Lab: Biological Databases
Course: Bioinformatics & Computational Biology

Date: 08/02/2024

Tushar Goyal

• How many different transcripts are listed for TEC? - 5

• How many encode a protein? - 3

• Of the protein coding transcripts how many are likely to be removed by nonsense mediated decay? - 2

• How long in both transcript base pairs and protein amino acids is the main splice form? 3661 bp for base pairs & protein amino acid length of 631.

• How close is TEC to the centromere of chr 4?

  1.6Mb

• Which genes flank TEC on either side? Answer this for both conventional protein coding genes and for small RNA genes.
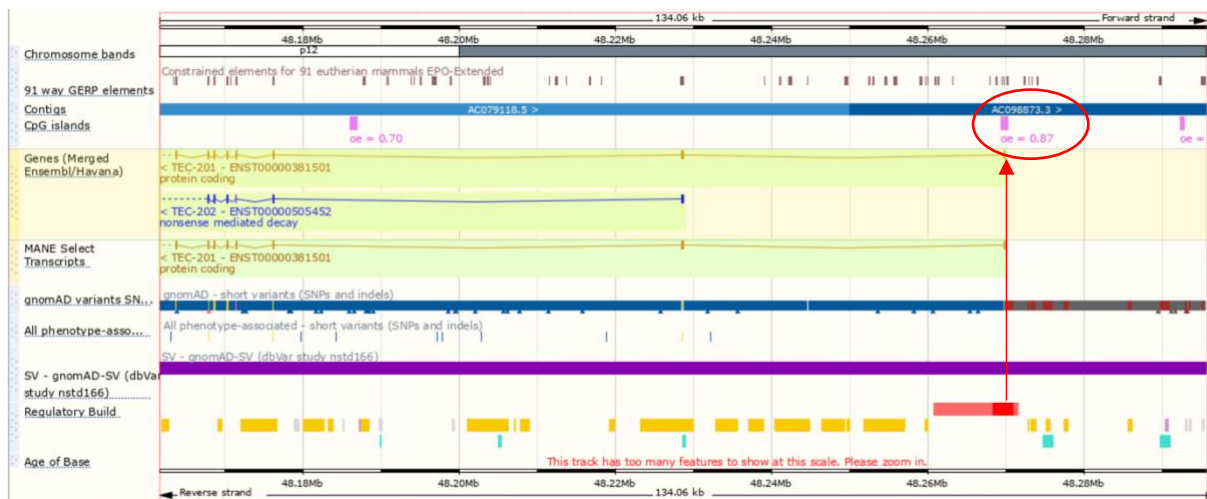
  Protein Coding - SLAIN, TXK,

  RNA Gene - U6, Y_RNA, RNU6-838P

• Look at the structure of the TEC transcripts. From which end (5¢ or 3¢) are most of the splice variants most truncated?
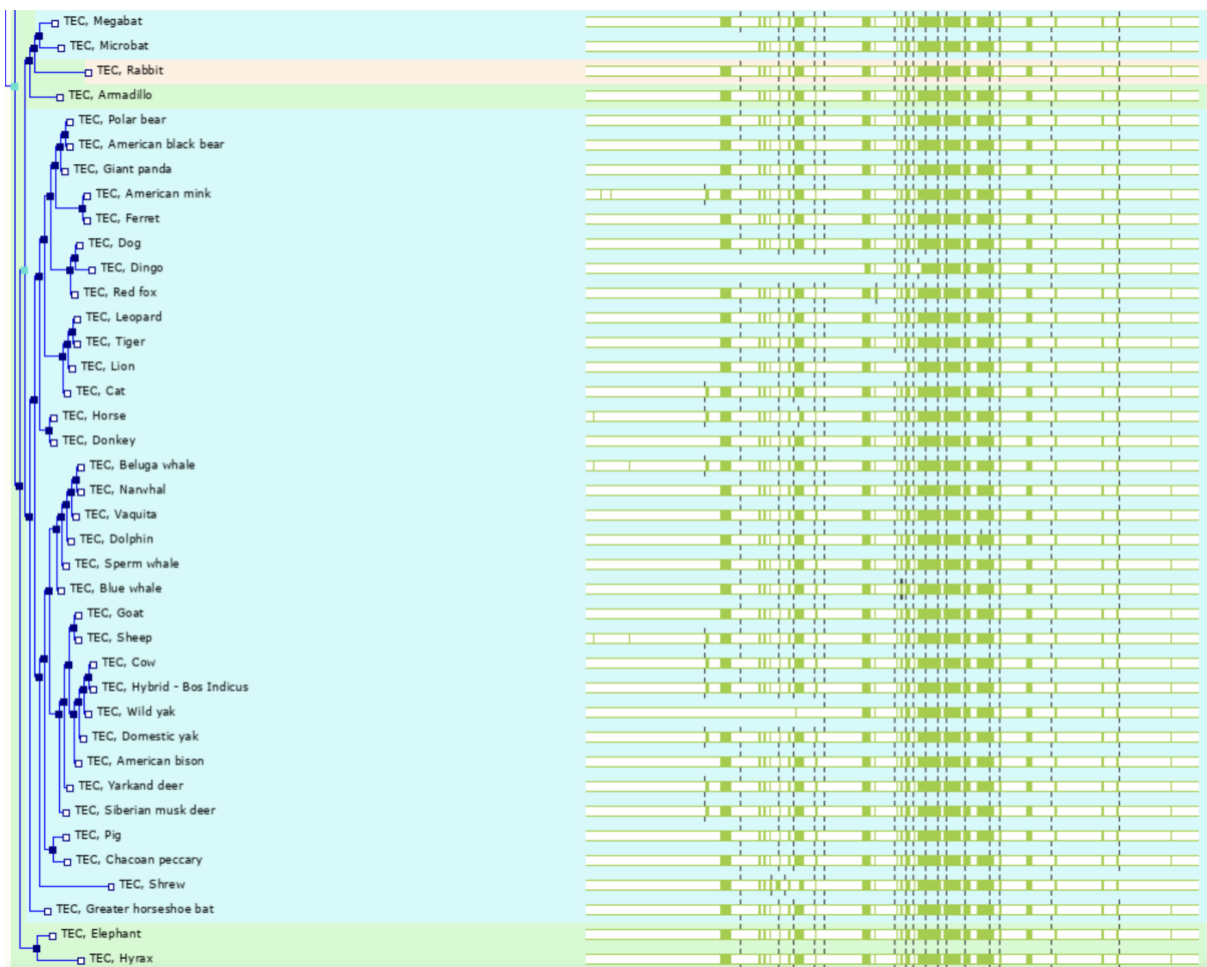
  3'

• Is there a CpG island in the promoter of TEC?

   Yes.

- In which placental mammals has TEC been observed? You will need to click on the placental group and expand the sub-tree to see the individual animals.

  39 Placental Mammals have TEC. Below is the expanded tree.

- Which phylogenetic group has a large deletion at the start of the gene (blank area on the exon structure picture)? Is this region missing from all animals in this group?

Jawed Vertebrates.
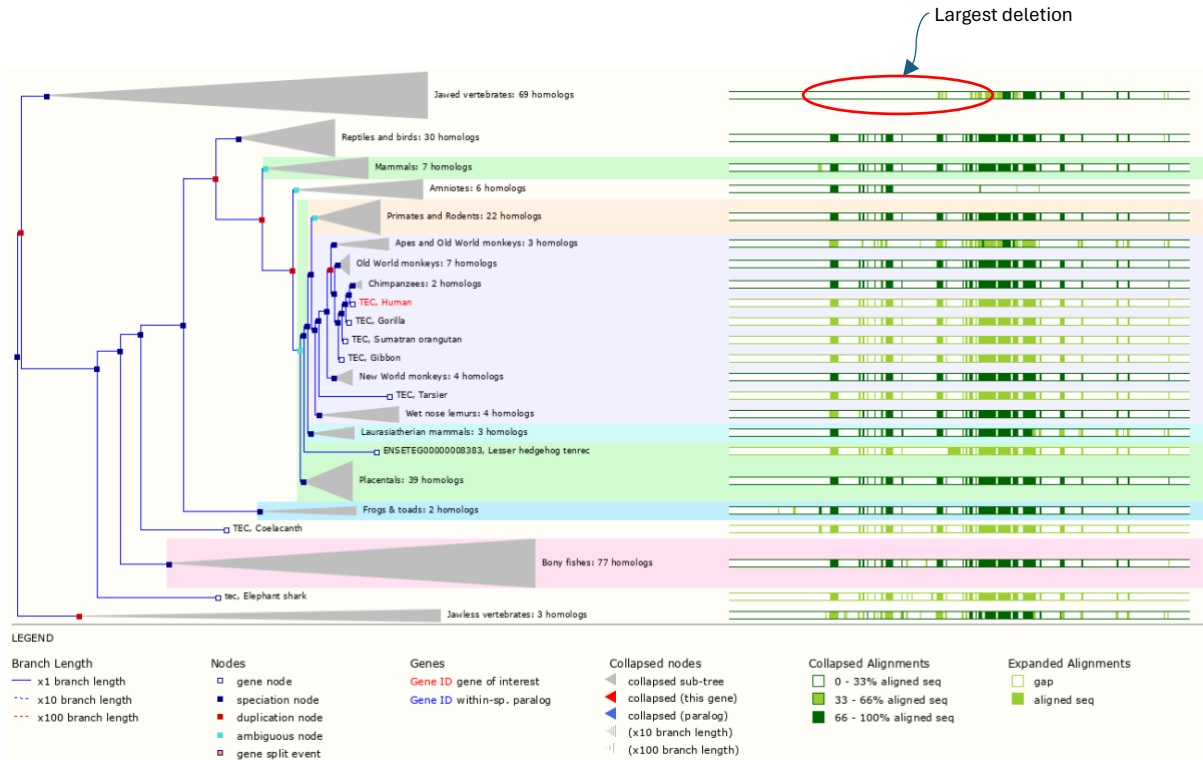
- What type of ortholog is this (1-to-1 or 1-to-many)?

  1-to-1

- How conserved is the DNA sequence (percentage identity)

  84.47 %

- Do all of the surrounding genes have orthologs in the two species?

No all the surrounding genes, don't have orthologs in the two species. For eg, ZAR1 is not present in Crocodile but is present in Human gene. Similarly SLAIN1 exists in both but is small in crocodile.
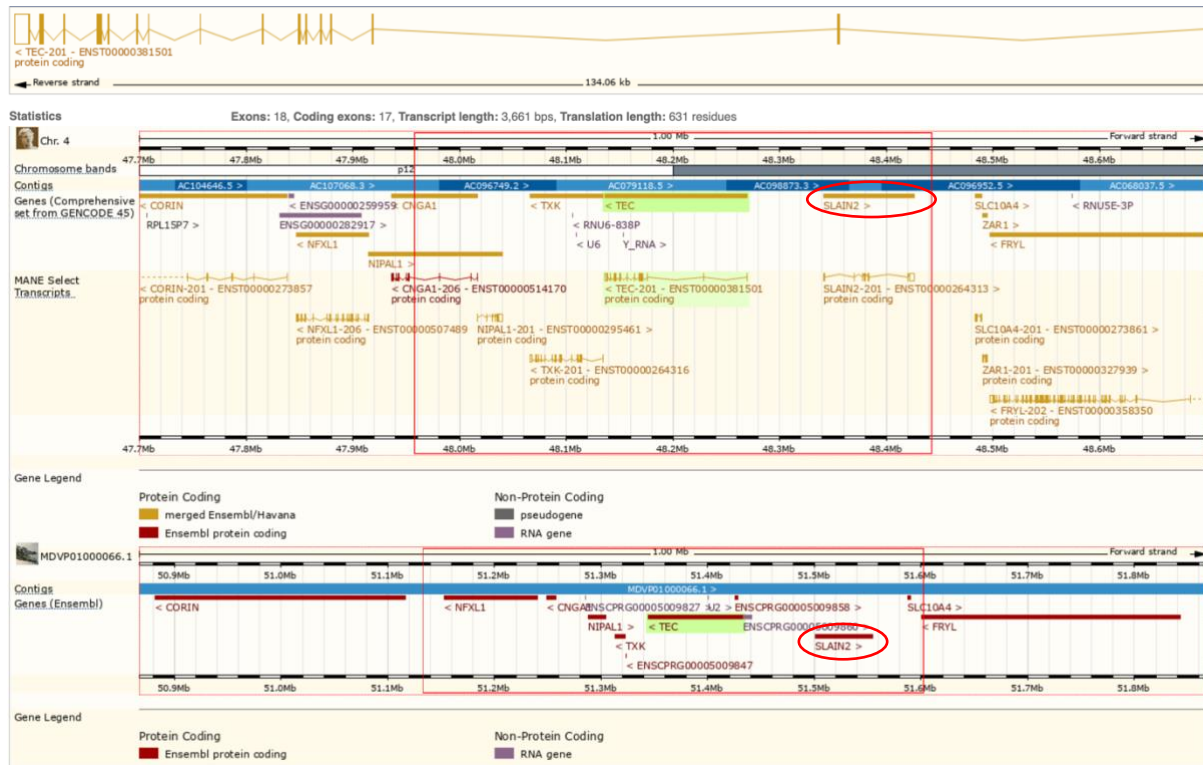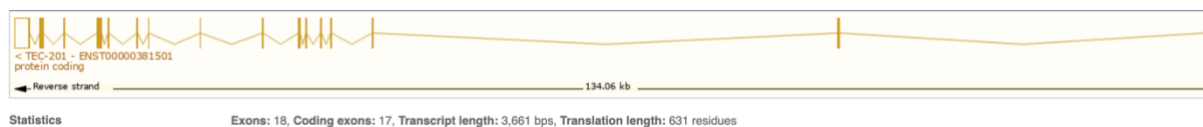

- How many exons does this transcript have?

  Exon: 18

- How many of them form part of the coding sequence?

  Coding Exons : 17

- Which is the longest intron, and how long is it?

Intron 2-3 -> Length: 52,290

• Download a fastA file of the TEC-201 transcript (turn off all of the other sequence types it offers to give you alongside the cDNA sequence)

```
>ENST00000381501.8 TEC-201 cdna:protein_coding
ACTCTGGGGCGCTAGGCTCCGGACTCCGCGGCGCAGACTGCACCTCGCAGTCTCCCCAGG
TCCGCCCAGCAGCCGCGCTTCAGCCAGAATACTGGGATCTTCAGTGGCAGGAGGAGTAAT
CAGAAGACGGAGATGAATTTTAACACTATTTTGGAGGAGATTCTTATTAAAAGGTCACAG
CAGAAAAAGAAGACATCGCCCTTAAACTACAAAGAGAGACTTTTTGTACTTACAAAGTCC
ATGCTAACCTACTATGAGGGTCGAGCAGAGAAGAAATACAGAAAGGGGTTTATTGATGTT
TCAAAAATCAAGTGTGTGGAAATAGTGAAGAATGATGATGGTGTCATTCCCTGTCAAAAT
AAGTATCCATTTCAGGTTGTTCATGATGCTAACACACTTTACATTTTTGCACCTAGTCCA
CAAAGCAGGGACCTGTGGGTGAAGAAGTTAAAAGAAGAAATAAAGAACAACAATAATATT
ATGATTAAATATCATCCTAAATTCTGGACAGATGGAAGTTATCAGTGTTGTAGACAAACT
GAAAAATTAGCACCCGGATGTGAAAAATACAATCTTTTTGAGAGCAGTATAAGAAAAGCA
CTACCTCCAGCACCAGAAACAAAGAAGCGAAGGCCTCCCCCACCAATTCCACTAGAAGAA
GAAGATAATAGTGAAGAAATCGTTGTAGCCATGTATGATTTCCAAGCAGCAGAAGGACAT
GATCTCAGATTAGAGAGAGGCCAAGAGTATCTCATTTTAGAAAAGAATGATGTTCATTGG
TGGAGAGCAAGAGATAAATATGGGAATGAAGGATATATCCCAAGTAATTACGTAACGGGA
AAGAAATCAAACAACTTAGATCAATATGAATGGTATTGCAGAAATATGAATAGAAGCAAG
GCAGAGCAACTCCTCCGCAGTGAAGATAAAGAAGGTGGTTTTATGGTAAGGGATTCCAGT
CAACCAGGCTTGTACACAGTCTCCCTTTATACCAAGTTTGGAGGAGAAGGTTCATCGGGT
TTTAGGCATTATCATATAAAGGAAACAACAACATCTCCAAAGAAGTATTACCTAGCTGAA
AAACATGCTTTTGGCTCCATTCCTGAGATTATTGAATATCATAAGCACAATGCAGCAGGA
CTTGTCACCAGGCTTCGGTACCCAGTTAGTGTGAAAGGGAAGAATGCACCCACCACTGCA
GGATTCAGCTATGAGAAATGGGAGATTAACCCTTCAGAACTGACCTTTATGAGGGAATTG
GGAAGTGGACTGTTTGGAGTGGTGAGGCTTGGCAAATGGCGAGCCCAGTACAAAGTCGCA
ATCAAAGCTATTCGGGAAGGTGCAATGTGCGAGGAGGACTTTATAGAAGAAGCTAAAGTG
ATGATGAAACTGACACACCCGAAGTTAGTGCAGCTTTATGGTGTGTGCACCCAGCAGAAA
CCAATATACATTGTTACTGAGTTCATGGAAAGGGGCTGCCTTCTGAATTTCCTCCGACAG
AGACAAGGTCATTTCAGTAGAGACGTACTGCTGAGCATGTGTCAGGATGTGTGTGAAGGG
ATGGAGTATCTGGAGAGAAACAGCTTCATCCACAGAGATCTGGCTGCCAGAAATTGTCTA
GTAAGTGAGGCGGGAGTTGTAAAAGTATCTGATTTTGGAATGGCCAGGTATGTTCTGGAT
GATCAGTACACAAGTTCTTCTGGTGCTAAGTTTCCTGTGAAGTGGTGTCCACCTGAAGTG
TTTAATTACAGCCGCTTCAGCAGCAAATCAGATGTCTGGTCATTTGGTGTTTTAATGTGG
GAAGTATTCACGGAAGGCAGAATGCCTTTTGAAAAATACACCAATTATGAAGTGGTAACC
ATGGTTACTCGAGGCCACCGACTCTACCAGCCGAAGTTGGCGTCCAACTATGTGTATGAG
GTGATGCTGAGATGTTGGCAGGAGAAACCAGAGGGAAGGCCTTCTTTCGAAGATCTGCTG
CGCACAATAGATGAACTAGTTGAATGTGAAGAAACTTTTGGAAGATAAGTGATGTGTGAC
CAGTGGCTCCCAGATTCCCAAGCACAAGGAAGGATGGGCATTTTGTGGCTTTTAATTTAT
TGAGCACTTGGACATGTAGATCATTTTACTTATACAGTGGAAACACATAAATAATTTGCT
TCTAGACCAGCCTCTGTCTAGACTTGCTTCTAGACAGAATCTCCCAGAGTGTGGAAATGT
TGCCTTAGAAATGGTGATTAAAATCACTCATTTCTATTCATTCCTCAGGCACTTGAGTGA
CAGTTGTTTACCAGGCACTGTGTGTAGCCCCAGGGTTTGGCCATTCAGGGGTGCACACAT
GGGACCATGTTAGCTGATGCCAGTTGAAGGCCAGGGTATTTGGGAAGGGGAAGGGTATTA
GAGTCATGACCAAGCAACCCTTCTTTTTCCCTTTGACTTCTACAGAAATCTGGGCCTGAG
ACATTGTCTACAATTGGGTTCTAGATACATCAGGAACCCATCTTGGATAAATAAATACCT
ATCTTTTGTTTTGAAAACATCTCAGTTTTCAAGACTGCTCTTAGTATTACATGAACAATA
TTTGTATGCTGTATATATTGTAAATATATATAATATATAAAGTTATATATTTATGAGAAA
CACGAATTGTCTTTTAATTGAAACTTTTAATCCTGTAGTATAGGAGTTCACCTTCTTAGG
ACTAGAGACTGTGCCTTATAGCTGTTAATTCATTTCCCCCTGAACATCAAATATGCCTGA
AGAGAAGAAAGTCTAGATTCTTCTATGAGTAACGCCCCCTCCTCACTCAGGTAAATGTGT
CTGGGGATGCCTGTCCAGCTTAACCACGTGCATTTGGCCTATGTAATCCTGCCCATGGTG
GCCGCAGCTAATCAGAATCAGATGGAAAATTAAACCGGGTAATCTACTTCTAAGCCTTAA
GAATATTCCCTGGGACACAGACACTATAATTGGAAGTGCTGAGCTCTGGGGCAGAAGGAT
CAGGTGACCTTCGCAACAAAGTTTGCCCCCACCTCACATAGGACCCGGAAGCAGCCTGAG
CTGTGGCGGAGGATCCAGGAAGCTACGGAGAGAAGCAGCCAGCATGGTGTTCCGTGCCTC
CCGGACGTTTTTCAGGAGGCCTGGTTGGACTTGGGTTCCTGGATGGTGGGATTGTTGTAC
AGCCTCTCAGGAGACCCTGCTGTCAAGACTGTGTGTGTGGATTTCTCACCCTTAGAAGCT
CTACTAAGACATCAACGGAATTAGGGCCTTCCTTTTTTGCCTTGTGAGCGCCAAGGAAAAG
AAACTATCTCGGTCACGTGAGCGCCAGCGAAAAGAAACTGTATCAGTCATCCAGAGACCG
TTTATTGCCCAACACGTTATTCTTGCTGTTGGTGGGGTAACTAGCCGAGGAAGACACAGC
GCCTTCCCTTCAGGAGTTGCGTCTCCTCTGCAGGCCACGATGGTCTGCTCTGGAGCATTG
GGTGAACACACAGGCTGGCTGCTCTGGGCAGCGCCTTCACTCTGACCCTGGAGAACCATT
TCATTTCATCCTGGTCAGTCTAGAGTCTGTGCACCAGGCAGTCCATCCACTGAAGGCTGT
GTTTATTCTTTTCCTGTGCCCCTCATAATGGAAGAAAGTAAACTGCTTATCCCGAGCCTT
A
```

- Which genes can you see?

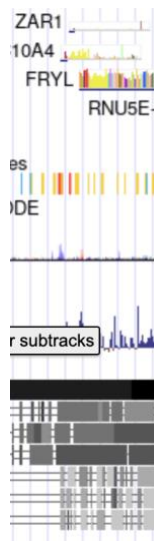 NIPAL1, SLAIN2, TXK, FRYL, CNGA1, NFXL1, ZAR1 etc.
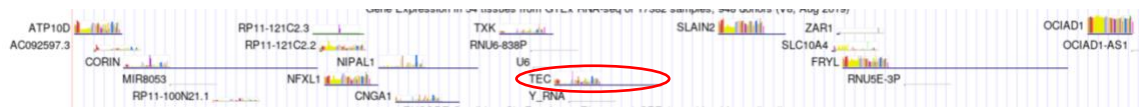
- How many splice variants do they have?
  - NIPAL1 – 6
  - SLAIN2 – 5
  - TXK – 6
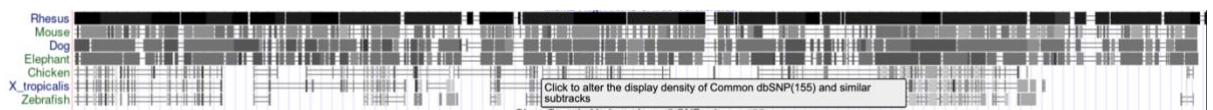 - FRYL – 21
  - CNGA1 – 7
  - NFXL1 – 8
  - ZAR1 – 1

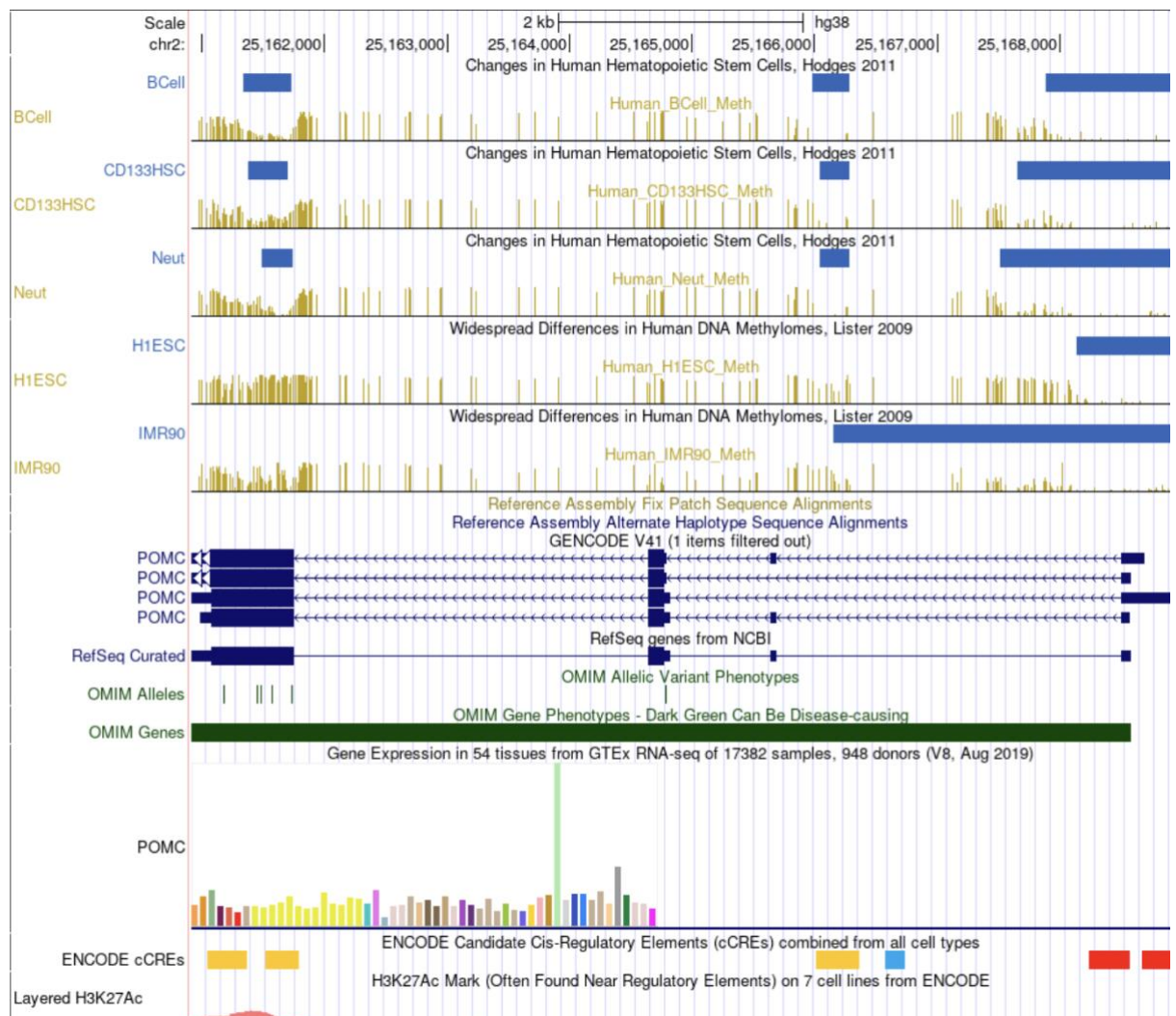- Which part of which gene has the highest overall conservation?

  ZAR1

- Do you observe a loss of conservation in intergenic regions (the bits between genes)?

  Yes. Evident in the screenshot below.

- Find the DNA Methylation hub and connect to it. Your session will restart, so re-find the TEC gene and you should see a load of new methylation tracks. - Yes

- Zoom out from TEC and see if you can find a region near the gene where the methylation level drops substantially.

  Yes.

- Is the extent of the unmethylated region the same in all of the tissues which are shown?

  Yes.
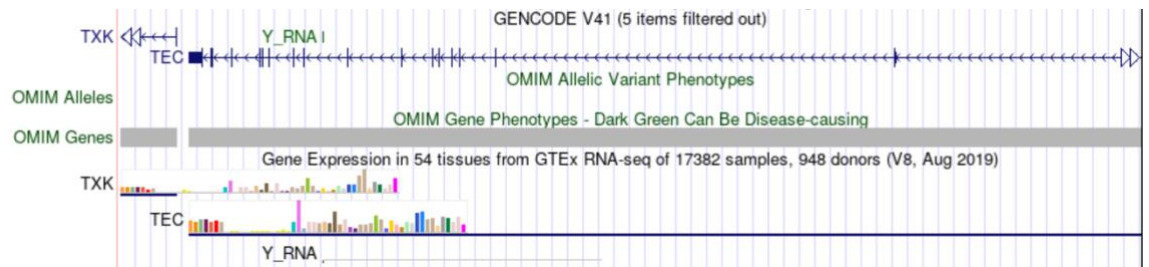
Biomart

Dataset 560 / 70711 Genes

Human genes (GRCh38.p14)

Filters

Chromosome/scaffold: 4
Gene type: protein_coding
Orthologous Platypus Genes:
Only