School of Medicine, University of Patras

Biology I – Bioinformatics Laboratory, 2026

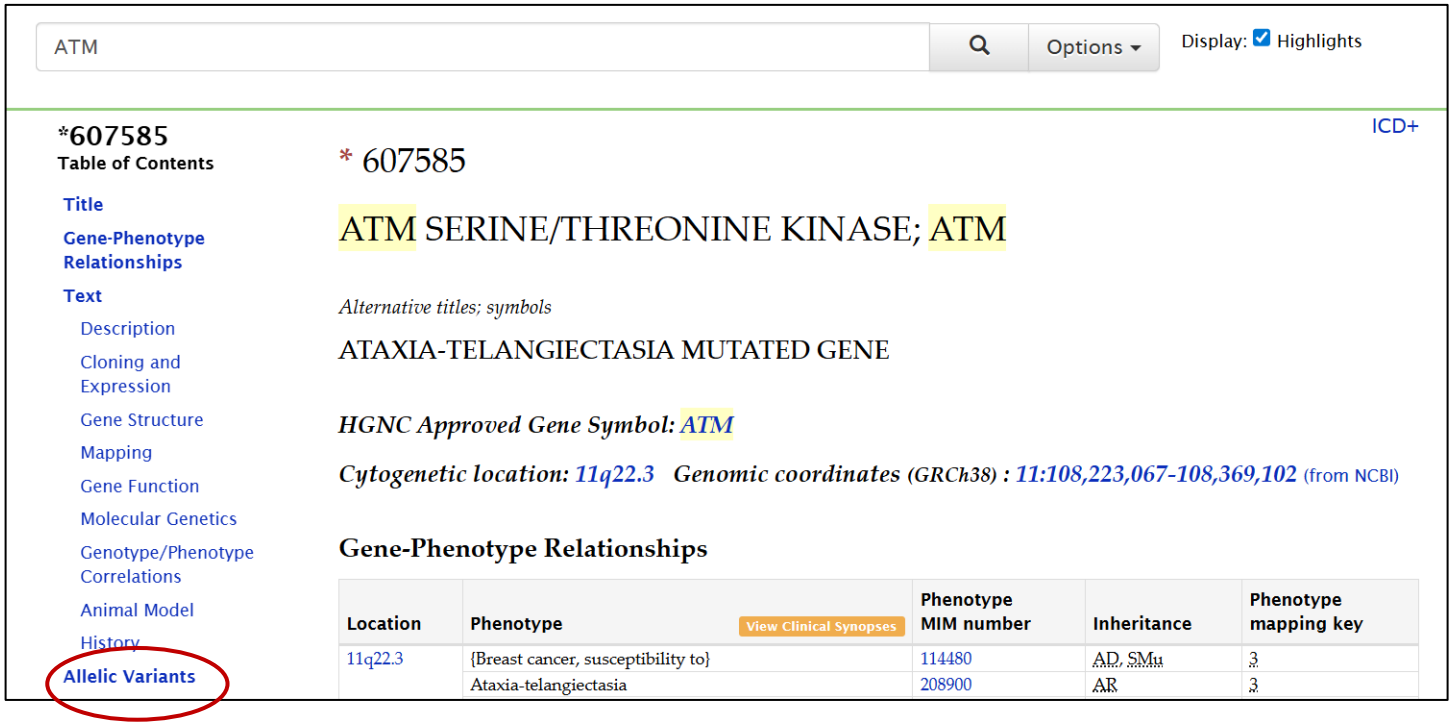
Laboratory Instructor: Yiannis Drosos

Guidelines for preparing the Bioinformatics Laboratory assignment

Step 1: Each student will choose one of the 20 genes in the list below for the assignment they have to submit. For the mutated sequence of the gene, go to the eclass folder Root directory › 3. Laboratory Exercises › Laboratory Exercise "Bioinformatics" › Genes_For_Bioinformatics_Analysis, and choose one file: this will be the gene and the mutation you will study.

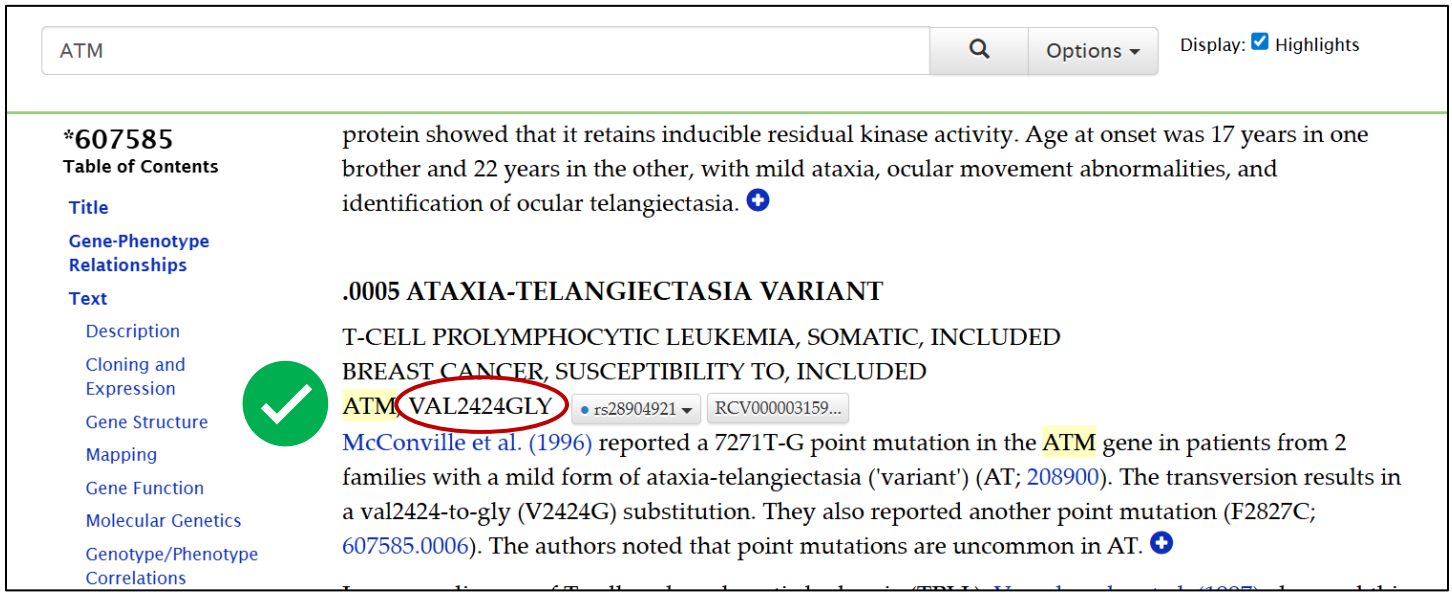
1. ATM
2. BRD4
3. CHD1
4. CHD3
5. CHD4
6. CHD5
7. CHD7
8. CREBBP
9. DNMT3A
10. DNMT3B
11. EP300
12. EZH2
13. KMT5B
14. MYST3
15. NSD1
16. PHF6
17. SMARCA2
18. SMARCA4
19. TAF1
20. TP53

Step 2: Next, go to the OMIM database and search for the gene you chose for your assignment. The example below shows the gene of the ATM kinase.

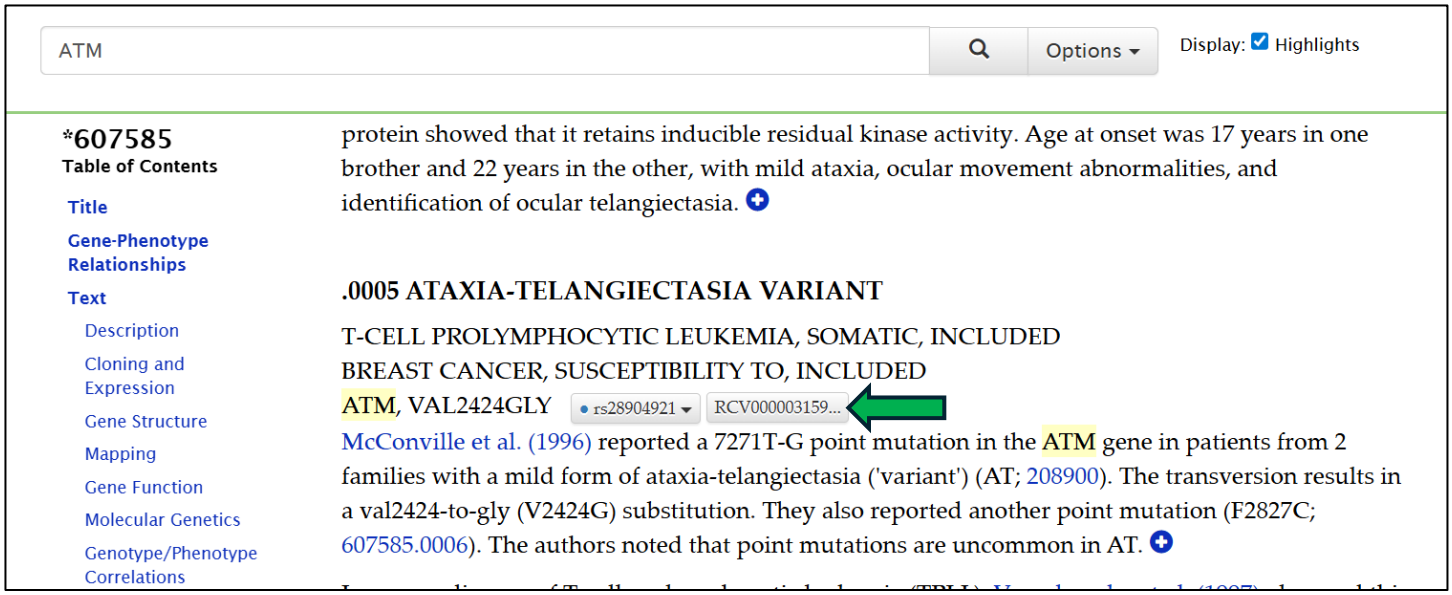


Step 3: From the list on the left, select "Allelic Variants" and scroll down to the bottom of the page, which lists all the mutations of the gene characterized so far.

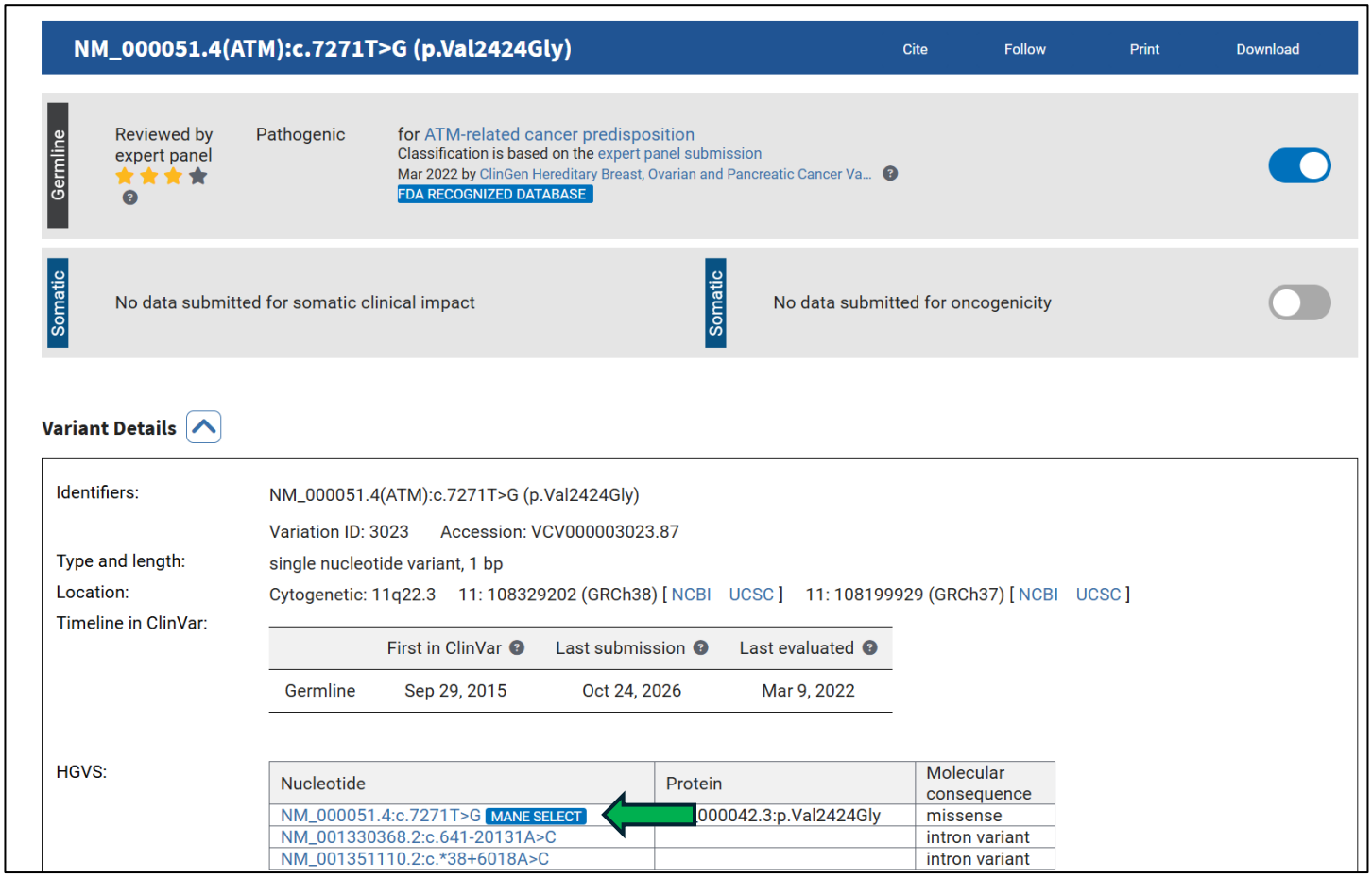
From the available mutations, select the one that corresponds to the mutated gene you chose from eclass. The example below shows the mutation VAL2424GLY, in which valine at position 2424 is replaced by glycine (VAL2424GLY).



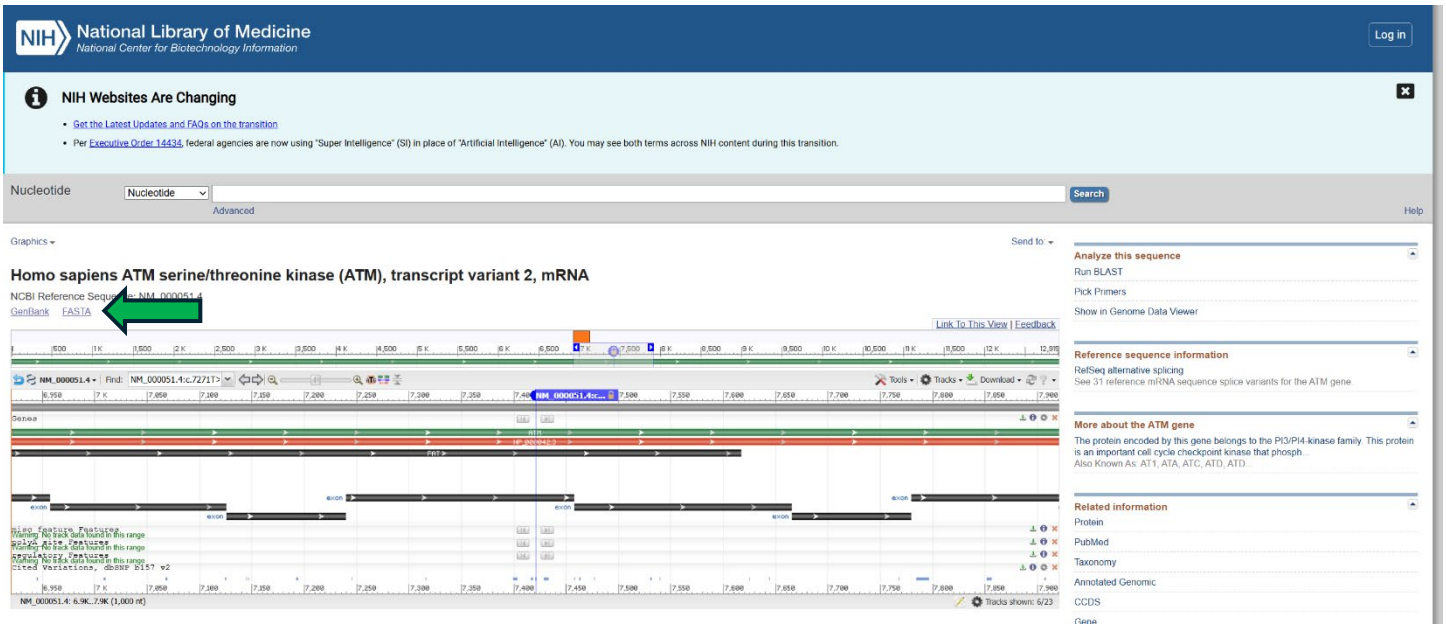
Step 4: Next, follow the "RCV..." link, as shown below for the substitution VAL2424GLY; it takes you to the website with the clinical data (ClinVar).



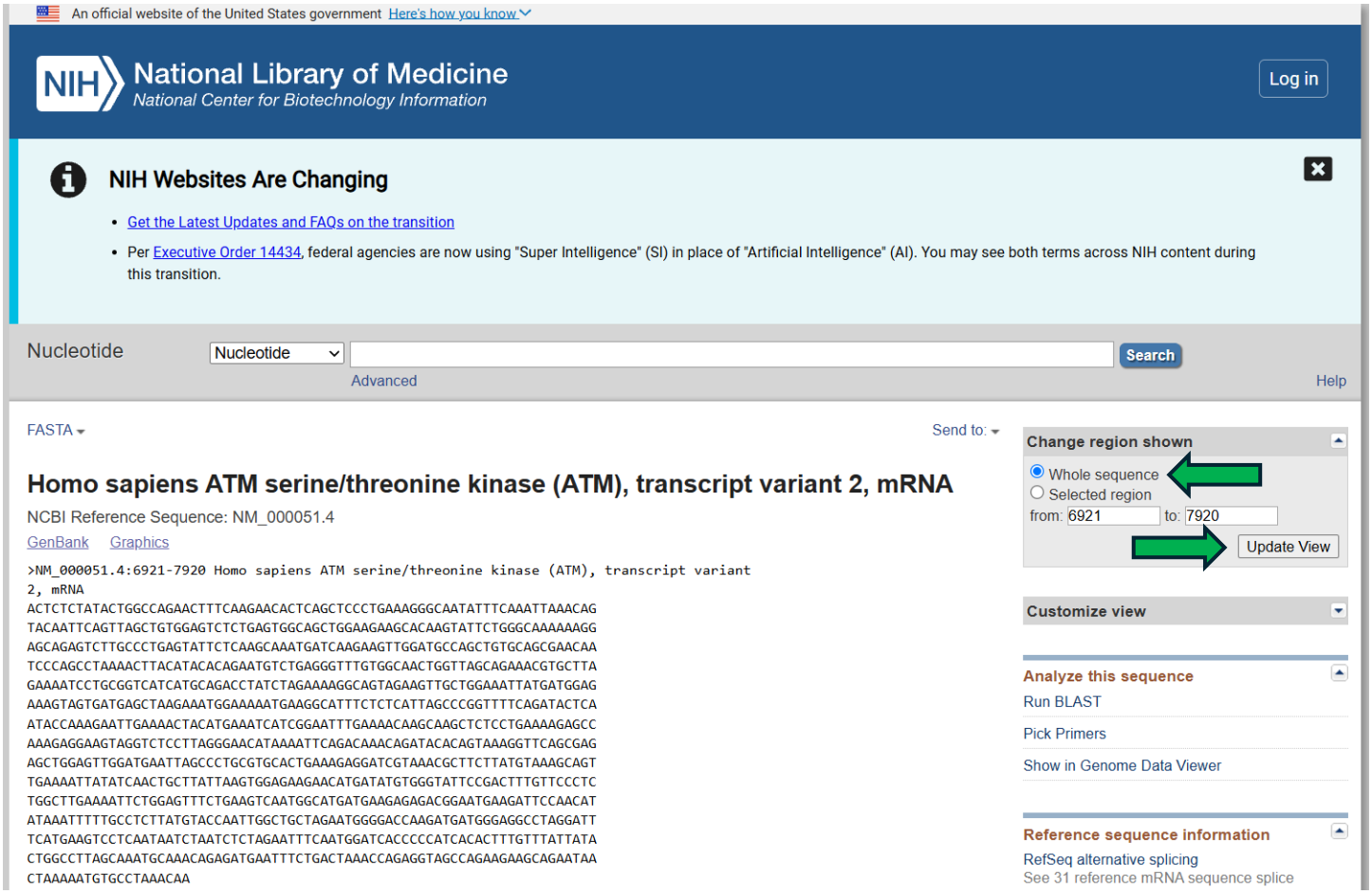
Step 5: Once you are on the ClinVar website, find the different transcripts of the gene and choose the one labeled "MANE", which is the "canonical" form of the normal gene.



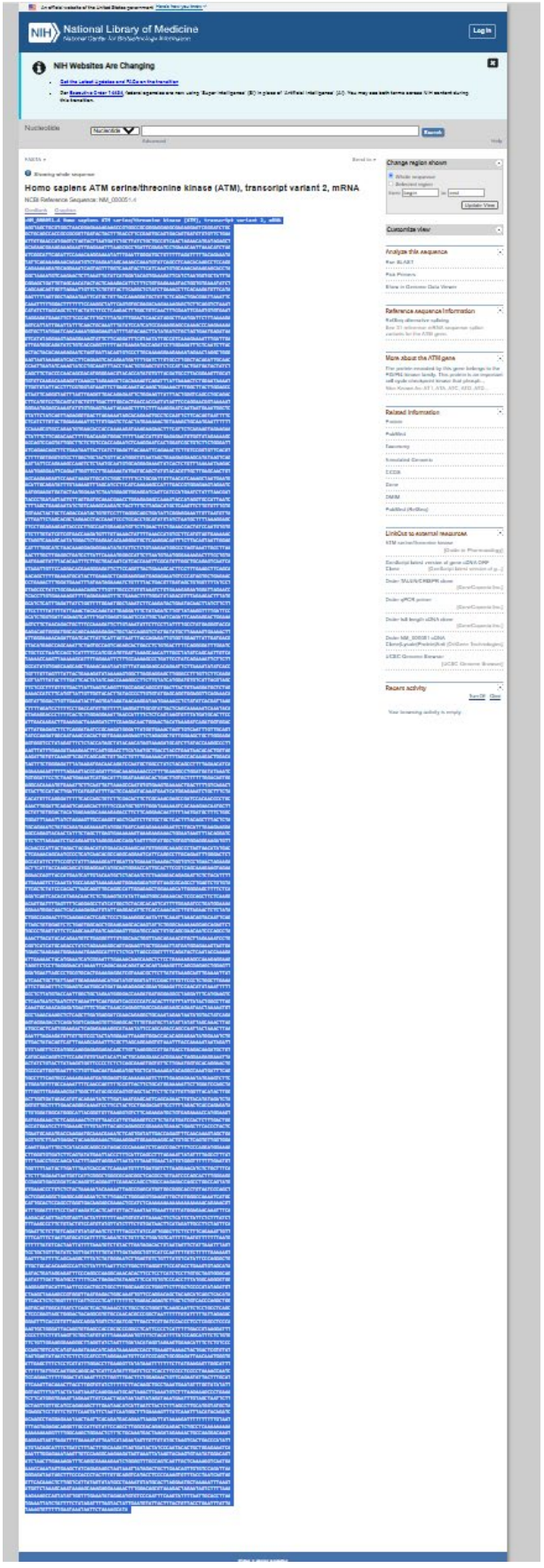
Step 6: Selecting the "MANE" transcript takes you to the NCBI/Nucleotide website, where you will find the reference sequence of the normal gene. Select the FASTA format.



Step 7: With the FASTA format selected, choose the whole sequence on the right of the page (Whole sequence→Update View)



Step 8: Select the entire FASTA sequence (including its name and the > symbol) and paste it wherever you need the sequence of the normal gene you chose.



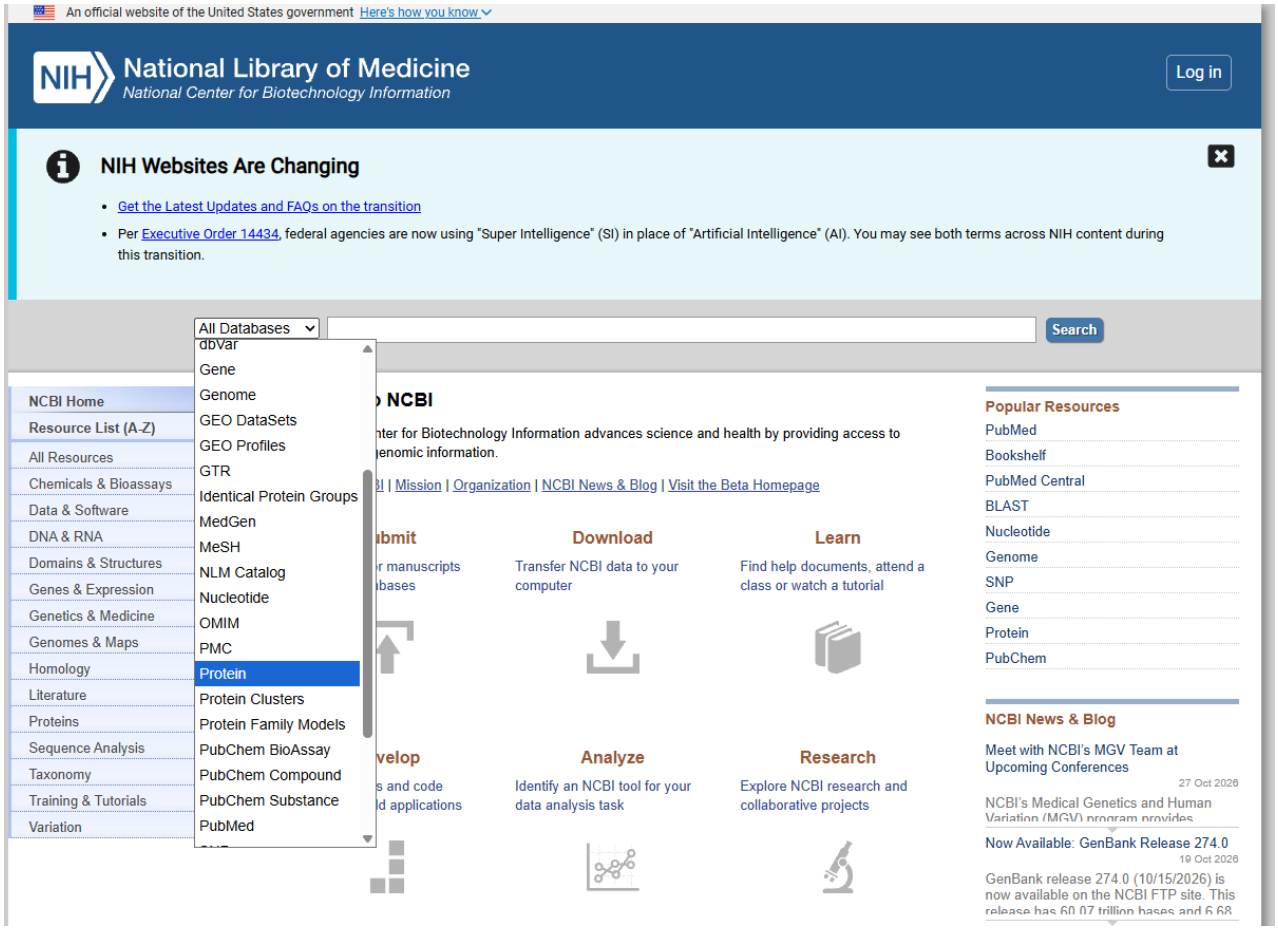
Step 9: Then, following the stages of the laboratory exercise you attended, answer the questions below about the mutated gene you chose.

Questions

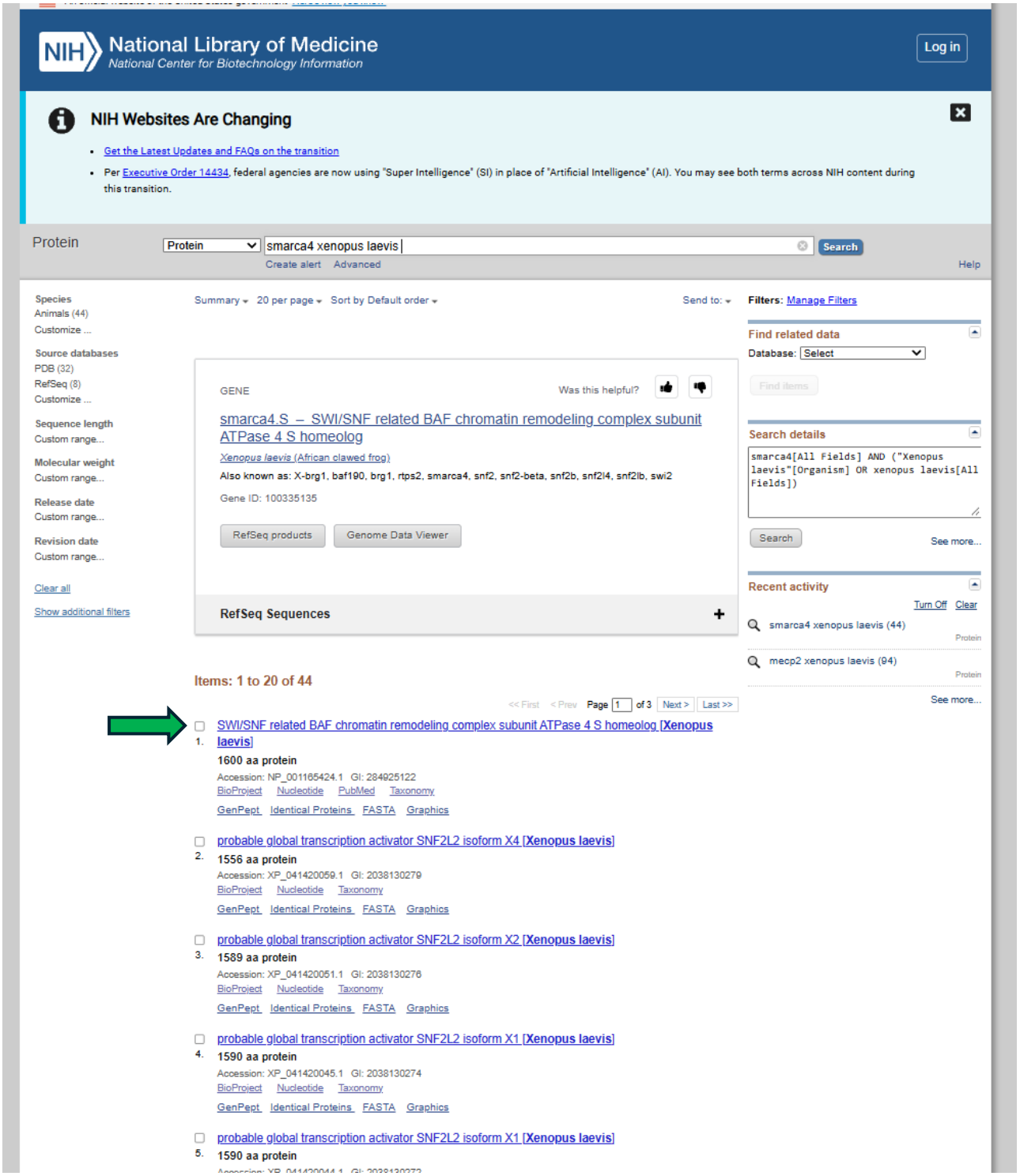
1. On which chromosome/locus is the gene you chose, and what is its mode of inheritance?
2. At how many positions does the mutated gene differ from the normal one?
3. How do the DNA changes alter the protein?
4. Are the amino-acid changes located within a functional region (domain)?
5. Are the altered amino acids conserved in other organisms?
6. Where are the changes located in the protein structure, and how dangerous does AlphaMissense consider them?

Note: To find the protein sequences from other organisms for question 5, follow these steps:

- Go to the NCBI website: https://www.ncbi.nlm.nih.gov/
- Select the "Protein" database.



- Search for the protein sequence in each of the organisms Mus musculus (mouse), Xenopus tropicalis (frog) and Danio rerio (fish).
- From the records that appear, choose the one with the largest number of amino acids (usually one of the first 10 records), as shown in the example of the frog SMARCA4 protein.



- If there are no proteins for the frog and the fish, search instead for the protein of Rattus norvegicus (rat).