Uniprot: enabling interpretation of protein variation effects

Where does Uniprot variant data come from?

Variant data from literature

Uniprot variant data is reflected in the UniprotKB database and on other online resources. Variants are captured from the scientific literature and manually reviewed for addition to UniprotKB/Swiss-Prot.

Large-scale variant data

Large-scale variant data is imported from a variety of resources to complement the set of variants captured from the literature.

How does Uniprot help with variant effect interpretation?

1. View variants in the context of other protein sequence data

The Uniprot feature viewer, ProtView, provides a graphical view of protein variants in the context of other sequence data such as domains, active sites and post-translational modifications. This helps to identify possible reasons for variant effects by highlighting the localisation of variants with important residues. The description of disease associated with genetic variations in a protein and links to variant resources can be found here: http://www.uniprot.org/downloads.

2. Integrate Uniprot data into genome browsers

Uniprot produces genome browser tracks which allow users to integrate Uniprot sequence information including protein variants into genome browsers such as Ensembl and UCSC browsers. You can couple Uniprot data with other available genomic information and also with your own sequencing data. This makes it easier to see how changes in the genome can contribute to altered protein function and lead to disruptive disease phenotypes.

3. Access data programatically via the Proteins API

The Proteins API provides programmatic access to protein and associated genetics data such as curated protein sequence positional annotations from UniprotKB as well as mapped variation and proteomics data from large-scale data sources.

Example use case

• Retrieve all UniprotKB records which have variants associated with Wilson disease, an autosomal recessive disorder of copper metabolism. This returns a single record, copper-transporting ATPase 2 from the ATP7B gene.

• Retrieve sequence annotations (features) to determine if any of the variants align to functionally important residues such as binding sites or active sites.

• The protein contains copper-binding sites and an active site and there are a number of variants from both reviewed and large-scale data, including variants which disrupt copper-binding sites as well as a variant from COSMIC which disrupts the active site.

• Unique proteomics peptides are found with the protein which can be used to identify it in mass spectrometry experiments.

Coming soon - PepVEP

Protein Variant Effect Predictor

Platform for interpreting the functional effects of genomic variants by integrating information available in resources such as Ensembl, Uniprot and PDB. It will combine genomic and protein data to define potential functional consequences of variants for a protein.

Learn more about UniProt

Read the Uniprot blog
http://insideuniprot.blogspot.co.uk/

Follow Uniprot on Twitter
@uniprot

Visit the Uniprot Facebook page

Watch videos on the Uniprot YouTube channel

Visit the EBI Training portal http://www.ebi.ac.uk/training

Funding

UniProt is funded by National Institutes of Health, European Molecular Biology Laboratory, Swiss Federal Government, British Heart Foundation, Parkinson's Disease United Kingdom and National Science Foundation