BI 188 Problem Set 1.

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

Question 1

1.1 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.

a) **What is the abbreviation that UCSC gives you for beta globin?**
b) **How many exons does beta globin have?**
c) **Approximately how long is beta globin from transcriptional start to transcriptional stop (answer in bp = base pairs or kb = kilo base pairs)?**
d) **What is the chromosomal location (i.e. chromosome: start postion - stop position) 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.**
e) **Zoom out from beta globin. What 10 annotated genes are nearest according to the UCSC Genes annotation?**
f) **Why would the tandem duplication events be preserved in evolution?**
g) **Where is alpha globin (either version) in hg19 (chromosomal coordinates)?**

1.2 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
```

a) **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.

1.3 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.

a) **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)?**
b) **Now BLAT this same sequence against 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?**

Question 2

Hemoglobin molecules are a tetramer consisting of 2 α-type chains and 2 β-type chains, for each of which several genes exist in the genome. The α-globins -ζ (HBZ), μ (HBM), α2 (HBA2), α1 (HBA1) and θ (HBQ) - are all situated in a cluster. Similarly, the β-globins are in a cluster of their own in the following order (5' to 3;): ε (HBE), γ2 (HBG2), γ1 (HBG1), δ (HBD) and β (HBB) (Note: when the term "β-globins" is used here, this refers to all these genes not just to the HBB β-globin gene). Both clusters are the result of duplications of an ancient globin gene. During embryonic and fetal development, $\zeta_2\varepsilon_2$, $\alpha_2\varepsilon_2$, $\zeta_2\gamma_2$, and $\alpha_2\gamma_2$ are used but in the adult the majority (97%) of the hemoglobin is of the $\alpha_2\beta_2$, with the rest being $\alpha_2\delta_2$. Numerous variants affecting the function or the expression of the globin genes cause diseases such as various hemoglobinopathies (caused by defective hemoglobin molecules) and thalassemias (caused by insufficient production of hemoglobin).

In this question, you will use the genome browser to examine the effects of structural variants on globin expression and function.

Go to the UCSC Genome Browser (hg19) and display the following tracks:

1. **"Base Position" from the "Mapping and Sequencing Tracks" section, in "dense" mode.**
2. **"UCSC Genes" from the "Genes and Gene Prediction Tracks" section, in "pack" mode**
3. **Set the "Conservation" supertrack in the "Comparative Genomics" section to "full" mode, and display only the "Element Conservation (phastCons)" scores from it, also in "full" mode**
4. **"RepeatMasker" in the "Variation and Repeats" section, in "full" mode**

Hint: Do not display any other tracks, if there are other tracks open, hide them.

2.1 In the table below, three structural variants are given together with their effect on b-globin expression.

	chromosome	left position	right position	Structural variant	Effect on b-globins expression
1	chr11	5301836	5302198	deletion	decreased expression of β-globin genes
2	chr11	5305485	5306559	deletion	decreased expression of β-globin genes
3	chr11	5293482	5309901	deletion	no expression of β-globin genes

Using the genome browser, can you produce a reasonable hypothesis about what the function of those regions is and why the structural variants have the effect they do?

2.2 In the table below, a number of structural variants are listed. Given what you learned in 2.1 and using the genome browser, what do you think their consequences would be for hemoglobin function? Why (i.e. what would happen on the molecular level)? You can give your predictions in terms of the potential disease (hemoglobinopathy or thalassemia) and its severity

	chromosome	left position	right position	Structural variant
4	chr11	5262712	5265196	deletion
5	chr11	5245576	5257624	deletion
6	chr11	5253210	5256700	deletion
7	chr11	5247781	5247838	deletion
8	chr11	5248095	5248130	deletion
9	chr11	5246957	5246962	deletion
10	chr11	5248305	5254017	inversion

Hint: It would be easier for you if you display the regions of the variants as a custom track on the browser.