— 2018年5月15日実施分 —

論点1~4

①4,589,774リードの結果ではPer base sequence contentでAの含有率が高く、実際にOverrepresented sequencesが一種(AAAAAAAAAAAAAAAAAA)見られた。この二つの結果は密接に関連しているので妥当、的なことを想定していましたが、それ以外にも適切に考察ができていればOKです。

感想やコメントについて

- ・FastQCの再起動の方法はダウンロードしたbatファイルを起動するやり方がよいのかどうかはわかりません。
- ・「各ツールの用途の違い(Galaxy)をもっと知りたい」についてですが、このあたりはほぼ趣味の違いだと思います。

— 2018年5月22日実施分 —

FaQCsがどのような処理を行うプログラムなのかを正しく理解できているかを問うています。Poly Aが正しくトリミングされた、配列長がオプションで与えた長さ以上に正しくなっている、クオリティスコア分布がFaQCs実行後に上がっている、などが書かれていればOK。

感想やコメントについて

- ・MinIONやPacBioは「トランスクリプトーム配列解析」で、Illuminaは「トランスクリプトーム発現解析」の分野で今後も利用されるのではと思われます。
- ・ショートリードとロングリードはデータの質が異なりますので(個人の感想です)、どのような違いがあ

るかというよりは（利用するプログラムも含めて）別物だと思うほうがよろしいかと思います。

―― 2018年5月29日実施分 ――

課題1

リード ID	マップされた領域	
	リファレンス配列名（例：chr1）	左端の位置（例：8番目）
chr3_3_37	chr5	3番目
chr3_3_37	chr3	3番目
chr3_1_35	chr5	1番目
chr3_1_35	chr3	1番目

課題2

Alignment score が高いほうを採用するのが妥当だから、ミスマッチが少ないほうを採用するのが妥当だから、的なことが書かれていればOK。信頼度が高い、という記述でもよい。

感想やコメントについて

・bowtie2の結果のSAMファイルに複数マップされた領域の情報を全て表示させる方法は、、、すみません私もわかりません。

―― アンケートの記述内容について ――

・BioLinuxをVMware player上でインストールした場合Bowtieは使えますか？ yes