

Using NCBI resources

Now that you know more about your protein, it becomes easier to search for it and to find it.

Search for your protein:

<https://www.ncbi.nlm.nih.gov/variation/view/>

You can see which keywords/accession codes are accepted by clicking on Search examples.

Which chromosome is it located on? Note its location.

What information is contained in dbVar?

What information is in dbSNP?

What information is in ClinVar?

You can filter by selecting

In ClinVar – yes

Most severe clinical significance – pathogenic

Molecular consequence – missense variant

How many variants do you have? (Preferably more than 3)

Click on each variant to see the mutation. Note this as the mutation (such as Trp318Cys). What type of residues are these residues? (acidic, basic, hydrophobic, polar). Note this as the mutation type (such as hydrophobic to polar). Fill out the following table for your favorite five variants.

Variant ID	Mutation	Mutation type	PDB ID	Location in protein (domain? Close to active/binding site?)	Polyphen 2 score	Prediction about impact?

Now click on the variant. It will take you to the dbSNP page, which contains a link to

“3D structure mapping” for some of the variants (Scroll down to NCBI resource links). Not all variants will have structure information. What is the PDB ID of the mutant protein? Note this in the table. The software Cn3D is invoked.

What is the PDB ID of the wild type protein?

Using iCn3D, make a structure figure for your protein of interest. If possible show the mutation site with iCn3D. You can also visualize the effect of the mutation using the structure viewer in PolyPhen 2.

Putting all the information together.

Go back to the PDB and find the structure entry with a recent publication record. You can choose the wildtype form and/or the mutant form. Go to the publication, download it and try to collect as much information about the protein, its structure, its conserved/critical structural elements as possible.

Some key biological questions are:

1. What does the protein do? What is its function?
2. Which disease state(s) is it implicated in?
3. How big is the protein?
4. Where is it located in the cell?

I would like you to put together the information you collected from the PDB, Uniprot, EBI and NCBI on the protein as well as its variants.

Structure figures are helpful in describing the effect of the variation. If there are no structures available for the mutant, predictions based on the wild type structure can be used. You can highlight individual residues using iCN3D. You can for example show the active site residues and the mutation site to describe the proximity/distance of the mutation site from the active site.

Prepare maximum 10 slides describing your findings. You will have 10 minutes to present your work as a group.

All of you have worked very hard this week and I look forward to hearing all about your protein on Friday.

Please prepare a report that brings together all of your findings. This report should include all of the different reports you have prepared so far and should have a good coherent flow.