GOA and QuickGO: Quick tour

George Georghiou [1], Sangya Pundir [2]

- Ontologies
- Proteins
- Beginner
- 0.5 hour

This quick tour provides a brief introduction to the Gene Ontology (GO), the Gene Ontology Annotation (GOA) project and QuickGO, a web-based browser for viewing GO annotations.

This course was published on 8 March 2018.

Learning objectives:

- Describe basic characteristics of GO
- List ways in which GO can help you find out more about your protein data and other gene products
- Search for GO using QuickGO
- Know where to find out more about GO

What is GO?

What is the Gene Ontology (GO)?

The Gene Ontology (GO) is a dynamic controlled vocabulary [3] made up of 3 ontologies that detail the biological processes, molecular functions and cellular components of a normal cell. It is arranged in a hierarchical fashion with terms connected to each by different relationships (Figure 1).
Figure 1: The Gene Ontology is structured hierarchy of terms.

The 3 ontologies of GO are described below (Figure 2):

**Biological process**: terms represent a series of molecular events or functions. Processes span all levels of granularity, covering subcellular [4] metabolic pathways, tissue- and organ-level chains of events, all the way up to organism-level processes such as behaviour and interactions with other organisms.

**Molecular function**: activities performed by individual gene products at the molecular level, for example, the catalysis of reactions or binding to certain chemicals or macromolecules.

**Cellular component**: describes the parts of the cell – subcellular structures and macromolecular complexes – and the extracellular [5] environment in which a gene product may be localised. Examples of intracellular environments include the cytoplasm, ribosome and translation [6] release factor complex

GO terms do not describe the gene products themselves, rather the activities that they perform and the locations where they are found. A description of a gene product is built by combining the annotations that describe its components, processes and functions. For example, you would not find ‘cytochrome’ in the ontology, but you would find the functions that cytochromes perform, such as electron carrier activity.
Why do we need GO?

To non-scientists, the language used by biologists can seem confusing; it is strewn with acronyms, jargon, and common words repurposed for scientific usage. Within the scientific community, this confusion is often compounded as terms are given organism-specific meanings – for example, ‘budding’ means one thing for plants and another for yeast – and different phrases are employed to refer to the same phenomenon. Even subtle differences in the use of terms and phrases can hamper communication between scientists, and can make automated data exchange prohibitively difficult. The Gene Ontology (GO) project is a major initiative to address this need for consistent descriptions of gene products (proteins and RNA).

What can I do with GO?

One of the most popular uses of GO is to find significant shared GO terms (or parents of those GO terms) that are annotated to genes in a particular query set (e.g. a set of genes that are overexpressed in a microarray [7] experiment). This process helps you to find out what those genes may have in common and is known as a GO enrichment analysis.

GO is also used for purposes as diverse as:

- integrating proteomic information from different organisms;
- assigning functions to protein domains;
- finding functional similarities in genes that are overexpressed or underexpressed in diseases and as we age;
- analysing groups of genes that are co-expressed during development;
• developing automated ways of deriving information about gene function from the literature;
• verifying models of genetic, metabolic and product interaction networks.

What is GOA?

Why do we need Gene Ontology annotation?

As proteomics [8] 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 Gene Ontology Annotation project?

The Gene Ontology Annotation (GOA [9]) project uses GO to describe proteins in UniProtKB [10]. The project assigns GO terms to all complete and incomplete proteomes that exist in UniProtKB, using a combination of manual and automatic annotation.

Manual annotation 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 that describes the evidence from the literature that supports the annotation. More information about our manual annotation procedure can be found on the GOA website [11].

Automatic annotation is a rapid way of assigning GO terms to gene products on a large scale, and the GOA project is the main producer of automatic annotations in the GO Consortium. Details of the different automatic annotation pipelines can be found on the GOA website.

GOA is updated on a 4 week basis, in accordance with the latest data released by UniProtKB, Ensembl [12], Ensembl Genomes [13] and InterPro [14]. GO annotations are also imported from other members of the GO Consortium and its collaborators [15].

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 GOA project has made an important contribution to this global effort.

What is QuickGO?

QuickGO [16] is a web-based browser for viewing and exploring the Gene Ontology [17] (GO), the Evidence and Conclusion Ontology (ECO), and GO annotations (Figure 3). Annotations created by GOA [9] curators and other member databases of the GO Consortium [18] can be accessed here.
Figure 3: The QuickGO homepage

What can I do with QuickGO?

- View GO terms and ECO terms and the hierarchical tree of where they fall in their respective ontologies
- Access functional information for gene products (proteins, protein complexes and RNA)
- Use GO Slims [19] to summarise the biological attributes of a proteome [20], compare proteomes or, for example, find out what proportion of a proteome has been found to be involved in apoptosis
- Find GO annotations to link between biological knowledge and high-throughput genomic or proteomic datasets

Searching and visualising data in QuickGO

Browsing and Searching

QuickGO [16] allows you to search for gene products (proteins, protein complexes, RNAs) as well as GO and ECO [21] terms using the search box on the QuickGo homepage (Figure 4). You can search for keywords and accession [22] IDs from UniProtKB [10], RNACentral, Complex Portal, GO, and ECO.
**Viewing GO and ECO**

QuickGO allows you to search and view the GO and ECO ontology terms, for example GO:0006925. Term definition (Figure 5), synonyms, ancestry trees (Figure 6), children terms and annotation restrictions/guidelines can be viewed on these pages. The ancestry chart on this page allows you to see where a particular term falls within the ontology and what other terms it is related to.
Figure 5: An example of a GO term page in QuickGO.
Figure 6: The ancestry chart for the gene ontology term GO:0006925, inflammatory cell apoptotic process.
Viewing GO annotations

You can also use QuickGO to search, browse and filter all GO annotations by going to 'View GO annotations' (Figure 7).

![QuickGO homepage](https://www.ebi.ac.uk/training/online)

**Figure 7**: View GO annotations on the QuickGO homepage.

At the time of publication, there are currently 494,877,648 annotations in GOA [9]. To help find the gene ontologies you are interested in, you can filter the annotations using the columns in the GO annotation table (Figure 8):

- **Taxon** - Filter for annotations just to human or mouse gene products
- **Gene product ID** - If you interested in annotations to a specific protein or a group of proteins and you have their accession IDs
- **Gene product type** - Filter for annotations to just proteins, protein complexes or RNAs. The largest portion of our annotations are to proteins.
- **GO term** - If you are interested in annotations to a particular term or a group of terms
- **GO aspect** - If you want to view annotations to only biological process, molecular function, or cellular component
- **Evidence** - Only see annotations you are interested in from manual curation [25] efforts or automatic annotation [26]
- **Qualifier** - Filter for annotations with a specific relationship between the gene product and GO term
- **Reference** - Whether you want to sort for annotations from papers or from a particular automatic annotation source
- **With/From** - Filter for annotations for gene products that interact with other proteins
- **Assigned By** - Filter for GO annotations from a specific database or group
Figure 8: Filters for the GO annotation table can be found at the top of the table. In addition, you can customize the columns you view and export annotations.

For more information and detail on how to use QuickGO, please watch our webinar on QuickGO and GO annotation [27].

What can I do with a GO slim?

What are GO slims?

GO slims are cut-down versions of the GO ontologies that contain a subset of terms from the GO resource. They give a broad overview of the ontology [23] content without the detail of the specific, fine-grained terms. GO slims can be created by users according to their needs and may be specific to species or to particular areas of the ontologies.

Where in QuickGO is the GO slim tool?

The GO slim tool is located on our homepage under 'Explore Biology' (Figure 9).
What is in a GO slim?

GO provides a generic GO slim, which is not species-specific and should be suitable for most purposes. There are also a number of GO slims which are focused on model organisms, or groups of organisms, such as plants and yeast. These are all listed under Predefined GO slims. (Figure 10)

When should I use GO slims?

GO slims are particularly useful for providing a summary of the GO annotation [24] of a genome, microarray [7], or cDNA collection when you require a broad classification of gene products.
GOA and QuickGO: Quick tour
Published on EMBL-EBI Train online (https://www.ebi.ac.uk/training/online)

You can find more information about GO slims from the GO slims webinar [28].

Getting data from GOA and QuickGo

Downloading annotations

Every month, GOA [9] produces gene association files containing all of the available GO annotations for UniProtKB [10] proteins. The 'UniProt' [29] 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 [20], such as human, mouse, Arabidopsis, yeast, zebrafish, etc.

The GOA gene association files are available from the download page on the GOA website [30], and previous versions of the gene association files are also available from the GOA ftp site [31].

QuickGO API

QuickGO can supply GO term information and GO annotation [24] data via RESTful web services [32]. Instructions on how to access these are on the QuickGO's API page [32].

Get help and support with GOA and QuickGo

Support

If you are in need of any help or have a question or comment on GOA [9] or QuickGo, please feel to get in touch with us at goa [at] ebi.ac.uk

You can also find us on Twitter to get the latest updates as @QuickGO_EBI [33]

References

Please feel free to check out our two webinars on QuickGO

QuickGO - Gene Ontology Annotation 2017 [27]

How to use QuickGO and GO Slims for your research [28]

Collaborators

The Gene Ontology Consortium [34]

The Evidence and Conclusion Ontology [35]

UniProtKB [36]

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

Contributors

George Georghiou

EMBL-EBI
Scientific Database - Protein Function Content

George Georghiou is a scientific database curator [37] for the Gene Ontology [17] Annotation [24] (GOA [9]) project at EMBL-EBI. His work focuses on the CAFA project, co-funded by the BBSRC and NSF. Prior to joining GOA, George obtained his PhD at Stony Brook University for his research on the characterisation of small molecule inhibitors of protein kinases using X-ray crystallography [38]. He went on to conduct his postdoctoral research at the University of Dundee, where he studied the Wnk signaling pathway using CRISPR/Cas9 gene editing technology.

Sangya Pundir

EMBL-EBI
User Experience Manager - Martin team: UniProt development

Sangya Pundir is a User Experience (UX) Manager in EMBL-EBI’s UniProt team, where she established a user-centred process for the redesign of the world’s leading protein resource. To make UniProt easy for researchers to explore, Sangya conducts usability testing and information-gathering methods such as card sorting, contextual studies and workshops. Before she came to EMBL-EBI, Sangya worked at a healthcare consultancy, designing bespoke management systems. She holds an MSc in Biotechnology, Bioprocessing and Business Management from the University of Warwick.

Source URL: https://www.ebi.ac.uk/training/online/course/goa-and-quickgo-quick-tour