Contextualising Human Missense Variation

Tutorial

CCPBioSim: September 19th 2023

James Stephenson

Tutorial Overview

This tutorial will guide you through some of the ways that variants can be explored in ProtVar. If anything is unclear then please ask.

a) Exploring variant input types
b) Mapping from coordinates to proteins
c) Functional annotations
d) Co-located variants
e) Variants in protein structure
f) Filtered download
g) ProtVar API

h) Further exploration

a) Exploring variant input types

ProtVar https://www.ebi.ac.uk/ProtVar/ accepts many different formats to help users. Explore the options using the below:

a1) Which genome assembly are these coordinates from (use auto detect)? Can you think how ProtVar can tell?

1-55505578-T-G
1-55505581-G-C
1-55505587-C-T
1-55505592-G-A
1-55505595-C-A
1-55509606-G-A
1-55509607-C-G
1-55509618-C-T
1-55509619-G-A
1-55509619-G-T

a2) Which nucleotide in the genome causes this amino acid change? Why are there two options?

P22309 71 Gly Arg

a3) What is the variant nucleotide for this ID? What does this tell us about using these IDs as variant identifiers?

rs587778656

b) Mapping from coordinates to proteins

Submit the following genomic locations:

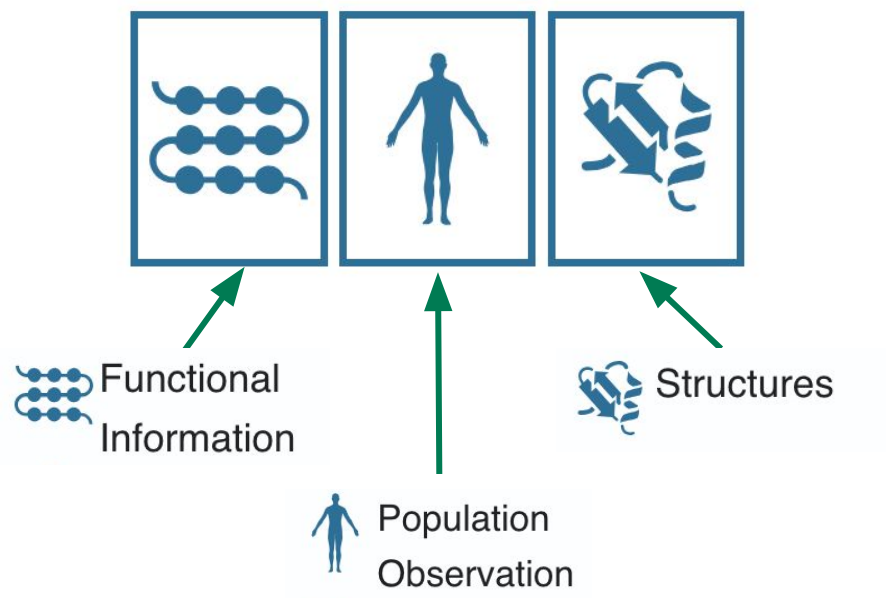
x 41345193 ID1 G/C
x 41345207 ID2 G/T
10 87933147 rs7565837 C/T

b1) Which proteins do these coordinates map to?
b2) What are the positions of the variants in each protein?
b2) Are all variants equally likely to be detrimental to the protein based on the CADD score and consequence?

f) Functional Annotations

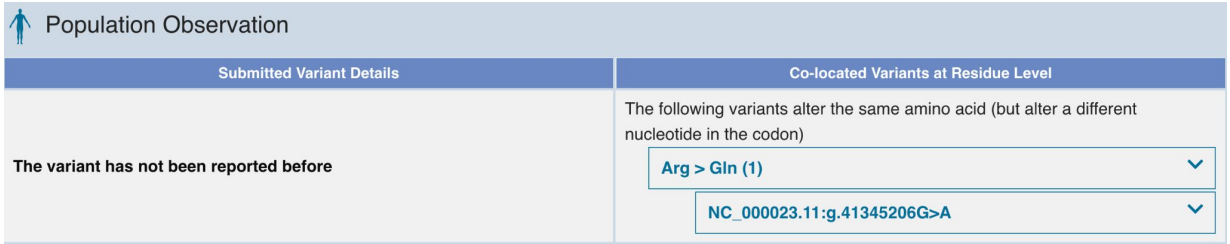
To the right of each entry row are three buttons which contain the annotations for the variant position. Use the first button (functional information) to answer the following questions):

c1) What variant specific information do we have for the first variant which we don't for the second?
c2) what is different about the regions containing the first two variant position?
c3) From the functional annotations would you consider these regions to be important to protein function?



g) Co-located Variants

The Population observation button contains information regarding the submitted variant in other databases. It also shows where variants have been reported at the same residue position (this could be different genomic positions).



d1) The second variant (351) is synonymous and therefore likely benign, it has never been reported before. However this does not mean that any variation in this position would be. What information can you find to suggest that changing this amino acid might be detrimental?

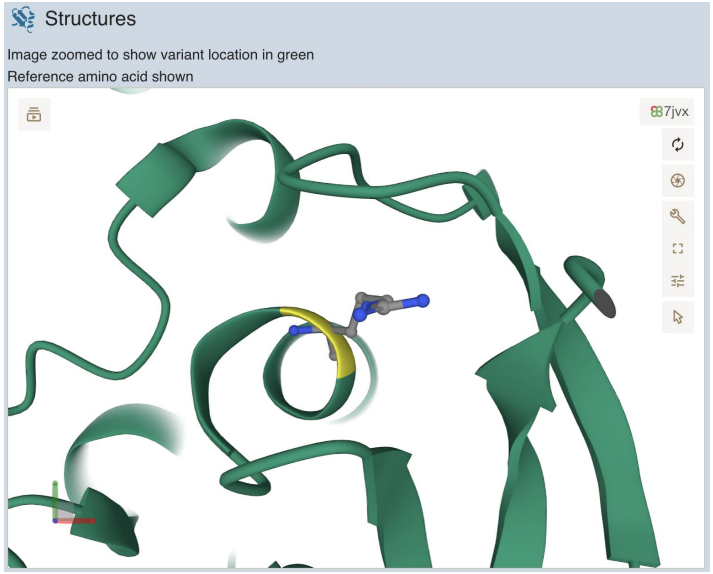
d2) List the databases that the the third variant (in PTEN) has been reported in.

d3) The PTEN variant has lots of reported variants at that amino acid position. What diseases are associated with altering this residue position?

e) Viewing Variant Positions in Protein Structure

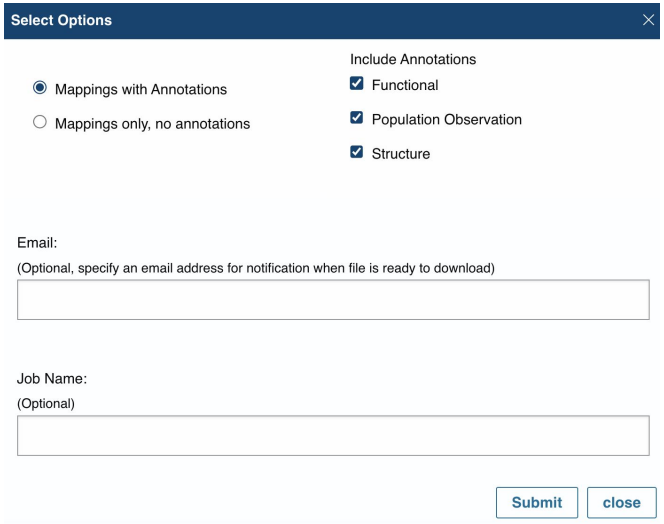
The third button (structures) shows your variant in the context of the protein structure.

e1) How many structures are available for the first variant? What is the range of their resolution?
e2) Click on the Alphafold model. Is the variant in a region where we can trust the structure? Click on some whitespace around the model to zoom out. Can you see regions where we had have less confidence about the structure?
a3) Is there anything special about the region containing the variant?
e4) how many different proteins interact with PTEN at the region containing our variant?



f) Filtered Download

All of the information from ProtVar can be downloaded. Find the download button near the top of the page. We would like all the annotations so all of the tick filters can stay selected. Download and open the file in excel.



Using the downloaded spreadsheet:

f1) What is the subcellular location for the protein (PTEN) affected by the third variant?

f2) Which family does protein containing the the first variant belong to?

f3) When was the canonical isoform sequence and the entry last updated for these proteins?

g) ProtVar API

Another useful way to access the data is via the ProtVar API. Go back to the ProtVar homepage and then select the ProtVar REST API button (it is to the top right of the screen).

Functional annotations, co-located variants and structures can all be queried programmatically here. The page helps you to see how to build queries.

g1) Go to the “Individual Amino Acid Annotations” section and expand the functional annotations (first) row. Here you can enter O00571 accession and the position 347. Execute the query.

g2) Copy the request URL generated and paste into your browser address bar. You will now see all the annotation data for this variant in your browser. The curl can also be used in your terminal to retrieve the same information.

h) Further Exploration

I hope that you enjoyed the session and now know slightly more about how to use UProtVar to help you understand coding variation.

If there are any questions please ask.

If you have completed all the above tasks please feel free to explore your favourite variants in ProtVar or explore the examples.