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

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
    ①   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.*

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

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:IPR001601 F
MGI     MGI:1918914     0610006F02Rik        GO:0008757    MGI:MGI:2429377    ISS    INTERPRO:IPR000051 F
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.

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

Left navigation menu:
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 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.

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 it's 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.*)

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 Rpgr) are annotated to a child of "eye morphogenesis."

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

The Amigo browser uses the gene-association files contributed by the member databases to enable searches across multiple species. Type in "eye morphogenesis" into the "Search GO:" box and hit return.

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

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, it's 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.

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.

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.

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	Ref(s)
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

Gene Ontology Classifications

Symbol **Dab2ip**
Name disabled homolog 2 (Drosophila) interacting protein
ID MGI:1916851

Category	Classification Term	Evidence	Inferred From	Ref(s)
Biological Process	biological_process unknown	ND		1
Cellular Component	cellular_component unknown	ND		1
Molecular Function	protein binding	IPI	SPTR:P97318	1

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.

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.

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):

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

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

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.

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.

```
            ┌─────────────────────────┐
            │ Node A                  │
            │ No direct annotations   │
            │ 4 inferred annotations  │
            └─────────────────────────┘
             ↙                      ↘
┌──────────────────────┐   ┌──────────────────────┐
│ Node B               │   │ Node C               │
│ 2 direct annotations │   │ 2 direct annotations │
└──────────────────────┘   └──────────────────────┘
```
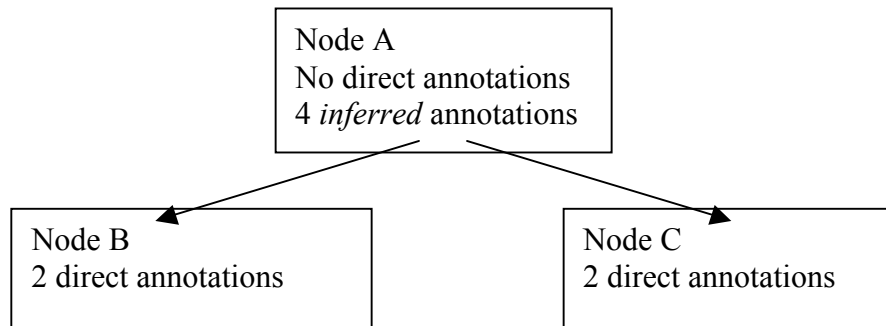
To determine significance, the TermFinder examines the group of genes to find GO terms to which a high proportion of the genes are associated compared to the number of times that term is associated with other genes in the genome. For example, if you search the process ontology and find that all of the 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 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, mgi_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:

Microsoft Excel - results.xls

File Edit View Insert Format Tools Data Window Help Acrobat

Arial   10

J2   =

	A	B	C	D	E	F	G	H	I	J	K
1	Ontology version										
2	Idate:	Mon Sep 08 16:10:56 BST 2003									
3	lversion: $Revision: 2.930 $										
4	Annotations version										
5	lsoftware version: $Revision: 1.12 $										
6	Idate: 09/09/2003 $										
7											
8	GOID	GO_term	Frequency	Genome frequency	P-value	Corrected P-value		Gene(s)			
9	GO:0006281	DNA repair	1.00000	0.00333	5.58843E-213	2.12360E-211		MGI:1261809,MGI:1197010,MGI:109765			
10	GO:0009719	response to endogenous stim	1.00000	0.00387	4.36646E-204	1.65926E-202		MGI:1261809,MGI:1197010,MGI:109765			
11	GO:0006974	response to DNA damage stin	1.00000	0.00387	4.36646E-204	1.65926E-202		MGI:1261809,MGI:1197010,MGI:109765			
12	GO:0006259	DNA metabolism	1.00000	0.01033	3.50666E-160	1.33253E-158		MGI:1261809,MGI:1197010,MGI:109765			
13	GO:0006950	response to stress	1.00000	0.02197	4.36099E-132	1.65718E-130		MGI:1261809,MGI:1197010,MGI:109765			
14	GO:0006139	nucleobase, nucleoside, nucle	1.00000	0.05933	4.14240E-97	1.57411E-95		MGI:1261809,MGI:1197010,MGI:109765			
15	GO:0008152	metabolism	1.00000	0.15150	6.71585E-65	2.55202E-63		MGI:1261809,MGI:1197010,MGI:109765			
16	GO:0007582	physiological processes	1.00000	0.24877	5.49464E-48	2.08796E-46		MGI:1261809,MGI:1197010,MGI:109765			
17	GO:0006310	DNA recombination	0.17949	0.00090	3.65804E-30	1.39006E-28		MGI:1261809,MGI:96995,MGI:1335098,			
18	GO:0000067	DNA replication and chromoso	0.23077	0.00327	6.23975E-29	2.37111E-27		MGI:1335098,MGI:1314881,MGI:134396			
19	GO:0006260	DNA replication	0.21795	0.00267	1.56017E-28	5.92864E-27		MGI:1335098,MGI:1314881,MGI:134396			
20	GO:0000084	S phase of mitotic cell cycle	0.21795	0.00270	1.97079E-28	7.48899E-27		MGI:1335098,MGI:1314881,MGI:134396			
21	GO:0006289	nucleotide-excision repair	0.14103	0.00053	7.59198E-26	2.88495E-24		MGI:103557,MGI:103582,MGI:95480,M(			
22	GO:0000278	mitotic cell cycle	0.23077	0.00547	1.16886E-24	4.44167E-23		MGI:1335098,MGI:1314881,MGI:134396			
23	GO:0007049	cell cycle	0.29487	0.01367	1.23385E-24	4.68864E-23		MGI:1335098,MGI:1314881,MGI:131667			
24	GO:0008283	cell proliferation	0.29487	0.01647	8.61203E-23	3.27257E-21		MGI:1335098,MGI:1314881,MGI:131667			
25	GO:0006284	base-excision repair	0.10256	0.00030	1.29533E-20	4.92227E-19		MGI:1261809,MGI:1097693,MGI:134111			
26	GO:0008151	cell growth and/or maintenanc	0.33333	0.08530	6.08409E-10	2.31196E-08		MGI:1335098,MGI:1314881,MGI:109292			
27	GO:0006298	mismatch repair	0.05128	0.00023	1.47081E-09	5.58907E-08		MGI:1343961,MGI:101816,MGI:101938,			
28	GO:0045005	maintenance of fidelity during	0.05128	0.00023	1.47081E-09	5.58907E-08		MGI:1343961,MGI:101816,MGI:101938,			
29	GO:0007126	meiosis	0.06410	0.00093	9.78190E-09	3.71712E-07		MGI:1100512,MGI:105393,MGI:97890,M			
30	GO:0000280	nuclear division	0.07692	0.00257	5.19932E-08	1.97574E-06		MGI:1100512,MGI:105393,MGI:97890,M			
31	GO:0000279	M phase	0.07692	0.00267	6.54923E-08	2.48871E-06		MGI:1100512,MGI:105393,MGI:97890,M			
32	GO:0006302	double-strand break repair	0.03846	0.00013	6.75031E-08	2.56512E-06		MGI:104779,MGI:1333799,MGI:104517			

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.

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:
Browse...

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

- ● Process
- ○ Function
- ○ Component

Step 3:

Search   Reset

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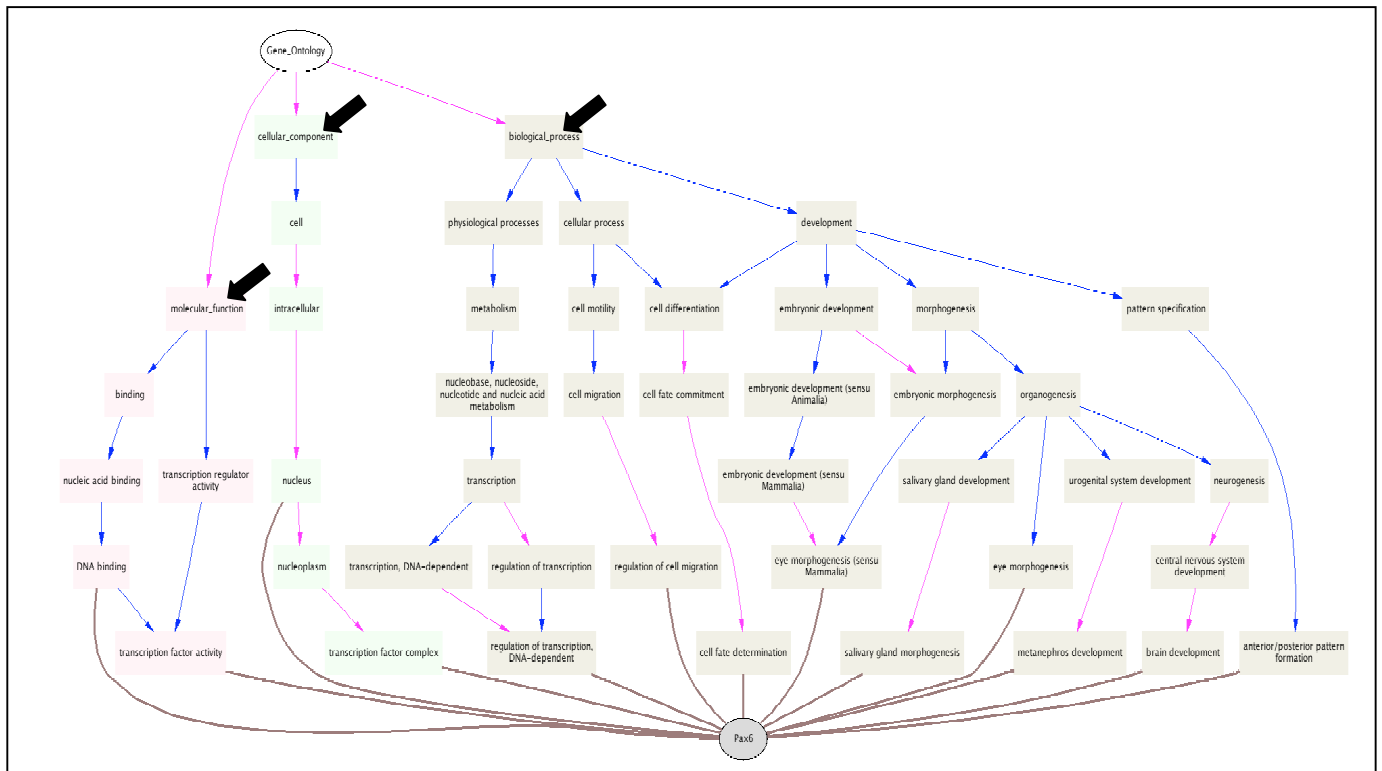
Again, you should expect to see results (more or less) consistent with the GO annotations on the Pax6 MGI gene detail page.

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

YKL114C
YGL163C
YML021C
YDR097C
YDL200C
YOL043C
YMR106C
YKL113C
YDL154W
YML095C
YOR005C
YAL015C
YOR346W
YDR030C
YHR120W
YOL090W
YMR167W
YML060W
YDL101C
YIL066C
YMR284W
YCR014C
YOR033C
YJR035W
YNL082W
YDR419W
YNL250W
YER142C
YMR137C
YER169W
YEL019C
YLR032W
YOR386W
YIL139C
YBL019W
YPL167C
YPL204W
YKR095W
YDR386W
YML058W-A
YGL021W
YKR035W-A
YDR263C
YML032C

YBR136W
YER095W
YDR076W
YDL059C
YDR004W
YDR369C
YCR092C
YML061C
YER162C
YBR114W
YEL037C
YJR052W
YGR258C
YMR201C
YPL022W
YCR066W
YGL087C
YLR383W
YER173W
YDR217C
YLR288C
YFR015C
YER133W
YLR258W
YJL137C
YOR178C
YNR032W
YKR058W
YPR160W
YER178W
YGL253W
YMR303C
YFR053C
YBR221C
YMR083W
YMR205C
YBR196C
YMR105C
YMR125W
YNL071W
YJR009C
YOR344C
YOR347C
YOL056W
YGL256W
YJL052W

YDR050C
YLR134W
YCR012W
YCL040W
YNL199C
YHR174W
YOL086C
YDR081C
YAL038W
YKL152C
YGR087C
YGR254W
YBR145W
YLR044C
YKL060C
YGR192C
YGR193C
YDL021W
YPL075W
YGR240C
YKL127

MGI sample set

MGI:1914689
MGI:1918961
MGI:1919580
MGI:1921585
MGI:1335098
MGI:2443149
MGI:1925860
MGI:1340806
MGI:1341112
MGI:88042
MGI:107202
MGI:103067
MGI:2145407
MGI:2384588
MGI:1924872
MGI:1351331
MGI:1314881
MGI:1355272
MGI:1277143
MGI:1202384
MGI:105393
MGI:95412
MGI:95413
MGI:95414
MGI:1354163
MGI:103582
MGI:95480
MGI:1277216
MGI:1345669
MGI:1338799
MGI:101789
MGI:109152
MGI:96977
MGI:101938
MGI:96995
MGI:97073
MGI:1100512
MGI:101816
MGI:1343961
MGI:1917853
MGI:1313275
MGI:1097693
MGI:104288
MGI:97740

MGI:1196391
MGI:1354947
MGI:1891457
MGI:1347081
MGI:1349767
MGI:1889000
MGI:104779
MGI:1316678
MGI:1890476
MGI:108016
MGI:105126
MGI:105128
MGI:109292
MGI:97890
MGI:2150020
MGI:1099436
MGI:1261809
MGI:101949
MGI:894697
MGI:1194912
MGI:2156841
MGI:97900
MGI:97901
MGI:1337131
MGI:1919401
MGI:108247
MGI:1197010
MGI:109352
MGI:99135
MGI:103557
MGI:99137
MGI:1927345
MGI:1333799
MGI:104517