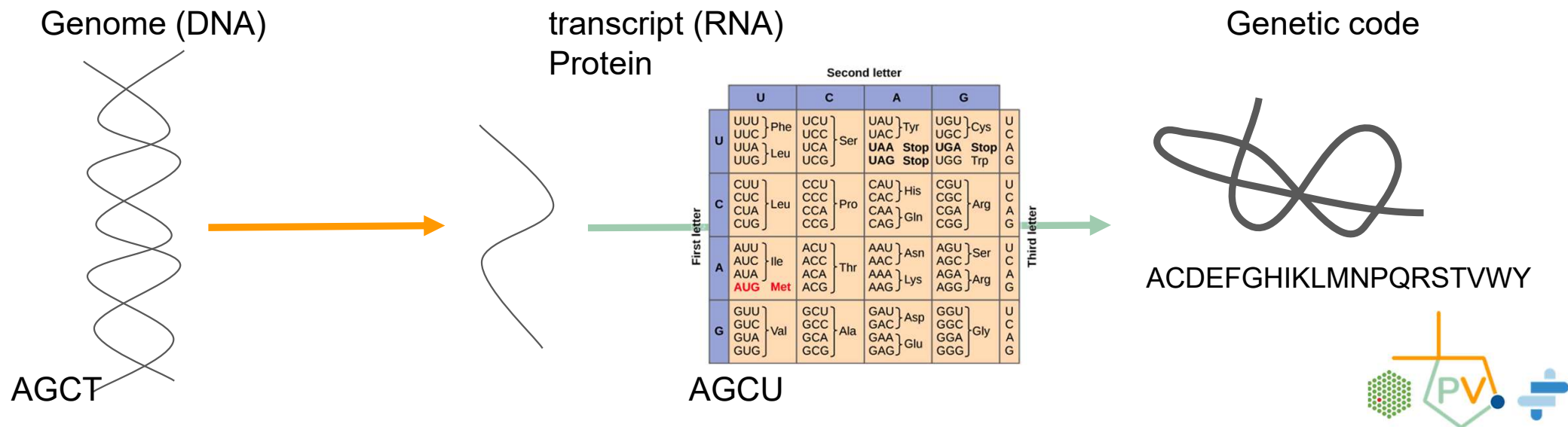




Contextualising Human Missense Variation

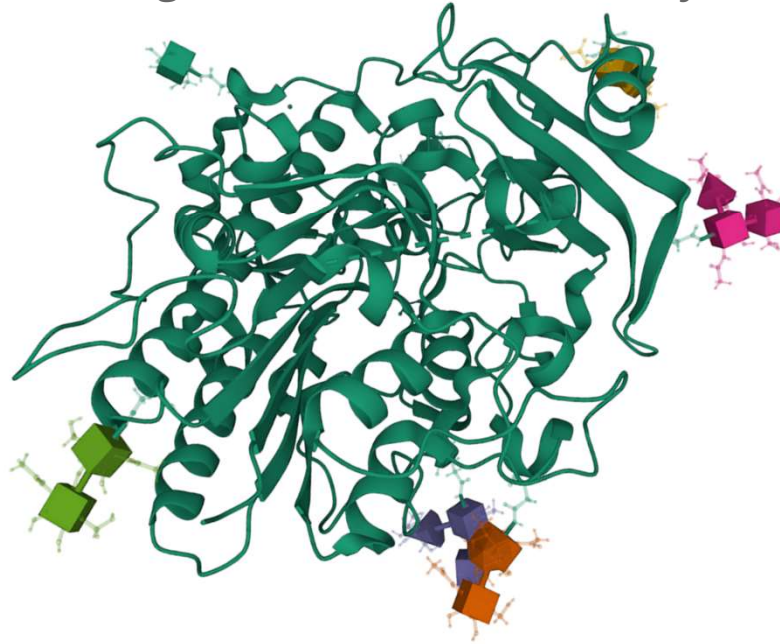
Genes, Proteins and Variation (abridged)

- **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

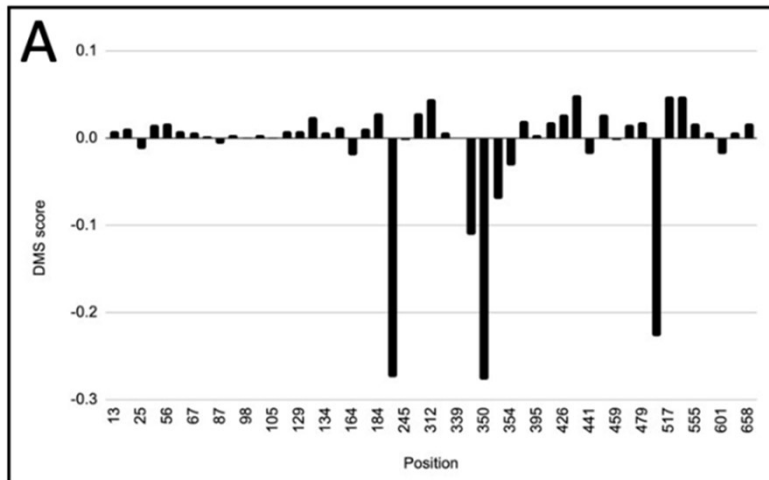
- Protein function is defined by its shape and chemistry which in turn is dictated by sequence.
- SNVs which perturb AA sequence (missense) can alter their shape and change how they work.
- The challenge is that the effect is position and context dependent.
- Interpretation requires a broad range of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



Contextualising human missense variation

[Contact](#) [API](#) [Help](#) [About](#)

Search

Results

My Downloads (0)

ProtVar (**Protein Variation**) is a resource to investigate missense variation in humans by presenting annotations which may be relevant to interpretation. Variants can be submitted below in genomic or protein formats, uploaded or accessed via our [API](#).

Search Variants - Please paste your variants below or upload your file

Paste variants here. Accepted formats are VCF, HGVS, dbSNP, protein position.
Mixed formats are allowed, however mixed genome assemblies are not.
Test inputs can be found to the right

X 149498202 . C G
X-149498202-C-G
NC_000023.11:g.149498202C>G
P22304 A205P
rs864622779

Examples:

[VCF](#) [gnomAD](#) [HGVS](#) [Protein Position](#) [dbSNP ID](#)

Reference Genome Assembly

☒ GRCh38 ☐ GRCh37 ☐ Auto-detect

Supported file formats

ProtVar accepts any files that are in plain text format (i.e. .txt, .csv)

 [Upload File](#)

 [Submit](#)

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

Further information can be found in the [ABOUT](#) section. Please [CONTACT](#) us with specific queries or suggestions.



Results Page



Contextualising human missense variation

[Contact](#) [API](#) [Help](#) [About](#)

Search

[< Prev](#) 1 / 1 [Next >](#)

25

[View Legends](#)

[Download Results](#)

Results

My Downloads (0)


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	<div></div>	<div><div></div><div></div><div></div></div>	
10	43118436	.	A	C	RET	aAc/aCc (+)	24.8	can P07949 ✓	Proto-oncogene tyros...	783	Asn/Thr	missense	<div></div>	<div><div></div><div></div><div></div></div>	
2	233760498	.	G	A	UGT1A1	Gga/Aga (+)	23.6	can P22309 ✓	UDP-glucuronosyltran...	71	Gly/Arg	missense	<div></div>	<div><div></div><div></div><div></div></div>	
14	89993420	.	A	G	TDP1	cAu/cGu (+)	26.1	can Q9NUW8 ✓	Tyrosyl-DNA phosphod...	493	His/Arg	missense	<div></div>	<div><div></div><div></div><div></div></div>	

[< Prev](#) 1 / 1 [Next >](#)

25



All mapping standardised regardless of user input type

Reference Function	
Variant Residue Position	Region Containing Variant Position
 <p>Alanine → Proline</p> <p>Conservation score: 0.918</p> <p>Foldx prediction</p>	<p>Curated observations from UniProt</p> <p>Chain-Iduronate 2-sulfatase 42 kDa chain</p> <p>Helix</p> <p>Predictions (Source: PubMed ID 15980494)</p> <p>Protein-protein interfaces containing variant</p>
General Protein Function (not specific to the variant) from UniProt	
<p>Lysosomal enzyme involved in the degradation pathway of dermatan sulfate and heparan sulfate</p> <p>PubMed : 10838181 11731225 28593992</p>	
Further General Protein Information (not specific to the variant) from UniProt	
<p>Recommended name: Iduronate 2-sulfatase</p> <p>Alternative name: Alpha-L-iduronate sulfate sulfatase</p> <p>Gene Name : IDS</p> <p>Synonyms: SIDS</p> <p>UniProtKB entry name: IDS_HUMAN</p> <p>Protein evidence: Evidence at protein level</p> <p>Entry last updated: 2023-05-03</p> <p>Sequence modified: 1991-08-01</p> <p>Sequence length: 550</p>	<p>Catalytic Activity</p> <p>Complex</p> <p>Subcellular Location</p> <p>PTM's</p> <p>Family</p> <p>Interactions</p>
Ensembl Gene ID	Translated Sequence and Transcript IDs for the Canonical Isoform
ENSG00000010404.18	ENSP000000339801.6 - ENST000000340855.11

Function Annotations

→ 183M precomputed variant free energy changes precalculated



Submitted Variant Details	Co-located Variants at Residue Level
<p>Genomic Location: NC_000023.11:g.149498202C>G</p> <p>Change: Ala>Pro</p> <p>Identifiers</p> <p>ClinGen : CA348757</p> <p>ClinVar : RCV000204533 Pathogenic</p> <p>UniProt : VAR_026935 Pathogenic</p> <p>Ensembl : rs864622779 Pathogenic</p> <p>dbSNP : rs864622779</p>	<p>The following variants alter the same amino acid (but alter a different nucleotide in the codon)</p> <p>Ala > Unk (1) ▼</p> <p>Ala > Thr (1) ▼</p> <p>NC_000023.11:g.149498202C>T ▼</p> <p>Identifiers</p> <p>ClinGen : CA414522339</p> <p>ClinVar : RCV001007857 Likely pathogenic</p> <p>Ensembl : rs864622779 Likely pathogenic</p> <p>dbSNP : rs864622779</p> <p>Associated Diseases from UniProt:</p> <p>Mucopolysaccharidosis, MPS-II (MPS2)-Mucopolysaccharidosis type II (MPS II; also known as Hunter syndrome) is an X-linked multisystem disorder characterized by glycosaminoglycan (GAG) accumulation.</p> <p>pubmed : 20301451 21863056 25071396</p> <p>ClinVar : RCV001007857</p>

Associated Diseases from UniProt
<p>Mucopolysaccharidosis 2 (MPS2)-An X-linked lysosomal storage disease characterized by intracellular accumulation of heparan sulfate and dermatan sulfate and their excretion in urine. Most children with MPS2 have a severe form with early somatic abnormalities including skeletal deformities, hepatosplenomegaly, and progressive cardiopulmonary deterioration. A prominent feature is neurological damage that presents as developmental delay and hyperactivity but progresses to intellectual disability and dementia. They die before 15 years of age, usually as a result of obstructive airway disease or cardiac failure. In contrast, those with a mild form of MPS2 may survive into adulthood, with attenuated somatic complications and often without intellectual disability.</p> <p>pubmed : 10215411 10220152 10447264 10671065 10838181 11015461 11683780 11731225 12655569 12794697 1284597 1303211 16699754 7581397 7599640 7728156 7866405 7887413 7981716 8281149 8566953 8664909 8830188 8940265 9222763 9266380 9375851 9452044 9501270 9660053 9762601 9875019 9921913 9950361</p>
<p>Mucopolysaccharidosis, MPS-II (MPS2)-Mucopolysaccharidosis type II (MPS II; also known as Hunter syndrome) is an X-linked multisystem disorder characterized by glycosaminoglycan (GAG) accumulation.</p> <p>pubmed : 20301451 21863056 25071396</p> <p>ClinVar : RCV000204533</p>

Population Observations

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

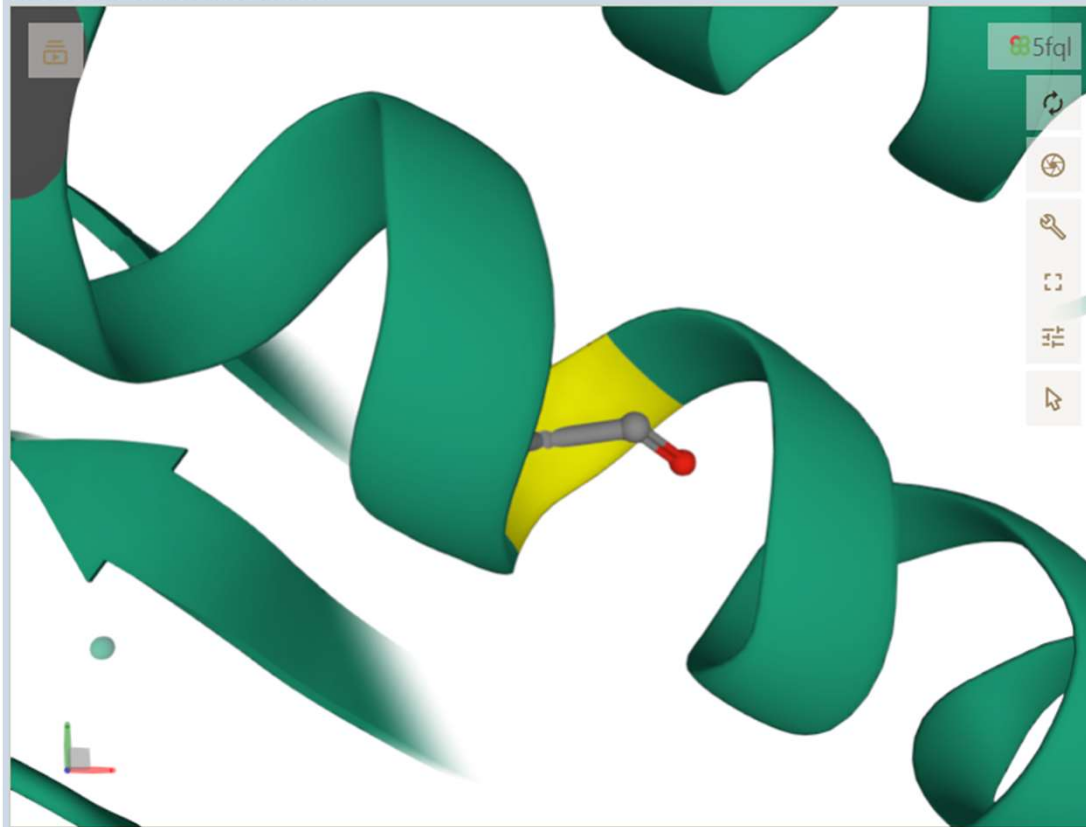


Structure Annotations

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

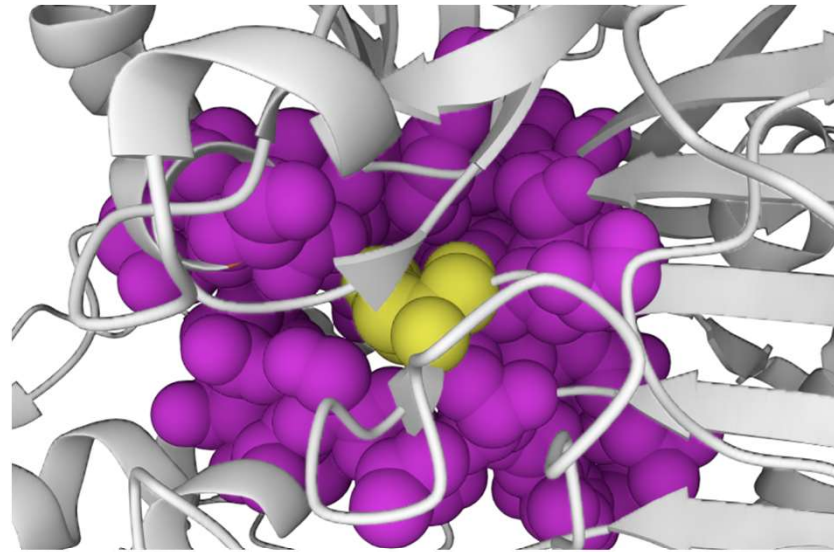
Source	Identifier	Position	Pockets
AlphaFold	AF-P22304-F1	205	N/A

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

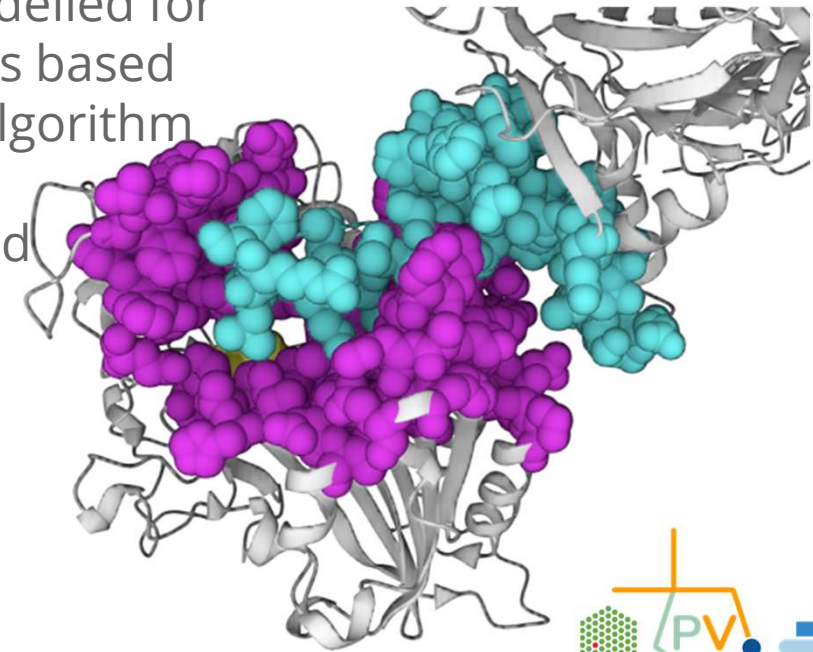


Predicted Pockets and Interfaces

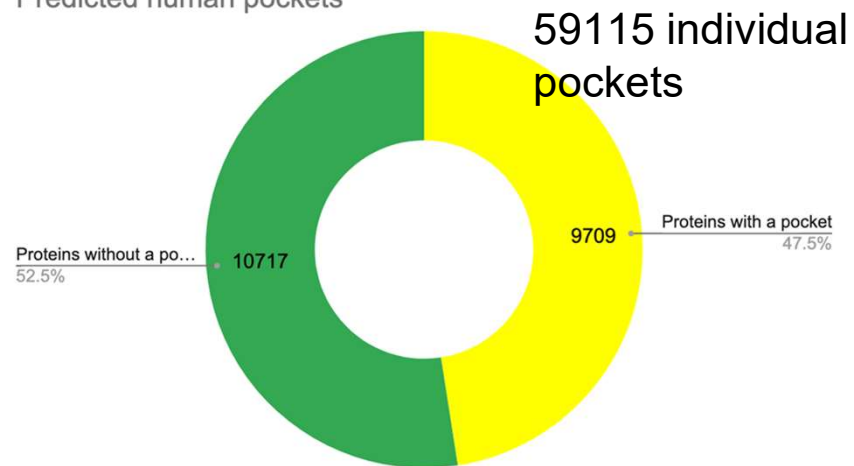


Protein pockets - predicted across all human proteins based on AlphaFold2 models

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



Predicted human pockets



Downloads

Results emailed
to user and stored
in their download
space

Select Options

☒ Mappings with Annotations

☐ Mappings only, no annotations

Include Annotations

☒ Functional

☒ Population Observation

☒ Structure

Email:
(Optional, specify an email address for notification when file is ready to download)

Job Name:
(Optional)

Submitclose

Search

Results

My Downloads (3)

3 downloads

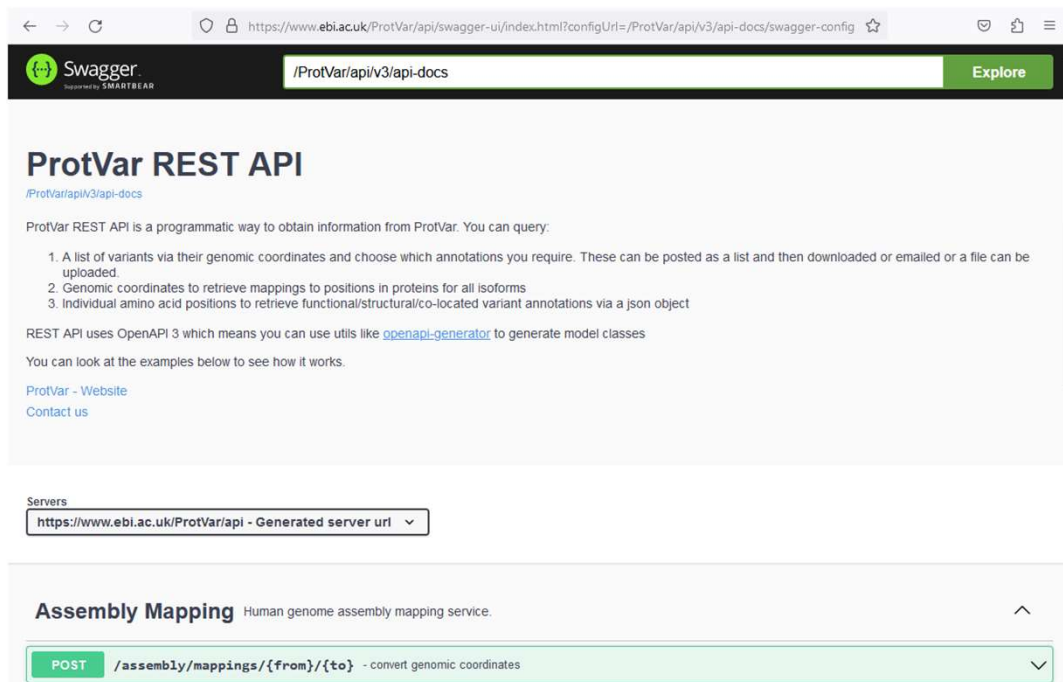
#	Requested	ID	Job name	Status	Download	Delete
1	2023-05-18T11:41:13.915797617	143c9c50-98e3-4079-ac68-16d0622f56e7		Ready	Download	Delete
2	2023-05-18T15:19:36.592689918	d3da59ed-942b-4755-9ea3-a7ef06b7810b		Ready	Download	Delete
3	2023-05-18T19:24:37.390554222	48a0d8fc-67f2-4493-8b11-fe6b10956792		Ready	Download	Delete



ProtVar Access

→ REST API

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



Swagger
/ProtVar/api/v3/api-docs Explore

ProtVar REST API

/ProtVar/api/v3/api-docs

ProtVar REST API is a programmatic way to obtain information from ProtVar. You can query:

1. A list of variants via their genomic coordinates and choose which annotations you require. These can be posted as a list and then downloaded or emailed or a file can be uploaded.
2. Genomic coordinates to retrieve mappings to positions in proteins for all isoforms
3. Individual amino acid positions to retrieve functional/structural/co-located variant annotations via a json object

REST API uses OpenAPI 3 which means you can use utils like [openapi-generator](#) to generate model classes

You can look at the examples below to see how it works.

[ProtVar - Website](#)
[Contact us](#)

Servers
https://www.ebi.ac.uk/ProtVar/api - Generated server url

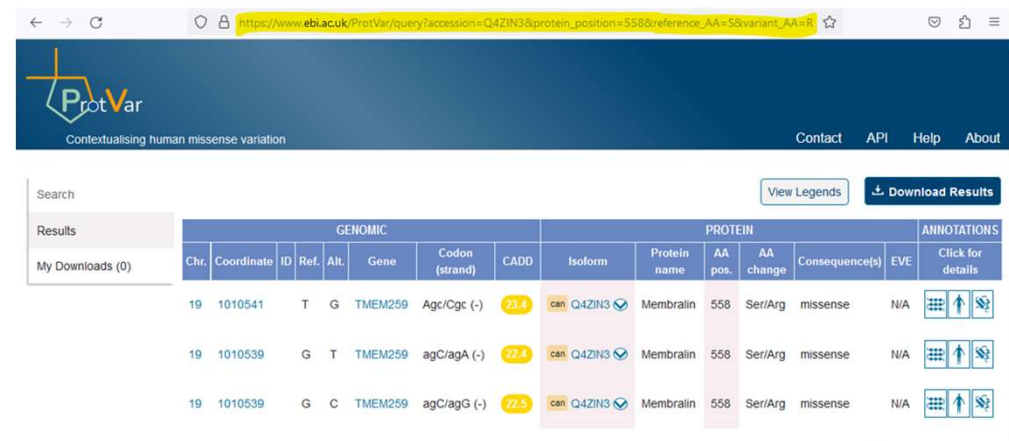
Assembly Mapping Human genome assembly mapping service.

POST /assembly/mappings/{from}/{to} - convert genomic coordinates

→ Direct Link

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

<https://www.ebi.ac.uk/ProtVar/query?search=rs864622779,P07949 asn783thr,2233760498 G A>












ProtVar
Contextualising human missense variation

Search

Results

My Downloads (0)

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
19	1010541	T	G		TMEM259	Agc/Cgc (-)	23.4	can Q4ZIN3	Membralin	558	Ser/Arg	missense	N/A	  
19	1010539	G	T		TMEM259	agC/agA (-)	22.4	can Q4ZIN3	Membralin	558	Ser/Arg	missense	N/A	  
19	1010539	G	C		TMEM259	agC/agG (-)	22.5	can Q4ZIN3	Membralin	558	Ser/Arg	missense	N/A	  

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



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



(protein function development)



→ Funded by

