Introduction to the Gene Ontology and GO annotation resources

Belo Horizonte
September 2014

Prudence Mutowo PhD
Embl EBI
GOA curator

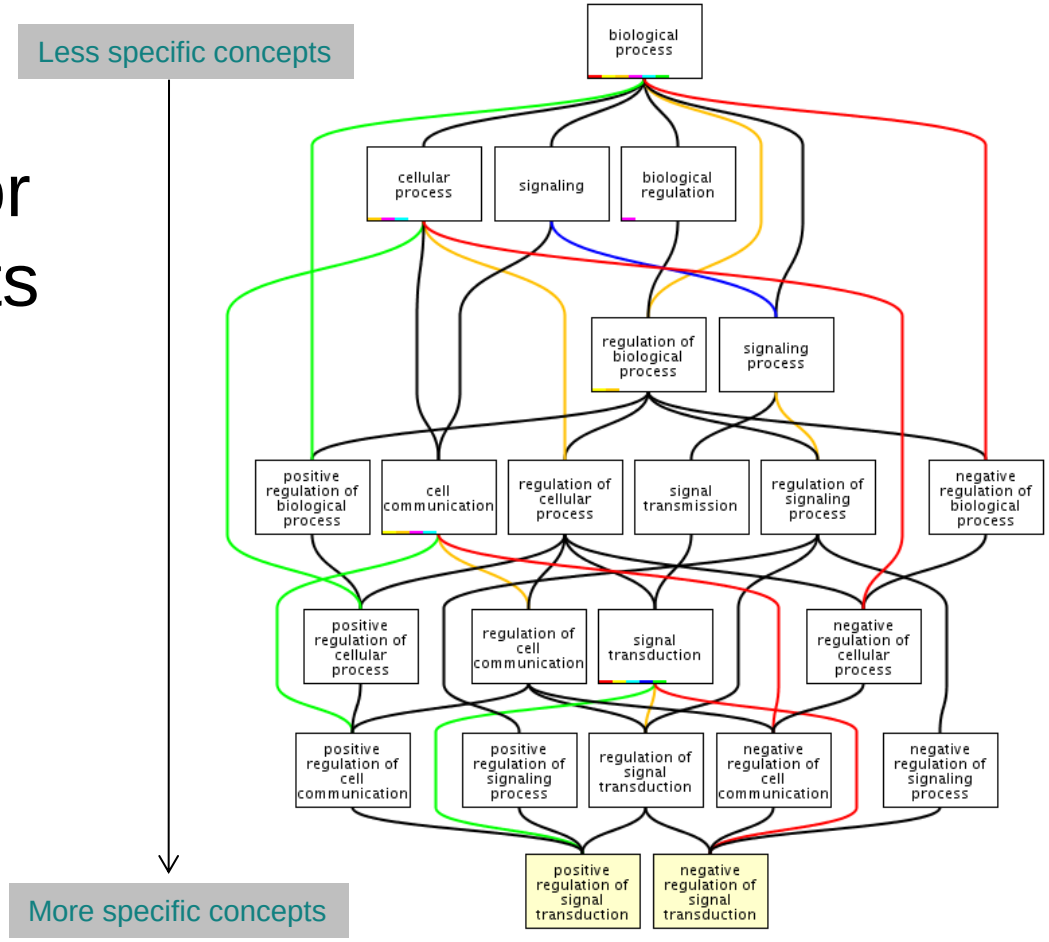
EMBL-EBI

What is GO?

The Gene Ontology

- A way to capture biological knowledge for individual gene products in a written and computable form

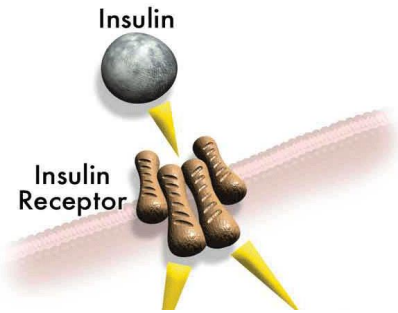
- A set of concepts and their relationships to each other arranged as a hierarchy



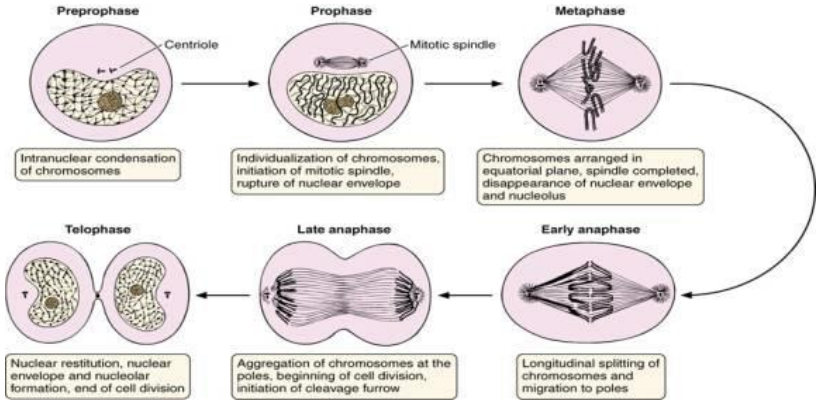
www.ebi.ac.uk/QuickGO

1. Molecular Function

An elemental activity or task or job



- protein kinase activity
- insulin receptor activity



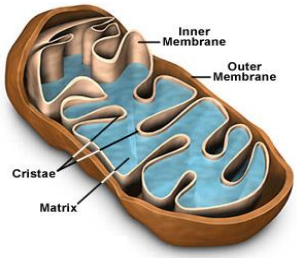
2. Biological Process

A commonly recognised series of events

- cell division

3. Cellular Component

Where a gene product is located



- mitochondrion
- mitochondrial matrix
- mitochondrial inner membrane

Anatomy of a GO term

Term Information | Ancestor Chart | Child Terms | Protein Annotation | Co-occurring Terms | Change Log

ID	GO:0005634
Name	nucleus
Ontology	Cellular Component
Definition	A membrane-bounded organelle of eukaryotic cells in which chromosomes are housed and replicated. In most cells, the nu organellar chromosomes, and is the site of RNA synthesis and processing. In some species, or in specialized cell types, RN
GONUTS	GO:0005634 Wiki Page

Unique identifier

Term name

Definition

Synonyms | Taxon Constraints | GO Slims | Cross-references | Replaces

Synonyms are alternative words or phrases closely related in meaning to the term name, with indication of the relationship betwee scope. Click on the icon for more details

Type	Synonym
exact	cell nucleus

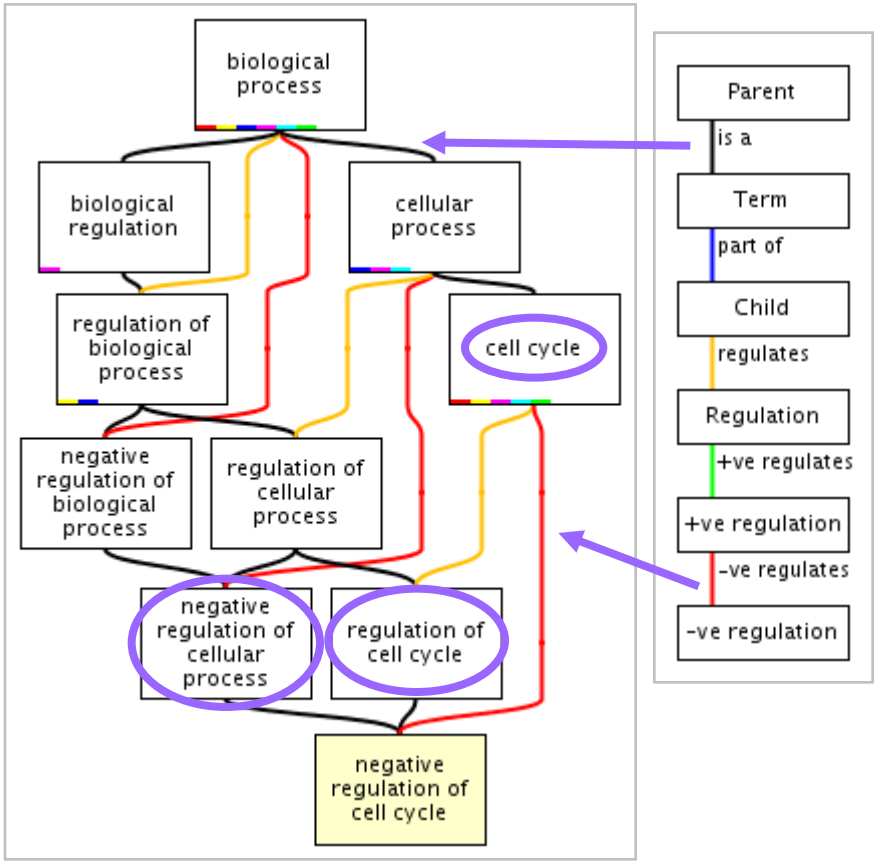
Synonyms

Database	ID
InterPro	IPR028782
InterPro	IPR028785
NIF_Subcellular	sao1702920020
UniProtKB-KW	Nucleus
UniProtKB-SubCell	Nucleus
Wikipedia	Cell_nucleus

Cross-references

Ontology structure

- Hierarchical
  - Terms can have more than one parent
  - Terms can have more than one child
- Terms are linked by relationships
  - is_a
  - part_of
  - regulates (and +/- regulates)
  - has_part
  - occurs_in



www.ebi.ac.uk/QuickGO

These relationships allow for complex analysis of large datasets

Why do we need GO?

Reasons for the Gene Ontology

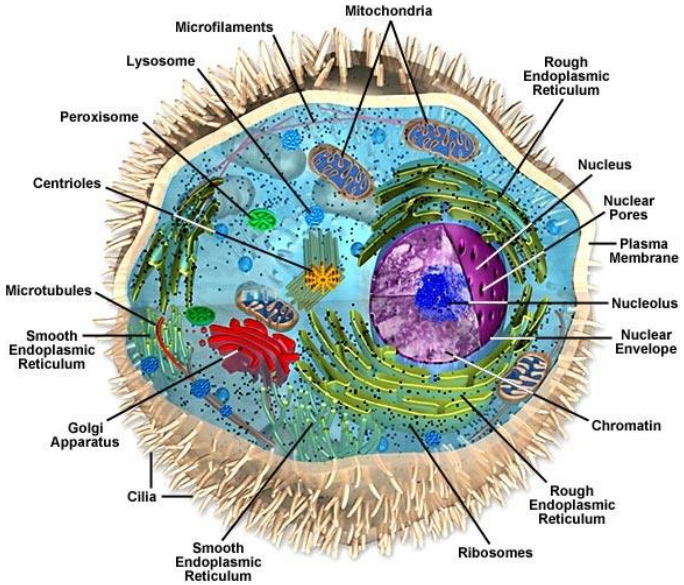
- Inconsistency in English language

Inconsistency in English languauge

- **Same** name for **different** concepts

Cell

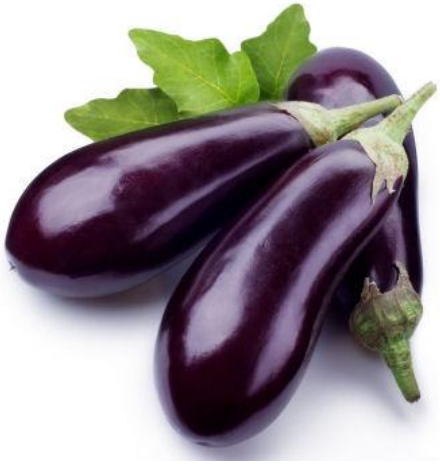
or



??

- **Different** names for the **same** concept

Eggplant



Brinjal

Aubergine

Melongene

Same for biological concepts

→ Comparison is difficult – in particular across species or across databases

Just one reason why the Gene Ontology (GO) is is needed…

Reasons for the Gene Ontology

- Inconsistency in English language
- Increasing amounts of biological data available
- Increasing amounts of biological data to come

Increasing amounts of biological data available

NCBI Resources How To

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed | DNA repair | Search

RSS Save search Limits Advanced

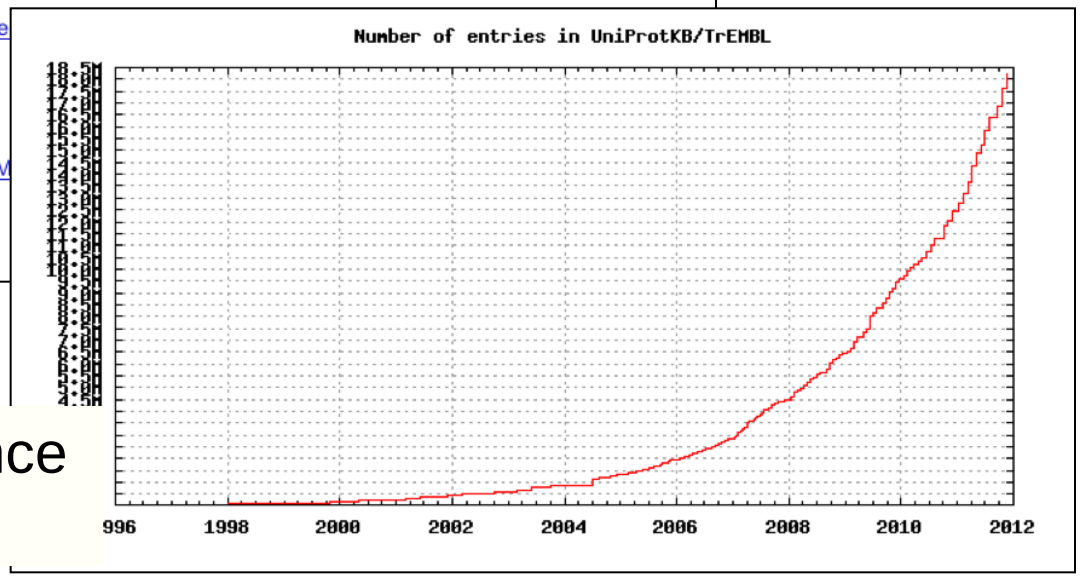
Display Settings: Summary, 20 per page, Sorted by Recently Added Send to:

Results: 1 to 20 of 64637

<< First < Prev Page 1 of 3232 Next > Last >>

1. Germline Stem Cell Gene PIWIL2 Mediates **DNA Repair** through Relaxation of Chromatin.
   Yin DT, Wang Q, Chen L, Liu MY, Han C, Yan Q, Shen R, He G, Duan W, Li JJ, Wani A, Gao JX.
   PLoS One. 2011;6(11):e27154. Epub 2011 Nov 16.
   PMID: 22110608 [PubMed - in process]
2. An MSI Tumor Specific Frameshift Mutation in a Coding Microsatellite of MSH3 Encodes Epitopes.
   Garbe Y, Maletzki C, Linnebacher M.
   PLoS One. 2011;6(11):e26517. Epub 2011 Nov 14.
   PMID: 22110587 [PubMed - in process]
3. A Review of ERCC1 Gene in Bladder Cancer: Implications for Carcinoge
   Kawashima A, Takayama H, Tsujimura A.
   Adv Urol. 2012;2012:812398. Epub 2011 Oct 26.
   PMID: 22110495 [PubMed - as supplied by publisher]
4. Polymorphisms in **DNA Repair** Gene XRCC1 and Skin Cancer Risk: A M
   Zhang H, Li W, Franklin MJ, Dudek AZ.
   Anticancer Res. 2011 Nov;31(11):3945-52.
   PMID: 22110224 [PubMed - in process]

Search on ‘DNA repair’... get almost 65,000 results



Expansion of sequence information

Reasons for the Gene Ontology

- Inconsistency in English language
- Increasing amounts of biological data available
- Increasing amounts of biological data to come
- Large datasets need to be interpreted quickly

GO provides a way of consolidating all the sequence and experimental information into one easily accessible resource

Who is GO?

dictyBase

GO Annotating cardiovascular genes

PomBase

MGI

UniProt

EnsemblGenomes

SGD

GENEONTOLOGY
Unifying Biology
http://www.geneontology.org

EcoliHub

FlyBase

WormBase
The Biology and Genome of C. elegans.

InterPro
Protein sequence analysis & classification

Ensembl

BERKELEY LAB

tair

GRAMENE

Reactome

AgBase

ZFIN

J. Craig Venter INSTITUTE

University of Maryland School of Medicine
INSTITUTE FOR GENOME SCIENCES

RGD

Aims of the GO project

- Compile the ontologies

  - currently over 38,000 terms
  - constantly increasing and improving

- Annotate gene products using ontology terms

  - around 40 groups provide annotations

- Provide a public resource of data and tools

  - regular releases of annotations
  - tools for browsing/querying annotations and editing the ontology

GO Annotation

UniProt-Gene Ontology Annotation (UniProt-GOA) project at the EBI

- Largest open-source contributor of annotations to GO
- Provide annotation for more than 441,000 species

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

Accession	*Name*	*GO ID*	*GO term name*	*Reference*	*Evidence code*
P00505	GOT2	GO:0004069	aspartate transaminase activity	PMID:2731362	IDA

New!

Annotation Extensions

Allows curators to combine GO terms with:

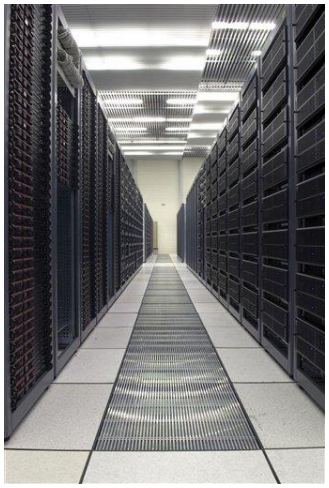
- other GO terms
- terms from external ontologies
- database identifiers representing genes or their products

"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

UniProt-GOA incorporates annotations made using two methods

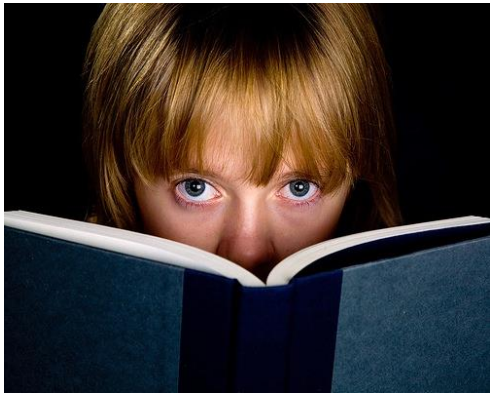


Electronic Annotation

- Quick way of producing large numbers of annotations
- Annotations use less-specific GO terms
- Only source of annotation for ~438,000 non-model organism species

Manual Annotation

- Time-consuming process producing lower numbers of annotations (~2,800 taxons covered)
- Annotations tend to use more specific GO terms
- Manual annotation is essential for creating predictions



Broad taxonomic coverage

We have annotations for species that may not have a dedicated curation effort;

e.g. for 1,400 *Solanacae* species' we have ~360,000 annotations for ~64,000 proteins

We provide annotation files for well-studied species…

Set of Proteins	Gene Association File	Identifier Cross-Reference File
UniProtKB	Gene association	UniProtKB ID mapping I Readme
Human	Gene association	UniProtKB ID mapping I Readme
Mouse	Gene association	UniProtKB ID mapping I Readme
Rat	Gene association	UniProtKB ID mapping I Readme
Arabidopsis	Gene association	UniProtKB ID mapping I Readme
Zebrafish	Gene association	UniProtKB ID mapping I Readme
Chicken	Gene association	UniProtKB ID mapping I Readme
Cow	Gene association	UniProtKB ID mapping I Readme
Proteomes	Gene association Browse	n/a
PDB	Gene association	n/a

…as well as less well-studied species that have;

- Complete proteome
- >25% GO annotation coverage

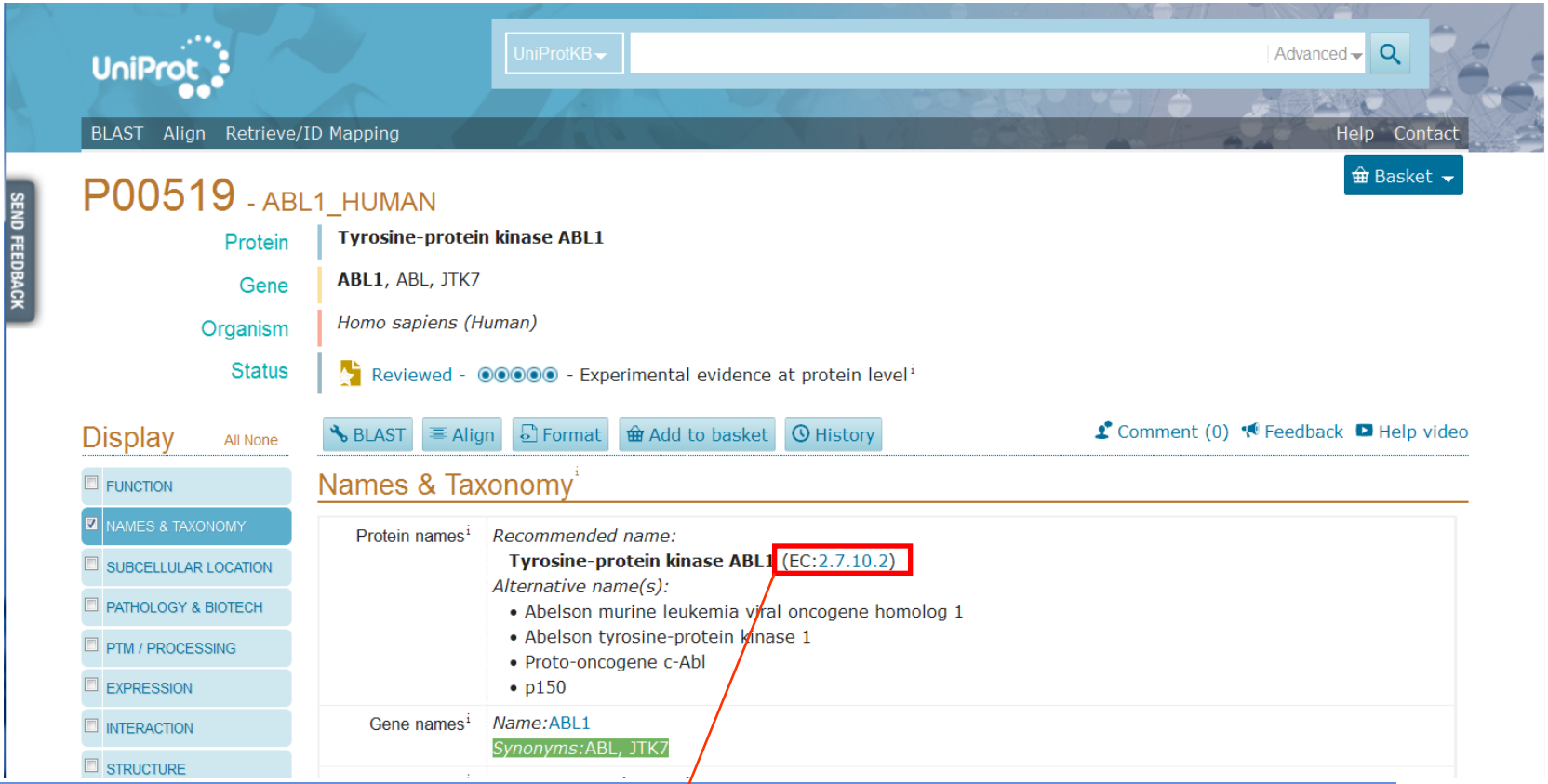
UniProt-GOA Proteome Sets 10-Dec-2013

Tax ID	Organism	File	Total entries	Entries Annotated	GO Annotations	Percentage coverage
11791	AKV murine leukemia virus	78984.A_murine_leukemia_virus.goa	3	3	66	100%
10498	ASFB7	31679.A.goa	149	89	533	59.7%
10816	Abutilon mosaic virus (isolate West India)	71348.A_mosaic_virus.goa	7	6	76	85.7%
329726	Acaryochloris marina	30122.A_marina.goa	8172	3581	20689	**43.8**%
522306	Accumulibacter phosphatis	34595.A_phosphatis.goa	4438	2811	18693	63.3%
634452	Acetobacter pasteurianus	34577.A_pasteurianus.goa	2906	1937	12764	66.7%
931626	Acetobacterium woodii	271017.A_woodii_DSM_1030.goa	3445	2449	16324	71.1%

Electronic annotation methods

1. Mapping of external concepts to GO terms

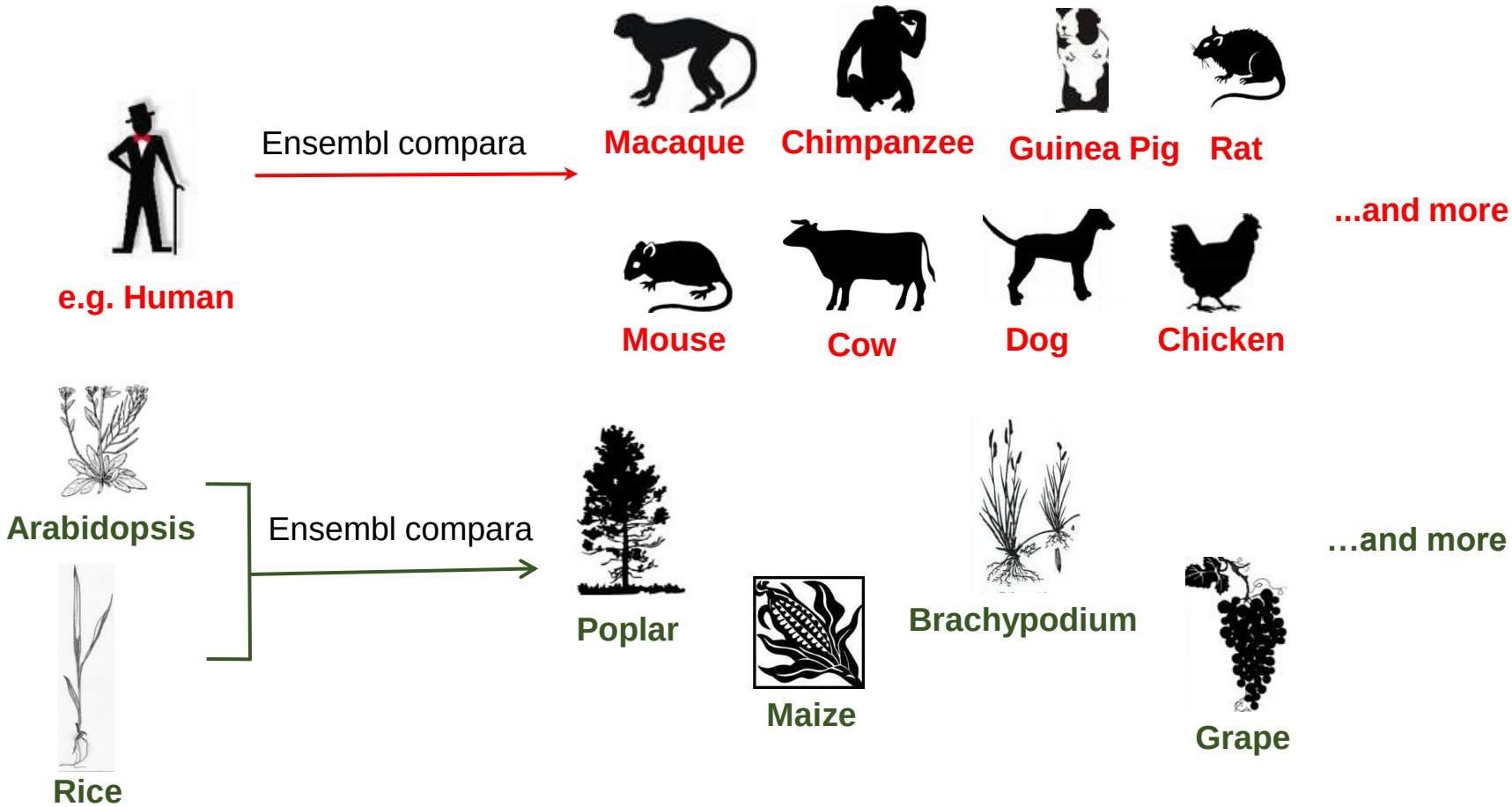
e.g. InterPro2GO, UniProt Keyword2GO, Enzyme Commission2GO



GO:0004715 ; non-membrane spanning protein tyrosine kinase activity

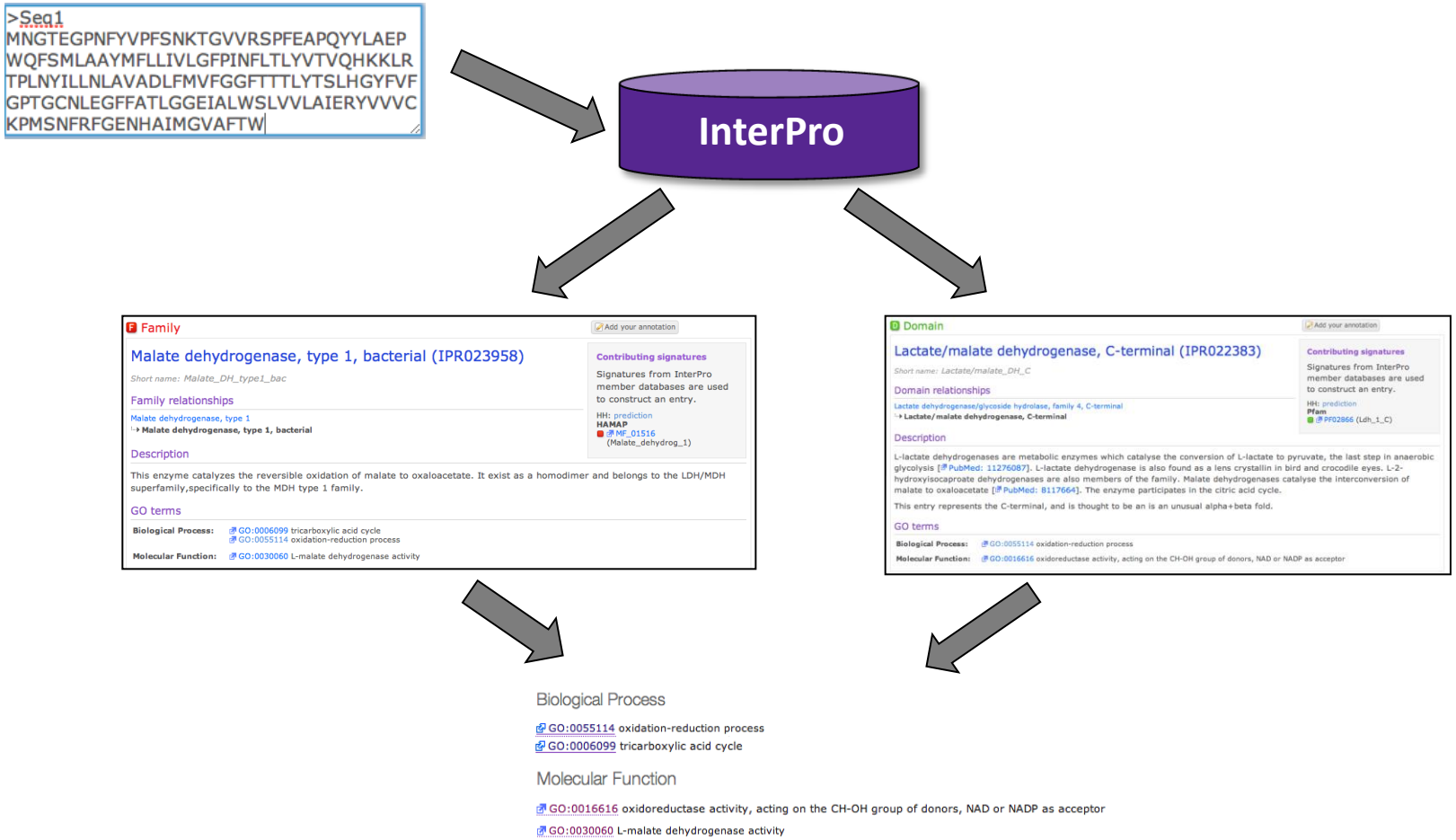
Electronic annotation methods

2. Automatic transfer of manual annotations to orthologs



Annotations are high-quality and have an explanation of the method (GO_REF)

3. Propagation of GO annotations to protein groups



Source of ~170 **million** GO mappings for **~50 million** distinct UniProtKB sequences

Considerations for mapping GO terms

GO mapping to domains:

Function of ***domain*** may not be function of ***protein***

Family members can be experimentally characterised as ***lacking function***:

P14210 - a serine protease homologue with no proteolytic activity

(proteins are reported to GOA to be blacklisted)

Broad families that are ***functionally diverse***:

The GHMP kinase superfamily includes

- Galactokinases (EC=2.7.1.6)
- Homoserine kinases (EC=2.7.1.39)
- Mevalonate kinases (EC=2.7.1.36)
- *Diphosphomevalonate decarboxylases* (EC 4.1.1.33)

Number of annotations in UniProt-GOA database

Electronic annotations	367,330,906
Manual annotations*	1,553,980

Sept 2014 Statistics

* Includes manual annotations integrated from external model organism and specialist groups

Manual annotation by UniProt

High–quality, specific annotations made using:

- Full text peer-reviewed papers
- A range of evidence codes to categorise the types of evidence found in a paper

  e.g. Inferred by Direct Assay (IDA),
  Inferred by Mutant Phenotype (IMP),
  Inferred from Physical Interaction (IPI)

How to access and use GO annotation data

Where can you find annotations?

UniProtKB

UniProt › UniProtKB

Search | Blast * | Align * | Retrieve | ID Mapping *

Search in: Protein Knowledgebase (UniProtKB) — Query

P00519 (ABL1_HUMAN) Reviewed, UniProtKB/Swiss-Prot

Last modified October 5, 2010. Version 162. History...

Clusters with 100%, 90%, 50% identity | Documents (7) | Third-party data

Names · Attributes · General annotation · Ontologies · Interactions · Alt products · Se

Customize order

Names and origin

Protein names	*Recommended name:* **Tyrosine-protein kinase ABL1** EC=2.7.10.2 *Alternative name(s):* Abelson murine leukemia viral oncogene homolog 1; Proto-oncogene c-Abl; p150
Gene names	Name: **ABL1** Synonyms: ABL, JTK7

Gene Ontology (GO)

Biological process

- DNA damage response, signal transduction resulting in induction of apoptosis — Traceable author statement. Source: ProtInc
- actin cytoskeleton organization — Inferred from sequence or structural similarity. Source: UniProtKB
- cell adhesion — Inferred from electronic annotation. Source: UniProtKB-KW
- mismatch repair — Traceable author statement. Source: ProtInc
- peptidyl-tyrosine phosphorylation — Inferred from direct assay (Ref.12). Source: UniProtKB
- positive regulation of oxidoreductase activity — Inferred from direct assay. Source: BHF-UCL
- regulation of transcription involved in S phase of mitotic cell cycle — Traceable author statement. Source: ProtInc
- signal transduction — Inferred from electronic annotation. Source: InterPro

Ensembl

Transcript-based displays
- Transcript summary
- Supporting evidence (24)
- Sequence
  - Exons (11)
  - cDNA
  - Protein
- External References
  - General identifiers (105)
  - Oligo probes (38)
  - Gene ontology (48)
- Genetic Variation
  - Population comparison
  - Comparison image
- Protein Information
  - Protein summary
  - Domains & features (43)
  - Variations (48)

Transcript: ABL1-001 (ENST00000318560)

c-abl oncogene 1, receptor tyrosine kinase [Source:HGNC Symbol;Acc:76]

Location Chromosome 9: 133,710,453-133,763,062 forward strand.

Gene This transcript is a product of gene ENSG00000097007 - There are 6 transcripts in this gene

Show/hide columns

Name	Transcript ID	Le
ABL1-001	ENST00000318560	
ABL1-002	ENST00000372348	
ABL1-003	ENST00000393293	
ABL1-201	ENST00000426772	
ABL1-202	ENST00000438426	
ABL1-203	ENST00000444970	

« Oligo probes — **Gene ontology** help

The following GO terms have been mapped to this entry via UniProt and/or RefSeq:

GO Accession	GO Term	Evidence Annotation Source
GO:0000115	regulation of transcription involved in S-phase of mitotic cell cycle	TAS
GO:0000166	nucleotide binding	IEA
GO:0000287	magnesium ion binding	IDA
GO:0003677	DNA binding	NAS
GO:0004672	protein kinase activity	IEA

Entrez gene

NCBI Resources How To

Entrez Gene — Genes and mapped phenotypes

Search: Gene — Limits Advanced search Help — Search

Display Settings: Full Report

ABL1 c-abl oncogene 1, non-receptor tyrosine kinase [*Homo sapiens*]

Gene ID: 25, updated on 5-Oct-2010

Summary

Official Symbol	ABL1 provided by HGNC
Official Full Name	c-abl oncogene 1, non-receptor tyrosine kinase provided by HGNC
Primary source	HGNC:76
Locus tag	RP11-83J21.1
See related	Ensembl:ENSG00000097007; HPRD:01809; MIM:189980
Gene type	protein coding
RefSeq status	REVIEWED

Gene Ontology provided by GOA

Function	Evidence	
	Evidence Code	Pubs
ATP binding	IDA	PubMed
DNA binding	NAS	PubMed
SH3 domain binding	IPI	PubMed
magnesium ion binding	IDA	PubMed
manganese ion binding	IDA	PubMed
non-membrane spanning protein tyrosine kinase activity	IEA	
nucleotide binding	IEA	
proline-rich region binding	IPI	PubMed
protein C-terminus binding	IPI	PubMed
protein binding	IPI	PubMed

Annotation Files

Contain all information for each annotation

the Gene Ontology

Search

gene or protein name go!

- Open menus
- Home
- FAQ
- **Downloads**
  - Ontologies
  - Annotations
  - Database
  - Mappings to GO
  - Teaching Resources

Current Annotations

Annotation Details and Downloads

- Filtered files
- Unfiltered files

gp2protein files

GO Consortium website

Filtered Annotation File Downloads

Species, Database	Gene Products Annotated	Annotations	Submission date *MM/DD/YYYY*	Download filtered files
Anaplasma phagocytophilum HZ JCVI	1289	3471 (3471 non-IEA)	8/27/2010	annotations [39.7 kb] README
Agrobacterium tumefaciens str. C58 PAMGO	83			
Arabidopsis thaliana TAIR	31339			
Aspergillus nidulans AspGD	11368			
Bacillus anthracis Ames JCVI	5280			

EMBL-EBI

Enter Text Here Find Help | Feedback

Databases | Tools | Research | Training | Industry | About Us | Help | Site Index

- UniProtKB-GOA Home
- New to UniProtKB-GOA?
- Who uses GO?
- Downloads
- Searching UniProtKB-GOA

EBI > Databases > Biological Ontologies > UniProtKB-GOA > UniProtKB-GOA Downloads

Search QuickGO

Query examples - apoptosis, GO:0006915, tropomyosin, P06727.

UniProtKB-GOA - Downloads

Set of Proteins	Gene Association File	Identifier Cross-Reference File
UniProtKB	Gene association	UniProtKB ID mapping I Readme
Human	Gene association	UniProtKB ID mapping I Readme
Mouse	Gene association	UniProtKB ID mapping I Readme
Rat	Gene association	UniProtKB ID mapping I Readme
Arabidopsis	Gene association	UniProtKB ID mapping I Readme
Zebrafish	Gene association	UniProtKB ID mapping I Readme
Chicken	Gene association	UniProtKB ID mapping I Readme
Cow	Gene association	UniProtKB ID mapping I Readme
Proteomes	Gene association Browse	n/a
PDB	Gene association	n/a

UniProt-GOA website

Numerous species-specific files →

http://www.ebi.ac.uk/GOA/downloads.html

EMBL-EBI

GO browsers

the Gene Ontology

Search | Browse | BLAST | More Tools | Help

Search GO ⊙ terms ○ genes or proteins ☐ exact match Submit Query

AmiGO

ABL1

Tyrosine-protein kinase ABL1

protein from *Homo sapiens* (human)

Term associations ⬇ Gene product information ➡ Peptide Sequence ➡ Sequence information ➡

Term Associations

Download all association information in: gene association format, RDF-XML

Select all | Clear all | Perform an action with this page's selected terms... | Go!

Accession, Term		Ontology	Qualifier	Evidence	Reference	Assigned by
GO:0030036 : actin cytoskeleton organization	2336 gene products view in tree	biological process		ISS With UniProtKB:P00520	GO REF:0000024	UniProtKB
GO:0007155 : cell adhesion	4735 gene products view in tree	biological process		IEA With SP_KW:KW-0130	GO REF:0000004	UniProtKB
GO:0008630 : DNA damage response, signal transduction resulting in induction of apoptosis	145 gene products view in tree	biological process		TAS	PMID:10391249	Proteome Inc.
GO:0006298 : mismatch repair						

EMBL-EBI | EB-eye Search | All Databases | Enter Text Here | Go | Reset | Advanced Search | Give us feedback

Databases | Tools | EBI Groups | Training | Industry | About Us | Help | Site Index

EBI > Databases > QuickGO

ABL1 Homo sapiens P00519

Quick GO | Click for example search | Search! | Web Services | Dataset | Your Terms: 16

Accession	P00519
Gene	ABL1
Taxonomy	Homo sapiens
Description	Tyrosine-protein kinase ABL1

Annotation

Results: 1 to 219 of 219 | Page size: 25 | Additional filters: None

Display | ID Mapping | Filter | Statistics | Download

Database ID	Symbol	Qualifier	GO Identifier	GO Term Name	Aspect	Evidence	Reference	With	Taxon	Date	Assigned By	Product Fo ID
				Process								
UniProtKB P00519	ABL1		GO:0000115	regulation of transcription involved in S phase of mitotic cell cycle	P	TAS	PMID:8242749		9606	20030904	PINC	
UniProtKB P00519	ABL1		GO:0007165	signal transduction	P	IEA	InterPro2GO	IPR020700	9606	20101009	InterPro	
UniProtKB P00519	ABL1		GO:0051353	positive regulation of oxidoreductase activity	P	IDA	PMID:12893824		9606	20071108	BHF-UCL	
UniProtKB P00519	ABL1		GO:0006468	protein amino acid phosphorylation	P	IEA	InterPro2GO	IPR001245	9606	20101009	InterPro	
UniProtKB P00519	ABL1		GO:0018108	peptidyl-tyrosine phosphorylation	P	IDA	PMID:7590236		9606	20100505	UniProtKB	
UniProtKB P00519	ABL1		GO:0006464	protein modification process	P	NAS	PMID:8242749		9606	20040707	UniProtKB	
UniProtKB P00519	ABL1		GO:0030036	actin cytoskeleton organization	P	ISS	GO_REF:0000024	Abl1 (ABL1_MOUSE)	9606	20061011	UniProtKB	
UniProtKB P00519	ABL1		GO:0006355	regulation of transcription, DNA-dependent	P	TAS	PMID:8242749		9606	20030904	PINC	
UniProtKB P00519	ABL1		GO:0006468	protein amino acid phosphorylation	P	IEA	InterPro2GO	IPR008266	9606	20101009	InterPro	
UniProtKB P00519	ABL1		GO:0018108	peptidyl-tyrosine phosphorylation	P	IEA	InterPro2GO	IPR020700	9606	20101009	InterPro	
UniProtKB P00519	ABL1		GO:0018108	peptidyl-tyrosine phosphorylation	P	IDA	PMID:9144171		9606	20061004	UniProtKB	
UniProtKB P00519	ABL1		GO:0006468	protein amino acid phosphorylation	P	IEA	InterPro2GO	IPR000719	9606	20101009	InterPro	
UniProtKB P00519	ABL1		GO:0006298	mismatch repair	P	TAS	PMID:10391249		9606	20030904	PINC	
UniProtKB P00519	ABL1		GO:0008630	DNA damage response, signal transduction resulting in induction of apoptosis	P	TAS	PMID:10391249		9606	20030904	PINC	

Quick GO

UniProt

The UniProt QuickGO browser

Search GO terms or proteins

Find sets of GO annotations

EMBL-EBI | EB-eye Search | All Databases | Enter Text Here | Go | Reset | Advanced Search | Give us feedback

Databases | Tools | EBI Groups | Training | Industry | About Us | Help | Site Index

- QuickGO
  - Help
  - Reference
  - FAQs
  - Web Services

EBI > Databases > QuickGO

QuickGO

QuickGO is a fast web-based browser for Gene Ontology terms and annotations, which is provided by the UniProtKB-GOA group at the EBI.

QuickGO | Click for example search | Search! | Web Services | Dataset | Your Terms: 16

Search and Filter GO annotation sets

Extensive filters are available from this page to allow the generation of specific subsets of GO annotations, mapped to sequence identifiers of your choice.

Investigate GO slims

GO slims are lists of GO terms that have been selected from the full set of terms available from the Gene Ontology project.

GO slims can be used to generate a focused view of part of the GO, or with annotation data they can be used to see how a set of proteins/genes can be broadly categorized (using annotation data and the relationships that exist between terms in the ontologies).

Further information on GO slims can be found at the GO Consortium web site.

Example queries

QuickGO can be queried for both **GO terms** and **proteins**:

- Search for terms by keyword or ID: apoptosis, GO:0006915
- Search for proteins by name or accession: tropomyosin, P06727
- List all terms in an ontology: Biological Process, Molecular Function, Cellular Component

QuickGO Tips

- QuickGO can help make customized GO slims, which can be used to 'map up' your gene/protein identifiers to broadly categorise their functions or subcellular locations. Click here for more information.
- QuickGO's search facility can be used to find information on GO terms as well as information on annotation to distinct proteins. Just type a partial GO term name/synonym or a gene symbol, protein name or sequence identifier into the Search box.

Tutorial

Try an interactive demo

Uses of GO annotation data

How scientists use the GO

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

Some examples…

Term enrichment

- Most popular type of GO analysis
- Determines which GO terms are more often associated with a specified list of genes/proteins compared with a control list or rest of genome
- Many tools available to do this analysis
- User must decide which is best for their analysis

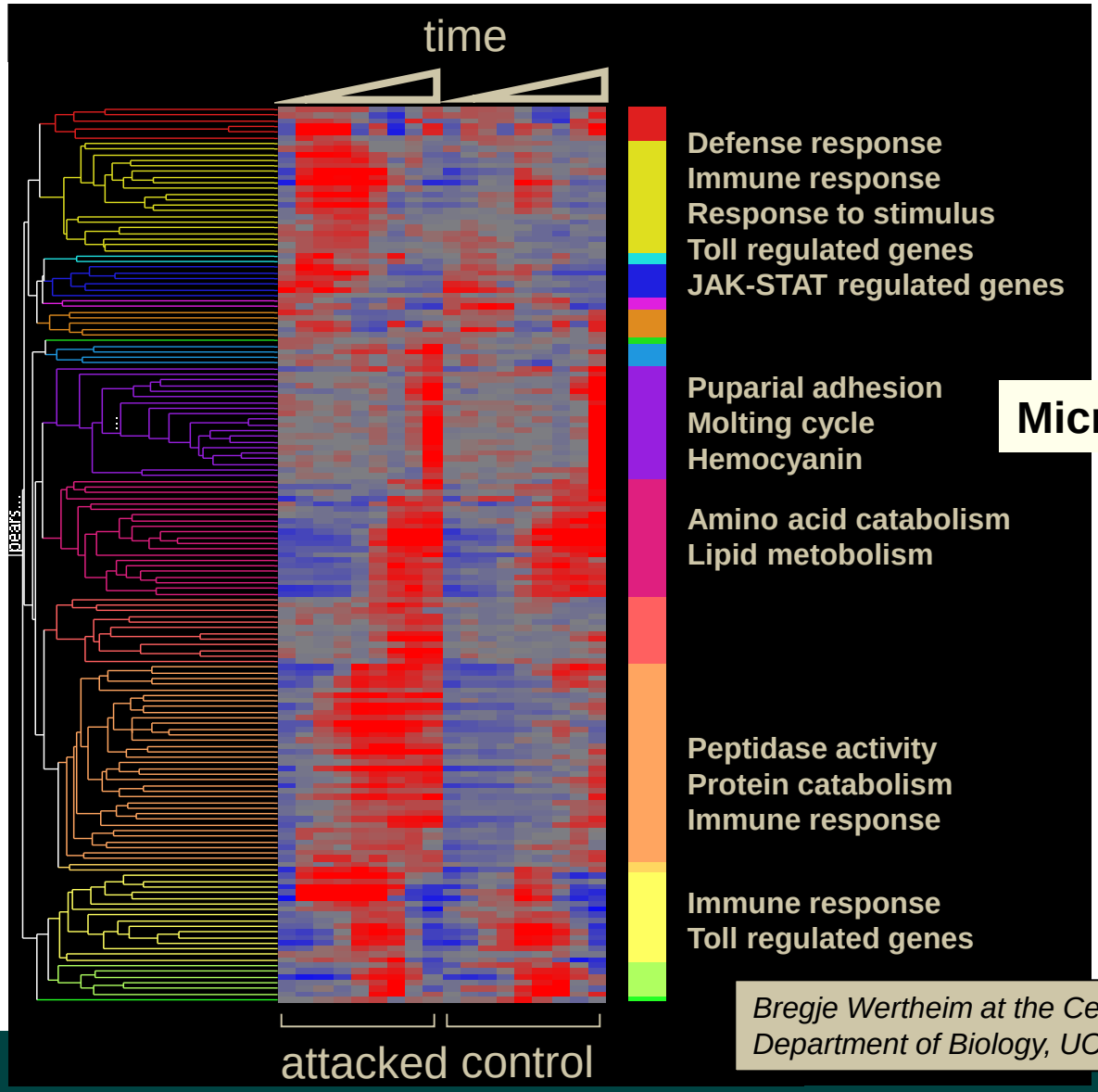
Numerous Third Party Tools

Statistical analysis: Table View

	Is part of	DefiningCitation	ExampleImage
Resource:Avadis	Strand Life Sciences	http://www.avadis-ngs.com	avadis® Access, Visualize, Analyze, Discover
Resource:Babelomics 4: Gene Expression and Functional Profiling Analysis Suite	Resource:CIPF Bioinformatics and Genomics Department	http://babelomics.bioinfo.cipf.es	BABELOMICS 4
Resource:BiNGO: A Biological Networks Gene Ontology tool	Ghent University; Ghent; Belgium	http://www.psb.ugent.be/cbd/papers/BiNGO/Home.html	BiNGO
Resource:Bioconductor	Resource:Fred Hutchinson Cancer Research Center	http://www.bioconductor.org/	Bioconductor OPEN SOURCE SOFTWARE FOR BIOINFORMATICS
Resource:Blast2GO	Resource:Principe Felipe Research Centre; Valencia; Spain	http://www.blast2go.com/b2ghome	b2g blast2go
Resource:CLASSIFI - Cluster Assignment for Biological Inference	Resource:UT Southwestern Medical Center Department of Pathology	http://www.utsouthwestern.edu/education/medical-school/departments/pathology/pathdb/classifi.html	
Resource:CLENCH	Resource:Stanford Center for Biomedical Informatics Research	http://www.stanford.edu/~nigam/cgi-bin/dokuwiki/doku.php?id=clench	CLENCH
Resource:CateGOrizer	Resource:NAGRP Bioinformatics Coordination Program	http://www.animalgenome.org/bioinfo/tools/catego/	CateGOrizer
Resource:ClueGO	National Institute of Health and Medical Research; Rennes; France	http://www.ici.upmc.fr/cluego/	ClueGO

http://neurolex.org/wiki/Category:Resource:Gene_Ontology_Tools

Analysis of high-throughput genomic datasets

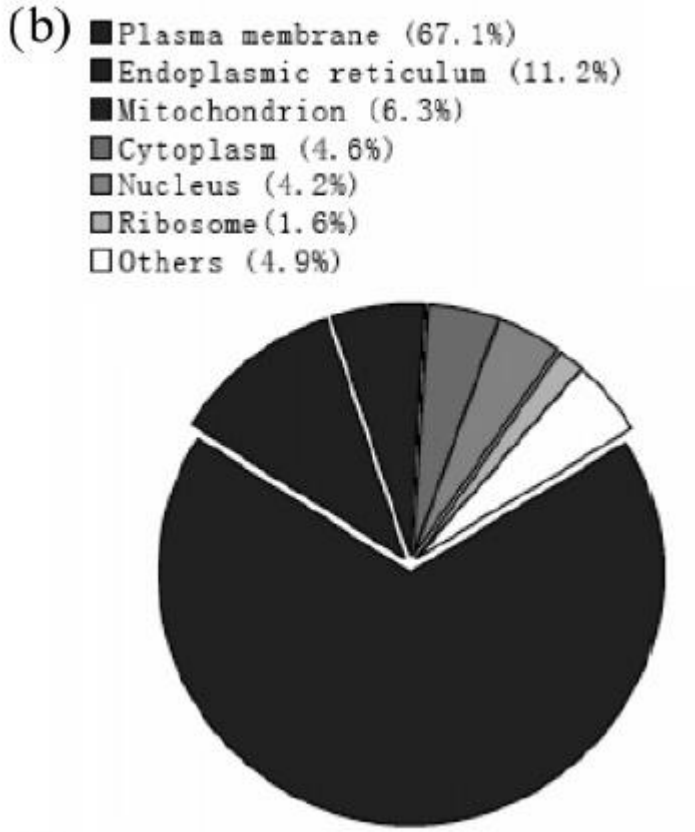


MicroArray data analysis

Bregje Wertheim at the Centre for Evolutionary Genomics, Department of Biology, UCL and Eugene Schuster Group, EBI.

EMBL-EBI

Validation of experimental techniques



Rat liver plasma membrane isolation

(Cao *et al*., Journal of Proteome Research 2006)

Annotating novel sequences

- BLAST can be used to find similar sequences with GO annotation which give an indication of the functional roles of the new sequence

- Two tools currently available;

AmiGO BLAST – searches the GO Consortium database

BLAST2GO – searches the NCBI database

AmiGO

b2q blast2go

* Note BLAST alone is not considered reliable enough to make automated functional predictions

Annotating novel sequences

- InterProScan for matching domains that are mapped to GO

Using the GO to provide a functional overview for a large dataset

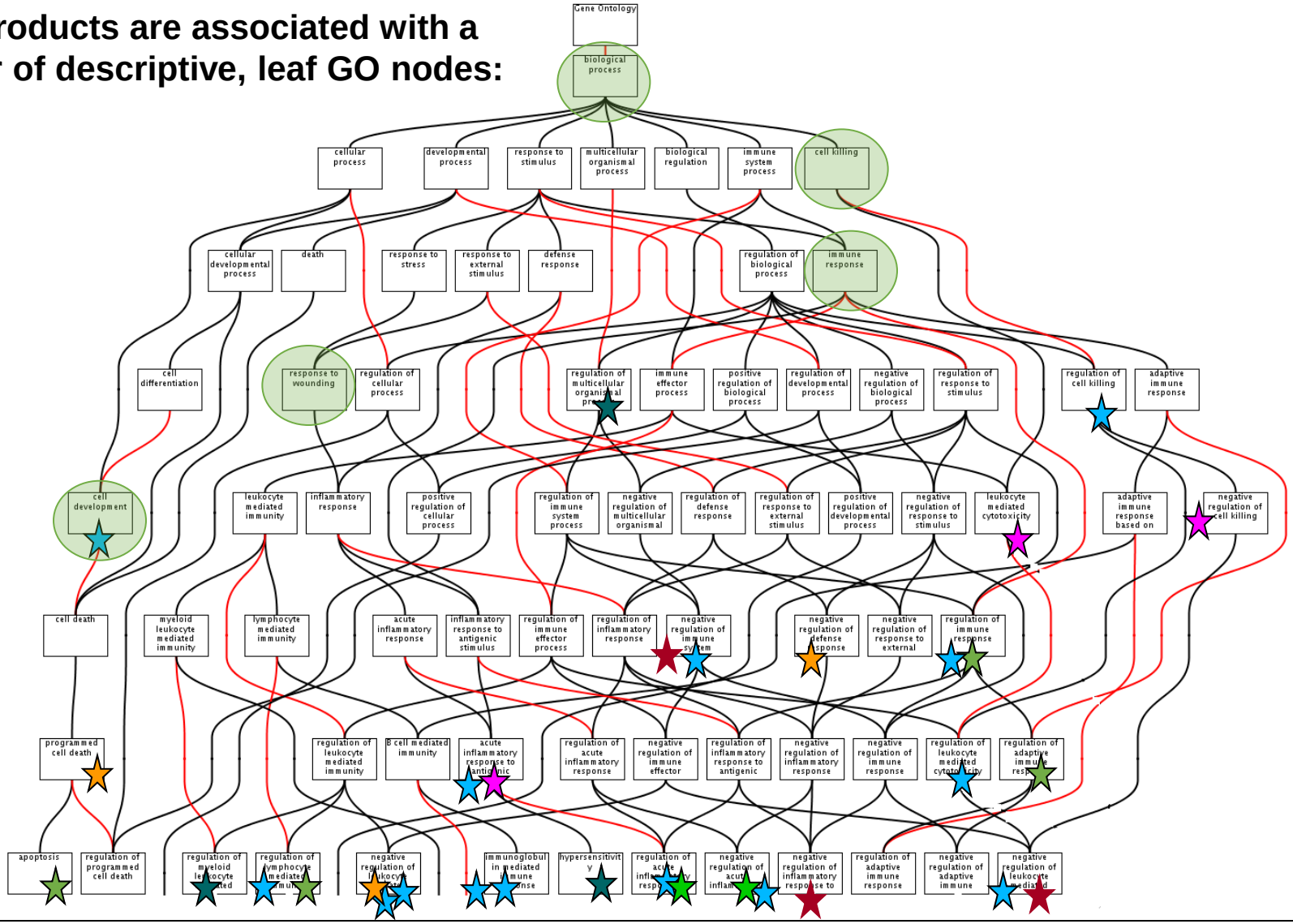
- Many GO analysis tools use GO slims to give a broad overview of the dataset

- GO slims are cut-down versions of the GO and contain a subset of the terms in the whole GO

- GO slims usually contain less-specialised GO terms

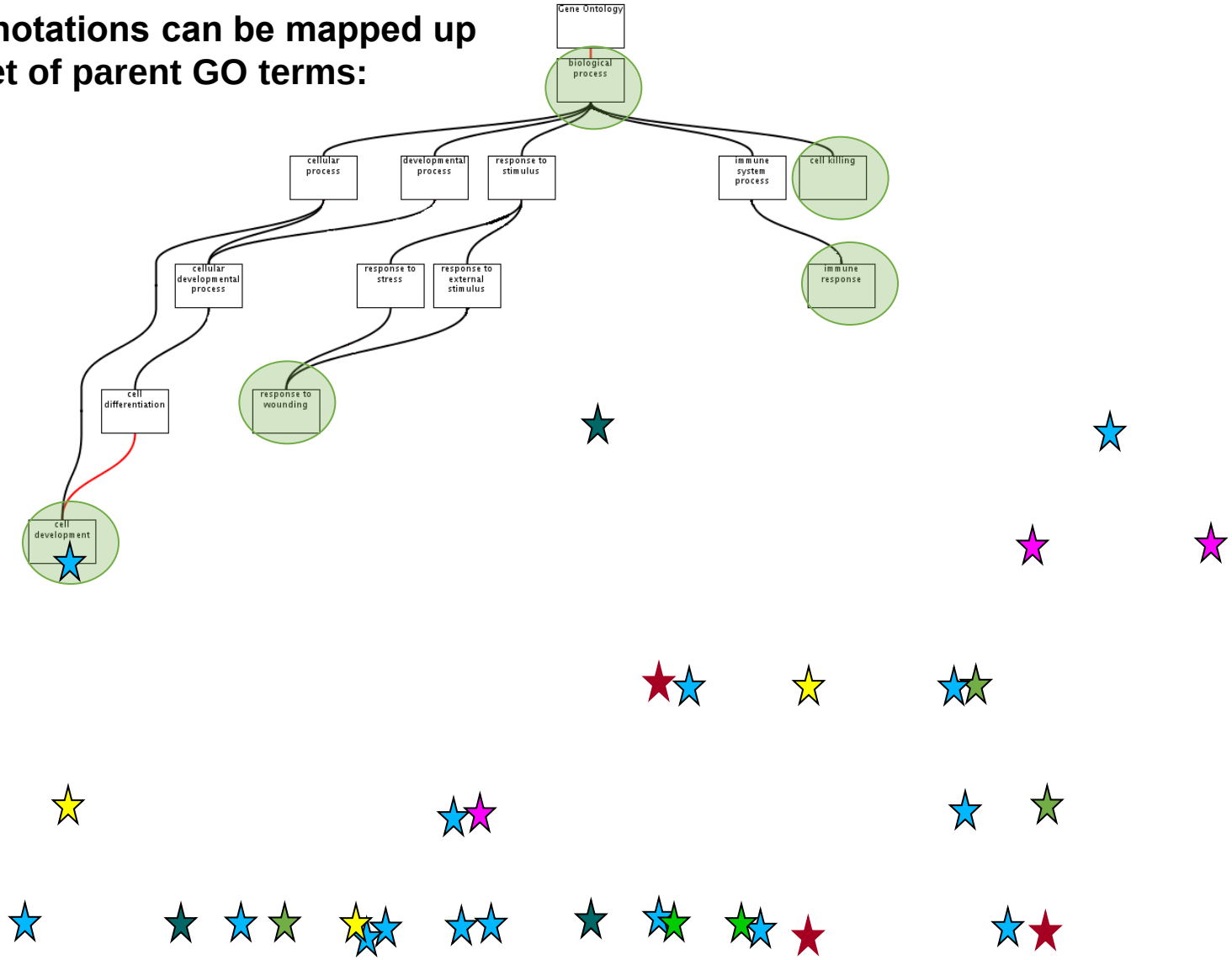
Slimming the GO using the 'true path rule'

Many gene products are associated with a large number of descriptive, leaf GO nodes:



Slimming the GO using the ‘true path rule’

…however annotations can be mapped up to a smaller set of parent GO terms:



GO slims

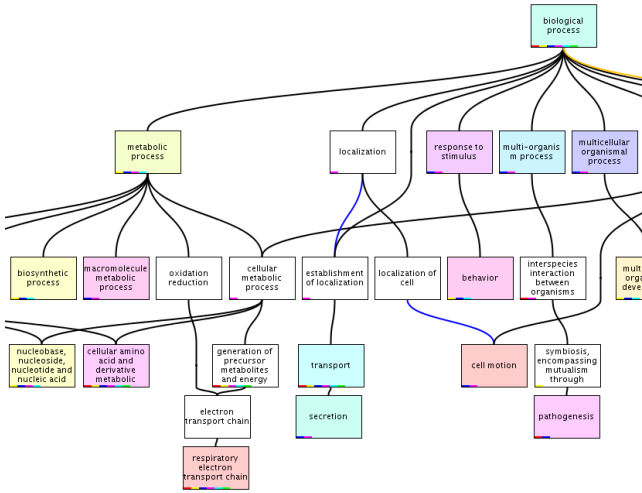
Custom slims are available for download;

http://www.geneontology.org/GO.slims.shtml

or you can make your own using;

- QuickGO
  http://www.ebi.ac.uk/QuickGO

- AmiGO's GO slimmer
  http://amigo.geneontology.org/cgi-bin/amigo/slimmer



The UniProt QuickGO browser

EMBL-EBI | EB-eye Search | All Databases | Enter Text Here | Go | Reset | Advanced Search | Give us feedback

Databases | Tools | EBI Groups | Training | Industry | About Us | Help | Site Index

- QuickGO
  - Help
  - Reference
  - FAQs
  - Web Services

EBI > Databases > QuickGO

QuickGO

Search GO terms or proteins

Questions on how to use QuickGO?
Contact goa@ebi.ac.uk

Find sets of GO annotations

Search and Filter GO annotation sets

Extensive filters are available from this page to allow the generation of specific subsets of GO annotations, mapped to sequence identifiers of your choice.

Map-up annotations with GO slims

Investigate GO slims

GO slims are lists of GO terms that have been selected from the full set of terms available from the Gene Ontology project.

GO slims can be used to generate a focused view of part of the GO, or with annotation data they can be used to see how a set of proteins/genes can be broadly categorized (using annotation data and the relationships that exist between terms in the ontologies).

Further information on GO slims can be found at the GO Consortium web site.

Example queries

QuickGO can be queried for both **GO terms** and **proteins**:

- Search for terms by keyword or ID: apoptosis, GO:0006915
- Search for proteins by name or accession: tropomyosin, P06727
- List all terms in an ontology: Biological Process, Molecular Function, Cellular Component

QuickGO Tips

- QuickGO can help make customized GO slims, which can be used to 'map up' your gene/protein identifiers to broadly categorise their functions or subcellular locations. Click here for more information.
- QuickGO's search facility can be used to find information on GO terms as well as information on annotation to distinct proteins. Just type a partial GO term name/synonym or a gene symbol, protein name or sequence identifier into the Search box.

Tutorial

Try an interactive demo

UniProt-GOA

Curators: Rachael Huntley
Prudence Mutowo
Aleksandra Shypitsyna

Software developers: Tony Sawford
Carlos Bonilla

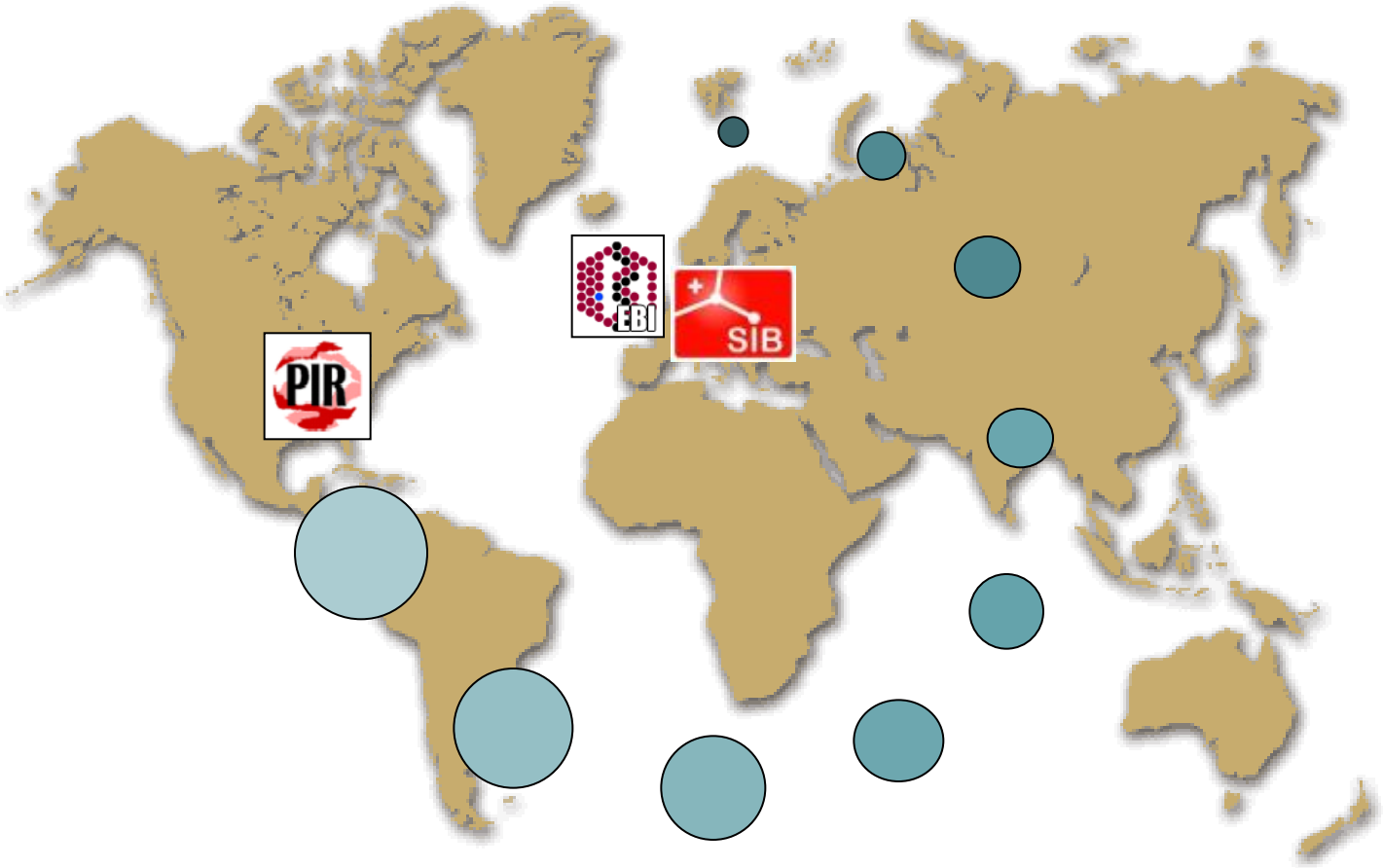
Team leaders: Claire O'Donovan
Maria Martin

Email: *goa@ebi.ac.uk*





UniProt



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**.

Acknowledgements

Members of;

InterPro

IntAct

HAMAP

Ensembl

Ensembl Genomes

GO Consortium

****EBI train online****

http://www.ebi.ac.uk/training/online/

Funding

Additional support for the EBI's involvement in UniProt comes from the NIH grant 2P41 HG02273, British Heart Foundation grant RG/13/5/30112 and EMBL.
Swiss-Prot activities at the SIB are supported by the Swiss Federal Government through the The State Secretariat for Education, Research and Innovation SERI and the European Commission contracts Gen2Phen (200754) and MICROME (222886).
PIR's UniProt activities are also supported by the NIH grants 5R01GM080646-07, 3R01GM080646-07S1, 5G08LM010720-03, and 8P20GM103446-12, and the National Science Foundation (NSF) grant DBI-1062520.