BI 188 Problem Set 1.

Answers

Turn in by 3pm April 8 for up to 3 points.

Question 1.

Globins are a family of proteins that reversibly bind oxygen. We focus on hemoglobins in these questions and self tutorial. Go to the UCSC genome browser for human assembly hg19 and search for "beta globin." Find the correct gene and view it in the genome browser.

1. What is the abbreviation that UCSC gives you for beta globin?

   HBB

2. How many exons does beta globin have?

   three

3. Approximately how long is beta globin from transcriptional start to transcriptional stop (answer in bp = base pairs or kb = kilo base pairs)?

   1500bp/1.5kb

4. What is the chromosomal location (i.e. chromosome:start bp-stop bp) of beta globin from transcriptional start to transcriptional stop? You can approximate this from the browser AND you can find the exact location from the gene information.

   chr11:5,246,696-5,248,301

5. Zoom out from beta globin. What 10 annotated genes are nearest?

   OR51V1, HBD, HBBP1, HBG1, HBG2, HBE1

6. Why would the tandem duplication events be preserved in evolution?

   To answer fully, address separately the protein coding function and the remainder of each gene.

7. Where is alpha globin (either version) in hg19 (chromosomal coordinates)?

   HBA1 chr16:226679-227520; HBA2 chr16:222846-223708

To extract the sequence of DNA for your region of choice in the UCSC genome browser, change the browser view to your region of interest. Next, click the "DNA" link at the very top of the page. This will take you to a page where you can modify the included region and the format of the result. The suggested format is for the sequence to be in upper case with repeats masked to lower case. When you select "get DNA," you will be taken to a page where the genome and regions are shown, followed by the sequence (50bp per line). Note that there is an option to get the reverse complement of the DNA sequence. Example output:

```
>hg19_dna range=chr11:5247294-5247429 5'pad=0 3'pad=0 strand=+ repeatMasking=none

AGCATTTTTTAAAATTACAAATGCAAAATTACCCTGATTTGGTCAATATG

TGTACACATATTAAAACATTACACTTTAACCCATAAATATGTATAATGAT

TATGTATCAATTAAAAATAAAAGAAAATAAAGTAGG
```

8. Sequence of beta globin. Chapter 2 of your book is relevant for this if you want review. What is the DNA sequence of beta globin's exons? Label each exon with its name (i.e. 'exon1', 'exon2') going in the direction of transcription. Include the untranslated parts. *Helpful hint: using your mouse at the very top of the genome browser picture, you can click and drag around an area to zoom in on those exact boundaries.*

    Answer: note that this is the + strand of DNA, but the gene goes right to left. Also note that the boundaries might be off by a few bp because I clicked and dragged.

Exon1:

```
>hg19_dna range=chr11:5248158-5248300 5'pad=0 3'pad=0 strand=+
repeatMasking=none

ACCTGCCCAGGGCCTCACCACCAACTTCATCCACGTTCACCTTGCCCCAC

AGGGCAGTAACGGCAGACTTCTCCTCAGGAGTCAGATGCACCATGGTGTC

TGTTTGAGGTTGCTAGTGAACACAGTTGTGTCAGAAGCAAATG
```

Exon2:

```
>hg19_dna range=chr11:5247807-5248031 5'pad=0 3'pad=0 strand=+
repeatMasking=none

CCTGAAGTTCTCAGGATCCACGTGCAGCTTGTCACAGTGCAGCTCACTCA

GTGTGGCAAAGGTGCCCTTGAGGTTGTCCAGGTGAGCCAGGCCATCACTA
```

AAGGCACCGAGCACTTTCTTGCCATGAGCCTTCACCTTAGGGTTGCCCAT

AACAGCATCAGGAGTGGACAGATCCCCAAAGGACTCAAAGAACCTCTGGG

TCCAAGGGTAGACCACCAGCAGCCT

Exon3:

>hg19_dna range=chr11:5246699-5246959 5'pad=0 3'pad=0 strand=+ repeatMasking=none

ATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTT

CATAATATCCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTT

AATAGAAATTGGACAGCAAGAAAGCGAGCTTAGTGATACTTGTGGGCCAG

GGCATTAGCCACACCAGCCACCACTTTCTGATAGGCAGCCTGCACTGGTG

GGGTGAATTCTTTGCCAAAGTGATGGGCCAGCACACAGACCAGCACGTTG

CCCAGGAGCTG

The UCSC genome browser also allows you to search for similar sequences of DNA. Once you have a DNA sequence, go to the "BLAT" link at the very top of the page. Next, paste your sequence into the white box (the fasta format like the one that is returned from the "get DNA" exercise above is okay) and click "submit." This will take you to a page where all of the locations in the current genome with similar sequence are listed, along with the percent similarity, span (amount of the original sequence included) and other statistics.

i)BLAT the sequence of the second exon (the most conserved one) of beta globin against hg19. How many other genes or pseudogenes have a similar sequence (with at least 60% of the sequence included)?

4 loci; HBB, HBG1/2, HBG1 (different part), and HBBP1

j) Now BLAT this same sequence in the mouse mm9 genome. How many other loci have a similar sequence (with at least 60% of each sequence included), and what genes are they affiliated with?

2 loci; 2 different places in HBB-b1

Question 2

Question 2 Answers

2.1

	chr	left position	right position	Structural variant	Effect on b-globins	Explanation
1	chr11	5301836	5302198	deletion	decreased expression of β-globin genes	This region is probably a transcriptional enhancer for the β-globin genes as it is conserved, is not an exon and its deletion results in decrease of β-globins expression
2	chr11	5305485	5306559	deletion	decreased expression of β-globin genes	This region is probably a transcriptional enhancer for the β-globin genes as it is conserved, is not an exon and its deletion results in decrease of β-globins expression
3	chr11	5293482	5309901	deletion	no expression of β-globin genes	This region contains both enhancers affected by deletions **1** and **2**. As its effect is more severe than that of **1** and **2** on their own, **1** and **2** probably work together or are redundant in the activation of β-globin expression

2.2

	chr	left position	right position	Structural variant	Effect on b-globins	Explanation
4	chr11	5262712	5265196	deletion	none	This deletion removes HBBP1, however HBBP1 is a pseudogene so its deletion is not expected to have phenotypic consequences
5	chr11	5245576	5257624	deletion	severe beta thalassemia	Both HBB and HBD are deleted, no β-globin produced in adults
6	chr11	5253210	5256700	deletion	none	HBD is deleted, however as HBD is usually only a small fraction of the total hemoglobin in blood, the effect is likely to be minor
7	chr11	5247781	5247838	deletion	Thalassemia	This deletion disrupts the splicing of the gene, causes a frameshift and deletes 10 amino acids, the protein is unlikely to be functional. Even though HBD is still present, some thalassemia phenotype is likely
8	chr11	5248095	5248130	deletion	none	This deletion is in the HBB intron and there are no conserved elements in it so one would postulate it will have no effect

9	chr11	5246957	5246962	deletion	hemoglobinopathy	This deletion occurs right at a splice site in HBB and it will likely result in aberrant splicing and a non-functional protein
10	chr11	5248305	5254017	inversion	thalassemia	Such an inversion would take the immediate upstream promoter region of HBB and place it 6kb away. This would probably result in disrupted HBB expression and thalassemia