University of Pennsylvania
BIOL4536 Fall 2023

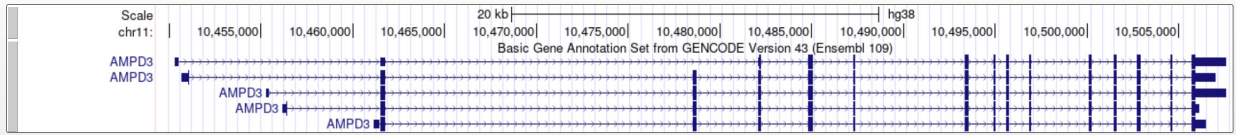
HW#3
(Genome Browser)

Assigned September 13th
Due September 20th, 3:30pm

Go to the Genome Browser Human Genome build GRCh38/hg38, turn on the "Base Position" and "GENCODE Versions" tracks and turn off all others. Surf to coordinates:

chr11:10,449,355-10,507,935

You should see the gene `AMPD3`.



(1) What DNA strand is the gene on?

(2) Is this a protein coding gene? Justify your answer.

(3) How many isoforms have an intron in the 5' UTR?

(4) How many isoforms have an intron in the 3' UTR?

(5) Does the CDS of all five isoforms end in the same place?

(6) Some amino acids in this protien are colored green. What's special about those amino acids?

Go to gene `SPTY2D1`. Turn on the Conservation track to "pacK" and configure it to show *only* "all birds" and "all fish" (turn off all other species).

(7) Which of the six exons (of the long isoform) is conserved far more between human and birds than between human and fish?

(8) Which of the six exons (of the long isoform) is least conserved across all shown species?

(9) Zooming in, give the DNA coordinates of an amino acid in this gene that is unchanged across all birds, but is highly variable in fish (with at least five different amino acids across fish at that position)

(10) Zoom way in and look at the first two and last two bases of the introns. What do you observe?

Turn off the conservation track and turn on the "dbSNP 155" track in the "Variation" section in pack mode. Click on the track itself in the graphic to toggle it ot show the SNP IDs (if they're not already showing). It will show a little hand when you're in the right position to click.

Zoom in on Exon 3 of SPTY2D1.

(11) Find the missense variant, what is its ID?

(12) What are the two possible amino acids at the location affected by this variant? You can find this on the info page for the variant.

(13) Looking at the info page for this variant, how many populations were genotyped at this location?

(14) A variant is called a SNP if each version occurs at least 1% of the population. Otherwise it's called a "mutation". So whether something is a SNP or a mutation depends on a reference population. Did it fail to be a SNP in any of the populations it was genotyped in?

Turn off the SNP track.

(15) What are the exact coordinates of intron #2 of `SPTYD1`?

(16) Make a custom track spanning exactly that intron and give it the name "Crab Leg". Show exactly what you entered in the custom track box.

(17) Download a PDF showing the custom track (zoom out to show the entire gene) and turn on only the one gene annotation track "GENCODE Versions". Include this PDF with your submission (can be included as a figure in your writeup or as a separate document).

(18) What tissue is `SPTYD1` most highly expressed in? *Note: You may have to turn off the GENCODE Versions track and turn on the GENCODE V43 track to get to this info*

(19) How many amino acids long is the protein sequence of `SPTYD1`?

(20) Get the protein sequence of and use the text search feature of the browser page to count the number of W's in the protein sequence.