

## ゲノム情報解析基礎（2014年4月23日）課題の回答とコメント

課題 1: シロイヌナズナの GFF3 形式ファイル(TAIR10\_GFF3\_genes.gff)を入力として下記 1~4 の解析結果を示せ。また、3 と 4 の結果の違いについて論ぜよ。

1. 染色体数(3 列目が"chromosome"となっている行数)

7 です。

2. mRNA の個数(3 列目が"mRNA"となっている行数)

35386 です。

3. 1 列目が"Chr1"および"ChrM"となっているトータルの行数

158,435 です。内訳は、Chr1 が 157,712、そして ChrM が 723 です。

4. 1 列目が"chromosome 1"および"ChrM"となっているトータルの行数

0 です。723 と間違っただけで回答しているヒトが結構いました。講義中にも述べましたが、ChrM のあとにスペースが入っています。このスペースをちゃんと反映させていないと ChrM 分の行数がカウントされてしまいます。

5. 3 と 4 の結果の違いについての考察

「プログラム上は文字列の違いを厳密に区別しているため」や「4.では、存在しない文字列を検索していたため」、みたいなことが書かれていればいいです。4. で ChrM のスペースを含めずに 723 個と間違っていたヒトでも、文字列として厳密に評価しているという類のことがわかっていると判断した場合には、ここは OK にしてあります。

課題 2: シロイヌナズナのゲノム配列ファイル(TAIR10\_chr\_all.fas)を入力として GC 含量解析

1. 染色体ごとの%GC 含量計算結果を示せ(小数点 2 ケタ程度でよい)

chr1: 35.87 (%)

chr2: 35.86 (%)

chr3: 36.33 (%)

chr4: 36.20 (%)

chr5: 35.94 (%)

ChrM (=44.77%)と ChrC (=36.29%)については評価の対象にはしていません。

2. 得られた結果について簡単に考察せよ。シロイヌナズナゲノム配列決定に関する原著論文中の GC 含量の数値との比較などでよい。

常識的なことがかかれていれば OK にしています。例えば、「GC 含量は全体的に大きくなっている」とか「染色体によっては 1%以上違う」みたいな考察ができていれば OK です。

課題 3: シロイヌナズナのゲノム配列ファイル(TAIR10\_chr\_all.fas)を入力として 2 連続塩基の出現頻度解析を行い、得られた結果を簡単に考察せよ。

常識的なことがかかれていれば OK にしています。例えば、「TAIR9 の結果と似ているとか同じ」とか「Chr1~5 において AA, AT, TT の出現頻度が他と比べて高い」とか「CpG の出現頻度は低いもののヒトゲノムほど極端ではない」みたいな考察ができていれば OK です。

感想やコメントなど (optional) :

講義スピードについての受け止め方は様々なようですね。しかし全般的にはちょうどよい、ついていくのは大変だったが...的な感想が多かったです。また、基礎系科目でもあるので、講義スピードはあんなものです。