Enabling interpretation of protein variation effects with UniProt

Introduction
Understanding the effect of genetic variants on protein function is crucial to thoroughly understand the role of proteins in disease biology. UniProt aims to support the scientific community, computational biologists and clinical researchers, by providing a comprehensive, high-quality and freely accessible resource of protein sequence and functional information. This includes a comprehensive catalogue of protein altering variation data coupled with information about how these variants affect protein function.

UniProt variant data sources

1. Variants in the context of other protein functional annotation data

The UniProt feature viewer, ProtVista, provides a graphical view of protein variants in the context of other functional annotations such as domains, active sites and post-translational modifications. Possible variant effects can be identified by investigating co-localised protein functional residues.

2. Integrate UniProt data into genome browsers

UniProt produces genome browser tracks which allow users to integrate UniProt sequence information including UniProt reviewed protein variants into genome browsers such as the 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 programmatically via the Proteins API

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

Interpretation protein variant effect with UniProt

1. Variants in the context of other protein functional annotation data

The UniProt feature viewer, ProtVista, provides a graphical view of protein variants in the context of other functional annotations such as domains, active sites and post-translational modifications. Possible variant effects can be identified by investigating co-localised protein functional residues.

2. Integrate UniProt data into genome browsers

UniProt produces genome browser tracks which allow users to integrate UniProt sequence information including UniProt reviewed protein variants into genome browsers such as the 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 programmatically via the Proteins API

The Proteins API provides programmatic access to protein and associated genomics 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

1. Use the ‘Variants’ end-point with the disease filter set to ‘Wilson’ to retrieve all UniProt variants that are annotated with Wilson disease and have associated variants. This returns a single record, copper transporting ATPase 2 from the ATP7B gene.
2. Using Wilson disease variants protein positions, retrieve sequence annotations via features end-point and determine if any of the variants align to functionally important residues such as binding sites or active sites.
3. The protein contains copper-binding sites and an active site; 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.
4. 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
A Platform for interpreting protein, structure and clinical information with genomic variants effect prediction from Ensembl variation.

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.