

Gene Ontology (GO) Tutorial

In this tutorial you will:

- visit the Gene Ontology (GO) site and see the principal components available to GO users;
- visit Mouse Genome Informatics (MGI) and explore the GO-related aspects of MGI;
- visit several other model organism sites and see how these show GO information;
- try some GO tools that extend the use of the GO from an annotation resource to a research/analysis tool.

Visiting the Gene Ontology (GO) site.

Go to the GO site: www.geneontology.org

GENE ONTOLOGY™ CONSORTIUM

[What is the Gene Ontology?](#) [Download the Ontologies](#)

The goal of the Gene Ontology™ (GO) Consortium is to produce a controlled vocabulary that can be applied to all organisms even as knowledge of gene and protein roles in cells is accumulating and changing. GO provides three structured [networks](#) of defined terms to describe gene product attributes. GO is one of the controlled vocabularies of the [Open Biological Ontologies](#).

- Submit new GO term suggestions via the [Curator Requests Tracker](#) at [SourceForge](#). [Help with new term submission](#) is available.
- Send comments and questions to: go@geneontology.org.

Search Terms and Annotations

This search uses the [AmiGO](#) browser. You can also use one of the many other [GO Browsers](#).

The table of contents on the left contains links to three objects of particular interest here:

- [Current Ontologies](#)
- [Current Annotations](#)
- [Documentation](#).

NOTE: It is important to distinguish the ontologies (developed as organism-independent structured vocabularies) from the annotations (organism-specific and describing gene product molecular function, biological process and cellular component using the GO). In some sense, the GO provides the tree on which biologists can hang their organism's gene products.

Exploring the ontologies that you can use.

The **Current Ontologies** are available at:

<http://www.geneontology.org/doc/index.shtml#downloads>

Download the Ontologies

The three organizing principles of GO are **molecular function**, **biological process** and **cellular component**. Links to the most up to date versions of the TEXT are provided below, adjacent to the descriptions of the scope of each ontology. Definitions of the terms within all three of the ontologies are contained in a single definitions file, available as a text file below. XML formatted versions, containing all three ontology files and all available definitions, are also available. On the first of each month, the current ontology flat files and XML format are saved to archives; the links are below. For more information, see the [general documentation](#). All of these files can also be downloaded from our anonymous [FTP site](#).

The gene ontologies in flatfile format	Molecular Function text file (7045 terms as of September 16, 2003) Biological Process text file (7763 terms as of September 16, 2003) Cellular Component text file (1335 terms as of September 16, 2003) ← Term Definition text file
XML format	On a monthly basis, XML Format Files are generated. Two files are available, one with gene associations and one without. Specific information on the file contents are available from the download page.
GO Database	API documentation, schema diagrams and full descriptions of all tables for the MySQL database developed and maintained by BDGP.

You can download the three gene ontologies (**Molecular Function; Biological Process, Cellular Component**) in different formats (**flatfile text; XML; MySQL database**) but all formats contain the same information. *Note that some of the content may vary slightly since these files are frequently updated by different sources but not on the same schedules.*

Try looking at the flatfile text version of the cellular component ontology, since it is the smallest (1335 terms as of September 16, 2003):

<http://www.geneontology.org/ontology/component.ontology>

Note the format:

```
!autogenerated-by:    DAG-Edit version 1.320
!saved-by:           gwg
!date:              Fri Oct 17 12:41:10 BST 2003
!version: $Revision: 2.402 $
!type: % ISA Is a
!type: < PARTOF Part of
$Gene_Ontology ; GO:0003673
<cellular_component ; GO:0005575
  %cell ; GO:0005623
  <bud ; GO:0005933
```

After some version information, there is an indication that

- “%” represents the “ISA” relationship of a child node to parent node and
- “<” represents the “PART OF” relationship.

The ontology proper begins with the root node “\$Gene_Ontology” with GO id GO:0003673. Then observe that “cellular component” with GO id number GO:0005575

is PARTOF “Gene_Ontology” and that “cell” with GO id number GO:0005623 ISA cellular component. “bud” is PART OF “cell” and so on.

Using the AmiGO Browser to observe the same ontology structure.



Another tool to explore the structure of the GO is the GO Browser, for example, AmiGO: <http://www.godatabase.org/cgi-bin/go.cgi>



AmiGO is an HTML-based browser, which you can use to browse and search both the gene ontologies and the gene associations (annotations, see below). Each GO term gets one horizontal line. The first thing on each line can be:

- + , which shows that the node can be expanded to show all children of the selected node
- , which can be used to collapse the node hiding its children
- (dot) icon, which shows that the node has no children.

The next thing on each line can be either a

-  which represents the PARTOF relationship
-  which represents the ISA relationship.

In the screenshot presented here, you see the same information as in the flatfile format that you looked at first, with “bud” a part of “cell” which is a “cellular component.” Following each term is a number in parentheses. This tells the total number of gene products (in the associated database) that have been annotated to this GO term or to its descendants, the more specific terms below this in the GO tree.

Similarly, MGI has a GO browser:

http://www.informatics.jax.org/searches/GO_form.shtml

that you can use to explore the GO and observe the current number of MGI annotations to various GO terms. Click the “?” icon at the top of the browser page for additional documentation *Using The Gene Ontology (GO) Browser*.

Gene Ontology Browser
Term Detail

GO term: **cell**
GO id: **GO:0005623**
Definition: **The basic structural and functional unit of all organisms. Includes the plasma membrane and any external encapsulating structures such as the cell wall and cell envelope.**

Number of paths to term: 1

Ⓞ denotes an 'is-a' relationship
Ⓟ denotes a 'part-of' relationship

Gene_Ontology
 @cellular component
 @cell [GO:0005623] (7814 genes, 14935 annotations)
 @bud +
 @cell fraction +

We will do more with GO browsers later in the tutorial.

Exploring the annotations that you can use.

The **Current Annotations** are available at:

<http://www.geneontology.org/doc/GO.current.annotations.shtml>

Current Annotations

What are IEA Codes? | View the Terms and Annotations

This table shows the number of gene products that have been annotated to the gene ontologies by each collaborating group. A gene product can have one or more molecular functions, be used in one or more biological processes and may be associated with one or more cellular components. Tab-delimited files of the associations between gene products and GO terms made by the member organizations are available from the FTP site or from the links in this table. The file format is described in the Annotation Guide. Any errors or omissions in annotations should be reported by writing to the GO mailing list: go@geneontology.org

Notes:
 1) "View" links usually open the file as a window in your browser; "Download" links should download the file to your disk. It is the same file in either case.
 2) For the CompuGen and GO Annotations at EBI files, please also see the appropriate README file

	Biological Process		Molecular Function		Cellular Component		Total Gene Products Associated	Total References Included as Evidence	TAB Delimited File of Associations & Last Update
	All codes	non-IEA codes	All codes	non-IEA codes	All codes	non-IEA codes			
SGD <i>Saccharomyces cerevisiae</i>	6446	6446	6434	6434	6435	6435	6448	4418	Download View Oct 2, 2003
FlyBase <i>Drosophila melanogaster</i>	4439	4428	6795	6789	3942	3918	7938	5923	Download View Aug 29, 2003
MGI <i>Mus musculus</i>	9594	5776	10523	6642	9691	7300	12694	3427	Download View Oct 17, 2003
TAIR <i>Arabidopsis thaliana</i>	6727	1983	7786	5454	13546	1894	18495	2103	Download View Oct 23, 2003
WormBase	2115	1550	5754	395	2055	550	5076	215	Download ...

The current annotations **for many organisms** are available here that you can use in conjunction with the ontologies. Notice that the total number of gene products associated varies significantly from organism to organism (column indicated by vertical arrow).

Why do you think this is the case?

For example, SGD *Saccharomyces cerevisiae* (yeast) shows 6448; MGI *Mus musculus* (mouse) shows 12694; ZFIN *Danio rerio* (zebrafish) shows only 983. These differences reflect both the differences in the number of genes in each organism and the state of annotation for that organism. The yeast genome has been essentially completely annotated to GO; the mouse genome is very well annotated for some 12,000 of its approximately 30,000 genes; the GO annotation work on the zebrafish is in its beginning stages.

Click to view (horizontal arrow) the MGI mouse annotations. This is a large file that takes a long time to download. You can stop the download after a few seconds in order and see only the beginning of the file.



```
!software version: $Revision: 1.130 $
!date: 10/16/2003 $
!
! from Mouse Genome Database (MGD) & Gene Expression Database (GXD)
!
MGI      MGI:1915562      0610005A07Rik      GO:0016740      MGI:MGI:2429377      ISS      SPTR:P08009      F      ←
MGI      MGI:1915562      0610005A07Rik      GO:0004364      MGI:MGI:2429377      ISS      SPTR:P08009      F
MGI      MGI:1915562      0610005A07Rik      GO:0006803      MGI:MGI:2429377      ISS      SPTR:P08009      P
MGI      MGI:1918914      0610006F02Rik      GO:0008168      MGI:MGI:2429377      ISS      INTERPRO:IPRO01601F
MGI      MGI:1918914      0610006F02Rik      GO:0008757      MGI:MGI:2429377      ISS      INTERPRO:IPRO00051F
MGI      MGI:1918920      0610006H10Rik      GO:0016787      MGI:MGI:1354194      IEA      F      RIKEN (
MGI      MGI:1914086      0610006I08Rik      GO:0016021      MGI:MGI:2429377      TAS      C      RIKEN (
MGI      MGI:1923502      0610006O14Rik      GO:0016021      MGI:MGI:2429377      TAS      C      RIKEN (
MGI      MGI:1915462      0610007H07Rik      GO:0016021      MGI:MGI:1354194      IEA      C      RIKEN (
MGI      MGI:1915462      0610007H07Rik      GO:0016021      MGI:MGI:2429377      TAS      C      RIKEN (
MGI      MGI:1918910      0610007L05Rik      GO:0005615      MGI:MGI:2429377      TAS      C      RIKEN (
```

Shown here is a part of the MGI annotations file. The first line shows that the gene product of the gene with MGI accession number MGI:1915562 is annotated to the molecular function ontology (F) term with GO id GO:0016740.

This GO id corresponds to the GO term “transferase activity.” You will not see the term itself here but you can find it on the MGI GO browser page http://www.informatics.jax.org/searches/GO_form.shtml if you enter the GO id or you can find it using the search tool on the main MGI page <http://www.informatics.jax.org/> if you enter the GO id and select accession Ids from the list.

Exploring the documentation that you can use.

An introduction to Gene Ontology: <http://www.geneontology.org/doc/GO.doc.html> is a good place to start in the Gene Ontology Documentation.

[Open all menus](#)
[Site map](#)
[Home](#) | [New](#) | [FAQ](#)

[Downloads](#)
[Current Ontologies](#)
[Current Annotations](#)
[GO Database](#)

[Documentation](#)
[FAQ](#)
[Introduction](#)
[Editorial Style Guide](#)
[File Format Guide](#)
[Synonym Guide](#)
[Annotation Guide](#)
[Evidence Codes](#)
[GNU CVS manual](#)
[Modifying GO](#)
[GO Tools](#)
[Mappings to GO](#)

[Archives](#)

An introduction to Gene Ontology

[What does the Gene Ontology Consortium do?](#) | [What GO is NOT](#) | [The ontologies](#) | [Gene products](#) | [Molecular function](#) | [Biological process](#) | [Cellular component](#) | [What do the ontologies look like?](#) | [Annotation and tools](#) | [File formats](#) | [Beyond GO](#) | [Cross-products](#) | [Mappings to other classification systems](#) | [Contributing to GO](#)

does the Gene Ontology Consortium do?

Biologists currently waste a lot of time and effort in searching for all of the available information about each small area of research. This is hampered further by the wide variations in terminology that may be common usage at any given time, and that inhibit effective searching by computers as well as people. For example, if you were searching for new targets for antibiotics, you might want to find all the gene products that are involved in bacterial protein synthesis, and that have significantly different sequences or structures from those in humans. But if one database describes these molecules as being involved in 'translation', whereas another uses the phrase 'protein synthesis', it will be difficult for you — and even harder for a computer — to find functionally equivalent terms.

The Gene Ontology (GO) project is a collaborative effort to address the need for consistent descriptions of gene products in different databases. The project began as a collaboration between three model organism databases: [FlyBase](#) (*Drosophila*), the [Saccharomyces Genome Database](#) (SGD) and the [Mouse Genome Database](#) (MGD) in 1998. Since then, the GO Consortium has grown to include many databases, including several of the world's major repositories for plant, animal and microbial genomes. See the [GO web page](#) for a full list of member organizations.

The GO collaborators are developing three structured, controlled vocabularies (ontologies) that describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner. There are three separate aspects to this effort: first, we write and maintain the ontologies themselves; second, we make associations between the ontologies and the

An important aspect of the GO is that the GO Consortium has grown to include many databases, including several of the world's major repositories for plant, animal and microbial genomes: <http://www.geneontology.org/doc/GO.consortiumlist.html>
 You will find more documentation on the use of the GO at many of those sites.

GO at MGI

Each Gene Detail page has a section that provides a brief summary of the GO annotation for a particular gene. Go to the MGI home page (www.informatics.jax.org) and search on the gene "Pax6". The Gene Detail page is seen in the figure on the right. The arrow points to the GO section. This gene product has multiple annotations in each of the three ontologies. There are actually 19 annotations for this gene. Clicking on the "19" brings up a summary page for the annotations (Gene Ontology Classifications).

Each annotation lists the ontology (Biological Process, Cellular Component, or Molecular Function), GO term (hyperlinked to the GO browser), evidence code, an optional "inferred from" field, and reference. There is a table at the bottom that indicates what the evidence codes stand for. The "inferred from" field is filled in for IMP, IGI, ISS, and IPI evidence codes.

- For IMP and IGI codes, the field contains an MGI accession number for another gene (IGI) or a mutant allele (IMP). In the latter case, this is hyperlinked to the mutant allele).
- When the evidence is an ISS, the "inferred from" field contains a Genbank or SwissProt accession number; if IPI, the field contains a SwissProt accession number.

In some instances, there are multiple annotations using the same term. That is because the evidence for these annotations is different; they are displayed individually (e.g., DNA binding or eye morphogenesis). Some annotations have more than one reference (e.g. brain development). This means that there are two references that use similar evidence for the annotation. To recap; Each complete annotation links a gene product to a GO term with a statement of evidence and a citation.

Category	Classification Term	Evidence	Inferred From	Refs
Biological Process	anterior/posterior pattern formation	IMP	MGI:1856155	1
Biological Process	brain development	IMP	MGI:1856155	2
Biological Process	cell fate determination	IMP	MGI:1939347	1
Biological Process	development	IEA		1
Biological Process	eye morphogenesis	IGI	MGI:108564	1
Biological Process	eye morphogenesis	IMP	MGI:1934348	1
Biological Process	eye morphogenesis (sensu Mammalia)	IMP	MGI:1939347	1
Biological Process	metanephros development	IMP		1
Biological Process	regulation of cell migration	IMP	MGI:1856155	1
Biological Process	regulation of transcription, DNA-dependent	IEA		1
Biological Process	regulation of transcription, DNA-dependent	TAS		1
Biological Process	salivary gland morphogenesis	IMP	MGI:1856158	1
Biological Process	transcription	IEA		1
Cellular Component	nucleus	IDA		1
Cellular Component	nucleus	IEA		1
Cellular Component	transcription factor complex	TAS		1
Molecular Function	DNA binding	IDA		1
Molecular Function	DNA binding	IEA		1
Molecular Function	transcription factor activity	TAS		1

Gene Ontology Evidence Code Abbreviations:

- IC Inferred by curator
- IDA Inferred from direct assay
- IEA Inferred from electronic annotation
- IGI Inferred from genetic interaction
- IMP Inferred from mutant phenotype
- IPI Inferred from physical interaction
- ISS Inferred from sequence or structural similarity
- ND No biological data available
- TAS Traceable author statement

Clicking the number in the reference column brings up the reference list for the annotation. Clicking on a link in that window brings up a reference detail page displaying the reference, abstract, and additional information.

Clicking on the term "eye morphogenesis" brings up the MGI GO Browser which displays the GO term with its GO_ID and definition, number of paths to the term (in this case 1), along with a view of its placement within the GO Process Ontology. The number of genes in MGI annotated to it or its children is also indicated. (*remember: The green "I" means that the term is a type its parent; a red "P" means that the term is a part of the process described by its parent term.*)

Gene Ontology Browser
Term Detail

GO term: **eye morphogenesis**
GO id: **GO:0001654**
Definition: **Formation and development of the eye, the organ of sight.**
Number of paths to term: **1**

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

Gene_Ontology
@biological_process
@development
@morphogenesis
@organogenesis
@adrenal_gland_development
@blood_vessel_development +
@determination_of_organ_boundary
@ear_morphogenesis +
@ectodermal_gut_morphogenesis +
@embryonic_development
@eye_morphogenesis (GO:0001654) (14 genes, 15 annotations)
@embryonic_eye_morphogenesis +
@eye_morphogenesis (sensu Drosophila) +
@eye_photoreceptor_cell_development +
@retinal_programmed_cell_death +
@fat_body_development +
@gonad_development +
@heart_development +
@hemopoiesis +

Clicking on the "14 genes..." brings up a list of all of the genes in MGI annotated to this term. Here we see the 15 annotations (note Pax6 is annotated twice). Some of the genes are annotated to the actual term "eye morphogenesis", whereas some, such as Gabrr2 or Rprgr are annotated to a child of "eye morphogenesis".

Gene Ontology Annotations
Query Results - Summary

15 matching items displayed

Searched Term: eye morphogenesis

Symbol, Name	Category	Annotated Term	Evidence	Ref(s)
Bhl1, blind	P	eye morphogenesis	TAS	1
Bmpr1b, bone morphogenetic protein receptor, type 1B	P	eye morphogenesis	IMP	1
Foxe3, forkhead box E3	P	eye morphogenesis	IMP	1
Gabrr2, gamma-aminobutyric acid (GABA-C) receptor, subunit rho 2	P	eye photoreceptor cell development	IMP	1
Mgr1, myeloid ectopic viral integration site-related gene 1	P	eye morphogenesis	IGI	1
Neurod1, neurogenic differentiation 1	P	eye morphogenesis	IMP	1
Neurod4, neurogenic differentiation 4	P	eye morphogenesis	IMP	1
Pax6, paired box gene 6	P	eye morphogenesis	IGI	1
Pax6, paired box gene 6	P	eye morphogenesis cell development	IMP	1
Rprgr, retinitis pigmentosa GTPase regulator	P	eye photoreceptor cell development	IMP	1
Rprgr1, retinitis pigmentosa GTPase regulator interacting protein 1	P	eye photoreceptor cell development	IMP	1
Shh, sonic hedgehog	P	eye morphogenesis	IDA	1
Six3, sine oculis-related homeobox 3 homolog (Drosophila)	P	eye morphogenesis	IDA	1
Tm4sf9, transmembrane 4 superfamily member 9	P	mystery cell fate differentiation (sensu Drosophila)	ISS	1
Wt1, Wilms tumor homolog	P	eye morphogenesis	IMP	1

Now, suppose you would like to know what genes in other organisms have been found to have something to do with "eye morphogenesis".

To do this, the best place to start might be at the GO database itself, using the Amigo Browser. <http://www.godatabase.org>

AmiGO: Your friend in the Gene Ontology.

Search GO: eye morphogenesis [Exact Match] [Submit]

Top Docs Gene Ontology GO Links GO Summary

GO:0001654 : Gene Ontology (103367)

- GO:0008150 : biological_process (68451)
- GO:0005575 : cellular_component (54946)
- GO:0003674 : molecular_function (75116)

Get this tree as RDF XML.
Get this data as a GO flat file.
Get a bookmarkable url of this tree.

Submit GO term or definition request.
Submit AmiGO bug report.

Copyright The Gene Ontology Consortium. All rights reserved.

The browser returns a page that lists a number of terms with the phrase "eye morphogenesis" in it.

AmiGO Search GO: eye morphogenesis [Exact Match] [Submit]

Top Docs Gene Ontology GO Links GO Summary

GO Term Name:	Definition:
compound eye morphogenesis (sensu Drosophila)	The stages of eye morphogenesis that are involved in forming the adult eye, a precise assem
embryonic eye morphogenesis	The stages of eye morphogenesis that take place in the embryo phase, and that are involved
embryonic eye morphogenesis (sensu Drosophila)	The stages of eye morphogenesis in Drosophila that take place in the embryo phase and that
eye morphogenesis	Formation and development of the eye, the organ of sight.
eye morphogenesis (sensu Drosophila)	Formation and development of the eye of Drosophila, the organ of sight in the fly.
eye morphogenesis (sensu Mammalia)	The formation and development of the mammalian eye.
post-embryonic eye morphogenesis	Post-embryonic morphogenesis of the eye, which is the organ of sight.
post-embryonic eye morphogenesis (sensu Drosophila)	Post-embryonic morphogenesis of the eye of Drosophila, the organ of sight in the fly.

[Check/Uncheck All] [Draw New Tree] [Submit]

If we click on the term "eye morphogenesis" we get a long page composed of several sections. The first section (1) details the GO term, its definition and its parents. There is a section that allows one to filter the annotations according to the database of origin, as well as by the evidence type (2). This section is then followed by a list of the annotations to the term (eye morphogenesis) and its children (3). A small section is shown.

We see that there are several entries from MGI as well as SwissProt/TrEmbl. Note that the entries for Pax6 from MGI look a bit different in several respects.

1. The Amigo browser does not display annotations with IEA evidence codes.
2. When there are multiple annotations with the same evidence code, only one is displayed (check the link for Pax6 using the IMP evidence code; only one of the two papers is displayed).
3. Instead of "inferred from", a "with" qualifier is used, but the display of associated content of the field is not yet implemented.

Clicking on a Gene Symbol brings up a summary page with all of the GO terms annotated to the gene or protein. Again, note the differences between this display and the GO summary page from MGI (for example, evidence code displayed but not the citations).

Going back to the gene list, clicking on the database link in an annotation line (such as "MGI" in Pax6) brings you to the database's detail page for that entry (see beginning of this section). If one clicks on the evidence code instead of the gene symbol, one is taken directly to an MGI page displaying the reference, abstract, and additional information.

eye morphogenesis

Accession:GO:0001654

Synonyms: GO:0042460

Definition: Formation and development of the eye, the organ of sight.

Term Lineage [Graph view](#)

GO:0003673 : Gene Ontology (103367)
 @ GO:0008150 : biological process (68451)
 @ GO:0007275 : development (8790)
 @ GO:0009633 : morphogenesis (4383)
 @ GO:0009887 : organogenesis (3590)
 @ GO:0001654 : eye morphogenesis (369)

External References

None.

Associated Genes Gene Filters: Filter by database: All, FlyBase, SGD, MGI

Filter by Evidence for Association: Curator Approved, Inferred from Mutant Phenotype, Inferred from Genetic Interaction, Inferred from Physical Interaction

Page 1

GO Term:	Gene Symbol:	Datasource:	Evidence:	Full name:
GO:0001654 : eye morphogenesis				
	<input type="checkbox"/> NDF4_MOUSE ^{03ae}	SPT	IMP	Neurogenic differentiation factor 4
	<input type="checkbox"/> WTL_MOUSE ^{03ae}	SPT	IMP	Wilms' tumor protein homolog
	<input type="checkbox"/> BMRB_MOUSE ^{03ae}	SPT	IMP	Bone morphogenetic protein receptor type 1B precursor
	<input type="checkbox"/> MEI2_MOUSE ^{03ae}	SPT	IGI	Homeobox protein Meis2
	<input type="checkbox"/> NDFL_MOUSE ^{03ae}	SPT	IMP	Neurogenic differentiation factor 1
	<input type="checkbox"/> SHH_MOUSE ^{03ae}	SPT	IDA	Sonic hedgehog protein precursor
	<input type="checkbox"/> SIX3_MOUSE ^{03ae}	SPT	IDA	Homeobox protein SIX3
	<input type="checkbox"/> Q8C765	SPT	IDA	Sonic hedgehog homolog
	<input type="checkbox"/> Q8CEI5	SPT	IGI	Paired box gene 6
	<input type="checkbox"/> Q8CEI7	SPT	IMP	Neurogenic differentiation 1
	<input type="checkbox"/> Q8VBX9	SPT	IGI	Pax6 paired-less isoform
	<input type="checkbox"/> Q8VBY9	SPT	IGI	Paired box protein
	<input type="checkbox"/> Q8VBZ1 ^{03ae}	SPT	IGI	Pax6 protein
	<input type="checkbox"/> Q921Q8 ^{03ae}	SPT	IGI	Unknown
	<input type="checkbox"/> EXE3_MOUSE ^{03ae}	SPT	IMP	Forkhead box protein E3
	<input type="checkbox"/> Bld	MGI	TAS	blind
	<input type="checkbox"/> Bmp1b ^{03ae}	MGI	IMP	bone morphogenetic protein receptor, type 1B
	<input type="checkbox"/> Foxe3 ^{03ae}	MGI	IMP -With	forkhead box E3
	<input type="checkbox"/> Mgl1 ^{03ae}	MGI	IGI -With	myeloid ectropic viral integration site-related gene 1
	<input type="checkbox"/> Neurod1 ^{03ae}	MGI	IMP	neurogenic differentiation 1
	<input type="checkbox"/> Neurod4 ^{03ae}	MGI	IMP -With	neurogenic differentiation 4
	<input type="checkbox"/> Pax6	MGI	IGI -With	paired box gene 6
	<input type="checkbox"/> Pax6	MGI	IMP -With	paired box gene 6
	<input type="checkbox"/> Shh ^{03ae}	MGI	IDA	sonic hedgehog
	<input type="checkbox"/> Six3 ^{03ae}	MGI	IDA	sine oculis-related homeobox 3 homolog (Drosophila)
	<input type="checkbox"/> Wt1 ^{03ae}	MGI	IMP -With	Wilms tumor homolog
GO:0001744 : optic placode formation (sensu Drosophila)				
	<input type="checkbox"/> eya	FlyBase	IMP	eya
	<input type="checkbox"/> so	FlyBase	IMP	so

Gene Product: Pax6

Full Name: paired box gene 6

Synonyms: None

Data Source: MGI

Associated To Terms:

Term Name	Association Evidence	Sequence Similarities
<input type="checkbox"/> GO:0009952 : anterior/posterior pattern formation	Tree View Inferred from Mutant Phenotype	1856155
<input type="checkbox"/> GO:0007420 : brain development	Tree View Inferred from Mutant Phenotype	1856155
<input type="checkbox"/> GO:0007230 : brain development	Tree View Inferred from Mutant Phenotype	1856155
<input type="checkbox"/> GO:0001709 : cell fate determination	Tree View Inferred from Mutant Phenotype	1939347
<input type="checkbox"/> GO:0003677 : DNA binding	Tree View Inferred from Direct Assay	
<input type="checkbox"/> GO:0001654 : eye morphogenesis	Tree View Inferred from Genetic Interaction	108564
<input type="checkbox"/> GO:0001654 : eye morphogenesis	Tree View Inferred from Mutant Phenotype	1934348
<input type="checkbox"/> GO:0001747 : eye morphogenesis (sensu Mammalia)	Tree View Inferred from Mutant Phenotype	1939347
<input type="checkbox"/> GO:0001656 : metanephros development	Tree View Inferred from Direct Assay	
<input type="checkbox"/> GO:0005634 : nucleus	Tree View Inferred from Mutant Phenotype	1856155
<input type="checkbox"/> GO:0003034 : regulation of cell migration	Tree View Traceable Author Statement	
<input type="checkbox"/> GO:0006355 : regulation of transcription, DNA-dependent	Tree View Traceable Author Statement	
<input type="checkbox"/> GO:0007435 : salivary gland morphogenesis	Tree View Inferred from Mutant Phenotype	1856158
<input type="checkbox"/> GO:0003700 : transcription factor activity	Tree View Traceable Author Statement	
<input type="checkbox"/> GO:0005667 : transcription factor complex	Tree View Traceable Author Statement	

No peptide sequence available.

Amigo draws on the gene association files of some 13 databases encompassing animals, plants, and various bacteria. Let's shorten the list by filtering the list (section 2). Select Fly, and MGI from the database dropdown (using either Ctrl-click (Windows) or Command-click (Mac), and select only those annotations coming from direct assays (IDA), and click on the "filter.." button. Amigo shows that only three entries satisfy this query: two genes from MGI, and one from Flybase.

GO Term:
Gene Symbol: **Datasource:** **Evidence:** **Full name:**
 GO:0001654 : eye morphogenesis
 Shh^{C3ae} MGI IDA sonic hedgehog
 Six3^{C3ae} MGI IDA sine oculis-related homeobox 3 homolog (Drosophila)
 GO:0007456 : eye morphogenesis (sensu Drosophila)
 fng FlyBase IDA fng
 Previous Page Next Page First Page All Gene Products

The GO BLAST SERVER:

The Amigo browser can also fetch sequences for those gene products whose databases have supplied appropriate SwissProt ID links to the gene products annotated. The product symbols that have a "GOst" icon next to them have sequence information associated with them. If you select these, there is an option at the bottom of the page to collect the selected sequences in a Fasta output. Also, clicking on the GOst icon for a single entry gives a Fasta output screen. Simultaneously, a Blast query is run against the database to find similar sequences. Clicking on the "retrieve your job" link brings up a page that displays the query sequence again, and a table containing gene products that are closely similar to the query sequence and their complete GO annotations. The page also has a standard Blast output with alignments.



AmiGO GOst New GOst Search Last Job Submitted

GOst Results:

Your query sequence:

```
>MGI|98297 symbol:Shh PDB:1VHH SPTR:Q62226 INTERPRO:IPR001657 INTERPRO:IPR001767 InterPro:II
M L L L A R C F L V I L A S S L L V C P G L A C G P G R G F G K R R H P K K L T P L A Y K Q F I P
N V A E K T L G A S G R Y E G K I T R N S E R F K E L T P N Y N P D I I F K D E N T G A D R L M T
Q R C K D K N A L A I S V M N Q P O V K L R V T G W E D G H S E E L S Y E G R A V D I T
T S D R D S K Y M L A R L A V E A G F D W V Y E S K A H I H C S V K A E N S V A A S G G C F
P G S A T V H L E Q G G T K V L D R P G D R V L A A D D Q R R L L Y S D F L T F L D R D E A G
K V F V I E T L E P R E R L L T A A H L L V A P H N D S G P T P G S A L F A S R V R P Q G R
V V V V A E R G D R R L L P A A V H S V T L R E E A G A Y A P L T A G T I L I N R V L A S C Y
A V I E E S W A H R A F A P F R L A H A L L A A L A P A R T D G G G G S I P A A Q S A T E A R G
A E P T A G I H W S Q L L Y H I G T W L L D S E T H P L G M A V K S S
```

Results:

Your job has been submitted to AmiGO. Your results should be ready shortly.

[Retrieve your job.](#)

High Scoring Gene Products:

Gene Product:	Datasource:	Associated To Terms:	Association Evidence:
<input type="checkbox"/> Shh ^{C3ae}	MGI	axon guidance axon guidance cell fate specification central nervous system development extracellular space eye morphogenesis eye morphogenesis (sensu Mammalia) male genital morphogenesis pattern specification regulation of cell proliferation signal transduction	Tree View Inferred from Direct Assay Tree View Inferred from Mutant Phenotype Tree View Inferred from Mutant Phenotype Tree View Traceable Author Statement Tree View Inferred from Direct Assay Tree View Inferred from Mutant Phenotype Tree View Inferred from Mutant Phenotype Tree View Inferred from Direct Assay Tree View Traceable Author Statement
<input type="checkbox"/> SHH_HUMAN ^{C3ae}	SPTtr	mesoderm cell fate determination ventral midline development	Tree View Traceable Author Statement Tree View Traceable Author Statement
<input type="checkbox"/> IHH_HUMAN ^{C3ae}	SPTtr	cell-cell signaling cholesterol binding extracellular patched binding	Tree View Non-traceable Author Statement Tree View Non-traceable Author Statement Tree View Non-traceable Author Statement
<input type="checkbox"/> hh	FlyBase	analia morphogenesis (sensu Holometabola) anterior/posterior lineage restriction, imaginal disc Bolwig's organ morphogenesis compartment specification compound eye morphogenesis (sensu Drosophila) cysteine-type endopeptidase activity cytoplasm determination of anterior/posterior axis, embryo endopeptidase activity epidermal differentiation extracellular extracellular	Tree View Traceable Author Statement Tree View Traceable Author Statement Tree View Inferred from Mutant Phenotype Tree View Traceable Author Statement Tree View Traceable Author Statement Tree View Inferred from Sequence Similarity Tree View Non-traceable Author Statement Tree View Non-traceable Author Statement Tree View Traceable Author Statement Tree View Non-traceable Author Statement Tree View Traceable Author Statement

```
>SPTR|Q15465 SPTtr:Q15465 symbol:SHH_HUMAN "Sonic hedgehog protein precursor"
SPTtr:Q15465 MEROPS:C46.002 InterPro:IPR000320 InterPro:IPR001767
InterPro:IPR003586 InterPro:IPR003587 InterPro:IPR002203
InterPro:IPR001657 Pfam:PF01079 Pfam:PF01085 PRINTS:PR00632
ProDom:PD003042 SMART:SM00305 SMART:SM0306 PROSITE:PS0017
HSP:Q62226 EMBL:L38518 EMBL:AC002484 Genew:HGNC:10848 MIM:600725
MIM:142945
Length = 462
Score = 1867 (662.3 bits), Expect = 6.3e-194, P = 6.3e-194
Identities = 366/399 (91%), Positives = 371/399 (92%)
Query: 2 L L L L A R C F L V I L A S S L L V C P G L A C G P G R G F G K R R H P K K L T P L A Y K Q F I P N V A E K T L G A S G 61
+ L L L A R C L + + L S S L I V C G L A C C P G R G F G K R R H P K K L T P L A Y K Q F I P N V A E K T L G A S G 60
Sbjct: 1 M L L A R C L L L V S S L L V C S G L A C C P G R G F G K R R H P K K L T P L A Y K Q F I P N V A E K T L G A S G 60
Query: 62 R Y E G K I T R N S E R F K E L T P N Y N P D I I F K D E N T G A D R L M T Q R C K D K N A L A I S V M N Q P O V 121
R Y E G K I + R N S E R F K E L T P N Y N P D I I F K D E N T G A D R L M T Q R C K D K N A L A I S V M N Q P O V 120
Sbjct: 61 R Y E G K I S R N S E R F K E L T P N Y N P D I I F K D E N T G A D R L M T Q R C K D K N A L A I S V M N Q P O V 120
Query: 122 K L R V T E G W E D G H S E E L S Y E G R A V D I T T S D R D S K Y M L A R L A V E A G F D W V Y E S K A H 181
K L R V T E G W E D G H S E E L S Y E G R A V D I T T S D R D S K Y M L A R L A V E A G F D W V Y E S K A H 180
Sbjct: 121 K L R V T E G W E D G H S E E L S Y E G R A V D I T T S D R D S K Y M L A R L A V E A G F D W V Y E S K A H 180
Query: 182 I H C S V K A E N S V A A S G G C F P G S A T V H L E Q G G T K V L D R P G D R V L A A D D Q R R L L Y S D F L T 241
I H C S V K A E N S V A A S G G C F P G S A T V H L E Q G G T K V L D R P G D R V L A A D D Q R R L L Y S D F L T 240
Sbjct: 181 I H C S V K A E N S V A A S G G C F P G S A T V H L E Q G G T K V L D R P G D R V L A A D D Q R R L L Y S D F L T 240
Query: 242 F L D R D G A K K V F V I E T P R E R L L T A A H L L V A P H N D S G P T P G S A L F A S R V R P Q G R 306
F L D R D + G A K K V F V I E T P R E R L L T A A H L L V A P H N D S G P G G G P G G 305
Sbjct: 241 F L D R D D G A K K V F V I E T P R E R L L T A A H L L V A P H N D S A T G E P E A S S G S G P P S G G A L G 286
Query: 287 P S A L F A S R V R P Q G R V V V A E R G D R R L L P A A V H S V T L R E E A G A Y A P L T A G T I L I N R V L 346
P A L F A S R V R P Q G R V V V A E R G D R R L L P A A V H S V T L R E E A G A Y A P L T A G T I L I N R V L 345
Sbjct: 301 P R A L F A S R V R P Q G R V V V A E R G D R R L L P A A V H S V T L R E E A G A Y A P L T A G T I L I N R V L 360
Query: 347 A S C Y A V I E E S W A H R A F A P F R L A H A L L A A L A P A R T D G G 385
A S C Y A V I E E S W A H R A F A P F R L A H A L L A A L A P A R T D G G 384
Sbjct: 361 A S C Y A V I E E S W A H R A F A P F R L A H A L L A A L A P A R T D R G G 399
```

Additional Comments: Other things you might see.

Quite often an experiment is performed whose results indicate that something is NOT involved in a particular process or does NOT have a particular activity. The annotation file structure allows for the use of a “NOT” qualifier. For example, since Pax6 is annotated to “nucleus”, we might want to look at other genes that are annotated to “nucleus” (it’s a big list). One of these is Cnbp. If you look at the Gene Ontology Classifications, you will notice that there are two annotations with an “IDA (inferred from direct evidence) code: one says it is in the nucleus; the other says it is not (a different reference). MGI includes both.

Gene Ontology Classifications				
Symbol	Cnbp			
Name	cellular nucleic acid binding protein			
ID	MGI:88431			
Category	Classification Term	Evidence	Inferred From	Refs)
Biological Process	cholesterol biosynthesis	ISS	SPTR:P20694	1
Biological Process	positive regulation of cell proliferation	IDA		1
Biological Process	positive regulation of transcription from Pol II promoter	IDA		1
Biological Process	regulation of transcription, DNA-dependent	IEA		1
Cellular Component	cytosol	IDA		1
Cellular Component	endoplasmic reticulum	IDA		1
Cellular Component	nucleus	IDA		1
Cellular Component	NOT nucleus	IDA		1
Molecular Function	DNA binding	IEA		1
Molecular Function	nucleic acid binding	IEA		1
Molecular Function	transcription factor activity	ISS	SPTR:P20694	1

Occasionally, one will come across gene products that have been annotated to Process, Function, and/or Component Unknown. These three terms are used when a curator has looked over the available literature and has found that none of it is adequate to assign a term. A good example is Dab2ip. As shown by it’s annotation summary, there was literature to demonstrate that it bound a protein, but no literature to suggest what process it participated in or where in the cell it was found. In some cases, there will be unknowns at the same time as IEA annotation. The IEA annotations in these cases were taken by data loads after the “unknowns” were applied to these genes. These are removed as literature becomes available.

Gene Ontology Classifications				
Symbol	Dab2ip			
Name	disabled homolog 2 (Drosophila) interacting protein			
ID	MGI:1916851			
Category	Classification Term	Evidence	Inferred From	Refs)
Biological Process	biological process unknown	ND		1
Cellular Component	cellular component unknown	ND		1
Molecular Function	protein binding	IPI	SPTR:P97318	1

Visiting several other model organism sites to see how these show GO information.
LocusLink: <http://www.ncbi.nlm.nih.gov/LocusLink/> is a database that provides a single query interface to curated sequence and descriptive information about genetic loci. It presents information on official nomenclature, aliases, sequence accessions, phenotypes, EC numbers, MIM numbers, UniGene clusters, homology, map locations, and related web sites.

Return to the MGI gene detail page for Pax6.

Gene Detail

Symbol Name Pax6
Name paired box gene 6
ID MGI:97490 [Nomenclature History](#)

Synonyms 1500038E17Rik, Dey, Dickie's small eye, Pax-6, Sey, small eye

Map position Chromosome 2
 58.0 cM
[Detailed Map ± 1 cM](#)
 Mapping data(25)

Mammalian orthology human; rat ([Mammalian Orthology](#))
 Comparative Map ([Mouse/Human Pax6 ± 2 cM](#))

Sequences Nucleotide [NM_013627](#) Protein [AAH36957](#)
 All sequences(32)

Other database links
 GXD literature index(211) cDNA source data(21)
 DoTS [DT_40139844](#), [DT_94398348](#), [DT_97362511](#)
 UniGene [3608](#)
 ENSEMBL [ENSMUSG00000027168](#)
 TIGR [TC816043](#), [TC816045](#), [TC854803](#), [TC854804](#), [TC900102](#)
 LocusLink [18508](#)

Protein InterPro ID Description

You will see that there are links to other databases; the last one listed is LocusLink. Click to go to the LocusLink page for mouse Pax6 (<http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=18508>):

NCBI LocusLink

Search: LocusLink | Display: Brief | Organism: All
 Query: | Go | Clear

View: Mm Pax6 | One of 1 Loci | Save All Loci

Click to Display mRNA-Genomic Alignments (spanning 28517 bps)
[PUB](#) [UNIGENE](#) [MAP](#) [VAR](#) [HOMOL](#) [MGI](#) [ef](#) [UCSC](#)

Pax6: paired box gene 6
 LocusID: 18308

Overview

Locus Type: gene with protein product, function known or inferred
Product: paired box gene 6
Alternate Symbols: Dey, Sey, Pax-6, 1500038E17Rik
Alias: small eye
 Dickie's small eye

Function [Submit GeneRIF](#) [\(All Pubs\)](#)
GeneRIF: Gene References into Function:
[11692186](#) • role in embryonic eye development

You can scroll down the page to find the gene ontology information

LocusLink Home Pax6 markedly increased in the latter

Gene Ontology:

Term	Evidence	Source	Pub
• DNA binding	IDA	MGI	pm
• transcription factor activity	TAS	MGI	pm
• brain development	IMP	MGI	pm
• transcription	IEA	MGI	
• regulation of transcription, DNA-dependent	TAS	MGI	pm
• development	IEA	MGI	
• transcription factor complex	TAS	MGI	pm
• nucleus	IDA	MGI	pm
• eye morphogenesis	IMP	MGI	pm
• metamorphosis development	IMP	MGI	pm
• cell fate determination	IMP	MGI	pm
• salivary gland morphogenesis	IMP	MGI	pm
• anterior/posterior pattern formation	IMP	MGI	pm
• regulation of cell migration	IMP	MGI	pm
• eye morphogenesis (sensu Mammalia)	IMP	MGI	pm

Relationships ?

Human Homology Maps:

NCBI vs. MGD	11p13	PAX6	Mm Hs
--------------	-------	----------------------	-----------------------

Map Information ?

and confirm that the annotation is consistent with the annotation given in the MGI gene detail page. Note that you may sometimes observe slight inconsistencies between sites due to the fact that information shared among databases is updated frequently but different databases update on different schedules.

Again from the MGI gene detail page for Pax6, click the link for Mammalian orthology to see if there is an orthologous human gene:

There is a human ortholog, PAX6, as the page below indicates:

Species	Symbol	Chr	AccID	Criteria			
				AA	CL	NS	NT
human	PAX6	11 (p13)	GDB 118997 (GDB-Australia) 3080 (LocusLink)	●	●	●	●
mouse, laboratory	Pax6	2 (58.0 cM)	MGI:97490	●	●	●	●
rat	Pax6	3	41394 (RATMAP)	●	●	●	●
Number of References				2	2	1	1

You can go to LocusLink to examine the information on the human PAX6 gene:

PAX6: paired box gene 6 (aniridia, keratitis)
LocusID: 3080

RefSeq Summary: PAX6 encodes paired box gene 6, one of many human homologues of the *Drosophila melanogaster* gene *prd*. In addition to the hallmark feature of this gene family, a conserved paired box domain, PAX6 also contains a homeo box domain. Both domains are known to bind DNA, and function as regulators of gene transcription. PAX6 is expressed in the developing nervous system, and in developing eyes. Mutations in PAX6 are known to cause aniridia as well as Peter's anomaly, both ocular diseases.

Locus Type: gene with protein product, function known or inferred

Product: paired box gene 6 isoform a
paired box gene 6 isoform b

Alternate Symbols: AN, AN2, MGDA, WAGR, D11S812E

Alias: Paired box homeotic gene-6
paired box homeotic gene 6 (aniridia, keratitis)

Function: Submit GeneRIF (All Pubs)

Phenotype: Aniridia, trvce II

and scroll to see what GO annotations have been made to human PAX6. Note that the gene ontology information is based on annotation from GOA, and that the most current GOA annotation are available from the GO site.

LocusLink Home

PAX6 Index

[Top of Page](#)

[Nomenclature](#)

[Overview](#)

[Function](#)

[Relationships](#)

[Map](#)

[RefSeq](#)

[Related Seqs](#)

[Links](#)

LocusLink

[Collaborators](#)

[Download](#)

[FAQ](#)

[Help](#)

[Statistics](#)

RefSeq

[About](#)

[Download](#)

[FAQ](#)

[Statistics](#)

pathfinding in a way opposite from ROCK (Rho kinase) and that it may act via Pax6 to modulate early neuronal development

Gene OntologyTM:

Term	Evidence	Source	Pub
transcription factor activity	TAS	GOA	pm
eye morphogenesis (sensu Drosophila)	TAS	GOA	pm
central nervous system development	TAS	GOA	pm
vision	TAS	GOA	pm
regulation of transcription, DNA-dependent	IEA	GOA	
histogenesis and organogenesis	TAS	GOA	pm
nucleus	IEA	GOA	
cell differentiation	IEA	GOA	

Relationships [?](#)

Mouse Homology Maps:

NCBI vs. MGD	2 58.00 cM	Pax6	Hs Mm
--------------	------------	----------------------	--------------

Map Information [?](#)

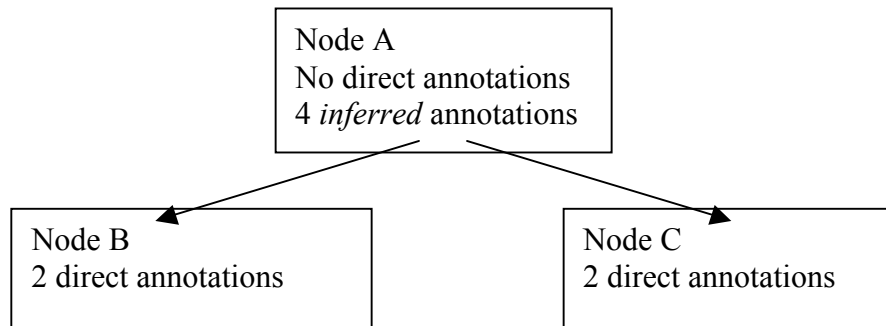
Chromosome: 11 **mv**

Trying some GO tools that extend the use of the GO from an annotation resource to a research/analysis tool.

So far, you have explored the use of the GO as a common resource for annotation of gene products for different organisms. It is also possible to use the GO as a research analysis tool by examining a set of genes, perhaps a set up-regulated in a microarray experiment. One such analysis tool is the GO TermFinder. The GO TermFinder attempts to determine whether an observed level of annotation for a group of genes is significant within the context of annotation for all genes within the genome.

Method/Algorithm Description

To provide the most detailed information available, genes are directly associated with GO terms that are as granular as possible. However, for many purposes, such as analyzing the results of microarray expression data, it is very useful to "calculate" on GO, moving up the GO tree from the specific terms used to annotate the genes in a list to find GO parent terms that the genes may have in common. Because the GO terms have structural relationships with each other, genes are also considered to be indirectly associated with all the parents of the granular terms to which they are directly associated.



To determine significance, the TermFinder examines the group of genes to find GO terms to which a high proportion of the genes are associated compared to the number of times that term is associated with other genes in the genome. For example, if you search the process ontology and find that all of the genes in the group were associated with "DNA repair", this term would be significant. However, since all genes in the genome (with GO annotations) are indirectly associated with the top level term "biological_process", it would not be significant if all the genes in the group were associated with this very high level term. The p-value is a measure of the significance, with smaller p-values indicating greater statistical significance.

Calculating with the MGI GO TermFinder Tool.

Go to the MGI GO Term Finder:

http://www.spatial.maine.edu/~mdolan/MGI_Term_Finder.html

The screenshot shows the 'MGI Gene Ontology Term Finder' web interface. It includes a title, a brief description of the tool's purpose, and four steps for user input. Step 1 involves entering gene names or uploading a file. Step 2 allows selecting an ontology (Process, Function, or Component). Step 3 allows selecting whether to include or exclude IEAs. Step 4 contains 'Search' and 'Reset' buttons.

MGI Gene Ontology Term Finder

This GO Term Finder tool searches for significant shared GO terms, or parents of the GO terms, used to describe the genes in common.

Step 1: Enter gene names*:
Input the gene names in the text box, as MGI.acclID or gene symbol (with carriage returns)
or
Select a file of gene names, as MGI.acclID or gene symbol (with carriage returns).

Step 2: Choose Ontology:
 Process
 Function
 Component

Step 3: Indicate whether to exclude evidence code IEA (Inferred from Electronic Annotation):
 Include IEAs
 Exclude IEAs

Step 4:

The query page allows you to enter the list of gene names and select the ontology that you want to search.

1. Enter gene names:
You can either type the name of the genes in the input box or upload a file that contains the genes names. Note that a long (more than 100 genes) list of genes might take awhile to return your results.
2. Choose ontology:
Select one of the three (biological process, molecular function, or cellular component) ontologies.
3. Indicate whether to include IEAs: The IEA code, Inferred from Electronic Annotation, means no human involvement in the assignment of the association and is thus lowest quality evidence code. IEA is the only code currently in use that does not require human judgment during the curation process. For more details, see the GO evidence code documentation (<http://www.geneontology.org/doc/GO.evidence.html>).
4. Click the Search button.

You will be provided a sample set of mouse genes, `mgf_sample.txt`, to test.

The results page displays in table form, the shared GO terms (or parents of GO terms) used to describe the set of genes. The table lists each GO term, the number of times the GO term is used to annotate genes in the list (or cluster) and the number of times that the term is used to annotate genes in the entire genome. In addition, the p-value is provided as well as a list of all the genes annotated, either directly or indirectly, to the term. You can download the results as an Excel spreadsheet as shown here:

Note: p-values show statistically significant numbers of genes in gene set are annotated (directly or indirectly through inference using DAG GO structure) to some GO classifications. Results are ordered by p-value, most significant results first.

GOLD	GO_term	Frequency	Genome frequency	P-value	Corrected P-value	Gene(s)
GO:0006281	DNA repair	1.00000	0.00333	5.58843E-213	2.12360E-211	MGI:1261809,MGI:1197010,MGI:109768
GO:0009719	response to endogenous stim	1.00000	0.00387	4.36646E-204	1.65926E-202	MGI:1261809,MGI:1197010,MGI:109768
GO:0006974	response to DNA damage stir	1.00000	0.00387	4.36646E-204	1.65926E-202	MGI:1261809,MGI:1197010,MGI:109768
GO:0006259	DNA metabolism	1.00000	0.01033	3.50666E-160	1.33253E-158	MGI:1261809,MGI:1197010,MGI:109768
GO:0006950	response to stress	1.00000	0.02197	4.36099E-132	1.65718E-130	MGI:1261809,MGI:1197010,MGI:109768
GO:0006139	nucleobase, nucleoside, nucle	1.00000	0.05933	4.14240E-97	1.57411E-95	MGI:1261809,MGI:1197010,MGI:109768
GO:0008152	metabolism	1.00000	0.15150	6.71585E-65	2.55202E-63	MGI:1261809,MGI:1197010,MGI:109768
GO:0007582	physiological processes	1.00000	0.24877	5.49464E-48	2.08796E-46	MGI:1261809,MGI:1197010,MGI:109768
GO:0006310	DNA recombination	0.17949	0.00090	3.65804E-30	1.39006E-28	MGI:1261809,MGI:96995,MGI:1335098,
GO:0000067	DNA replication and chromos	0.23077	0.00327	6.23975E-29	2.37111E-27	MGI:1335098,MGI:1314881,MGI:134396
GO:0006260	DNA replication	0.21795	0.00267	1.56017E-28	5.92864E-27	MGI:1335098,MGI:1314881,MGI:134396
GO:0000084	S phase of mitotic cell cycle	0.21795	0.00270	1.97079E-28	7.48899E-27	MGI:1335098,MGI:1314881,MGI:134396
GO:0006289	nucleotide-excision repair	0.14103	0.00053	7.59198E-26	2.88495E-24	MGI:103557,MGI:103582,MGI:95480,M
GO:0000278	mitotic cell cycle	0.23077	0.00547	1.16886E-24	4.44167E-23	MGI:1335098,MGI:1314881,MGI:134396
GO:0007049	cell cycle	0.29487	0.01367	1.23385E-24	4.68864E-23	MGI:1335098,MGI:1314881,MGI:131667
GO:0008283	cell proliferation	0.29487	0.01647	8.61203E-23	3.27257E-21	MGI:1335098,MGI:1314881,MGI:131667
GO:0006284	base-excision repair	0.10256	0.00030	1.29533E-20	4.92227E-19	MGI:1261809,MGI:1097693,MGI:134117
GO:0008151	cell growth and/or maintenanc	0.33333	0.08530	6.08409E-10	2.31196E-08	MGI:1335098,MGI:1314881,MGI:109292
GO:0006298	mismatch repair	0.05128	0.00023	1.47081E-09	5.58907E-08	MGI:1343961,MGI:101816,MGI:101938,
GO:0045005	maintenance of fidelity during	0.05128	0.00023	1.47081E-09	5.58907E-08	MGI:1343961,MGI:101816,MGI:101938,
GO:0007126	meiosis	0.06410	0.00093	9.78190E-09	3.71712E-07	MGI:1100512,MGI:105393,MGI:97890,M
GO:0000280	nuclear division	0.07692	0.00257	5.19932E-08	1.97574E-06	MGI:1100512,MGI:105393,MGI:97890,M
GO:0000279	M phase	0.07692	0.00267	6.54923E-08	2.48871E-06	MGI:1100512,MGI:105393,MGI:97890,M
GO:0006307	double-strand break repair	0.03846	0.00013	6.75031E-08	2.56512E-06	MGI:104779,MGI:1333799,MGI:104517

A similar Term Finder tool is available at the SGD yeast database:
<http://db.yeastgenome.org/cgi-bin/SGD/GO/goTermFinder>

SGD Gene Ontology Term Finder Help

This GO Term Finder tool searches for significant shared GO terms, or parents of the GO terms, used to describe the genes in your list to help you discover what the genes may have in common.

- You could input the gene/ORF names in the text box provided or select a file of gene/ORF names.
- Choose from *only one* of the 3 ontologies at a time.

Step 1: Enter Gene/ORF names:
(separate by return)

OR

Select a file of Gene/ORF names:

Step 2: Choose Ontology (Choose from only one of the 3 ontologies at a time):

Process

Function

Component

Step 3:

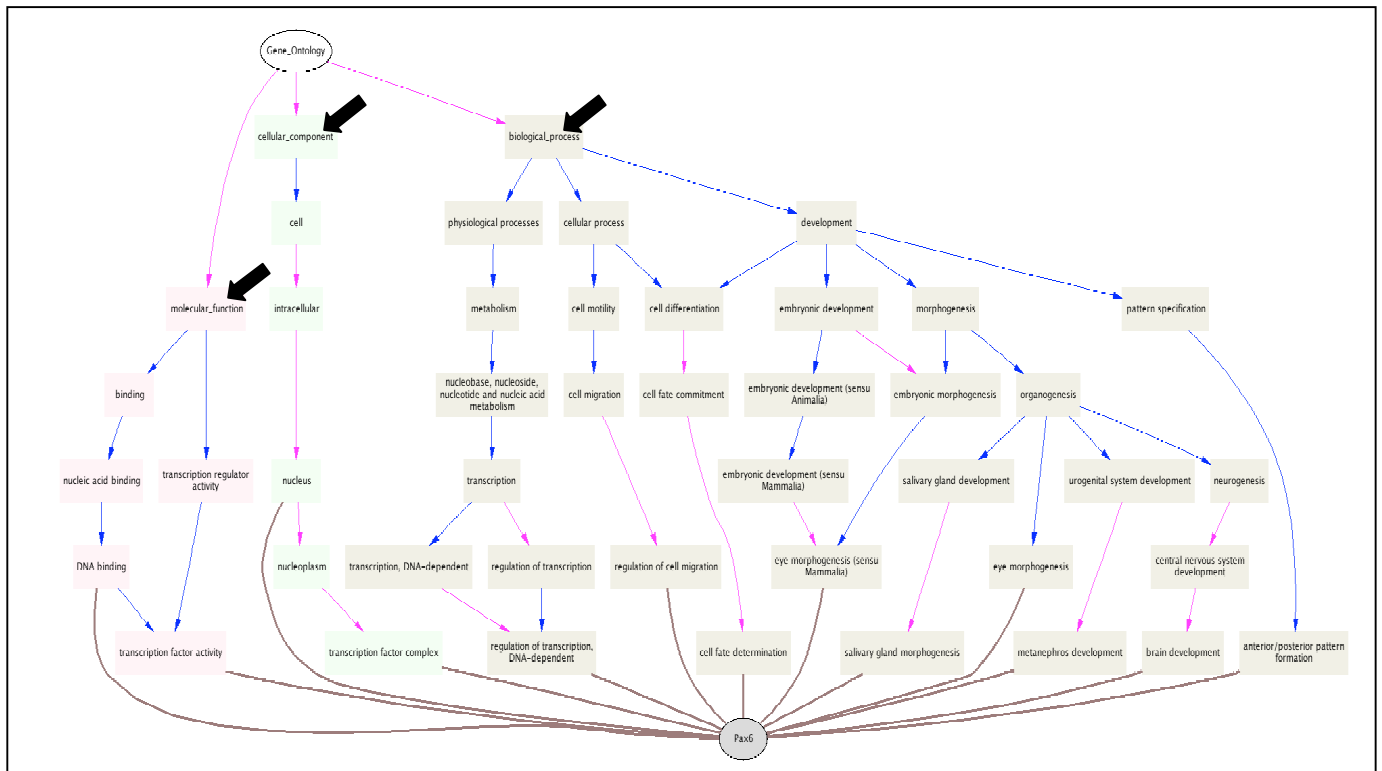
You will be provided a sample set of yeast genes, `sgd_sample.txt`, to test. Note that the results are similar to the MGI TermFinder results but, in addition, the statistically significant nodes are presented in the context of the ontology, displayed as a graph, and color-coded according to p-value.

Visualizing the GO structure.

Being able to visualize the ontology as a graph is very useful in understanding the relationship of different GO terms, as you may have seen by looking at your results using the SGD Term Finder. A tool that provides such a visualization of a gene is GenNav: <http://etbsun2.nlm.nih.gov:8000/perl/gennav.pl>



You can try out GenNav by entering Pax6 in the Term box, selecting Gene/Gene product, and selecting Species restriction Mus musculus. The resulting display will include the GO annotations for mouse Pax6 indicated by thicker lines in the graph but, in addition, the graph structure that shows how those annotations occur in the three ontologies (arrows), as shown here:



Again, you should expect to see results (more or less) consistent with the GO annotations on the Pax6 MGI gene detail page.

The next following pages contain samples lists used in this tutorial.
The first set is a list of SGD ids to be used at SGD TermFinder, the second is for the MGI
GoTermFinder demo

SGD sample set	YBR136W	YDR050C
	YER095W	YLR134W
YKL114C	YDR076W	YCR012W
YGL163C	YDL059C	YCL040W
YML021C	YDR004W	YNL199C
YDR097C	YDR369C	YHR174W
YDL200C	YCR092C	YOL086C
YOL043C	YML061C	YDR081C
YMR106C	YER162C	YAL038W
YKL113C	YBR114W	YKL152C
YDL154W	YEL037C	YGR087C
YML095C	YJR052W	YGR254W
YOR005C	YGR258C	YBR145W
YAL015C	YMR201C	YLR044C
YOR346W	YPL022W	YKL060C
YDR030C	YCR066W	YGR192C
YHR120W	YGL087C	YGR193C
YOL090W	YLR383W	YDL021W
YMR167W	YER173W	YPL075W
YML060W	YDR217C	YGR240C
YDL101C	YLR288C	YKL127
YIL066C	YFR015C	
YMR284W	YER133W	
YCR014C	YLR258W	
YOR033C	YJL137C	
YJR035W	YOR178C	
YNL082W	YNR032W	
YDR419W	YKR058W	
YNL250W	YPR160W	
YER142C	YER178W	
YMR137C	YGL253W	
YER169W	YMR303C	
YEL019C	YFR053C	
YLR032W	YBR221C	
YOR386W	YMR083W	
YIL139C	YMR205C	
YBL019W	YBR196C	
YPL167C	YMR105C	
YPL204W	YMR125W	
YKR095W	YNL071W	
YDR386W	YJR009C	
YML058W-A	YOR344C	
YGL021W	YOR347C	
YKR035W-A	YOL056W	
YDR263C	YGL256W	
YML032C	YJL052W	

MGI sample set	MGI:1196391
	MGI:1354947
MGI:1914689	MGI:1891457
MGI:1918961	MGI:1347081
MGI:1919580	MGI:1349767
MGI:1921585	MGI:1889000
MGI:1335098	MGI:104779
MGI:2443149	MGI:1316678
MGI:1925860	MGI:1890476
MGI:1340806	MGI:108016
MGI:1341112	MGI:105126
MGI:88042	MGI:105128
MGI:107202	MGI:109292
MGI:103067	MGI:97890
MGI:2145407	MGI:2150020
MGI:2384588	MGI:1099436
MGI:1924872	MGI:1261809
MGI:1351331	MGI:101949
MGI:1314881	MGI:894697
MGI:1355272	MGI:1194912
MGI:1277143	MGI:2156841
MGI:1202384	MGI:97900
MGI:105393	MGI:97901
MGI:95412	MGI:1337131
MGI:95413	MGI:1919401
MGI:95414	MGI:108247
MGI:1354163	MGI:1197010
MGI:103582	MGI:109352
MGI:95480	MGI:99135
MGI:1277216	MGI:103557
MGI:1345669	MGI:99137
MGI:1338799	MGI:1927345
MGI:101789	MGI:1333799
MGI:109152	MGI:104517
MGI:96977	
MGI:101938	
MGI:96995	
MGI:97073	
MGI:1100512	
MGI:101816	
MGI:1343961	
MGI:1917853	
MGI:1313275	
MGI:1097693	
MGI:104288	
MGI:97740	