Welcome - webinar instructions

- GoToTraining works best in **Chrome** or on Linux, **Firefox**

- All **microphones will be muted** while the trainer is speaking

- If you have a question please use the **chat box** at the bottom of the GoToTraining box

- Please complete the **feedback survey** which will launch at the end of the webinar
An Introduction to QuickGO

A tutorial on how to use our online browser for GO annotations

George Georghiou
Curator

www.ebi.ac.uk/QuickGO
What is the gene ontology (GO)?

• A way to capture biological knowledge in a written and computable form

• A set of concepts and their relationships to each other arranged as a hierarchy
What aspects are covered by the GO?

Three structured ontologies that allow describing gene products in terms of their biological processes, molecular functions, and cellular components.
How is the GO structured?

- Terms are linked by relationships:
  - Is a
  - Part of
  - Regulates
  - Positively regulates
  - Negatively regulates
  - Occurs in
  - Capable of
  - Capable of part of

See the GO wiki for more details;
**GO relationship definitions**

*Is a*

'urea cycle' *is a* type of 'urea metabolic process' and *is a* type of 'amide biosynthetic process'

*Part of*

Photosynthetic dark and light reactions are *part of* 'photosynthesis'
**GO relationship definitions**

**Regulates**
‘regulation of immune response’
regulates ‘immune response’

**Positively regulates**
‘immune response’ is positively regulated by ‘positive regulation of immune response’

**Negatively regulates**
‘immune response’ is negatively regulated by ‘negative regulation of immune response’
GO relationship definitions

Occurs in
'Intracellular transport' occurs in the 'intracellular' compartment

Capable of
The ‘SAGA complex’ is capable of ‘thiol-dependent ubiquitin-specific protease activity’
**GO relationship definitions**

**Capable of part of**
The ‘synapsis initiation complex’ is **capable of** the ‘regulation of synaptonemal complex assembly’, and ‘homologous recombination’ and is **part of** the ‘synapsis’,
What is the aim of the GO?

- GO terms aim to describe the ‘normal’ functions/processes/locations that gene products are involved in.

- It does not capture pathological processes, experimental conditions or temporal information.
What is a gene product?

- Any molecule that is synthesized by what a gene encodes
  - RNA
  - Proteins
  - Protein Complexes
A GO annotation is …

…a statement that a gene product;

1. has a particular molecular function
   or is involved in a particular biological process
   or is located within a certain cellular component

2. as determined by a particular method

3. as described in a particular reference

<table>
<thead>
<tr>
<th>Accession</th>
<th>Name</th>
<th>GO ID</th>
<th>GO term name</th>
<th>Reference</th>
<th>Evidence code</th>
</tr>
</thead>
<tbody>
<tr>
<td>P00505</td>
<td>GOT2</td>
<td>GO:0004069</td>
<td>aspartate transaminase activity</td>
<td>PMID:2731362</td>
<td>IDA</td>
</tr>
</tbody>
</table>
What are evidence codes?

- indicate how the annotation to a particular GO term is supported
- based on Evidence and Conclusion Ontology (ECO)

GO term for protein X:
alcohol dehydrogenase activity (NAD)

source → provides data → supports

IDA
PMID:12345656
evidence: experimental
method: manual

evaluated by curator
What classes of evidence code are used?

<table>
<thead>
<tr>
<th>Evidence Class</th>
<th>ECO ID</th>
<th>GO Evidence Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct assay</td>
<td>ECO:0000314</td>
<td>IDA</td>
</tr>
<tr>
<td>Mutant phenotype</td>
<td>ECO:0000315</td>
<td>IMP</td>
</tr>
<tr>
<td>Physical interaction</td>
<td>ECO:0000353</td>
<td>IPI</td>
</tr>
<tr>
<td>Genetic interaction</td>
<td>ECO:0000316</td>
<td>IGI</td>
</tr>
<tr>
<td>Expression pattern</td>
<td>ECO:0000270</td>
<td>IEP</td>
</tr>
<tr>
<td>High-throughput direct assay</td>
<td>ECO:0007005</td>
<td>HDA</td>
</tr>
<tr>
<td>High-throughput mutant phenotype</td>
<td>ECO:0007001</td>
<td>HMP</td>
</tr>
<tr>
<td>High-throughput genetic interaction</td>
<td>ECO:0007003</td>
<td>HGI</td>
</tr>
<tr>
<td>High-throughput expression pattern</td>
<td>ECO:0007007</td>
<td>HEP</td>
</tr>
<tr>
<td>Traceable author statement</td>
<td>ECO:0000304</td>
<td>TAS</td>
</tr>
<tr>
<td>Non(traceable) author statement</td>
<td>ECO:0000303</td>
<td>NAS</td>
</tr>
<tr>
<td>Sequence similarity</td>
<td>ECO:0000250</td>
<td>ISS</td>
</tr>
<tr>
<td>Genomic context</td>
<td>ECO:0000317</td>
<td>IGC</td>
</tr>
<tr>
<td>Loss of key residues in orthologues</td>
<td>ECO:0000320</td>
<td>IKR</td>
</tr>
<tr>
<td>Computational combinatorial evidence</td>
<td>ECO:0000245</td>
<td>RCA</td>
</tr>
<tr>
<td>Curator inference</td>
<td>ECO:0000305</td>
<td>IC</td>
</tr>
<tr>
<td>No Data available</td>
<td>ECO:0000307</td>
<td>ND</td>
</tr>
<tr>
<td>Electronic annotation</td>
<td>ECO:0000256</td>
<td>IEA</td>
</tr>
</tbody>
</table>
What classes of evidence code are used?

<table>
<thead>
<tr>
<th>Evidence Class</th>
<th>ECO ID</th>
<th>GO Evidence Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct assay</td>
<td>ECO:0000314</td>
<td></td>
</tr>
<tr>
<td>Mutant phenotype</td>
<td>ECO:0000315</td>
<td></td>
</tr>
<tr>
<td>Physical interaction</td>
<td>ECO:0000353</td>
<td></td>
</tr>
<tr>
<td>Genetic interaction</td>
<td>ECO:0000316</td>
<td></td>
</tr>
<tr>
<td>Expression pattern</td>
<td>ECO:0000270</td>
<td></td>
</tr>
<tr>
<td>High-throughput direct assay</td>
<td>ECO:0007005</td>
<td></td>
</tr>
<tr>
<td>High-throughput mutant phenotype</td>
<td>ECO:0007001</td>
<td></td>
</tr>
<tr>
<td>High-throughput genetic interaction</td>
<td>ECO:0007003</td>
<td></td>
</tr>
<tr>
<td>High-throughput expression pattern</td>
<td>ECO:0007007</td>
<td></td>
</tr>
<tr>
<td>Traceable author statement</td>
<td>ECO:0000304</td>
<td></td>
</tr>
<tr>
<td>Non-traceable author statement</td>
<td>ECO:0000303</td>
<td></td>
</tr>
<tr>
<td>Sequence similarity</td>
<td>ECO:0000250</td>
<td></td>
</tr>
<tr>
<td>Genomic context</td>
<td>ECO:0000317</td>
<td></td>
</tr>
<tr>
<td>Loss of key residues in orthologues</td>
<td>ECO:0000320</td>
<td></td>
</tr>
<tr>
<td>Computational combinatorial evidence</td>
<td>ECO:0000245</td>
<td></td>
</tr>
<tr>
<td>Curator inference</td>
<td>ECO:0000305</td>
<td></td>
</tr>
<tr>
<td>No Data available</td>
<td>ECO:0000307</td>
<td></td>
</tr>
<tr>
<td>Electronic annotation</td>
<td>ECO:0000256</td>
<td></td>
</tr>
</tbody>
</table>

**Children of ECO:0000314:**
- Kinase assays (ECO:0001202)
- *In vitro* reconstitution (ECO:0007111)
- Immunofluorescence (ECO:0001249)
- Cell fractionation (ECO:0006042)
What is a GO term qualifier?

• Specifies the relationship between the gene product and the GO term in an annotation

• Available qualifiers for each GO aspect:

<table>
<thead>
<tr>
<th>Molecular Function</th>
<th>Cellular Component</th>
<th>Biological Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>enables</td>
<td>part_of</td>
<td>involved_in</td>
</tr>
<tr>
<td>contributes_to</td>
<td>colocalizes_with</td>
<td>acts_upstream_of</td>
</tr>
<tr>
<td></td>
<td>is_active_in</td>
<td>acts_upstream_of_positive_effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>acts_upstream_of_negative_effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>acts_upstream_of_or_within_positive_effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>acts_upstream_of_or_within_negative_effect</td>
</tr>
</tbody>
</table>

• Any qualifier can be preceded with ‘NOT|’ (e.g. NOT|enables) to indicate that the gene product does not do what you might expect it to do.

http://wiki.geneontology.org/index.php/Annotation_Relations
What is a annotation extension?

- Allows curators to combine GO terms with:
  - other GO terms
  - terms from external ontologies
  - database identifiers representing genes or their products

“Pap1 is a transcription factor that regulates transcription of several oxidative stress response genes, including trr1, pmd1, obr1, bfr1 and ctt1”

“TMEM115 is located in the nucleus that is part of an epithelial cell that is part of the cervix epithelium”

Simple GO annotation | Extended GO annotation
How can scientists use the gene ontology?

- Access gene product functional information
- Analyse high-throughput genomic or proteomic datasets
- Validation of experimental techniques
- Get a broad overview of a proteome
- Obtain functional information for novel gene products
What is QuickGO?

- QuickGO is our online browser for the GO, GO annotations, and the ECO.
  - Can search for GO and ECO terms, as well as gene products
  - Can filter the annotation sets for what you are interested in
  - Can create or use pre-set GO slims

- Annotations are updated once a week, the ontology daily
What is QuickGO?

Gene Ontology and GO Annotations

Search

e.g. apoptosis; GO:0006915; ECO:0000314; tropomyosin

View GO Annotations

Explore biology

Use sets of GO terms (slims) that describe your area of interest

GO version 2018-09-30
Annotation set created on 2018-09-21 15:52

https://www.ebi.ac.uk/QuickGO
Let’s start with the QuickGO search bar
Search will list terms and gene products related to your query.
Example of term search page for ‘kinase activity’

Filters for ontology aspects and type

Link to GO term definition, option to add term to basket, or view the term in the GO ancestry tree

View all annotations to this GO term and its children
An example of a GO term page

Add GO term to your basket

All annotations to this GO term

Navigation bar
An example of a GO term page

Child terms and relation to the parent GO term are listed here

Guidelines on how this GO term should be used

Restrictions on how this GO term is assigned with electronic annotations are found here
An example of a GO term page

### Cross-references

<table>
<thead>
<tr>
<th>Database</th>
<th>ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>InterPro</td>
<td>IPR000040</td>
<td>Acute myeloid leukemia 1 protein (AML1)/Runt</td>
</tr>
<tr>
<td>InterPro</td>
<td>IPR000185</td>
<td>Protein translocase subunit SecA</td>
</tr>
<tr>
<td>InterPro</td>
<td>IPR000194</td>
<td>ATPase, F1/4/1 complex, alpha/beta subunit, nucleotide-binding domain</td>
</tr>
<tr>
<td>InterPro</td>
<td>IPR002006</td>
<td>RNA-directed RNA polymerase, favirnus</td>
</tr>
<tr>
<td>InterPro</td>
<td>IPR002122</td>
<td>DNA helicase, UvrD/RE type</td>
</tr>
<tr>
<td>InterPro</td>
<td>IPR002391</td>
<td>GPCR kinase</td>
</tr>
<tr>
<td>InterPro</td>
<td>IPR003300</td>
<td>SNF2-related, N-terminal domain</td>
</tr>
<tr>
<td>InterPro</td>
<td>IPR003333</td>
<td>Ser/Thr protein kinase, TGFbeta receptor</td>
</tr>
<tr>
<td>InterPro</td>
<td>IPR003881</td>
<td>Sulphonyluroa receptor</td>
</tr>
<tr>
<td>InterPro</td>
<td>IPR003932</td>
<td>NidA/flox family</td>
</tr>
</tbody>
</table>

### Cross-Ontology Relations

<table>
<thead>
<tr>
<th>Relation</th>
<th>Other Ontology</th>
<th>ID</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>has_input</td>
<td>CHEBI</td>
<td>CHEBI:15422</td>
<td>ATP</td>
</tr>
</tbody>
</table>

### Co-occurring Terms

<table>
<thead>
<tr>
<th>Co-occurring Term</th>
<th>PR</th>
<th>%</th>
<th>#Together</th>
<th>#Compared</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0005524 ATP binding</td>
<td>10.20</td>
<td>100.0</td>
<td>9,113,332</td>
<td>9,113,332</td>
</tr>
<tr>
<td>GO:000166 nucleotide binding</td>
<td>8.67</td>
<td>67.27</td>
<td>6,956,936</td>
<td>8,184,691</td>
</tr>
</tbody>
</table>

What other terms are commonly annotated alongside this GO term.
An example of a GO term page

History of changes to this GO term are found here.
Example of gene product search page for ‘kinase activity’

- Filters for manual or electronic entry in UniProtKB
- Filters for gene product type
- Filters for species
- Filters for gene products part of reference proteome

- Links to database entry for gene product
- View all annotations to gene product
Let’s move onto the annotation viewer.
How to filter annotations in QuickGO
Taxon allows you to filter by species.

GO annotations

Taxon IDs must be used. If you are unsure of the taxon ID, use the UniProt Taxonomy page (link in the filter)
Gene Products allows to filter for protein, RNAs, or protein complexes.

Selecting proteins brings up additional filters, allowing you look for proteins only in the reference proteome of every species or proteins that have been manually curated in UniProt (Swiss-Prot).
GO term allows you filter for a particular term or set up a GO slim from the annotation page.

By clicking on ‘Options’, you can filter for GO terms and their child terms, as well as select the relationships you want for particular terms.
References filter allows filtering by specific reference.
Aspect allows filtering for annotations from a specific aspect of GO.
Evidence allows filtering for annotations with specific evidences
Extension allows filtering for annotations with specific annotation extensions.
There are also filters for qualifiers, interaction partners or orthologs, and who generated the annotation.
How to setup a GO slim

Gene Ontology and GO Annotations

Search

Input:
e.g. apoptosis, GO:0006915, ECO:0000314, tropomyosin

View GO Annotations

Explore biology

Use sets of GO terms (slims) that describe your area of interest

Choose your terms

Get GO slim annotations

GO version 2018-09-30
Annotation set created on 2018-09-21 15:52
What is a GO slim and why are they useful?

• GO slims are cut-down versions of the GO ontologies containing a subset of the terms in the whole GO. They give a broad overview of the ontology content without the detail of the specific fine grained terms.

• GO slims are particularly useful for giving a summary of the results of GO annotation of a genome, microarray, or cDNA collection when broad classification of gene product function is required.

• GO slims are created by users according to their needs, and may be specific to species or to particular areas of the ontologies.
Select a predefined GO Slim or make your own!

Explore Biology
Make a selection of terms below (maximum 600) and submit to see their associated annotations

Select terms

Additional Options

• You can refine the search by searching for select gene products and selecting the species you are interested in.
Terms in a slim should be independent of each other
Example: Flybase GO Slim annotations

You can view all the annotations that map to the GO term used in the slim
You can also view statistics for the slim
Example: FlyBase GO Slim statistics

- Statistics report allows to see how many annotations and gene products mapped to your slim
- Can also see how many unique GO terms, which GO aspect is best represented, the types of evidence, the species and who assigned them
Tips for creating a GO Slim

1. When creating a slim for the entire genome, you should try to make sure that it covers as many annotated genes in your set as possible.
   - You should be aware of how many genes are annotated but not in your slim, and how many are "unknown" (i.e., annotated only to the root node).

2. For display purposes you usually want to keep the number of terms as small as possible to convey your results. However, you should ensure that the terms you include are specific enough to capture biologically relevant information.
   - Many terms (e.g. metabolic process or cellular process are too general for the purpose of most slim-based analyses).

These tips are courtesy of PomBase
Tips for creating a GO Slim

3. On a related note, if you are using your slim for data analysis (e.g. to summarize an enrichment), you should ensure that the terms are specific enough to demonstrate their relevance to the biological topic of interest.

  • For example, lumping all genes involved in transport may mask overrepresentation of transmembrane transport vs. underrepresentation of vesicle-mediated transport in your results set, so you need to ensure that the slim has categories to represent your results effectively.

4. Most current implementations of software to create "GO slims" include the \textit{regulates} relationship by default, so that (for example) genes involved in \textit{regulation of cytokinesis} will be included with the set of genes annotated to \textit{cytokinesis}.
How to use the basket for useful GO analysis

Add GO Terms to your Basket to generate GO term comparisons and custom GO Slims
How to use the basket for useful GO analysis

• Add GO Terms to your Basket to generate GO term comparisons and custom GO Slims
What can you do with terms in your basket?

• Adding terms can allow you to compare GO terms on the ancestry tree as well as export all annotations to the GO terms in the basket
Comparing GO terms from the basket
What if I need help?

- Visit the FAQ page or contact us for help with your query
Thanks for listening!

- Feel free to contact with any questions or comments via the website or our email
  - https://www.ebi.ac.uk/QuickGO
  - https://www.ebi.ac.uk/GOA
  - goa@ebi.ac.uk
  - Twitter - @QuickGO_EBI
Upcoming webinars

See the full list of upcoming webinars at https://www.ebi.ac.uk/training/webinars

Don’t forget!

Please fill in the survey that launches after the webinar – thanks!