Annotating Gene Products to the GO

Harold J Drabkin
Senior Scientific Curator
The Jackson Laboratory
Mouse Genome Informatics
Bar Harbor, ME

http://www.geneontology.org/GO.annotation.html

The Jackson Laboratory

What **is** an annotation?

GO Term

An annotation is a statement that a gene product …

…has a particular **molecular function**

…is involved in a particular **biological process**

…is located within a certain **cellular component**

…as determined by a particular method

…as described in a particular reference.

Evidence Code

Reference

Smith et al. determined by a direct assay that Abc2 has protein kinase activity, is involved in the process of protein phosphorylation, and is located in the cytoplasm.

Anatomy of an annotation

- ***Object (previously mentioned)***
- ***GO Term from most recent GO***
  - *GO Term Qualifier (optional)*
    - NOT, Co_localizes with, or Contributes_to
- ***Evidence Code :* IDA, IPI, IMP, IEP, IGI, ISS, IEA, TAS, NAS, or IC**
  - *Evidence Code Qualifier (required for some codes)*
    - Used in combination with IPI, IMP, IGI, and ISS
      - Seq_ID or DB_ID required.
- ***Reference: literature or database specific reference***
  - DB_ID or PMID

Getting the GO

GO Evidence Codes

Code	Definition	
IEA	**I**nferred from **E**lectronic **A**nnotation	
NAS	**N**on-traceable **A**uthor **S**tatement	
TAS	**T**raceable **A**uthor **S**tatement	
ND	**N**o **D**ata	Use with annotation to unknown
IDA	**I**nferred from **D**irect **A**ssay	
*IPI	**I**nferred from **P**hysical **I**nteraction	
*IGI	**I**nferred from **G**enetic **I**nteraction	
IMP	**I**nferred from **M**utant **P**henotype	
IEP	**I**nferred from **E**xpression **P**attern	
*IC	**I**nferred from **C**urator	
*ISS	**I**nferred from **S**equence **S**imilarity	

Manually annotated

Annotation Strategies

- Electronic (IEA)
  - Good for first pass
    - Usually based on some sort of sequence comparisons (but use ISS if paper based)
      - IP2GO (InterPro to GO
      - SPTR2GO (SwissProt to GO)
- Manual (literature)

Literature selection

- A paper is selected for GO curation of a mouse gene product if:
  - A paper gives ***direct*** experimental evidence for the ***normal*** function, process, or cellular location of a *mouse** gene product (IDA, IMP, IGG, IPI).
  - A paper gives *direct* experimental evidence for the *normal* function, process, or cellular location of a **non-mouse gene** product **AND** the paper presents homology data to a mouse gene product (ISS)

Annotation process

- **READ** the full papers!
  - Abstracts alone can be very misleading
    - Quite often, the species are not specified. Sometimes a paper uses human, mouse and rat interchangeably , or uses human for one gene and mouse for a different one.

Example Annotations

Direct inhibition of Bruton's tyrosine kinase by IBtk, a Btk-binding protein

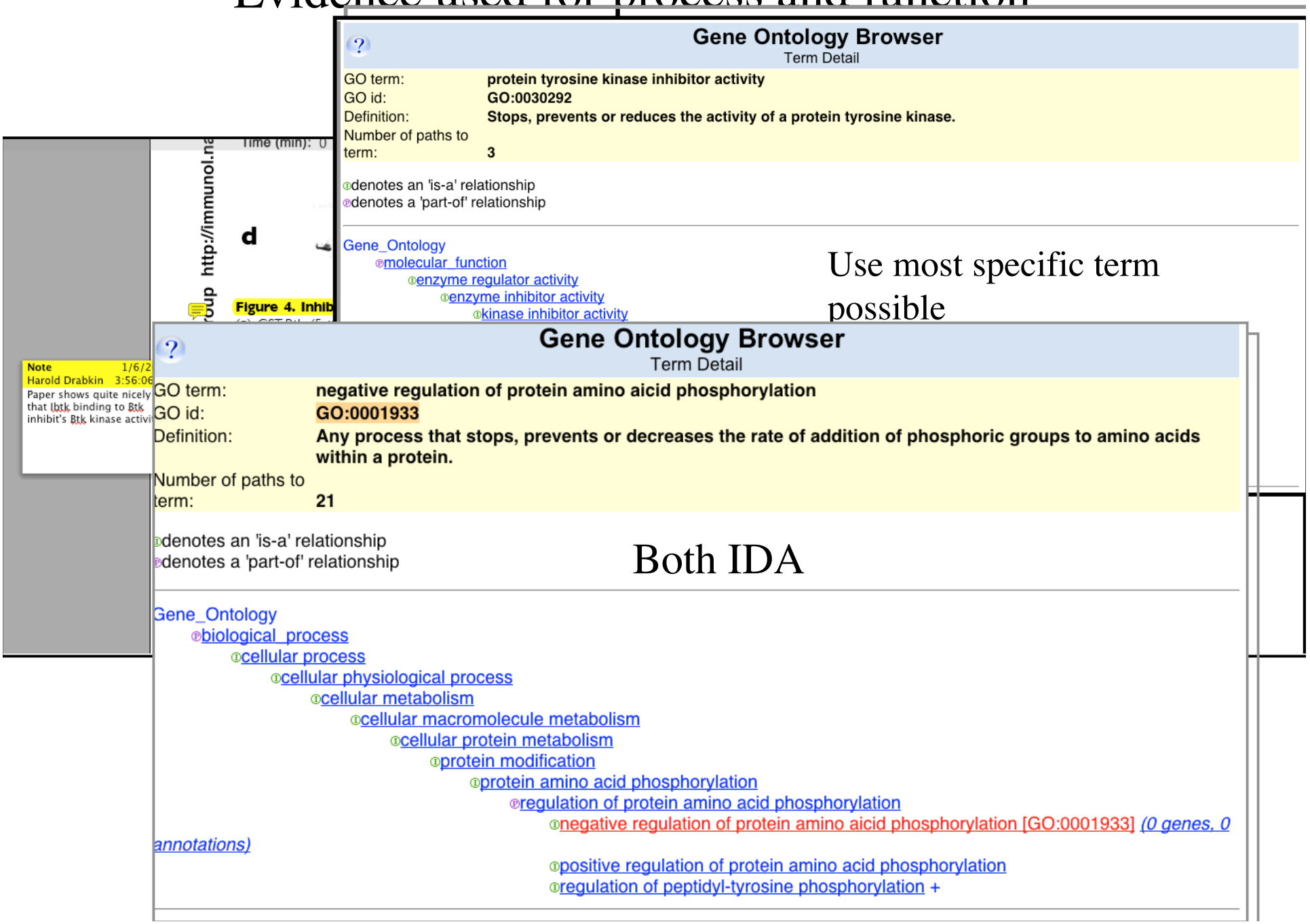
Weimin Liu[1], Ileana Quinto[1,2], Xueni Chen[1], Camillo Palmieri[1], Ronald L. Rabin[3], Owen M. Schwartz[4], David L Nelson[5] and Giuseppe Scala[1,2]

Published online: 1 October 2001, DOI: 10.1038/ni718

Bruton's tyrosine kinase (Btk) is required for human and mouse B cell development. Btk deficiency causes X-linked agammaglobulinemia (XLA) in humans and X-linked immunodeficiency in mice. Unlike Src proteins, Btk lacks a negative regulatory domain at the COOH terminus and may rely on cytoplasmic Btk-binding proteins to regulates its kinase activity by *trans*-inhibitor mechanisms. Consistent with this possibility, IBtk, which we identified as an inhibitor of Btk, bound to the PH domain of Btk. IBtk downregulated Btk kinase activity, Btk-mediated calcium mobilization and nuclear factor-κB–driven transcription. These results define a potential mechanism for the regulation of Btk function in B cells.

- Abstract suggests that this paper demonstrates that Ibtk
  - –Binds to a protein kinase
  - –Inhibits kinase activity
  - –Inhibits calcium mobolization
  - –Inhibits transcription

Evidence used for process and function



Gene Ontology Browser

Term Detail

GO term:	**protein kinase binding**
GO id:	**GO:0019901**
Definition:	**Interacting selectively with a protein kinase, any enzyme that catalyzes the transfer of a phosphate group, usually from ATP, to a protein substrate.**
Number of paths to term:	**1**

ⓘdenotes an 'is-a' relationship
ⓟdenotes a 'part-of' relationship

Gene_Ontology
- ⓟmolecular_function
  - ⓘbinding
    - ⓘprotein binding
      - ⓘenzyme binding
        - ⓘkinase binding
          - ⓘprotein kinase binding [GO:0019901] *(24 genes, 26 annotations)*
            - ⓘJUN kinase binding
            - ⓘmitogen-activated protein kinase binding +
            - ⓘneuronal Cdc2-like kinase binding
            - ⓘprotein kinase A binding
            - ⓘprotein kinase C binding

Both Btk and iBtk have protein binding activity to each other, IPI evidence code

were tested as a positive controls[21]. C63 denotes the positive clone selected by screening a lymphoblastoid B cell cDNA library with a region of Btk encompassing the PH, TH, SH3 and SH2 domains [Btk-(PH-SH2)] used as a bait. [b]β-galactosidase activity is expressed as β-Gal units.

Figure 5. IBtk down-regulates the intracellular Ca^{2+} fluxes in B cells after BCR cross-linking. DT40 WT cells were transfected with 20 μg of either p
encoding
(**b**). Cells
of M4 an
were tra
pGFP (**c**
were stir
μg/ml) t
antibody
Indo-1-lo
tinuously
Transfect
by FCS
(right pa

942

Gene Ontology Browser

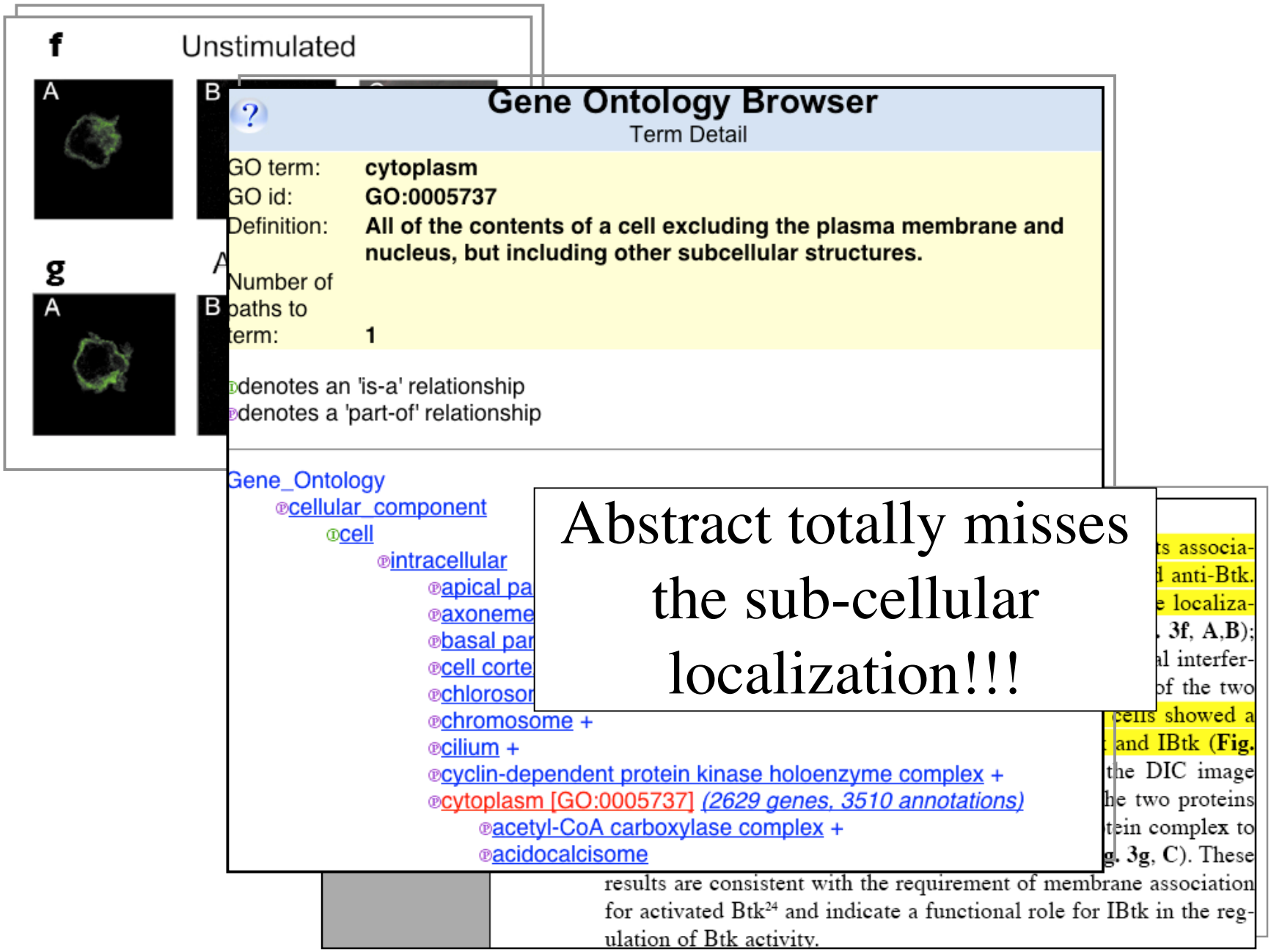
Term Detail

GO term:	**release of sequestered calcium ion into cytoplasm**
Synonym:	**calcium ion (Ca2+) mobilization**
Synonym:	**calcium mobilization**
Synonym:	**release of sequestered calcium ion**
Synonym:	**release of sequestered calcium ion (Ca2+)**
Synonym:	**release of stored calcium ion (Ca2+)**
GO id:	**GO:0051209**
Definition:	**The process by which calcium ions sequestered in the endoplasmic reticulum or mitochondria are released into the cytoplasmic compartment.**
Number of paths to term:	**8**

ⓘdenotes an 'is-a' relationship
ⓟdenotes a 'part-of' relationship

Gene_Ontology
ⓟbiological_process
ⓘcellular process
ⓘcellular physiological process
ⓘcell homeostasis
ⓘcell ion homeostasis
ⓘcation homeostasis
ⓘdi-, tri-valent inorganic cation homeostasis
ⓘcalcium ion homeostasis
ⓟlight-induced release of internally sequestered calcium ion (Ca2+)
ⓟrelease of sequestered calcium ion into cytoplasm [GO:0051209] *(0 genes, 0 annotations)*
ⓟsequestering of calcium ion
ⓘvacuolar calcium ion homeostasis

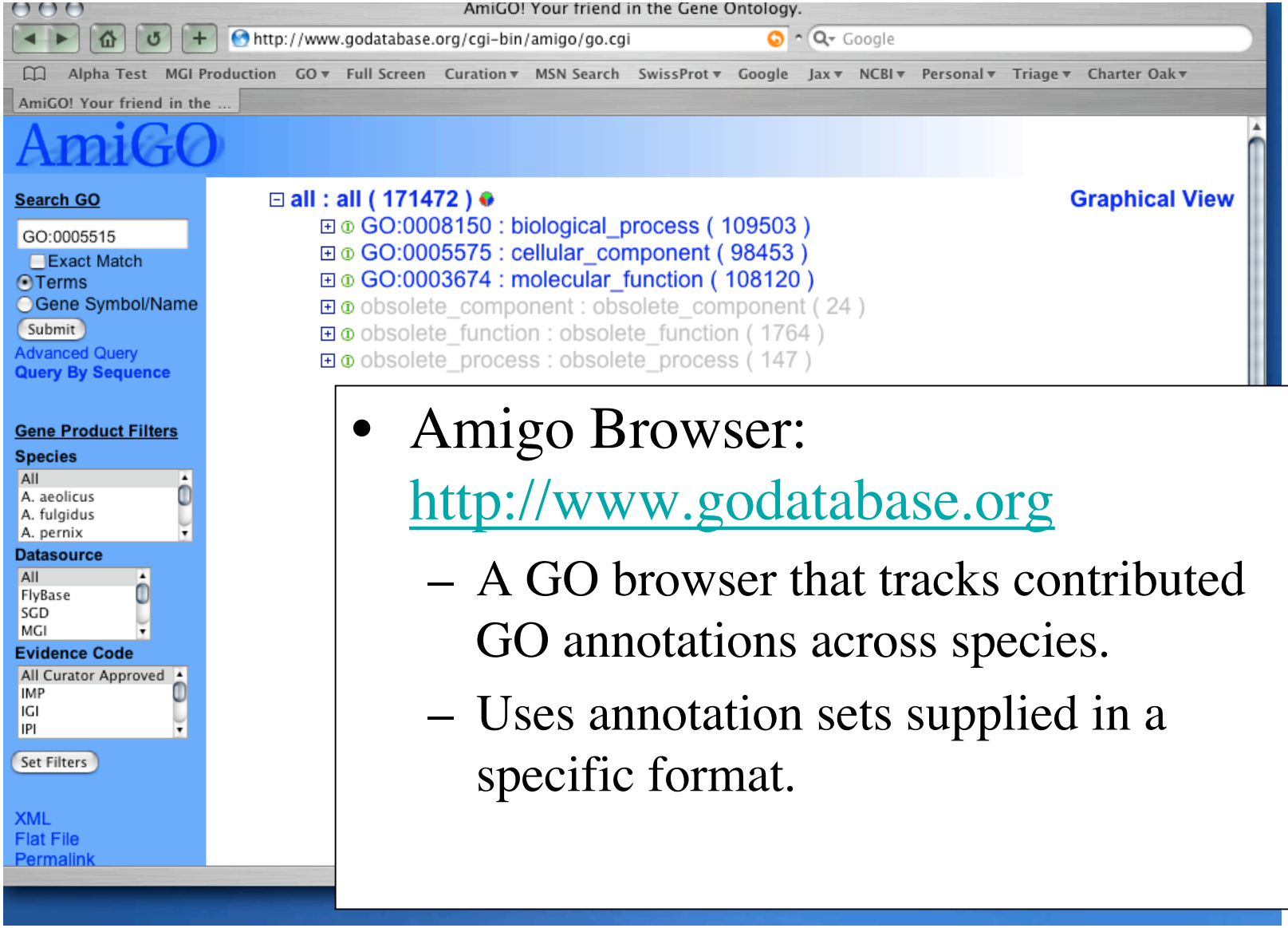
IDA evidence code


f
Unstimulated
A
B
g
A
B
Gene Ontology Browser
Term Detail
GO term: cytoplasm
GO id: GO:0005737
Definition: All of the contents of a cell excluding the plasma membrane and nucleus, but including other subcellular structures.
Number of paths to term: 1
denotes an 'is-a' relationship
denotes a 'part-of' relationship
Gene_Ontology
cellular_component
cell
intracellular
chromosome +
cilium +
cyclin-dependent protein kinase holoenzyme complex +
cytoplasm [GO:0005737] (2629 genes, 3510 annotations)
acetyl-CoA carboxylase complex +
acidocalcisome
Abstract totally misses the sub-cellular localization!!!
results are consistent with the requirement of membrane association for activated Btk[24] and indicate a functional role for IBtk in the regulation of Btk activity.

Sharing Annotations

The Gene Association File

Annotation Sharing



- Amigo Browser: http://www.godatabase.org
  - A GO browser that tracks contributed GO annotations across species.
  - Uses annotation sets supplied in a specific format.

The Gene Association files

gene_association.mgi

```
!software version: $Revision: 1.21 $
!date: 01/07/2005 $
!
! from Mouse Genome Database (MGD) & Gene Expression Database (GXD)
!
MGI     MGI:1918914     0610006F02Rik           GO:0016740      MGI:MGI:1354194 IEA             F       RIKEN cDNA 0610006F02 gene              gene    taxon:10090     20041227        SWALL
MGI     MGI:1918914     0610006F02Rik           GO:0008168      MGI:MGI:2429377 ISS     INTERPRO:IPR001601      F       RIKEN cDNA 0610006F02 gene              gene    taxon:10090     20030221        MGI
MGI     MGI:1918914     0610006F02Rik           GO:0008757      MGI:MGI:2429377 ISS     INTERPRO:IPR000051      F       RIKEN cDNA 0610006F02 gene              gene    taxon:10090     20030221        MGI
MGI     MGI:1914086     0610006I08Rik           GO:0016021      MGI:MGI:2429377 TAS             C       RIKEN cDNA 0610006I08 gene              gene    taxon:10090     20030221        MGI
MGI     MGI:1923502     0610006O14Rik           GO:0016021      MGI:MGI:2429377 TAS             C       RIKEN cDNA 0610006O14 gene              gene    taxon:10090     20030221        MGI
MGI     MGI:1918918     0610007C21Rik           GO:0016021      MGI:MGI:2429377 TAS             C       RIKEN cDNA 0610007C21 gene      HSPC013|p18     gene    taxon:10090     20030221        MGI
MGI     MGI:1915462     0610007H07Rik           GO:0016021      MGI:MGI:2429377 TAS             C       RIKEN cDNA 0610007H07 gene              gene    taxon:10090     20030221        MGI
MGI     MGI:1915578     0610007N03Rik           GO:0005525      MGI:MGI:1354194 IEA             F       RIKEN cDNA 0610007N03 gene      B230212B15Rik   gene    taxon:10090     20041227        SWALL
MGI     MGI:1915578     0610007N03Rik           GO:0005525      MGI:MGI:2152098 IEA             F       RIKEN cDNA 0610007N03 gene      B230212B15Rik   gene    taxon:10090     20041227        SWALL
MGI     MGI:1915578     0610007N03Rik           GO:0015031      MGI:MGI:1354194 IEA             P       RIKEN cDNA 0610007N03 gene      B230212B15Rik   gene    taxon:10090     20041227        SWALL
MGI     MGI:1915578     0610007N03Rik           GO:0015031      MGI:MGI:2152098 IEA             P       RIKEN cDNA 0610007N03 gene      B230212B15Rik   gene    taxon:10090     20041227        SWALL
MGI     MGI:1915578     0610007N03Rik           GO:0007264      MGI:MGI:2152098 IEA             P       RIKEN cDNA 0610007N03 gene      B230212B15Rik   gene    taxon:10090     20041227        SWALL
MGI     MGI:1923501     0610007P08Rik           GO:0004386      MGI:MGI:1354194 IEA             F       RIKEN cDNA 0610007P08 gene      1700019D06Rik   gene    taxon:10090     20041227        SWALL
MGI     MGI:1923501     0610007P08Rik           GO:0003676      MGI:MGI:2429377 ISS     INTERPRO:IPR001650|INTERPRO:IPR002464   F       RIKEN cDNA 0610007P08 gene      1700019D06Rik   gene    taxon:10090     20030221        MGI
MGI     MGI:1923501     0610007P08Rik           GO:0008026      MGI:MGI:2429377 ISS     INTERPRO:IPR001410|INTERPRO:IPR002464   F       RIKEN cDNA 0610007P08 gene      1700019D06Rik   gene    taxon:10090     20030221        MGI
MGI     MGI:1923501     0610007P08Rik           GO:0003677      MGI:MGI:2429377 ISS     INTERPRO:IPR000330      F       RIKEN cDNA 0610007P08 gene      1700019D06Rik   gene    taxon:10090     20030221        MGI
MGI     MGI:1915571     0610007P14Rik           GO:0016021      MGI:MGI:1354194 IEA             C       RIKEN cDNA 0610007P14 gene      1190004E09Rik|ORF11     gene    taxon:10090     20041227        SWALL
MGI     MGI:1915571     0610007P14Rik           GO:0016021      MGI:MGI:2152098 IEA             C       RIKEN cDNA 0610007P14 gene      1190004E09Rik|ORF11     gene    taxon:10090     20041227        SWALL
MGI     MGI:1915568     0610008A10Rik           GO:0016021      MGI:MGI:2445234 ISS     UniProt:Q8WW43  C       RIKEN cDNA 0610008A10 gene      APH-1b  gene    taxon:10090     20031002        MGI
MGI     MGI:1915568     0610008A10Rik           GO:0005515      MGI:MGI:2152098 IEA             F       RIKEN cDNA 0610008A10 gene      APH-1b  gene    taxon:10090     20041227        SWALL
MGI     MGI:1915568     0610008A10Rik           GO:0008233      MGI:MGI:2445234 ISS     UniProt:Q8WW43  F       RIKEN cDNA 0610008A10 gene      APH-1b  gene    taxon:10090     20031002        MGI
MGI     MGI:1915568     0610008A10Rik           GO:0016485      MGI:MGI:2445234 ISS     UniProt:Q8WW43  P       RIKEN cDNA 0610008A10 gene      APH-1b  gene    taxon:10090     20031002        MGI
MGI     MGI:1915568     0610008A10Rik           GO:0001656      MGI:MGI:2675249 IMP             P       RIKEN cDNA 0610008A10 gene      APH-1b  gene    taxon:10090     20031002        MGI
MGI     MGI:1915568     0610008A10Rik           GO:0043085      MGI:MGI:2152098 IEA             P       RIKEN cDNA 0610008A10 gene      APH-1b  gene    taxon:10090     20041227        SWALL
MGI     MGI:1922056     0610008L10Rik           GO:0016021      MGI:MGI:2429377 TAS             C       RIKEN cDNA 0610008L10 gene              gene    taxon:10090     20030221        MGI
MGI     MGI:1913300     0610009B22Rik           GO:0005783      MGI:MGI:2429377 ISS     UniProt:O14582  C       RIKEN cDNA 0610009B22 gene              gene    taxon:10090     20030221        MGI
MGI     MGI:1913300     0610009B22Rik           GO:0005794      MGI:MGI:2429377 ISS     UniProt:O14582  C       RIKEN cDNA 0610009B22 gene              gene    taxon:10090     20030221        MGI
MGI     MGI:1913300     0610009B22Rik           GO:0006810      MGI:MGI:2429377 ISS     UniProt:O14582  P       RIKEN cDNA 0610009B22 gene              gene    taxon:10090     20030221        MGI
MGI     MGI:1913300     0610009B22Rik           GO:0016192      MGI:MGI:2429377 ISS     UniProt:O14582  P       RIKEN cDNA 0610009B22 gene              gene    taxon:10090     20030221        MGI
MGI     MGI:1913300     0610009B22Rik           GO:0001501      MGI:MGI:2429377 ISS     UniProt:O14582  P       RIKEN cDNA 0610009B22 gene              gene    taxon:10090     20030221        MGI
MGI     MGI:1913300     0610009B22Rik           GO:0006888      MGI:MGI:2429377 ISS     UniProt:O14582  P       RIKEN cDNA 0610009B22 gene              gene    taxon:10090     20030221        MGI
MGI     MGI:1913305     0610009D07Rik           GO:0005634      MGI:MGI:1354194 IEA             C       RIKEN cDNA 0610009D07 gene      6030419K15Rik   gene    taxon:10090     20041227        SWALL
MGI     MGI:1913305     0610009D07Rik           GO:0003723      MGI:MGI:1354194 IEA             F       RIKEN cDNA 0610009D07 gene      6030419K15Rik   gene    taxon:10090     20041227        SWALL
MGI     MGI:1913305     0610009D07Rik           GO:0003723      MGI:MGI:2152098 IEA             F       RIKEN cDNA 0610009D07 gene      6030419K15Rik   gene    taxon:10090     20041227        SWALL
MGI     MGI:1913305     0610009D07Rik           GO:0006397      MGI:MGI:1354194 IEA             P       RIKEN cDNA 0610009D07 gene      6030419K15Rik   gene    taxon:10090     20041227        SWALL
```

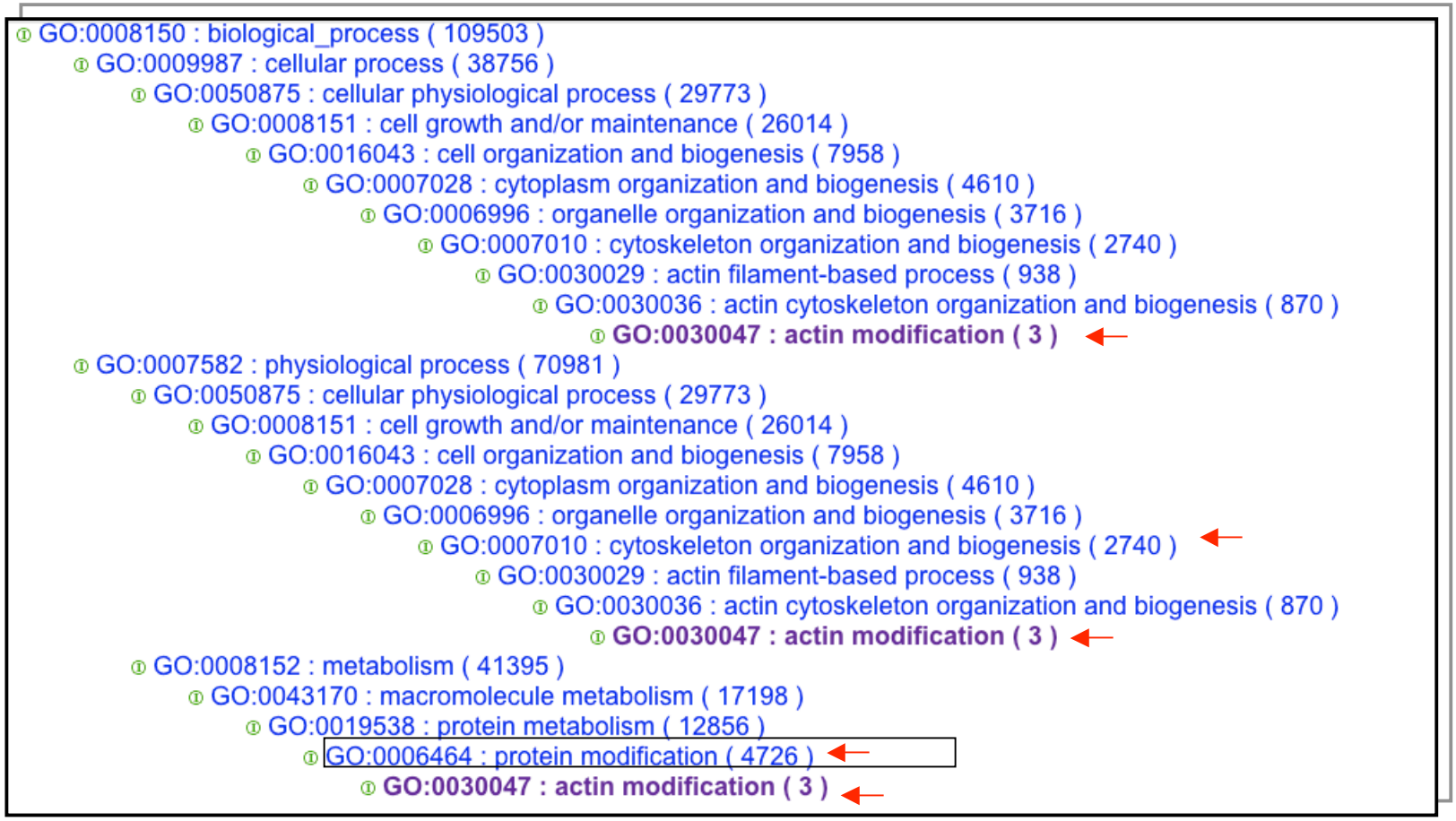
15 column tab delimited text file

Anatomy of a gene association file

Column	Content	Example
1	DB	SGD, MGI
2	**DB_Object ID**	**MGI:1234568**
3	DB_Object_Symbol	Gras3
4	GO_ID Qualifier	NOT, co_localizes_with, contributes_to
5	GO_ID	GO:0001515
6	DB_Ref	PMID:234567
7	Evidence_Code	IDA, etc.
8	With/From	
9	GO_aspect	P (process), C (component) F (function)
10	DB_Object_Name	Grasshopper 3 homlog
11	DB_Object_Synonym	Locust III, 0122345E12Rik
12	DB_Object_Type	Gene, transcript, or protein
13	Taxon	taxon:4932
14	Date	20050101
15	Assigned_by	DB (usually same as column 1)

Some Special Cases

Annotate to finest granularity



Annotating to GO:0030047 automatically annotates to all of its parents; thus a product is annotated to both protein modification AND cytoskeleton organization

GO Does not annotate substrates

- A gene product that has protein kinase activity is also involved in the process of protein phosphorylation
- The protein that gets phosphorylated is NOT involved in the process of protein phosphorylation.

Qualifiers

- GO Term Qualifiers
  - "NOT"
    - Can be used with any term
  - "contributes_to"
    - Used for molecular function
  - "co_localizes with"
    - Used with cellular component
- Evidence Code Qualifiers
  - Sequence ID (for ISS)
  - Protein ID (for IPI and protein binding)
  - Mutant ID (for IMP)
  - Gene (for IGI)
  - GO ID (for IC)

The "not" GO Term Qualifier

ALS2_1018-1657 (final 0–1.6 μM) (Fig. 4B). Both ALS2_1018-1657 and human ALS2CL revealed the protein concentration-dependent Rab5-GEF activities with approximately eightfold lower dissociation constant with human ALS2CL (ALS2_1018-1657; ~25 nM vs. human ALS2CL; ~200 nM). However, mouse ALS2CL did not show any significant Rab5-GEF activities at any concentrations of the protein (Fig. 4B).

3.6. Interaction of ALS2CL and Rab5

To examine whether the ALS2CL proteins directly interact with Rab5, we conducted the in vitro binding assays using the FLAG-M2 pull-down experiments. The amino-terminally FLAG-tagged ALS2_1018-1657, human ALS2CL, and mouse ALS2CL were immunoprecipitated in the presence of re-combinant Rab GDP or

Note 11/15/2004
Harold Drabkin 1:43:47 PM

Authors show that although the human ALS2CL has Rab. GEF activity (GO:0017112), the mouse protein does not. Therefore, a NOT GO:0017112 can be assigned based on direct assay (IDA).

The ‘contributes_to’ qualifier

Contributes_to: An individual gene product that is part of a complex can be annotated to terms that describe the action (function or process) of the complex.

This practice is colloquially known as annotating 'to the potential of the complex‘.

This qualifer allows us to distinguish the individual subunit from complex functions e.g. contributes_to ribosome binding when part of a complex but does not perform this function on its own.

All gene products annotated using 'contributes_to' must also be annotated to a cellular component term representing the complex that possesses the activity.

Only used with GO Function Ontology

The Qualifier documentation:
http://www.geneontology.org/GO.annotation.html

GO:0005515 Protein Binding

GO term:	**protein binding**
Synonym:	**protein amino acid binding**
GO id:	**GO:0005515**
Definition:	**Interacting selectively with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules).**

- Used to annotate a gene product as being able to bind another protein
  - If the target protein is known, then use the IPI evidence code and the UniProt identifier in the "with" field.
  - If the target is not known, then use the IDA evidence code.
- The gene product being annotated does not have to be a protein itself: eg: Rpph1, ribonuclease P RNA component H1, has protein binding activity (GO:0005515)

MGI Accession ID: MGI:2156126
J Number: J:73639
Other Accession IDs:

- 21634895 (MEDLINE)
- 11687586 (PubMed)

Title: RGC-32 increases p34CDC2 kinase activity and entry of aortic smooth muscle
Authors: Badea T; Niculescu F; Soane L; Fosbrink M; Sorana H; Rus V; Shin ML; Rus
Journal: J Biol Chem
Volume: 277
Issue: 1
Date: 2002 Jan 4
Year: 2002
Pages: 502-8
Review Status: Peer Reviewed

Author states Orthology

med

in the non-mouse organism

experiment are

paper

MGI Accession ID: MGI:2154458
J Number: J:73065
Title: Gene Ontology Annotation by the MGI Curatorial Staff
Authors: Mouse Genome Informatics Scientific Curators
Journal: Not Applicable
Volume:
Issue:
Date: 2001 Dec
Year: 2001
Pages:
Review Status: Reviewed by MGI Editorial Staff

Abstract:

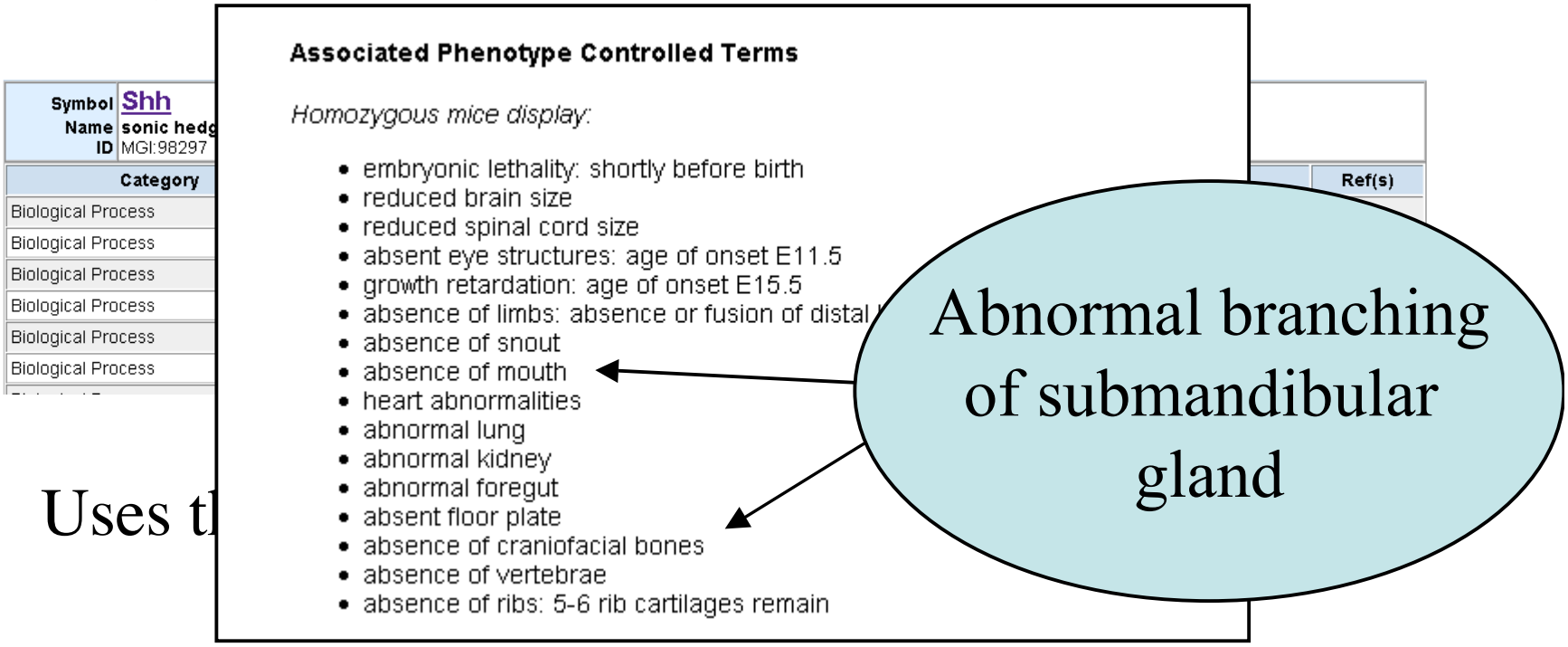
The sequence conservation that permits the establishment of orthology between mouse and rat or mouse and human genes is a strong predictor of the conservation of function for the gene product across these species. Therefore, in instances where a mouse gene product has not been functionally characterized, but its human or rat orthologs have, Mouse Genome Informatics (MGI) curators append the GO terms associated with the orthologous gene(s) to the mouse gene. Only those GO terms assigned by experimental determination to the ortholog of the mouse gene will be adopted by MGI. GO terms that are assigned to the ortholog of the mouse gene computationally (i.e. IEA), will not be transferred to the mouse ortholog. The evidence code represented by this citation is Inferred by Sequence Similarity (ISS.)

MGI curates Orthology

195178 | 1

IMP:Inferred from mutant phenotype

- Mostly used in inferring function from knock-out mice



- Uses t

Inferred from Curator (IC)

Used where an annotation is not supported by any evidence, but can be reasonably inferred by a curator from other GO annotations, for which evidence is available.

The 'with' field is required, and is populated by a GO id using the same reference

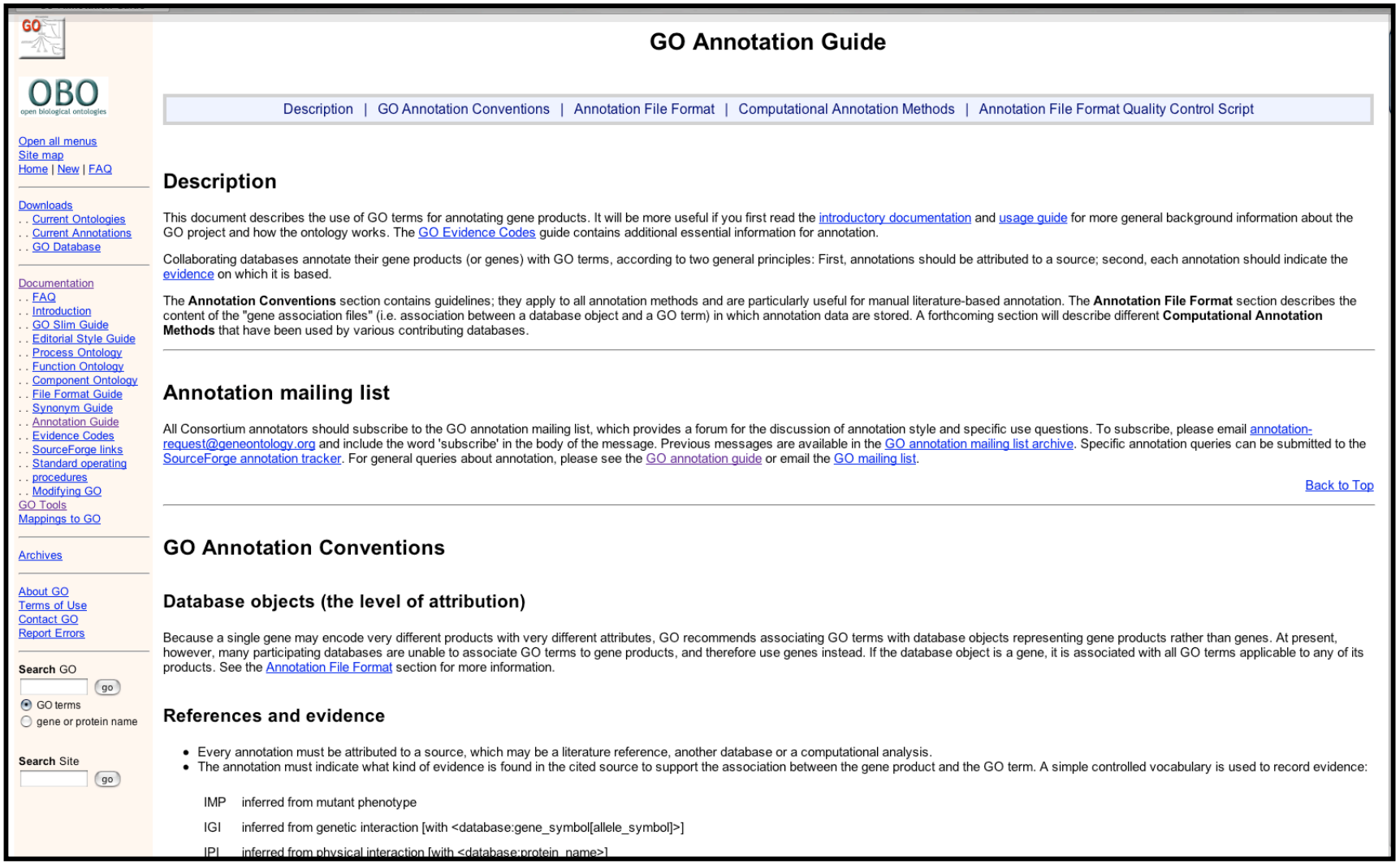
Example: Ref. 1 shows that a gene product has chloride channel activity (GO:0005254:) by direct assay (IDA). A curator can then add the component annotation 'integral to membrane' (GO:0016021) using the IC evidence code and put GO:0005254 in the "with" field.

Caution: The IC evidence code should not be used for something obvious. For example, if a gene product is being annotated to the function "protein kinase activity" (GO:0004672) by IDA, then it is also involved in the process "protein amino acid phosphorylation" (GO:0006468) by the same experiment (IDA).

Unknown v.s. Unannotated

- GO has three terms to be used when the curator has determined that there is no existing literature to support an annotation.
  - Biological_process unknown GO:0000004
  - Molecular_function unknown GO:0005554
  - Cellular_component unknown GO:0008372
- These are NOT the same as having no annotation at all.
  - No annotation means that no one has looked yet.

http://www.geneontology.org/GO.annotation.html



GO

OBO
open biological ontologies

Open all menus
Site map
Home | New | FAQ

Downloads
. . Current Ontologies
. . Current Annotations
. . GO Database

Documentation
. . FAQ
. . Introduction
. . GO Slim Guide
. . Editorial Style Guide
. . Process Ontology
. . Function Ontology
. . Component Ontology
. . File Format Guide
. . Synonym Guide
. . Annotation Guide
. . Evidence Codes
. . SourceForge links
. . Standard operating
. . procedures
. . Modifying GO
GO Tools
Mappings to GO

Archives

About GO
Terms of Use
Contact GO
Report Errors

Search GO
go
GO terms
gene or protein name

Search Site
go

GO Annotation Guide

Description | GO Annotation Conventions | Annotation File Format | Computational Annotation Methods | Annotation File Format Quality Control Script

Description

This document describes the use of GO terms for annotating gene products. It will be more useful if you first read the introductory documentation and usage guide for more general background information about the GO project and how the ontology works. The GO Evidence Codes guide contains additional essential information for annotation.

Collaborating databases annotate their gene products (or genes) with GO terms, according to two general principles: First, annotations should be attributed to a source; second, each annotation should indicate the evidence on which it is based.

The **Annotation Conventions** section contains guidelines; they apply to all annotation methods and are particularly useful for manual literature-based annotation. The **Annotation File Format** section describes the content of the "gene association files" (i.e. association between a database object and a GO term) in which annotation data are stored. A forthcoming section will describe different **Computational Annotation Methods** that have been used by various contributing databases.

Annotation mailing list

All Consortium annotators should subscribe to the GO annotation mailing list, which provides a forum for the discussion of annotation style and specific use questions. To subscribe, please email annotation-request@geneontology.org and include the word 'subscribe' in the body of the message. Previous messages are available in the GO annotation mailing list archive. Specific annotation queries can be submitted to the SourceForge annotation tracker. For general queries about annotation, please see the GO annotation guide or email the GO mailing list.

Back to Top

GO Annotation Conventions

Database objects (the level of attribution)

Because a single gene may encode very different products with very different attributes, GO recommends associating GO terms with database objects representing gene products rather than genes. At present, however, many participating databases are unable to associate GO terms to gene products, and therefore use genes instead. If the database object is a gene, it is associated with all GO terms applicable to any of its products. See the Annotation File Format section for more information.

References and evidence

- Every annotation must be attributed to a source, which may be a literature reference, another database or a computational analysis.
- The annotation must indicate what kind of evidence is found in the cited source to support the association between the gene product and the GO term. A simple controlled vocabulary is used to record evidence:

IMP	inferred from mutant phenotype
IGI	inferred from genetic interaction [with <database:gene_symbol[allele_symbol]>]
IPI	inferred from physical interaction [with <database:protein_name>]