

生体情報工学演習6

北海道大学情報エレクトロニクス学科

生体情報コース3年

学生番号: 02180144 茂木貴紀

作成日: 2020/11/14

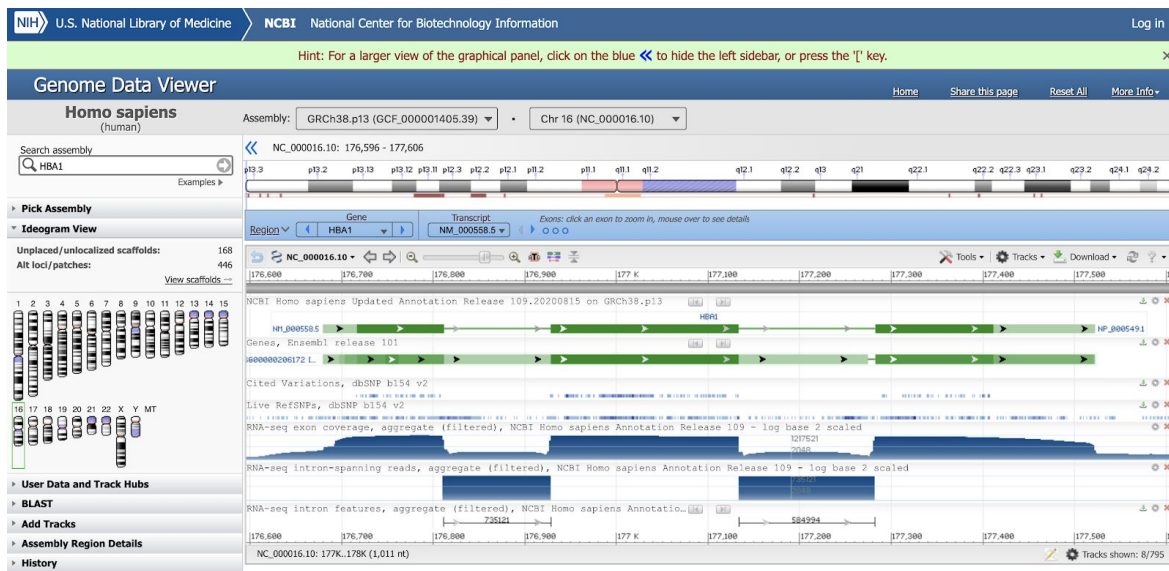
このレポートは、講義「生体情報工学演習Ⅱ」のテーマ6の課題レポートである。指定された課題に準じてレポートを作成した。

1

a)

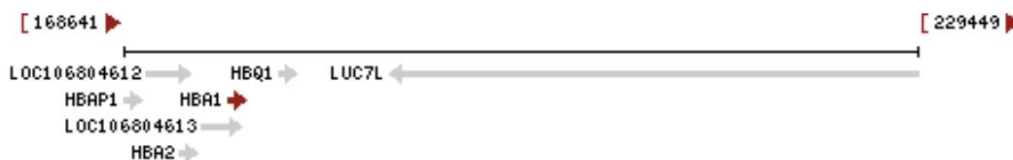
ゲノムアセンブリとは、解読されたDNAの断片配列をコンピュータ上で組み合わせる（マッピングする）することである。その際に、どの配列を基準として参照するか、というのがゲノムアセンブリのバージョンに対応する。

b)



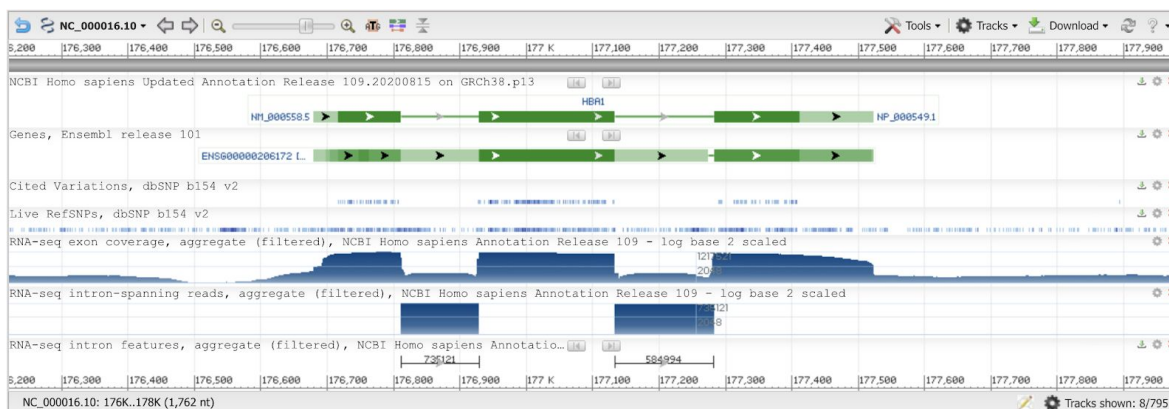
HBA1遺伝子は、16番染色体の短腕の先端(16p13.3)に位置する。

c)



転写の向きは、短腕から長腕の方向である。そもそも、転写は2本鎖の一方がアンチセンス鎖となり、もう一方のセンス鎖が複製される（5末端から3末端へと伸長していくため）。この時、どこにRNAポリメラーゼが結合し、どちらの鎖をセンス鎖とするかはプロモーターと呼ばれるものである。つまり、プロモーターによってどちらに転写するかが決められている。

d)



エキソンとイントロンがおおよそ半分ずつの割合で存在していることから、スプライシングされると半分くらいの塩基長になると考えられる。

e)



上図のように、HBA1遺伝子の周りにHBA2遺伝子とHBQ1遺伝子が存在することが確認できる。これら3つはほぼ同じ塩基長で、かつエキソン・イントロンも似たような構造をとっている。（ただ、下から2番目のグラフより、HBQ1遺伝子は周りの遺伝子として作用していない塩基配列とほぼ同じような翻訳をされており、あまり遺伝子としての機能が期待できていないことも確認される。）

ここから、これらの遺伝子は全て同じようなタンパク質を発現し、同じような機能を持っていると考えることができる。特に、HBA1遺伝子とHBA2遺伝子はほぼ同じような構造をとり、発見されたタイミングや発見した人物など、遺伝子の機能とは全く異なる理由で名前が異なっているとも考えることができる。

f)

塩基配列を取得し、作成した以下のコードを実行してみる。実行結果は以下のようになった。

https://github.com/manattan/_c/blob/main/No6/countLen.py

>同じ塩基は 650 つ, 違う塩基は 185 つ, 塩基長の差は 8 つ.

ここから、少なくとも塩基配列はよく似ていることがわかる。そして、それぞれの遺伝子についてのWEBサイトを参照すると¹²³⁴、以下のようなことがわかった。

- ncbiにて、要約には同じ内容が記載されている
- これらの遺伝子は5末端非翻訳領域とイントロンで多少異なるほか、3末端非翻訳領域では大幅に異なる。コード配列（エキソン）は全く同一である。ここから、Pythonのプログラムで明らかになった異なる塩基は、これらのものであることがわかる。
- これらが発現するタンパク質は、それぞれHbA1とHbA2であり、これらはヘモグロビンのαサブユニットを構成するタンパク質である。コード配列が同じであることから、タンパク質は同じものである。
- > 発見されたタイミングや発見した人物など、遺伝子の機能とは全く異なる理由で名前が異なっているとも考えることができる

これに関しては、「翻訳領域は同等であったが、非翻訳領域は一部異なり、機能としては同じであるものの遺伝子として考えると同じではない」というのが結論である。

g)

前提として、Refseq塩基配列にポリA領域は存在しない（スプライシングを受けてから付加されるものであるからである）。実際、塩基配列を見てもAが羅列している箇所は存在していない。



この図から、5末端非翻訳領域は一番左のイントロンの部分（40塩基ぐらい）、3末端は一番右のイントロンの部分（100塩基ぐらい）。コード領域は絵既存の部分である。強い断定を持たせるために、実際に

¹ <https://www.ncbi.nlm.nih.gov/gene/3039>

² <https://www.ncbi.nlm.nih.gov/gene/3040>

³ https://en.wikipedia.org/wiki/Hemoglobin,_alpha_1

⁴ https://en.wikipedia.org/wiki/Hemoglobin,_alpha_2
