UniProt at EMBL-EBI's role in CTTV: contributing to improved disease knowledge

Introduction

The mission of UniProt is to provide the scientific community with a comprehensive, high quality and freely accessible resource of protein sequence and functional information. The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. As much annotation information as possible is added to each UniProtKB record and this includes widely accepted biological ontologies, classifications and cross-references, and clear indications of the quality of annotation in the form of evidence attribution of experimental and computational data.

UniProt's disease expert curation

This section provides information on the disease(s) associated with genetic variations in a given protein. The information is extracted from the scientific literature and diseases that are also described in the OMIM database are represented with a controlled vocabulary.

If known, we describe the amino acid change, the abbreviation of the associated disease and the effect(s) of the variation on the protein.

UniProt's contribution to CTTV

UniProt curators at EMBL-EBI affiliated with the CTTV project are currently involved in updating the UniProtKB entries containing target proteins associated with two inflammatory bowel disease (IBD) conditions: Crohn’s disease and ulcerative colitis (Fig.1). Curation efforts aim to provide the most up-to-date experimental evidence for the protein’s involvement in the disease, including single amino acid substitutions causing the disease or susceptibility to the disease, and any other additional protein-related information that may be significant.

With every UniProt release the newly added information is made available to the public and shared with the CTTV platform.

UniProtKB’s disease and natural variant information contributes to the genetic associations score in CTTV. Other information from UniProtKB can also be found in the target profile page on the CTTV website (Fig.2).

CTTV’s integration of multiple resources, including UniProt, is possible thanks to mapping of the disease and phenotype terms from each source to the Experimental Factor Ontology (EFO) (Fig.3).

UniProt curators at EMBL-EBI have contributed to the creation of the disease associations by manually mapping over 1000 rare diseases to EFO.