

University of Pennsylvania

BIOL4536 Fall 2023

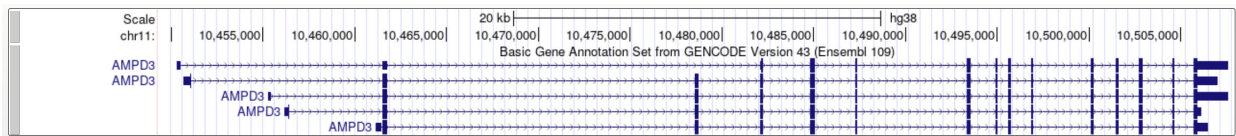
HW#3 (SOLUTIONS)

(Genome Browser)

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?

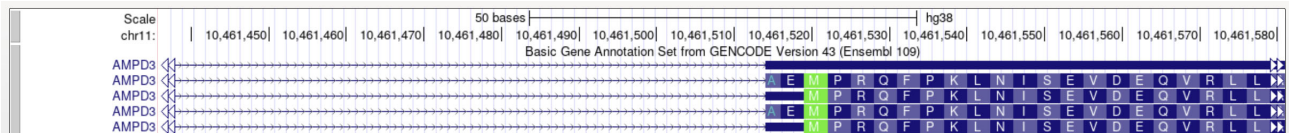
ANSWER: The forward strand.

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

ANSWER: Yes, they are indicated by the thicker parts of the exons.

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

ANSWER: Since this is on the forward strand, the 5’ UTR is on the left. It is obvious that the first isoform does and the 2nd and 4th do not have introns in the 5’ UTR. To see the other two you have to zoom way in. From this we see that isoforms 3 and 5 also do have an intron in the 5’ UTR.

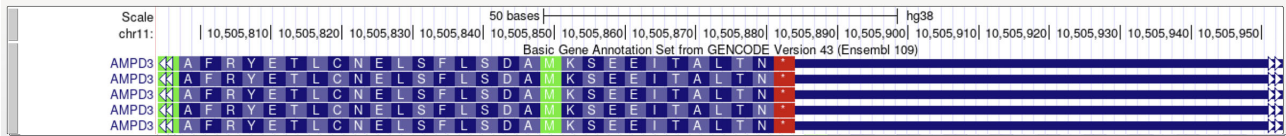


(4) How many isoforms have an intron in the 3’ UTR?

ANSWER: The 3’ UTR is on the right. Clearly none have an intron since the final exon has both thick and thin parts.

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

ANSWER: You have to zoom way in to see that they indeed all do end in the same place.



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

ANSWER: The green ones are Methionine which are colored green because that's also the Start Codon.

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?

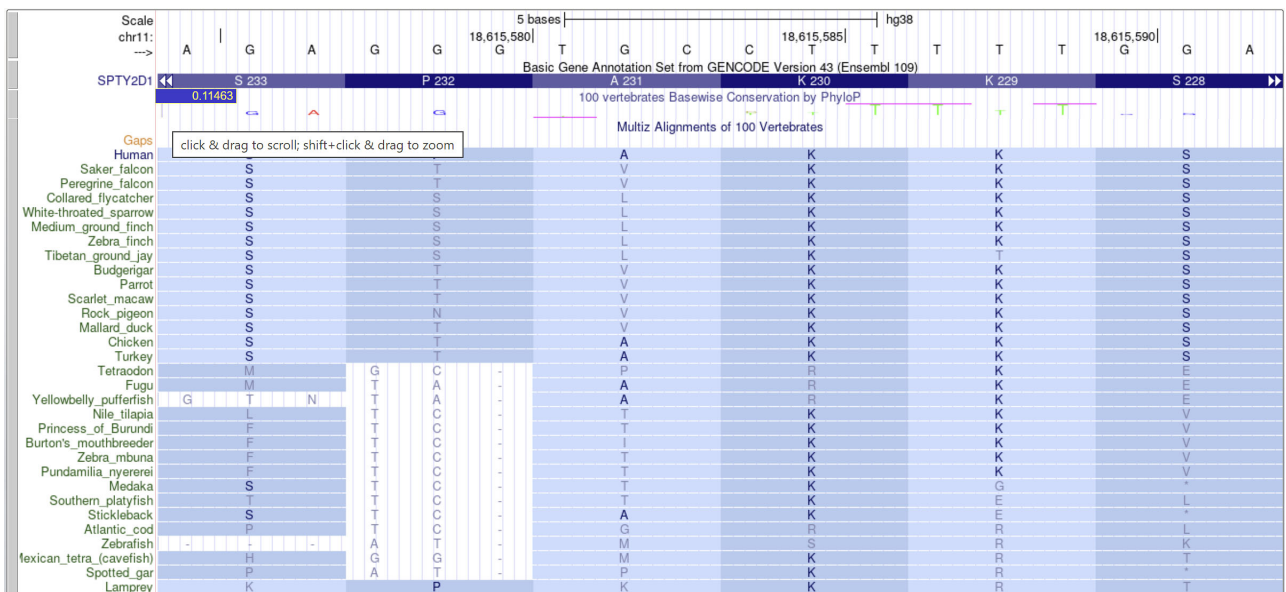
ANSWER: Exon 6 (at the far left, since this gene is on the reverse strand).

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

ANSWER: Exon 1 (at the far right)

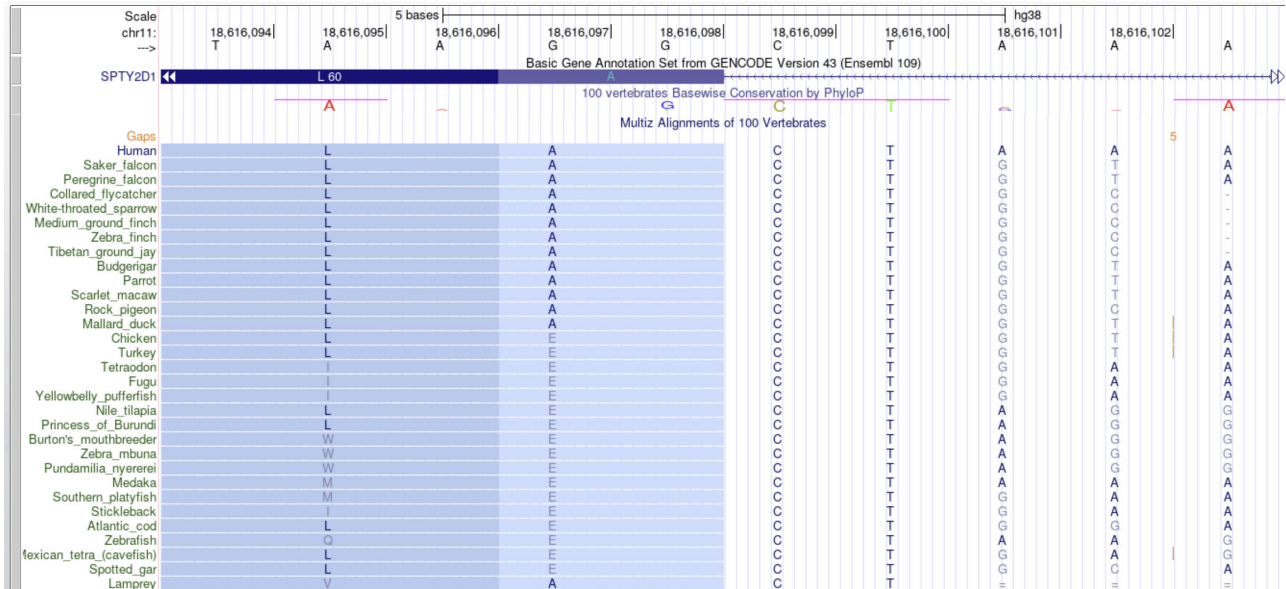
(9) Zooming in, give the DNA coordinates of an amino acid in this gene that is unchanged across all birds, but is highly variable in fish (with at least five different amino acids across fish at that position)

ANSWER: You really have to zoom in to see those amino acids. Then it takes a bit of poking around to find one, but there are two in the graphic below, one at the far left and one at the far right. The left has coordinates chr11:18,615,575-18,615,577 and the right has coordinates chr11:18,615,590-18,615,592.



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

ANSWER: Lo and behold, the first two are always GT and the last two are always AG (*note: this gene is on the reverse strand*). It turns out these are highly conserved splicing signals that vary little across all genes in all eukariotes.

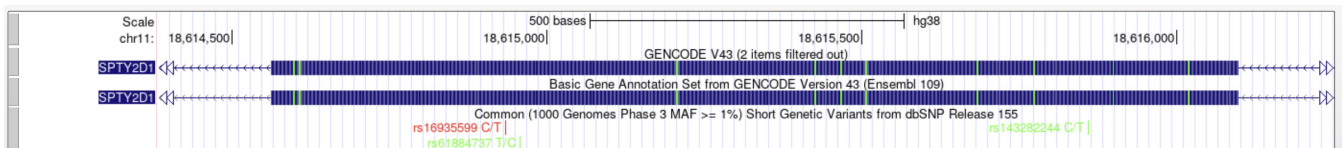


Turn off the conservation track and turn on the “dbSNP 155” track in the “Variation” section in pack mode. Click on the track itself in the graphic to toggle it or show the SNP IDs (if they’re not already showing). It will show a little hand when you’re in the right position to click.

Zoom in on Exon 3 of SPTY2D1.

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

ANSWER: It’s that red one, rs16935599:



(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.

ANSWER: You have to click on it to pull up its info page. Look for this about a page down, this says amino acid R is replaced by Q:

UCSC's predicted function relative to selected gene tracks:
 GENCODE V43 SPTY2D1 (ENST00000536336.5) [nc_transcript_variant](#)
 GENCODE V43 SPTY2D1 (ENST00000536334.6) [missense_variant](#) R (CGA) --> Q (CAA)

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

ANSWER: A lot, counting there are 21.

Allele	1000Genomes	dbSNP_PopFreq	TOPMED	SGDP_PRJ	Utah	NorthernSweden	Siberian	TWINSUK	ALSPAC	GENOME_DK	GnomAD	GoNL	Estonian	HapMap	PAGE_STUDY	MGP	PRJEB37584	GoESP	ExAC	GnomAD_exomes	FINRISK
C	4897/008 (0.977828)	28377/202414 (0.922274)	264659/264990 (0.962103)	14/20 (0.499997)	1892/16 (0.827290)	6/1/000 (0.920000)	5/10 (0.500000)	3448/3708 (0.930695)	3676/3854 (0.952814)	28/40 (0.850000)	13426/140286 (0.877009)	939/998 (0.940822)	4120/4480 (0.919433)	1717/1770 (0.972299)	7876/78700 (0.972299)	696/634 (0.947569)	790/792 (0.997475)	12438/12984 (0.957948)	11616/120734 (0.961932)	241424/259112 (0.947603)	288/204 (0.937600)
T	111/508 (0.022155)	14637/300414 (0.048726)	13031/59490 (0.037897)	16/30 (0.533333)	17/216 (0.078704)	39/600 (0.065000)	5/10 (0.500000)	190/378 (0.543150)	179/384 (0.441196)	2/40 (0.050000)	9031/140286 (0.042991)	59/968 (0.059118)	353/4480 (0.080257)	53/1770 (0.029944)	1444/78700 (0.024701)	28/534 (0.052434)	2/792 (0.002525)	140/12984 (0.042352)	4068/120734 (0.038167)	1688/251912 (0.006719)	19/204 (0.009250)

(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?

ANSWER: Only one, called PRJEB37584

PRJEB37584
790/792 (0.997475)
2/792 (0.002525)

Turn off the SNP track.

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

ANSWER: chr11:18616099-18616873

(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.

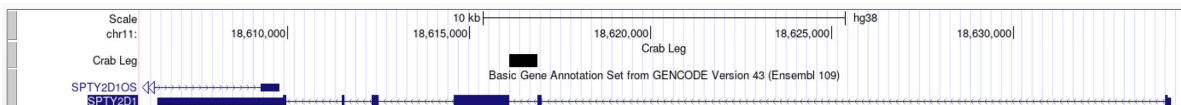
ANSWER: You have to subtract one from the left coordinate:

Paste URLs or data: Or upload: No file chosen

```
track name="Crab Leg"
chr11 18616098 18616873
```

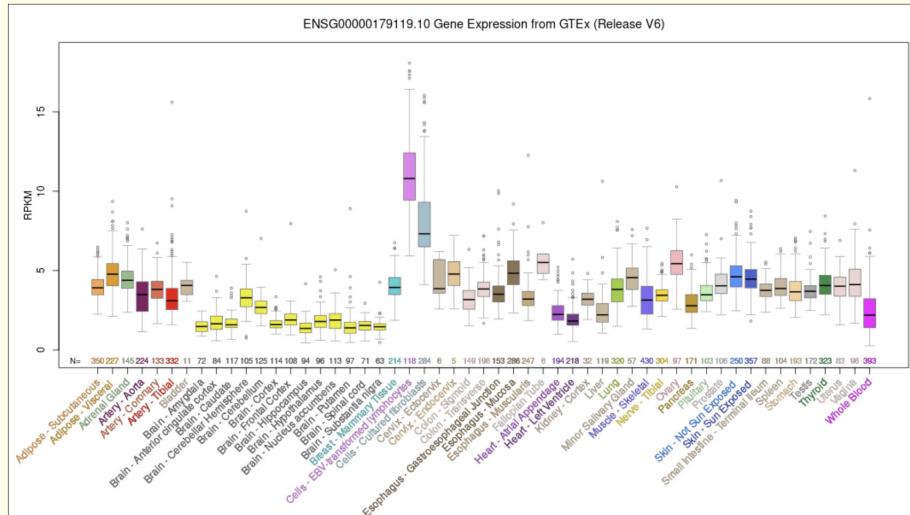
(17) Download a PDF showing the custom track (zoom out to show the entire gene). Include this PDF with your submission (can be included as a figure in your writeup or as a separate document).

ANSWER: You have to subtract one from the left coordinate:



(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*

ANSWER: Make sure to turn on the GENCODE V43 track and click on the gene identifier in the browser graphic, that will take you to its info page. Scroll down to the expression tissue profile figure and it shows the greatest expression in "Cells - EBV-transformed lymphocytes"



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

ANSWER: On the same info page, near the top is a table that indicates Protein (685 aa).

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

ANSWER: The thing from the previous question “Protein (685 aa)” is a live link to the protein sequence.

```
>ENST00000336349.6 (SPTY2D1) length=685
MDFREILMIASKGGQGVNVPKRYSLAVGPPKDPKVKGVQSAVQAFLLKRRKEELRRLKAL
EKRKREELVKKRIELKHDKKARAMAKRTKDNFHGYNGIPIEEKSKRQATESHTSQGTD
REYEMEEENEFL EYNHAESEQEEQEPKVKVPLKSAPPMPNF TDLRLAEKKQ
FEPVEIKVVKSEERPMTAEELREREFLEKRRKKLETGKLPPTVSKKAPSQKESVGT
KLSKSGDRHPSSKGMPLPHEAKSRPSEMANEKHLALSSSKSMPGERIKAGSGNSQPSL
REGHDKPVFNAGAKPHSSTSSPSVPKTSASRTQKSAVEHKAKKLSLHPSHSRPGPMVTPH
NKAKSPGVRQPGSSSSSAPGQPSGVARTVSSGVPVRRQNGSSSSGPERISGSKKPTN
DSNPSRRTVSGTCGPGQPASSSGGPRPISGVS SARPLGSSRGPGRPVSPHELRRPV
GLGPPGRSVPGRSISGSIPAGRTVNSVGRPVSSLGPGQTVSSSGPTIKPKCTVSE
TISKNIISRSNGQMGMPPLSGYRAAQQRLLPPTGYKRQREYEEEDDDDEYDSE
MEDFIEDEGEPEEISKHIREIFGYDRKRYKDESDYALRYMESSWKEQQKEAKSLRLGM
QEDLEEMRREEEEMORRRAKLKR
```

Searching this page there is only one W.

```
>ENST00000336349.6 (SPTY2D1) length=685
MDFREILMIASKGGQGVNVPKRYSLAVGPPKDPKVKGVQSAVQAFLLKRRKEELRRLKAL
EKRKREELVKKRIELKHDKKARAMAKRTKDNFHGYNGIPIEEKSKRQATESHTSQGTD
REYEMEEENEFL EYNHAESEQEEQEPKVKVPLKSAPPMPNF TDLRLAEKKQ
FEPVEIKVVKSEERPMTAEELREREFLEKRRKKLETGKLPPTVSKKAPSQKESVGT
KLSKSGDRHPSSKGMPLPHEAKSRPSEMANEKHLALSSSKSMPGERIKAGSGNSQPSL
REGHDKPVFNAGAKPHSSTSSPSVPKTSASRTQKSAVEHKAKKLSLHPSHSRPGPMVTPH
NKAKSPGVRQPGSSSSSAPGQPSGVARTVSSGVPVRRQNGSSSSGPERISGSKKPTN
DSNPSRRTVSGTCGPGQPASSSGGPRPISGVS SARPLGSSRGPGRPVSPHELRRPV
GLGPPGRSVPGRSISGSIPAGRTVNSVGRPVSSLGPGQTVSSSGPTIKPKCTVSE
TISKNIISRSNGQMGMPPLSGYRAAQQRLLPPTGYKRQREYEEEDDDDEYDSE
MEDFIEDEGEPEEISKHIREIFGYDRKRYKDESDYALRYMESSWKEQQKEAKSLRLGM
QEDLEEMRREEEEMORRRAKLKR
```

W 1/1 ^ v x