

The GO Annotation (GOA) Project

Rachael Huntley, Emily Dimmer, Daniel Barrell, David Binns and Rolf Apweiler.



EMBL-EBI



www.ebi.ac.uk/GOA/

GOA is a member of the GO Consortium and provides manual and electronic GO annotations to almost 190,000 species

Barrell *et al.* (2009) NAR 37 The GOA database in 2009 - an integrated Gene Ontology Annotation resource.

Manual annotation

- Manual assignment of GO terms by ~40 curators in UniProtKB using published literature
- Each annotation is given one of 11 evidence codes which describes the supporting evidence
- GOA integrates manual annotation from 20 external databases and specialist groups providing a comprehensive set of annotations for all species
- GOA currently has a total of over 520,000 manual annotations*

Finding annotations in a paper

...for B. napus PERK1 protein (Q9ARH1)

In this study, we have shown that like other plant RLKs, the kinase domain of PERK1 has **serine/threonine kinase activity**, in addition, the location of a PERK1-GTP fusion protein to the **plasma membrane** supports the prediction that PERK1 is an integral membrane protein...these kinases have been implicated in early stages of **wound response**...

PubMed ID: 12374299

Function: protein serine/threonine kinase activity GO:0004674
 Component: plasma membrane GO:0005886
 Process: response to wounding GO:0009611

Electronic annotation

- GOA applies six methods for automatic propagation of annotation
- Almost 38 million electronic annotations have been created using these methods*
- Electronic annotation is the **ONLY** annotation available for 170,000 species
- All electronic annotation is given the 'IEA' evidence code

InterPro2GO

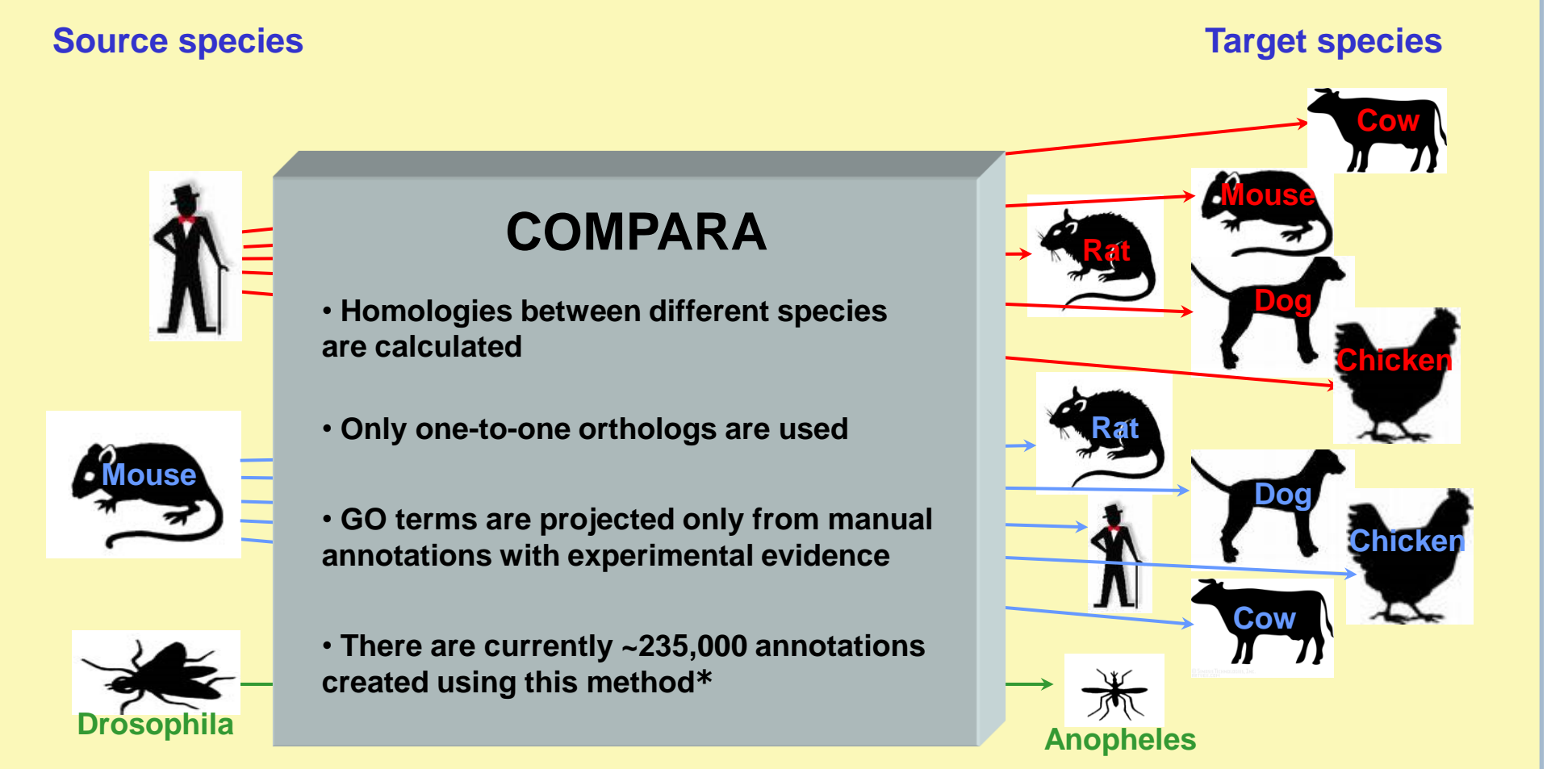
<http://www.geneontology.org/external2go/interpro2go>

InterPro:IPR000020 Anaphylatoxin/fibulin > GO:extracellular region ; GO:0005576
 InterPro:IPR000021 Hok/gef cell toxic protein > GO:membrane ; GO:0016020
 InterPro:IPR000022 Carboxyl transferase > GO:ligase activity ; GO:0016874
 InterPro:IPR000023 Phosphofructokinase > GO:6-phosphofructokinase activity ; GO:0003872
 InterPro:IPR000023 Phosphofructokinase > GO:glycolysis ; GO:0006096

- InterPro integrates protein signature databases, e.g. Pfam, ProSite etc. in order to classify proteins into families and identify domains
- InterPro2GO mapping is done manually but using a computer-based tool
- There are currently ~25.6 million annotations created using this method*

Ensembl Compara

GO term projection via homology



Resulting annotation:

Protein accession	GO term	Reference	Evidence	With information	Acknowledged DB
Q9ZD7	GO:0003696	GOA:comparaGO_REF:0000019	IEA	Ensembl:ENSP00000249910	Ensembl

HAMAP2GO

<http://www.geneontology.org/external2go/hamap2go>

HAMAP:MF_00092 > GO:molecular_function ; GO:0003674
 HAMAP:MF_00092 > GO:DNA binding ; GO:0003677
 HAMAP:MF_00092 > GO:ATP binding ; GO:0005524
 HAMAP:MF_00093 > GO:transmission release factor activity, codon specific ; GO:0016149
 HAMAP:MF_00093 > GO:translational termination ; GO:0006415

- The HAMAP (High-Quality Automated and Manual Annotation of Microbial Proteomes) project at the Swiss Institute of Bioinformatics aims to annotate proteins originating from bacterial and archaeal genome sequencing projects
- HAMAP identifiers are manually assigned to GO terms
- A protein is provided with a HAMAP2GO annotation when it has one or more matches to a HAMAP identifier
- There are currently ~530,000 annotations created using this method*

SPSL2GO

<http://www.geneontology.org/external2go/spsl2go>

SP_SL:SL-0045 Cellular thylakoid membrane > GO:thylakoid membrane ; GO:0042651
 SP_SL:SL-0046 Centriole > GO:centriole ; GO:0005814
 SP_SL:SL-0047 Centromere > GO:chromosome, centromeric region ; GO:0000775
 SP_SL:SL-0048 Centromere > GO:centromere ; GO:0005813
 SP_SL:SL-0049 Chloroplast > GO:chloroplast ; GO:0009507
 SP_SL:SL-0050 Chloroplast envelope > GO:chloroplast envelope ; GO:0009941

- UniProtKB maintain a list of 367 defined subcellular locations 93% of which have been manually mapped, by GOA curators, to GO terms
- A protein is provided with a SPSL2GO annotation when it has a match to a subcellular location term
- There are currently ~670,000 annotations created using this method*

Information is taken from UniProtKB entries

Reviewed UniProtKB/Swiss-Prot Q9ZJR1 (RNH2_HELPJ)
 Last modified January 20, 2009. Version 52. History...

Names and origin: Protein attributes: General annotation (Comments): Ontologies: Sequence annotation (Features): Sequences: Reference

Protein names: Recommended name: Ribonuclease H1; Short name: RNase H1; EC:3.1.26.4

General annotation (Comments):

Function: Endonuclease that specifically degrades the RNA of RNA-DNA hybrids (By similarity)

Catalytic activity: Endonucleolytic cleavage to 5'-phosphonucleoside (HAMAP MF_00025)

Cofactor: Manganese or magnesium. Binds 1 divalent metal ion per monomer in the absence of substrate binding (By similarity)

Subcellular location: Cytoplasm (Eukarya)

Sequence similarities: Belongs to the RNase H1 family.

Ontologies:

Keywords: Cellular component: Cytoplasm; Ligand: Manganese, Metal-binding; Molecular function: Endonuclease, Hydrolase, Nuclease

Cross-references:

Family and domain databases:

HAMAP: MF_00052 (Tree)

InterPro: IPR001352 RNase_H1/H1 (Graphical view)

PANTHER: PTHR10954 RNase_H1/H1 1 hit

Pfam: PF01351 RNase_H1 1 hit (Graphical view)

ProteinNet: Search...

*March 2009 GOA release

EC2GO

<http://www.geneontology.org/external2go/ec2go>

EC:2 > GO:transferase activity ; GO:0016740
 EC:2.1 > GO:transferase activity, transferring one-carbon groups ; GO:0016741
 EC:2.1.1 > GO:methyltransferase activity ; GO:0008168
 EC:2.1.1.1 > GO:nicotinamide N-methyltransferase activity ; GO:0008112
 EC:2.1.1.10 > GO:homocysteine S-methyltransferase activity ; GO:0008898

- Enzyme Commission (EC) numbers are manually curated into the description line of Swiss-Prot entries and added automatically to TrEMBL entries
- Mappings between EC and GO are made using the EC cross-references in the GO function_ontology files
- There are currently ~740,000 annotations created using this method*

SPKW2GO

<http://www.geneontology.org/external2go/spkw2go>

SP_KW:KW-0067 ATP-binding > GO:ATP binding ; GO:0005524
 SP_KW:KW-0937 Abscisic acid biosynthesis > GO:abscisic acid biosynthesis ; GO:0009688
 SP_KW:KW-0938 Abscisic acid signaling pathway > GO:abscisic acid mediated signaling ; GO:0009738
 SP_KW:KW-0005 Acetoin biosynthesis > GO:acetoin biosynthesis ; GO:0045151
 SP_KW:KW-0006 Acetoin catabolism > GO:acetoin catabolism ; GO:0045150

- UniProtKB maintain a list of 951 defined keywords, 74% of which have been manually mapped, by GOA curators, to a corresponding GO term
- A protein is provided with a SPKW2GO annotation when it has a match to one or more of these keywords
- There are currently ~10 million annotations created using this method*

QuickGO

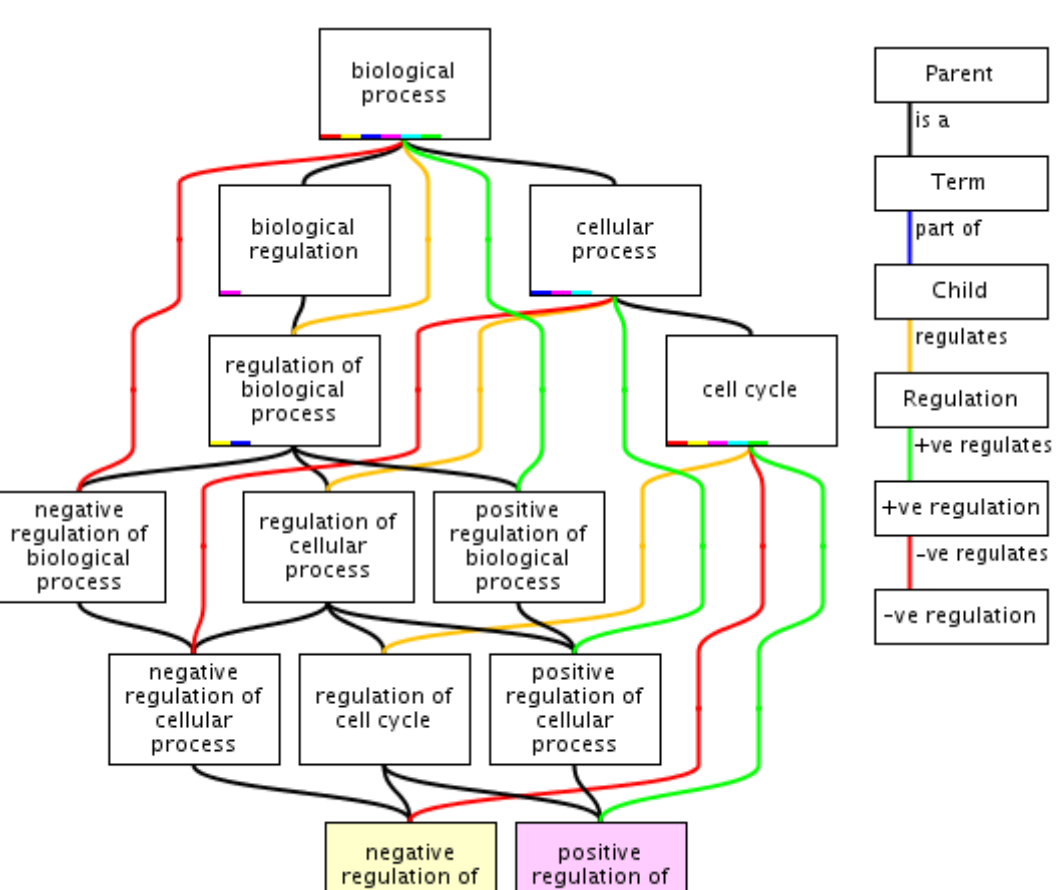


Fig. 2. Hierarchical display of GO terms

www.ebi.ac.uk/QuickGO

- QuickGO is the only GO browser to display both manual and electronic annotations (Fig. 1)
- QuickGO allows you to search various gene/protein identifiers for associated GO annotation
- Custom sets of annotation can be made using extensive filtering options
- Annotations can be categorised in QuickGO using GO slims
- QuickGO offers several download options (Fig. 1)
- GO terms can be viewed in the context of the GO hierarchy (Fig. 2)

Fig. 1. QuickGO display of annotations to protein BLOS3

Accession	GO term	Reference	Evidence	With information	Acknowledged DB
Q9ZD7	GO:0003696	GOA:comparaGO_REF:0000019	IEA	Ensembl:ENSP00000249910	Ensembl

GOA produces multi-species and species-specific annotation files which can be downloaded from www.ebi.ac.uk/GOA/

Rachael Huntley
 Scientific Database Curator
 PANDA Group
huntley@ebi.ac.uk

GOA is supported by:
 US National Institutes of Health
 EMBL
 Kidney Research UK
 British Heart Foundation

EMBL-EBI
 Wellcome Trust Genome
 Campus
 Hinxton
 Cambridge
 CB10 1SD
 UK

T +44 (0) 1223 494 444
 F +44 (0) 1223 494 468
<http://www.ebi.ac.uk>