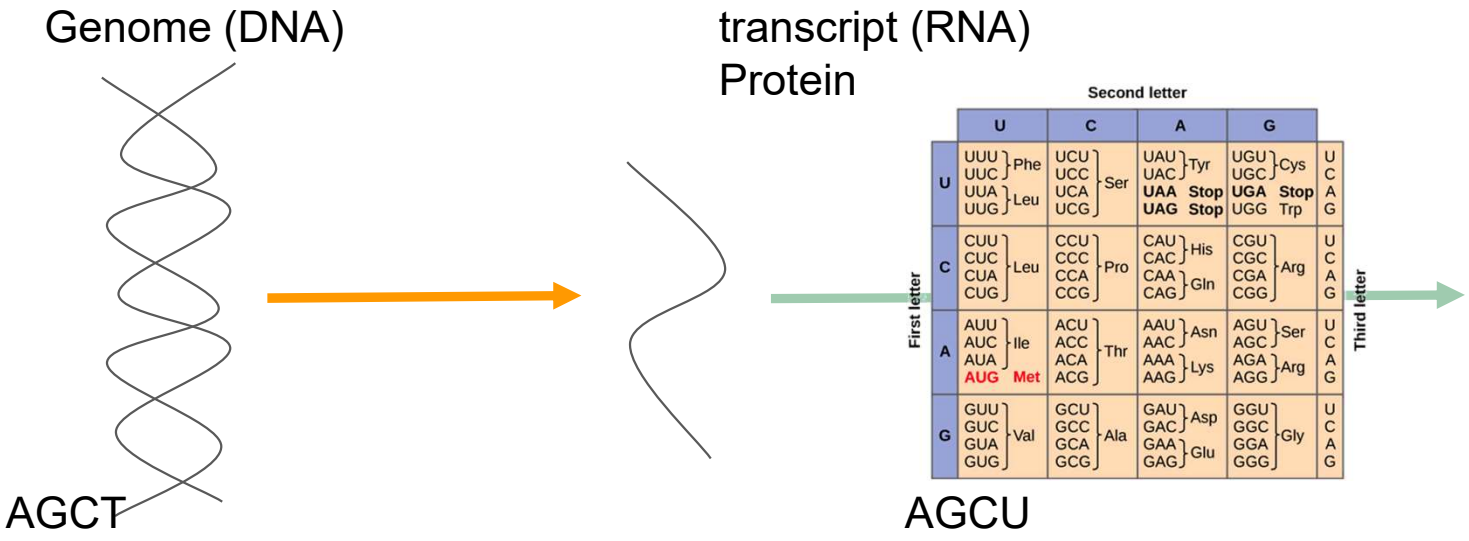
Contextualising Human Missense Variation

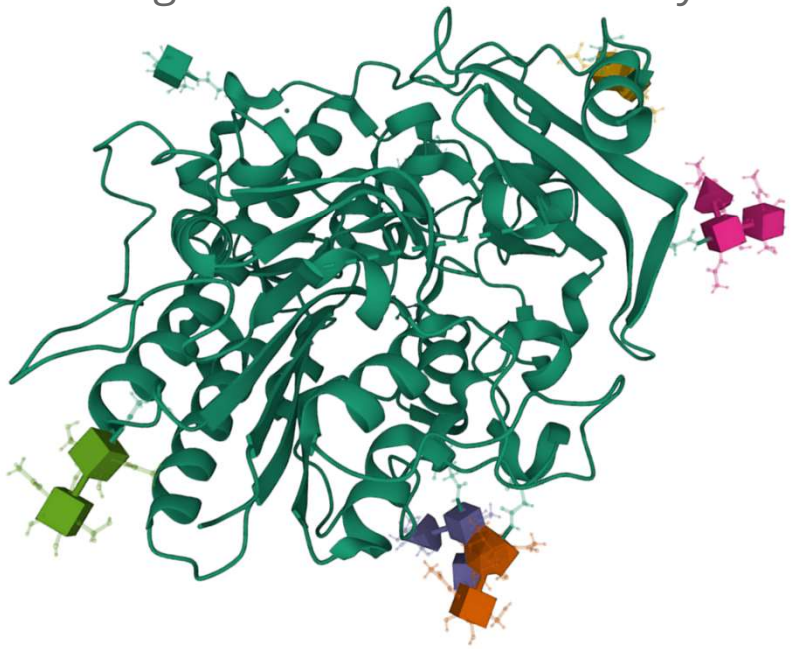
Genes, Proteins and Variation (abridged)

- **Genome**: All the genetic information for an organism in terms of the sequence of nucleotides.
- **Gene**: A sub-string of nucleotides which is transcribed. (a copy made of a short section of the genome). Some encode proteins.
- **Protein**: Large molecule made from a string of amino acids which do most of the work in a cell.
- **Variation**: the difference between a genetic sequence and the reference. This can be one nt. change (single nucleotide variant or SNV) or many. Sometimes this will affect the protein sequence.



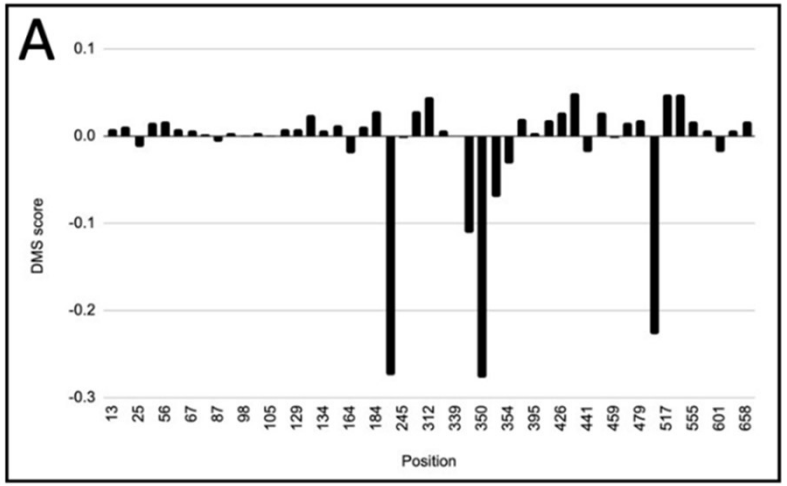
Missense Variation

- Protein function is defined by its shape and chemistry which in turn is dictated by sequence.
- SNVs which perturb AA sequence (missense) can alter their shape and change how they work.
- The challenge is that the effect is position and context dependent.
- Interpretation requires a broad range of data to be coherently assimilated.



Interpreting Missense Variation

1) **Does the variant significantly affect the protein function?**
    a) What is the reference function at each residue?
    b) How likely is the variant at that location to affect that role?



Alanine to Aspartic acid change effect across different positions in the same protein.

1) **Does that change significantly affect the organism?**
    a) What role did the reference protein play?
    b) How will a change affect the organism?

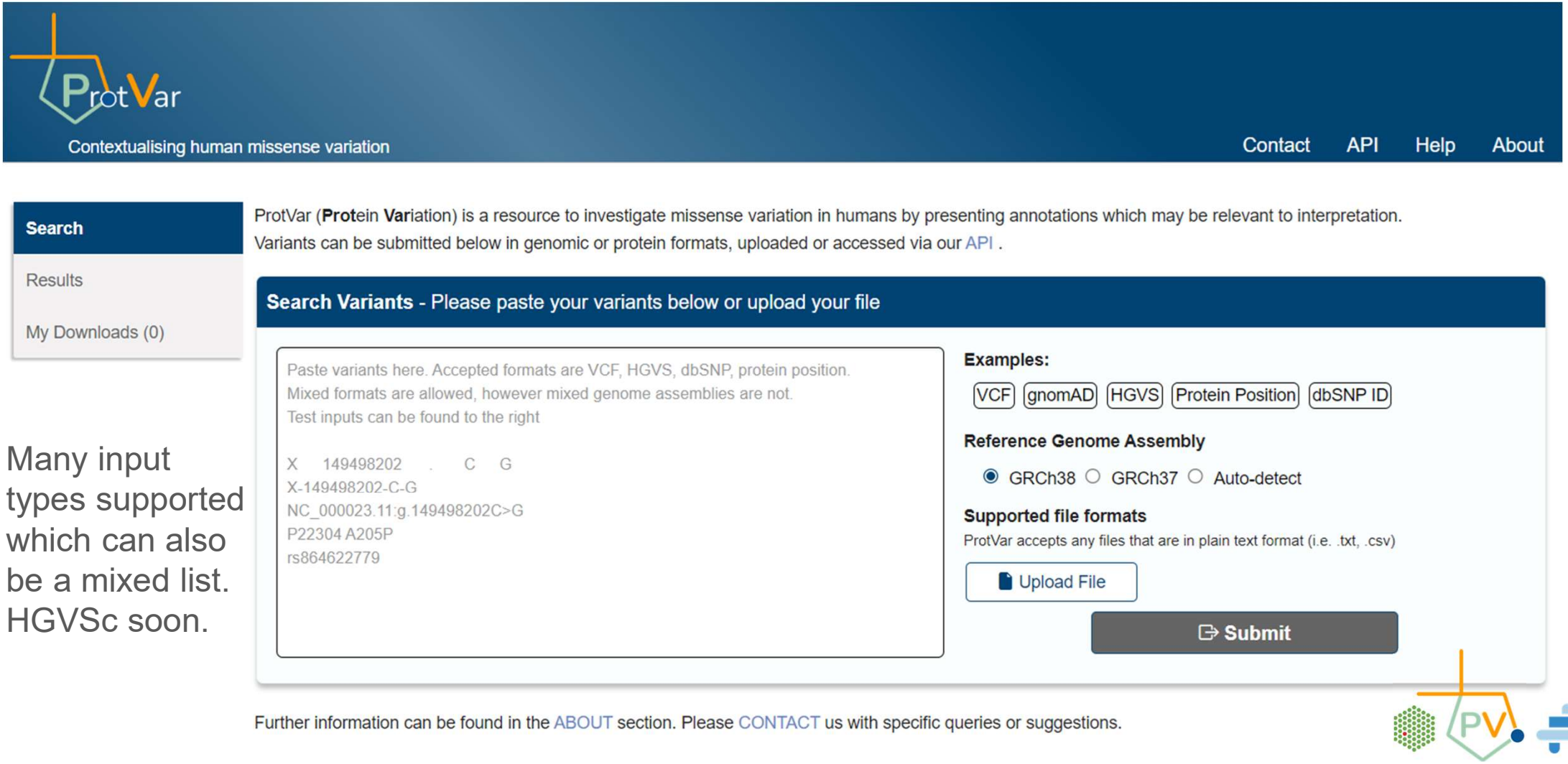
ProtVar

- **ProtVar** is hosted at EMBL-EBI and maintained by UniProt developers to help users to collect all the information they might need to assess how likely each variant is to affect the function of a protein and ultimately human health.

- It is free to everyone and is currently used by clinical geneticists, drug designers, synthetic biologists, protein researchers....

ebi.ac.uk/protvar

Home Page



Many input types supported which can also be a mixed list. HGVSc soon.

Results Page

ProtVar

Contextualising human missense variation

Contact API Help About

Search

Results

My Downloads (0)

〈 Prev 1 / 1 Next 〉 25

View Legends | Download Results

GENOMIC								PROTEIN					ANNOTATIONS
Chr.	Coordinate	ID	Ref.	Alt.	Gene	Codon (strand)	CADD	Isoform	Protein name	AA pos.	AA change	Consequence(s)	EVE	Click for details
X	149498202	.	C	G	IDS	Gcc/Ccc (-)	26.0	can P22304	Iduronate 2-sulfatas...	205	Ala/Pro	missense		
10	43118436	.	A	C	RET	aAc/aCc (+)	24.8	can P07949	Proto-oncogene tyros...	783	Asn/Thr	missense		
2	233760498	.	G	A	UGT1A1	Gga/Aga (+)	23.6	can P22309	UDP-glucuronosyltran...	71	Gly/Arg	missense		
14	89993420	.	A	G	TDP1	cAu/cGu (+)	26.1	can Q9NUW8	Tyrosyl-DNA phosphod...	493	His/Arg	missense		

〈 Prev 1 / 1 Next 〉 25

All mapping standardised regardless of user input type

Function Annotations

Reference Function

Variant Residue Position	Region Containing Variant Position

Alanine Proline

Conservation score: 0.918

Foldx prediction

Curated observations from UniProt

Chain-Iduronate 2-sulfatase 42 kDa chain

Helix

Predictions

(Source: PubMed ID 15980494)

Protein-protein interfaces containing variant

General Protein Function (not specific to the variant) from UniProt

Lysosomal enzyme involved in the degradation pathway of dermatan sulfate and heparan sulfate

PubMed : 10838181 | 11731225 | 28593992

Further General Protein Information (not specific to the variant) from UniProt

Recommended name: Iduronate 2-sulfatase
Alternative name: Alpha-L-iduronate sulfate sulfatase
Gene Name : IDS
Synonyms: SIDS
UniProtKB entry name: IDS_HUMAN
Protein evidence: Evidence at protein level
Entry last updated: 2023-05-03
Sequence modified: 1991-08-01
Sequence length: 550

Catalytic Activity

Complex

Subcellular Location

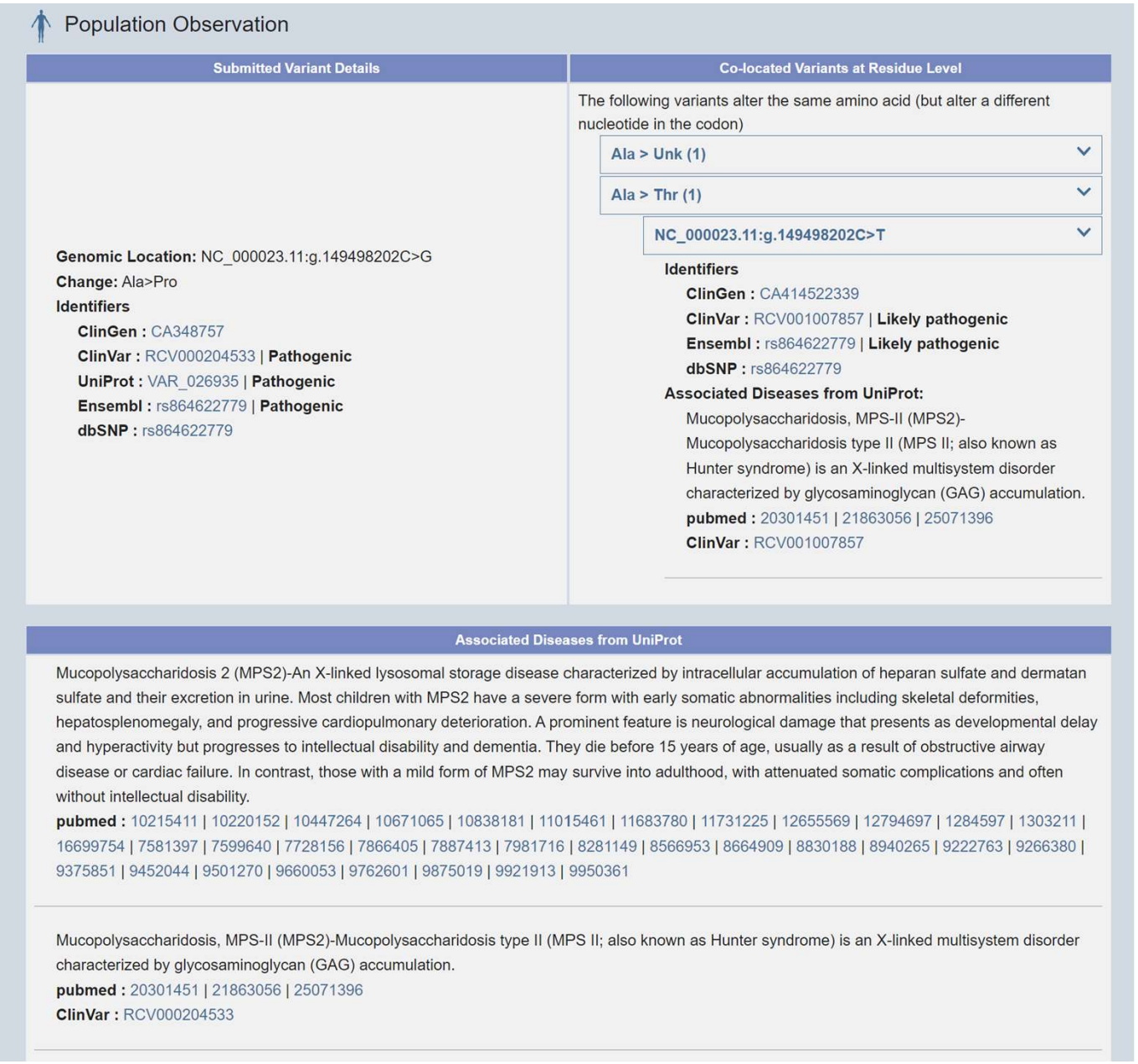
PTM's

Family

Interactions

Ensembl Gene ID	Translated Sequence and Transcript IDs for the Canonical Isoform
ENSG00000010404.18	ENSP00000339801.6 - ENST00000340855.11

→ 183M precomputed variant free energy changes precalculated



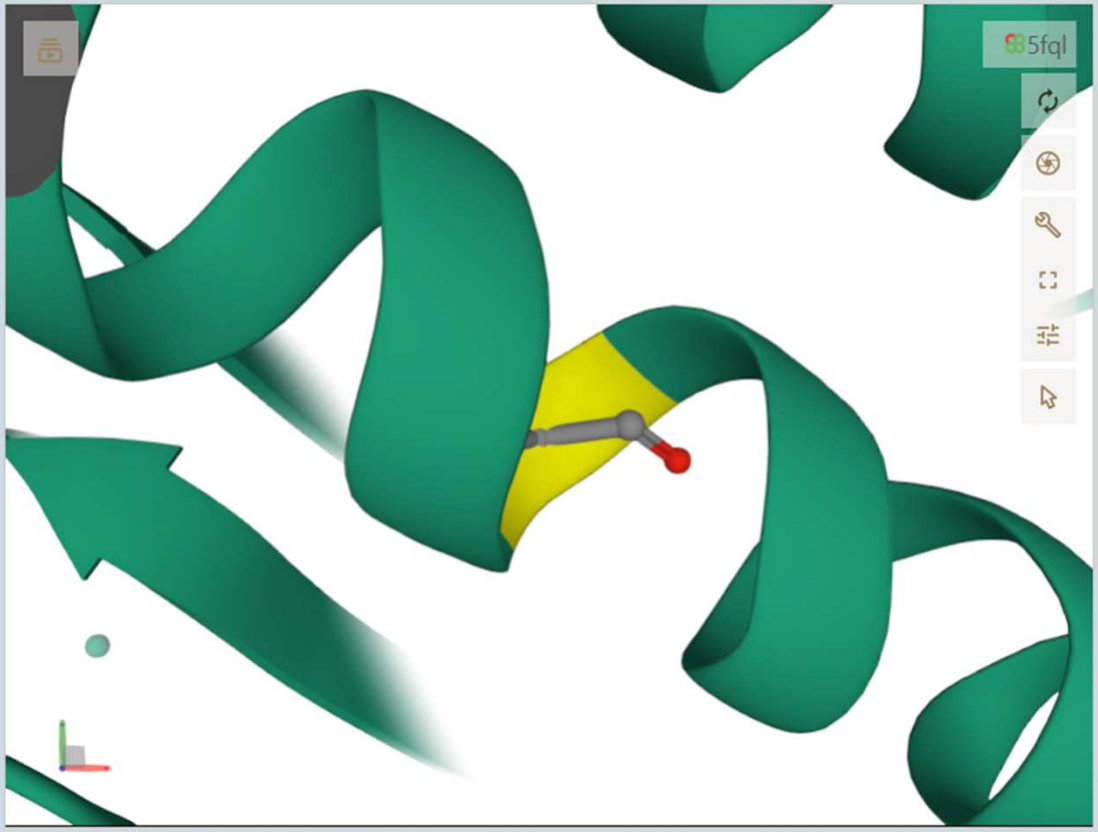
Population Observations

- Co-located at the amino acid level, not necessarily genomic position

Structure Annotations

Structures

Image zoomed to show variant location in green
Reference amino acid shown



Click variant to see surrounding residues
Click white space to zoom out to whole structure

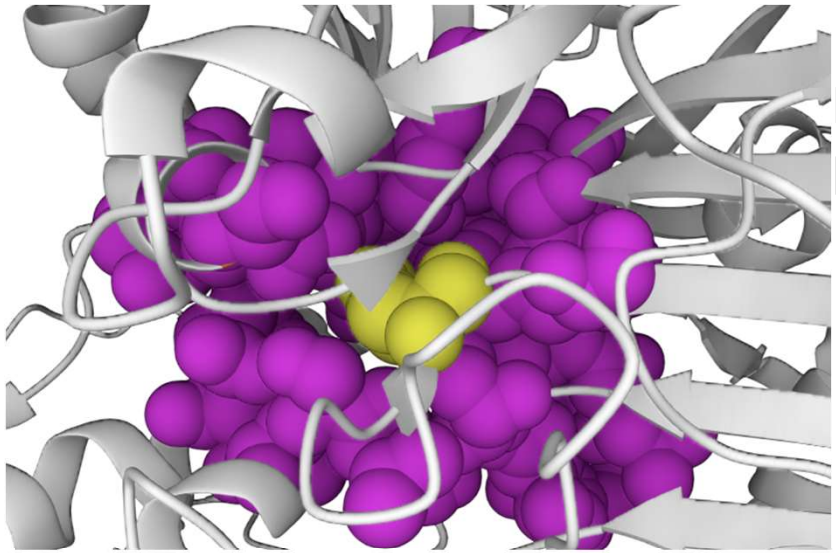
Further information from PDBeKB

Experimental Structure - PDBe				
PDB ID	Chain	PDB pos.	Resolution (Å)	Method
5fql	A	180	2.3	X-ray diffraction
6ioz	A	172	3.1	X-ray diffraction

Predicted Structure			
Source	Identifier	Position	Pockets
AlphaFold	AF-P22304-F1	205	N/A

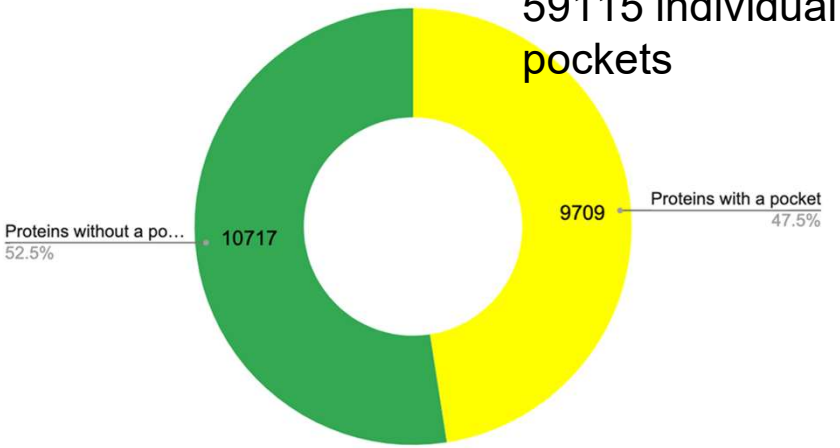
Predicted Interacting Structure				
A	Residues	B	Residues	pDockQ
P22304	95,98-101,10...	Q9BWS9-3	86-89,112-11...	0.601
P22304	106-118,121,...	P35475	184-186,224-...	0.601
P22304	124,129-134,...	Q9NWM8	33-40,49,51,...	0.445
O14657	49-58,129,13...	P22304	124,127-133,...	0.300

Predicted Pockets and Interfaces

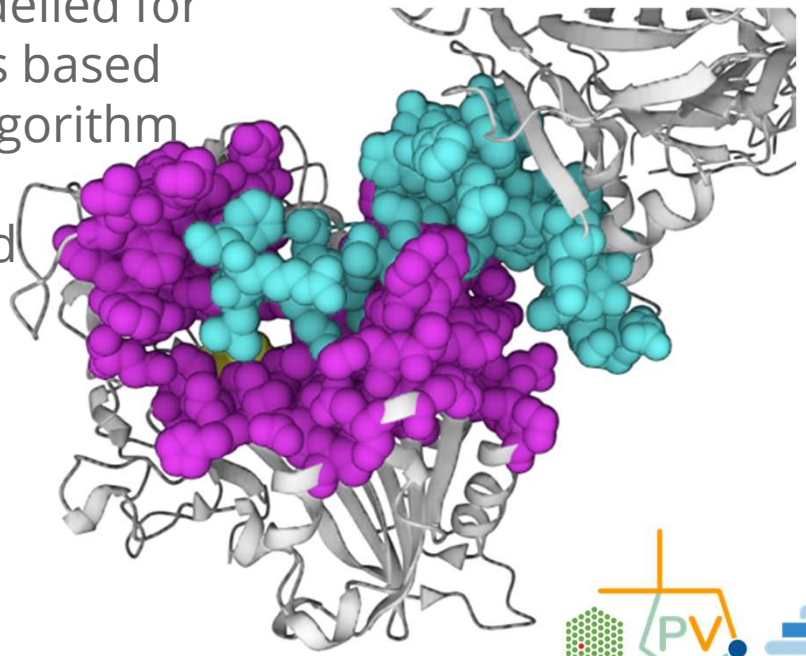


Protein pockets - predicted across all human proteins based based on Alphafold2 models

Predicted human pockets

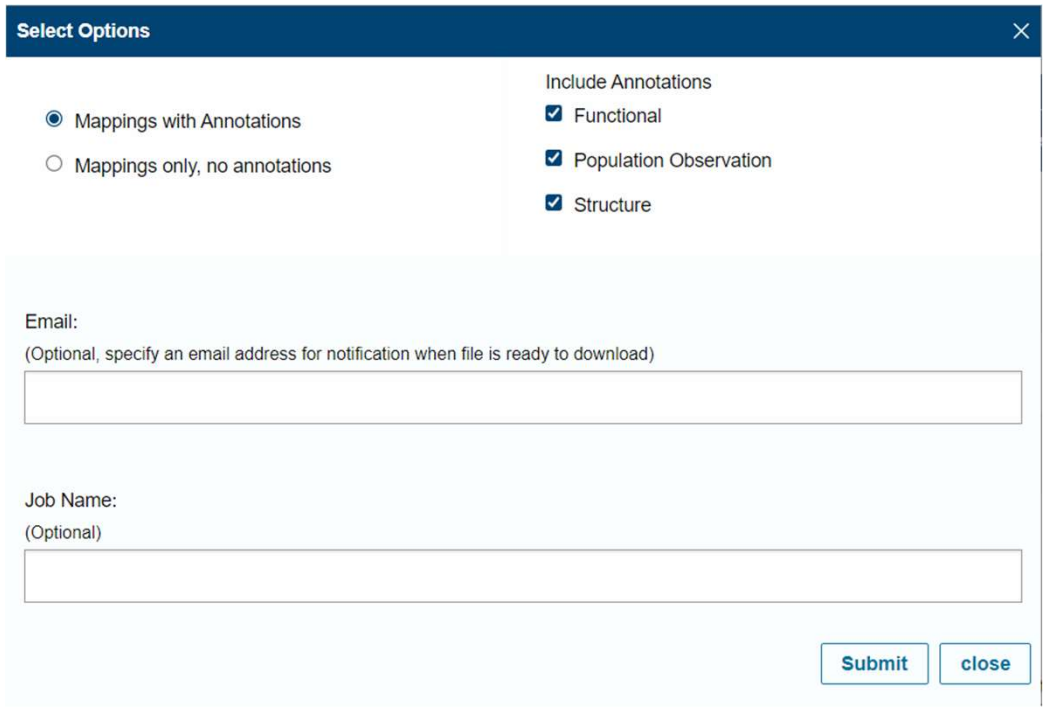


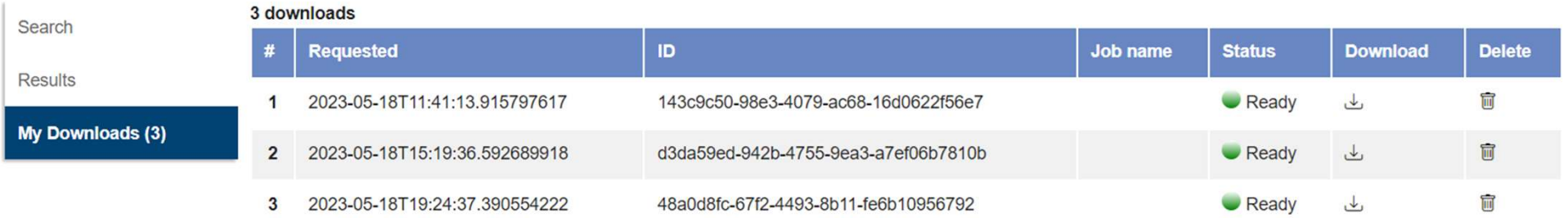
Protein-Protein Interfaces - Modelled for 500k interactions based on AF2 folding algorithm of two chains concurrently and docking score minimisation



Downloads

Results emailed to user and stored in their download space

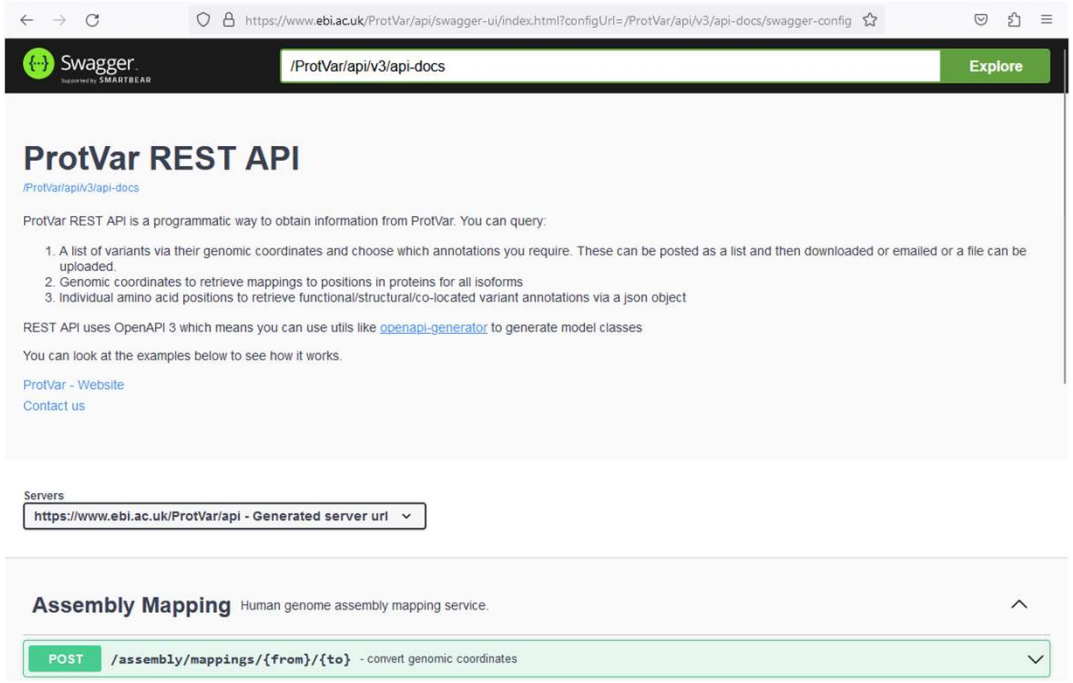




ProtVar Access

➔ REST API

https://www.ebi.ac.uk/ProtVar/api/swagger-ui/index.html?configUrl=/ProtVar/api/v3/api-docs/swagger-config



https://www.ebi.ac.uk/ProtVar/api/function/O00571/347

➔ Direct Link

E.g. Using genomic coordinates, protein position and/or IDs

https://www.ebi.ac.uk/ProtVar/query?search=rs864622779,P07949 asn783thr,2 233760498 G A

ProtVar Summary

- Fast and Flexible access and retrieval with high coverage
- Broad range of annotations and predictions
- Co-located variants at the AA level
- Interactive visualisations of experimental and predicted structures
- Programmatic access
- Maintained, updated and actively developed

ProtVar aims to offer an intuitive and efficient resource for the interpretation of missense variation and we hope it can help you in your work

ProtVar is developed by:

- James Stephenson
  Prabhat Totoo

ProtVar

UniProt

(protein function development)

EMBL-EBI

- Funded by

EMBL

European Molecular Biology Laboratory

Open Targets