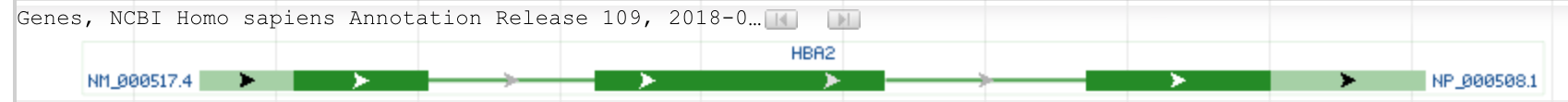
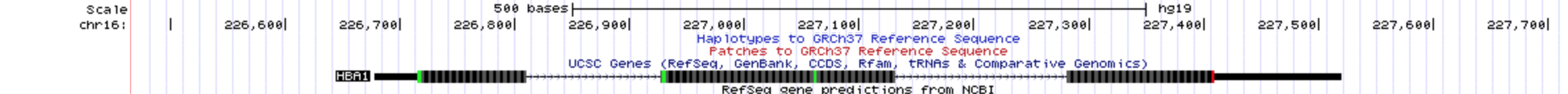
Pairwise Alignment Homework

Globin genes form a large and critically important gene family in mammals. The alpha- and beta-globin genes produce two proteins which form a heterotetramer (two copies of each protein) which binds heme (an iron containing biomolecule) and is responsible for transport of oxygen in the blood. Mutations in the alpha- and beta-globin genes cause the most common inherited diseases of man, including sickle cell disease and the thalassemias.

Interestingly, genes for alpha-globin and beta-globin are located in different genomic regions and their transcription is regulated in different ways. Both alpha and beta globin genes are located in gene clusters, which may be the result of tandem genomic duplications. We will use dotplots and Smith-Waterman alignments to compare some of these gene sequences.

<http://bioinfo.nhri.org.tw/cgi-bin/emboss/>  
<http://genome.ucsc.edu>

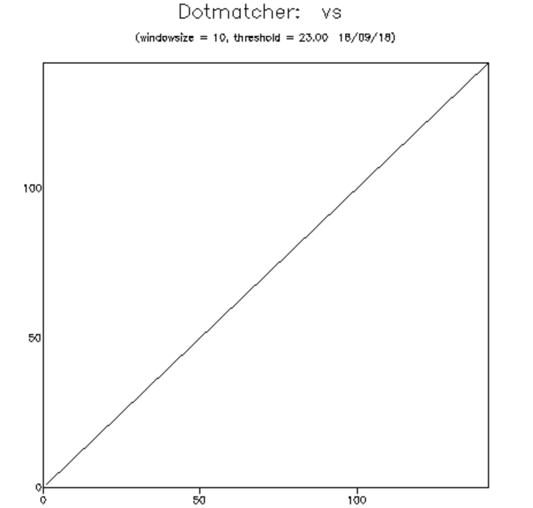
Humans have two hemoglobin alpha genes, alpha-1 (HBA1) and alpha-2 (HBA2).

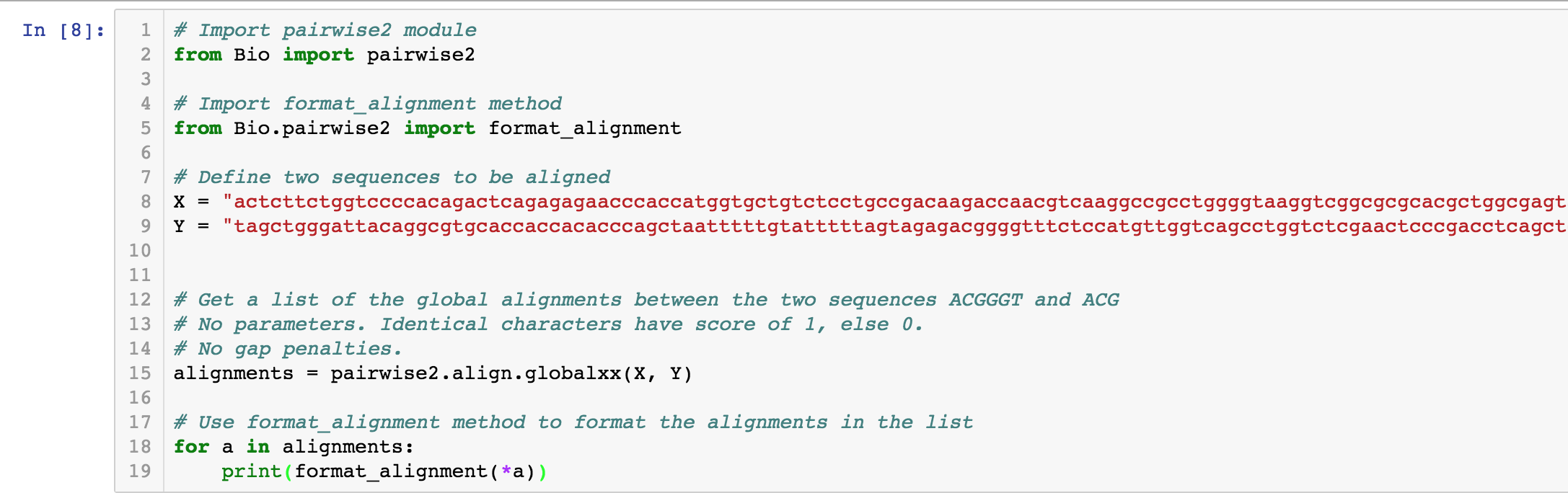


1. Where on the genome (chromosome position) are HBA1 and HBA2 located?

chr16:226,679-227,520 842 bp

chr16:222,846-223,709 864 bp.

1. Are they transcribed in the same direction?   
   yes
2. Get the coding sequences of HBA1 and HBA2 and compare on a dotplot.   
   
3. Now compare the full gene sequences (introns, 5' and 3' UTRs).
4. Use a dotplot to look for boundaries of a possible tandem duplication of DNA between HBA1 And HBA2. How large a chunk of DNA appears to be duplicated? How similar are the flanking sequences 1000 bp upstream and downstream of the two genes? You will need to adjust the window and threshold of the dotmatcher to get a nice plot.
5. How similar are the mRNA sequences? Are there many synonymous substitutions? What is the % sequence identity in the 5'UTR and in the 3' UTR.
6. Repeat the comparison of HBA1 and HBA2 plus 1Kb of 5' & 3' flanking sequence using Smith-Waterman alignment. Now can you define the exact region of genomic duplication?
7. HBZ is another member of the alpha-globin family located in the same gene cluster but with a somewhat different function and completely different expression pattern (embryonic). Extract the entire genomic sequence for HBZ plus 1Kb of 5' & 3' flanking sequence and compare to HBA1 with both a dotplot and Smith-Waterman alignment. What regions of similarity are seen?



1. How similar are the protein sequences of the two genes?
2. Compare DNA sequences of alpha globin HBA1 to the beta globin gene HBB. Try the full mRNA sequence (CDS plus UTRs), and then just the protein coding portion. Is there similarity in the 1000 base 5' sequence.
3. Now compare the protein sequences.
4. HBB is in its own tandem cluster of beta-globin genes. Compare the DNA sequences of HBB and HBD, including flanking sequences, in the same way as we did for HBA1 and HBA2.

