UniProt-GOA: Quick tour

Rachael Huntley [1]

- Ontologies
- Beginner
- 0.5 hour

This quick tour provides a brief introduction to UniProt-GOA, the Gene Ontology Annotation resource. For information about the Gene Ontology, see the GO quick tour [2].

Learning objectives:

- Basic understanding of UniProt-GOA and how it can help you to find out more about your protein data
- Know where to find out more about UniProt-GOA

What is UniProt-GOA?

Why do we need Gene Ontology annotation?

As proteomics [3] research gains momentum, biologists need new ways to access and analyse information on proteins. To exploit the potential of these data fully, we need to capture all the available biological information related to each protein, including consistent descriptions of protein function. The process of adding extra biological information to the data is called annotation.

What is the UniProt [4]-Gene Ontology Annotation project?

The Gene Ontology (GO [5]) is a dynamic controlled vocabulary [6] that details the biological processes, molecular functions and cellular components of a generic cell. The UniProt-Gene Ontology Annotation (UniProt-GOA [7]) project uses GO to describe proteins in UniProtKB [8]. The project has assigned GO terms to all complete and incomplete proteomes that exist in UniProtKB, using a combination of manual and automatic annotation [9]. For more information about the Gene Ontology, see the GO quick tour [2].

Manual annotation [10] is carried out by curators who directly assign GO terms to proteins based on evidence from the scientific literature. GO associations that are determined manually are allocated an evidence code [11] that describes the evidence from the literature that supports the annotation. More information about our manual annotation procedure can be found on the UniProt-GOA website [12].

Automatic annotation is a rapid way of assigning GO terms to gene products on a large scale, and the UniProt-GOA project is the main producer of automatic annotations in the GO Consortium [13]. Details of the different automatic annotation pipelines can be found on the UniProt-GOA website [14].
UniProt-GOA is updated on a monthly basis, in accordance with the latest data released by UniProtKB, Ensembl [15], Ensembl Genomes [16] and InterPro [17] (Figure 1). GO annotations are also imported from other members of the GO Consortium and its collaborators.

By annotating all characterised proteins with GO terms and helping to transfer this knowledge to similar uncharacterised proteins, we hope to contribute to a better understanding of all proteomes. The success of GO can be measured by the number of databases that use it to annotate and exchange biological knowledge. The UniProt-GOA project has made an important contribution to this global effort.

Figure 1. Sources and flow of data for the UniProt-GOA resource. UniProt-GOA contains protein sequences from UniProtKB, which are annotated (both manually and automatically) using GO terms. Data can be downloaded as complete files or filtered annotation sets. Data can be searched using QuickGO. All data are exchanged with the GO Consortium.

What can I do with UniProt-GOA?

UniProt-GOA [18] allows you to:

- access functional information for proteins in UniProtKB [8] or Ensembl [15], either by
searching for individual proteins using the QuickGO [19] or Ensembl browsers, or by
downloading and parsing one of our gene association files;
• use GO Slims [20] to summarise the biological attributes of a proteome [21], compare
proteomes or, for example, find out what proportion of a proteome has been found to be
involved in apoptosis;
• incorporate GO annotations into your own database to enhance the functional information
available to your user community;
• use GO annotations to link between biological knowledge and high-throughput genomic or
proteomic datasets;
• generate automated GO annotations to new genomic or protein sequences, using the
InterProScan [22] tool;
• find the GO terms for genes of interest using Ensembl [23].

You can find out more about the applications of GO annotation [24] by looking at the studies listed
on the GO bibliography page [25].

Searching and visualising data from UniProt-GOA

Accessing the data in UniProt-GOA

Browsing and searching

The UniProt-GOA [18] website [7] features a QuickGO [19] search bar at the top of the page, which
allows you to search for GO terms or for UniProtKB [8] proteins and their associated GO annotations.
QuickGO is a web-based browser developed by the UniProt-GOA project (Figure 2).

QuickGO also allows you to view annotations using GO Slim [20], a customised subset of GO, which
can provide the user with an overview of the biological attributes for a list of proteins of interest. This
facility is accessible from the home page of QuickGO (Figure 2).

QuickGO has an array of filtering [26] options, accessible from the annotation [24] page, which allow
you to search for custom sets of annotations. For example, you can filter by taxon, GO term, GO
evidence code [27], or you can search for annotations to a list of protein accessions. Various
statistics are calculated for the annotation sets, which allow you to instantly view, for example, how
many proteins have been associated with a particular GO term, or how many annotations were
derived from experimental evidence.

Downloading annotation sets from QuickGO

Sets of annotations obtained using QuickGO can be downloaded to your computer in a variety of
formats, including: gene association files, FASTA [28] and protein lists. There is a 'Download' button
on the annotation page.
QuickGO is a fast web-based browser for Gene Ontology terms and annotations, which is provided by the UniProtKB-GOA group at the EBI.

Figure 2. The QuickGO web browser. For searching Gene Ontology [29] terms and proteins from UniProtKB.

QuickGO term information

When viewing an individual GO term, all of the information associated with that particular term is displayed. This includes the definition, synonyms, child terms, etc. (Figure 3).
Figure 3. The QuickGO term information page displays the biological information related to individual GO terms, including the definitions and synonyms. This example is for GO term 'GO:0045787' ('positive regulation of cell cycle').

QuickGO ancestor chart

You can also view the GO term in the context of its related terms by using the ancestor chart view (Figure 4).

Figure 4. The QuickGO ancestor chart displays the GO term in the context of its related terms. In this example, 'positive regulation of cell cycle' is related to 'regulation of cell cycle', 'positive regulation of cellular processes', and so on.
QuickGO annotation

Additionally, you can search for GO annotations for a chosen individual or set of proteins (Figure 5).

Figure 5. QuickGO annotation page for a single protein, angiomotin (Q4VCS5). Lists the biological information associated with this protein, based on evidence from the scientific literature.

Getting data from UniProt-GOA

Downloading annotations

Every month, UniProt-GOA [18] produces gene association files containing all of the available GO annotations for UniProtKB [8] proteins. The 'UniProt' file contains GO annotations for proteins from all species in UniProtKB. There are also a number of species-specific files which contain GO annotations for certain species that have a complete proteome [21], such as human, mouse, Arabidopsis, yeast, zebrafish, etc.

The UniProt-GOA gene association files are available from the download [31] page on the UniProt-GOA website, and previous versions of the gene association files are also available from the UniProt-GOA ftp site [32].

Web services [33]

QuickGO can supply GO term information and GO annotation [24] data via REST [34] web services. Instructions on how to access these are on the QuickGO's web services [35] page.

Your feedback
Please tell us what you thought about this course. Your feedback is invaluable and helps us to improve our courses and thus enhance your learning experience.

Get help and support on GOA

Support and find out more

- For support on the UniProt-GOA resource, email the support team [36].

References


Funding

UniProt-GOA is supported by the European Molecular Biology Laboratory [41], the US National Institutes of Health [42] (NIH), the British Heart Foundation [43] and Kidney Research UK [44].

Contributors

Rachael Huntley [1]

EMBL-EBI
Project Leader, UniProt Gene Ontology Annotation project

Rachael Huntley is the project leader for the UniProt-Gene Ontology Annotation (UniProt-GOA) project at EMBL-EBI. She studied for a PhD in Plant Biochemistry at the University of Cambridge, working on plant growth hormones in Oil Palm. She completed post-doctoral studies into the
regulation of the plant cell cycle by D-type cyclins and the contribution of the pinhead gene to growth of the Arabidopsis shoot apical meristem. In her current position at the EBI she manages UniProt's Gene Ontology Annotation project, which is responsible for contributing and maintaining GO annotations for proteins in UniProt. She is also involved with teaching GO to both curators and researchers.

Source URL: http://www.ebi.ac.uk/training/online/course/uniprot-goa-quick-tour