Introduction to Gene Ontology

Presenter: Wayne Xu, Ph.D
Computational Genomics Consultant,
Supercomputing Institute

Email: wxu@msi.umn.edu

Phone: (612) 624-1447

Help: help@msi.umn.edu
(612) 626-0802

April.13, 2006

Outline

- Introduction
- Gene Ontology and GO Consortium
- GO data descriptive vocabularies
- GO annotation
- GO Databases
- GO Tools

Introduction

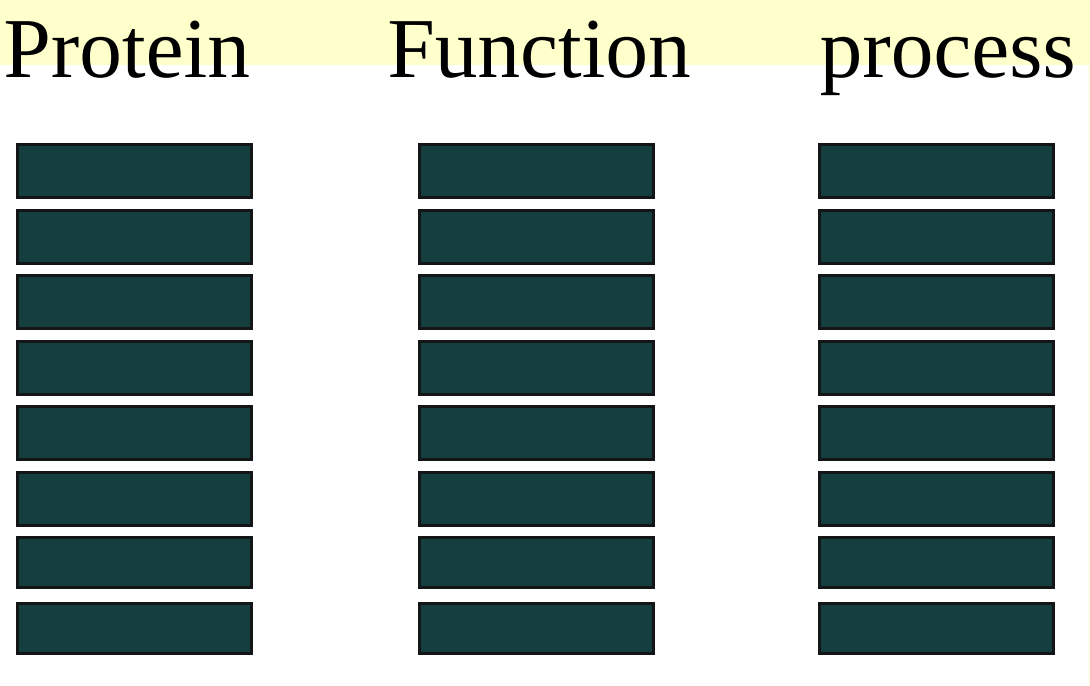
Motivation

- Explosively-increasing amount of sequence data leads the creation of many databases for the data management
  - Domain-specific: PIR,PDB,GenBank,TIGR, UniProt, …
  - Organism-specific: AceDB, FlyBase, SGD, MGI,…
- But limitation in data integration:
  - Can list a gene product P53 in all organisms and what it does in these organisms?
  - Can list all “receptor signaling protein tyrosine kinase activity” proteins in all organisms?
  - Can list all “defense response to pathogenic bacteria” proteins in all organisms?
  - Even within the same organism, how do you classify a group of proteins?

Solutions

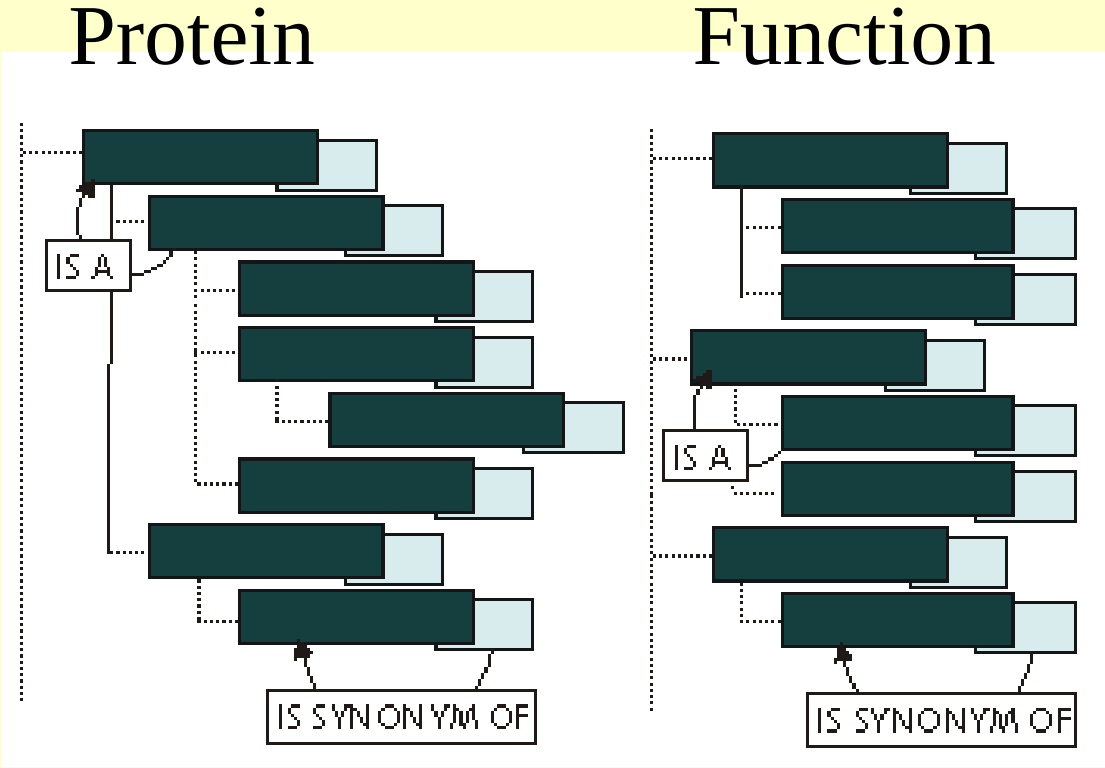
- The most fundamental questions for the biologists served by these databases revolve around the genes
  - Describe the genes or gene products
  - Genes have relationships to others
  - Gene product has multiple features
- So, the challenge is to develop one common data description schema for all organisms and all databases
- What is a best way?
  - Description
    - Location, function, process
  - Presentation:
    - List
    - Taxonomy
    - Ontology

List



- No relationships within the same type of concepts
- Very useful for simplest applications

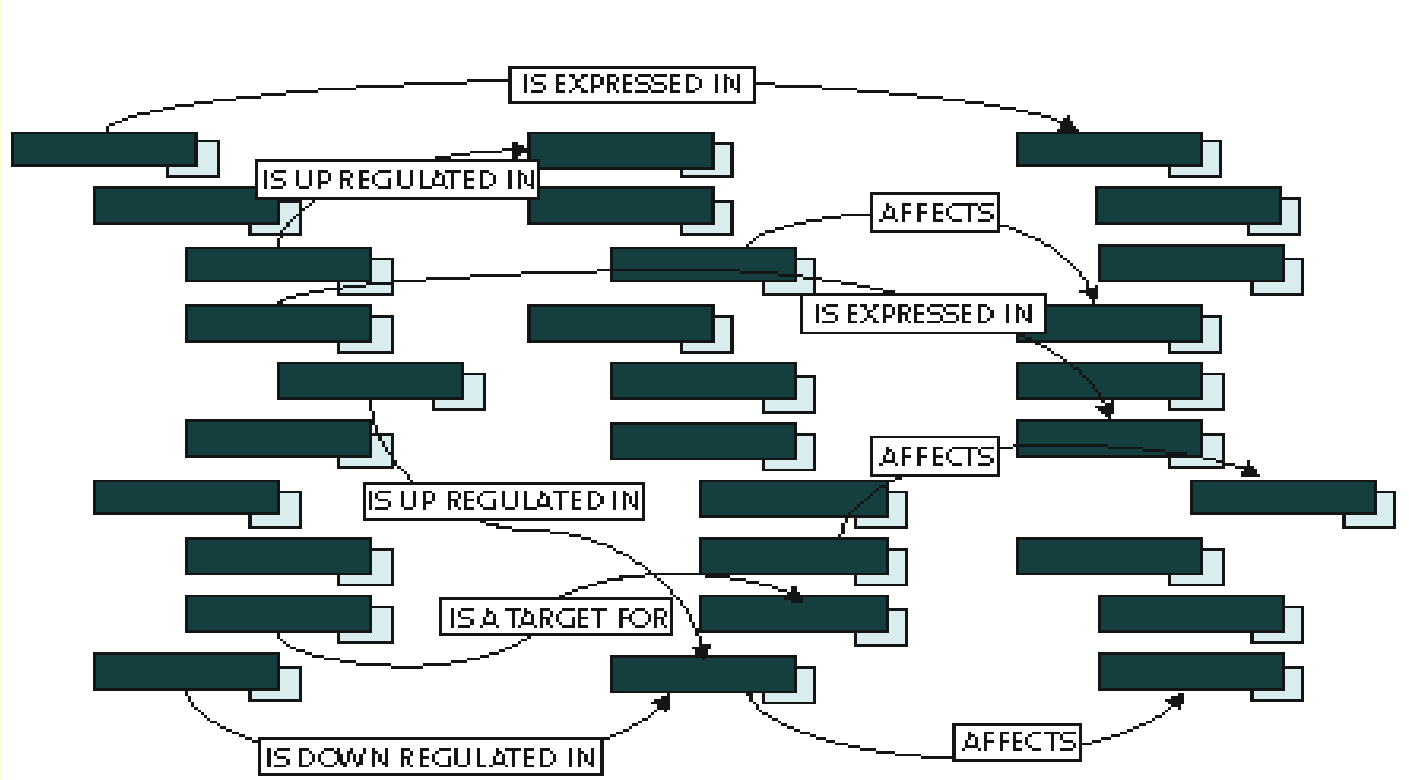
Taxonomy



- Hierarchical relationship among the same type of concept
- But 1:1 relationship between concepts, not the case in genes

Ontology

Protein Function Location



- Include much richer and more descriptive relationships between concepts

Gene Ontology and GO Consortium

Gene Ontology

- In July 1998, at the Montreal International Conference on Intelligent Systems for Molecular Biology (ISMB) bio-ontologies Workshop
  - Michael Ashburner presented a simple hierarchical controlled vacabulary as Gene Ontology
  - It was agreed by three model databases: FlyBase (Suzanna E Lewis), SGD (Steve Chervitz), and MGI (Judith Blake)
  - The Gene Ontology Consortium was founded

Ontologies

- Ontology is derived from the Greek meaning “a description of what exists”.
- An ontology is used now a description of the concepts and relationships that exist for a community of agents
- Practically write an ontology as a set of definitions of formal vocabulary
- For the purpose of enabling knowledge sharing and reuse
  - Plant ontology (PO): a controlled vocabulary for plant structure (anatomy) and growth stages
  - Trait ontology (TO): a controlled vocabulary to describe each trait as a distinguishable feature, characteristic, quality or phenotypic feature of a developing or mature individual. Examples are glutinous endosperm, disease resistance, plant height, photosensitivity, male sterility, etc.
  - Mammalian Phenotye Ontology
  - Mouse ontology
  - Cell type ontology
  - Sequence Ontology
  - Gene Ontology
  - …

GO Consortium

- Three major goals:
  - To develop a set of controlled, structured vocabularies – gene ontology (GO) – to describe key domains of molecular biology, gene
  - To apply GO terms in the annotation of genes in biological databases
  - To provide a centralized public resource allowing universal access to the GO, annotation data sets and software tools developed for use with GO data

GO Data Descriptive Vocabularies

GO Vocabularies (Terms)

- Define all gene products by the three organizing GO principles:
  - **molecular function**
  - **biological process**
  - **cellular component**
- Eukaryotes and virus share a same data description schema (controlled vocabularies)
  - problem?

GO Molecular Function

- Describes activities, such as catalytic or binding activities, at the molecular level
- Examples:
  - Broad molecular function terms:
    - catalytic activity,
    - transporter activity,
    - binding;
  - Narrower molecular function terms
    - Adenylate cyclase activity
    - Toll receptor binding

GO Biological Process

- Series of events accomplished by one or more molecular functions
- Examples:
  - Broad biological process terms
    - cellular physiological process
    - signal transduction,
  - Narrower biological process terms:
    - pyrimidine metabolism
    - alpha-glucoside transport.
- Distinguish between a biological process and a molecular function, but the general rule is that a process must have more than one distinct steps
- A biological process is not equivalent to a pathway.

GO Cellular Component

- A component of a cell such as part of some larger object
- Examples:
  - an anatomical structure (e.g. **rough endoplasmic reticulum** or **nucleus**)
  - a gene product group (e.g. **ribosome**, **proteasome** or a protein dimer)

GO Vocabularies (Terms)

- A gene product has one or more molecular functions and is used in one or more biological processes; it might be associated with one or more cellular components.
- Example, the gene product cytochrome c can be described by
  - the molecular function term **oxidoreductase activity**,
  - the biological process terms **oxidative phosphorylation** and **induction of cell death**,
  - and the cellular component terms **mitochondrial matrix** and **mitochondrial inner membrane**.

Define GO Terms

- Controlled Vocabularies,
- Explore into all the three principles and their hierarchical relationships
- must use our extensive domain knowledge of biology
  - GO Consortium
  - Many Curator interest groups

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

GO Terms

[Term] id: GO:0000002
name: mitochondrial genome maintenance
namespace: biological_process
def: "The maintenance of the structure and integrity of the mitochondrial genome." [GOC:ai]
is_a: GO:0007005 ! mitochondrion organization and biogenesis

[Term] id: GO:0000003
name: reproduction
namespace: biological_process
Alt_id: GO:0019952
def: "The production by an organism of new individuals that contain some portion of their genetic material inherited from that organism." [GOC:go_curators, ISBN:0198506732]
subset: goslim_generic
subset: goslim_plant
subset: gosubset_prok
is_a: GO:0008150 ! biological_process

GO Annotation

GO Gene Annotation

- All GO collaborating databases annotate their gene products (or genes) with GO terms
  - Source
    - Literature
    - another database
    - computational analysis
  - Evidence codes:
    - IMP
    - IGI
    - IPI
    - ISS
    - IDA
    - IEP
    - IEA
    - TAS
    - NAS
    - ND
    - IC

Annotation File Format

- Gene associate file or Mysql gene associate table
  - Link between term and gene or gene product (transcript or protein)
- 15 columns:

1. DB
2. DB_Object_ID
3. DB_Object_Symbol
4. NOT
5. GO ID
6. DB:Reference
7. Evidence
8. With (or) from
9. Aspect
10. DB_Object_Name
11. DB_Object_Synonym
12. DB_Object_Type
13. Taxon
14. Date
15. Assigned_by

GO Database

Address http://www.geneontology.org/ Go Link

Google gene ontology Search 2346 blocked Check AutoLink AutoFill Options gene ontology

the Gene Ontology

Search go!
gene or protein name

- Open menus
- Home
- **Downloads**
  - Ontologies
  - Annotations
  - Database
  - Mappings to GO
  - Teaching Resources
  - Monthly Reports
- **GO Tools**
- **Documentation**
- **About GO**
- **GO Editor Guides**
- Contact GO
- Site Map

Gene Ontology Home

The Gene Ontology project provides a controlled vocabulary to describe gene and gene product attributes in any organism. *Read more...*

Popular Links

Search the Gene Ontology Database

GO!

gene or protein name GO term or ID

This search uses the browser AmiGO. Browse the Gene Ontology using AmiGO.

GO website

- GO downloads: including ontology files, annotations and the GO database
- Tools for using GO
- Request new terms or ontology changes via the SourceForge tracker system; help with new term submission is available.
- Documentation on all aspects of the GO project and the FAQ
- Gene Ontology mailing lists and contact details

Back to top

News
Funding and Acknowledgements
Usage Statistics

News

Gene Ontology Database Downloads

Downloads

The GO monthly releases are available, either as RDF XML or as a MySQL database dump.

Monthly release contains data as of **2006-02-01** : [HTTP or FTP].

Weekly release prepared on **2006-02-26** : [HTTP or FTP].

Daily release prepared on **2006-03-02** : [HTTP or FTP].

View and download all daily, weekly and monthly releases.

Release Cycle

The GO Database is built from the data publically available as flatfiles from the main GO website. The database is not used for data management, only for querying, either with AmiGO, the go-db-perl modules or with MySQL.

The GO database follows a monthly cycle. Each monthly release takes a few days to build, and requires manual QC. Although the timing of the release is irregular, it always corresponds to the data in the main GO CVS repository and FTP site as of midnight on the first of every month.

Documentation

General Documentation

Consult the GO Software and Databases webpage for software and API details, or the main GO website

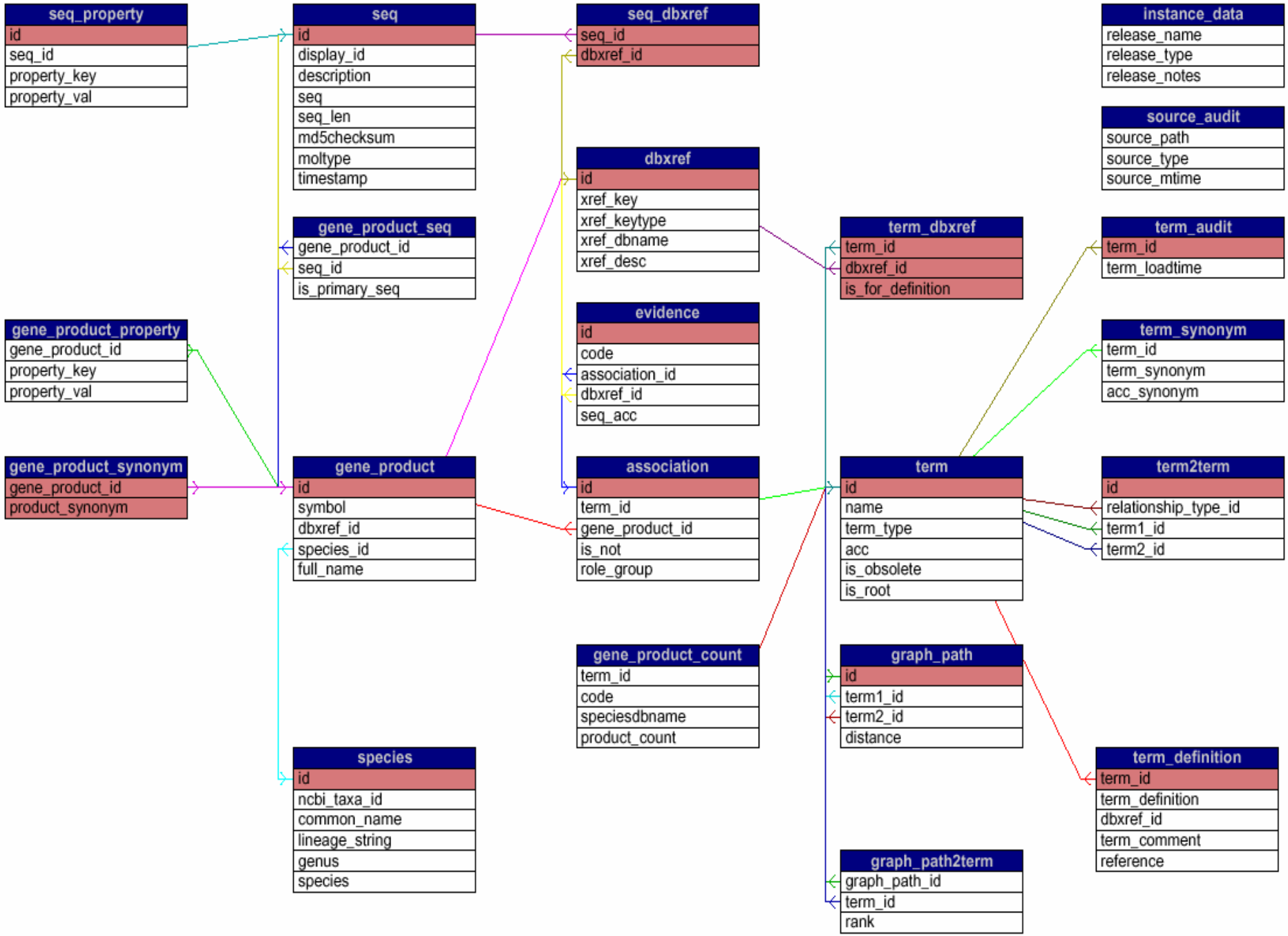
Schema Documentation

GO Database Schema Description
Main documentation on the GO Database, and the table creation SQL

Example queries
Examples of common queries you might want to ask the GO database

Hyperlinked HTML Tables
A web page showing the tables and columns in the GO database. You can traverse the foreign key relationships as hyperlinks. **Note:** there are no embedded comments on this autogenerated page; also, the tables are not arranged into their modular structure. For full documentation, please refer to the full documentation

Contributed Diagrams
Diagrams showing the structure of the GO database - these are contributed by third parties, and may be incorrect or out of date with respect to the most


seq_property
id
seq_id
property_key
property_val
seq
id
display_id
description
seq
seq_len
md5checksum
moltype
timestamp
seq_dbxref
seq_id
dbxref_id
instance_data
release_name
release_type
release_notes
source_audit
source_path
source_type
source_mtime
dbxref
id
xref_key
xref_keytype
xref_dbname
xref_desc
gene_product_seq
gene_product_id
seq_id
is_primary_seq
term_dbxref
term_id
dbxref_id
is_for_definition
term_audit
term_id
term_loadtime
evidence
id
code
association_id
dbxref_id
seq_acc
gene_product_property
gene_product_id
property_key
property_val
term_synonym
term_id
term_synonym
acc_synonym
gene_product_synonym
gene_product_id
product_synonym
gene_product
id
symbol
dbxref_id
species_id
full_name
association
id
term_id
gene_product_id
is_not
role_group
term
id
name
term_type
acc
is_obsolete
is_root
term2term
id
relationship_type_id
term1_id
term2_id
gene_product_count
term_id
code
speciesdbname
product_count
graph_path
id
term1_id
term2_id
distance
species
id
ncbi_taxa_id
common_name
lineage_string
genus
species
term_definition
term_id
term_definition
dbxref_id
term_comment
reference
graph_path2term
graph_path_id
term_id
rank

Recursive Querying

- **Find all *DNA binding* genes**
- **term2term** table to iterate through the graph, but this requires multiple SQL calls
- *precompute* the path from every node to all of its ancestors.This goes in the **graph_path** table, which also holds the distance between terms

Query GO Database

- **Direct MySQL queries**
  - use the mysql command line interface to issue queries
- **Query via the perl API**
  - need go-db-perl for this
- **Local copy of AmiGO**
  - install AmiGO as a local CGI script, and issue web queries
- **Query via your own code**
  - write your own code to query the db, using a database driver such as DBI or JDBC
- **Query via DBStag**
  - use the stag module for issuing queries to the GO db and getting back XML. query with arbitrary SQL, or use the stag templates provided (see README).

SQL Command Line

Login db1.msi.umn.edu

. /usr/local/mysql/mysql_client

mysql -h 127.0.0.1 -P 9903 -u geneontology -p

Enter password:

```
mysql> show tables;
+------------------------+
| Tables_in_geneontology |
+------------------------+
| assoc_rel              |
| association            |
| association_qualifier  |
| db                     |
| dbxref                 |
| evidence               |
| evidence_dbxref        |
| gene_product           |
| gene_product_count     |
| gene_product_property  |
| gene_product_seq       |
| gene_product_synonym   |
| graph_path             |
| graph_path2term        |
| instance_data          |
| seq                    |
| seq_dbxref             |
| seq_property           |
| source_audit           |
| species                |
| term                   |
| term2term              |
| term_audit             |
| term_dbxref            |
| term_definition        |
| term_synonym           |
+------------------------+
26 rows in set (0.00 sec)
```

```
mysql> select name from db;
+--------------------+
| name               |
+--------------------+
| AgBase             |
| CGD                |
| DDB                |
| FB                 |
| GDB                |
| GeneDB_Lmajor      |
| GeneDB_Pfalciparum |
| GeneDB_Spombe      |
| GeneDB_Tbrucei     |
| GOA                |
| GR                 |
| HGNC               |
| IntAct             |
| MGI                |
| PINC               |
| Reactome           |
| RGD                |
| SANGER             |
| SGD                |
| TAIR               |
| TIGR               |
| UniProt            |
| WB                 |
| ZFIN               |
+--------------------+
24 rows in set (0.04 sec)
```

SQL Command Line

Say we want to find the total number of gene products that are BOTH GTP binding (GO:0005525) and immune response (GO:0006955)

```
SELECT count(DISTINCT a1.gene_product_id)
 FROM term AS t1
      INNER JOIN graph_path AS p1 ON (t1.id=p1.term1_id)
      INNER JOIN association AS a1 ON (a1.term_id=p1.term2_id)
      INNER JOIN term AS t2 ON (t2.id=p2.term1_id)
      INNER JOIN graph_path AS p2 ON (a2.term_id=p2.term2_id)
      INNER JOIN association AS a2 ON (a2.gene_product_id=a1.gene_product_id)
  WHERE t1.acc = 'GO:0005525' AND t2.acc = 'GO:0006955';
|
```

```
+-----------------------------------------------+
| count(DISTINCT a1.gene_product_id)  |
+-----------------------------------------------+
|                                          16 |
+-----------------------------------------------+
```

GO-DB-Perl Handler

http://www.godatabase.org/dev/

```
#!/usr/local/bin/perl
use GO::AppHandle;

my $dbname = "geneontology";
my $mysqlhost = "127.0.0.1:9903";
my $user = "geneontology";
my $passwd = “gois_here";

$apph = GO::AppHandle->connect(-dbname=>$dbname, -dbhost=>$mysqlhost, -dbuser=>$user, -
      dbauth=>$passwd);

$product =$apph->get_product({symbol=>"Cyp1a1"});
  printf "Product; name=%s  Acc=%s\n",
          $product->full_name(),
          $product->acc();
```

- -bash-3.00$ ./symbol.pl
- Product; name=cytochrome P450, family 1, subfamily a, polypeptide 1 Acc=MGI:88588

GO Tools

GO Tools

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

- Consortium Tools:
  - AmiGO
  - DAG-Edit
- Non-Consortium Tools:
  - Search and browse
    - GOFish, QuickGO, ….
  - Annotation
    - Manatee, GeneTools,…
  - Gene expression
    - BiNGO, GeneMerge, GOArray, GO Term Finder, …
  - Others
    - Blast2GO, Generic GO term Mapper, GO SLIM Mapper, …

Google gene ontology Search 2419 blocked Check AutoLink AutoFill Options gene ontology

AmiGO

Search GO

p53

Exact Match

Terms

Gene Symbol/Name

Submit Query

Advanced Query

Query By Sequence

Gene Product Filters

Species

All
A. japonica
A. niger

Datasource

All
FlyBase
SGD

Evidence Code

All Curator Approved
IC
IMP

Set Filters

Query Summary

Your Query
p53

Exact Match
no

Target
Gene Products

Fields

Acp53C14c, **Acp53C14c**

gene from *Drosophila melanogaster*, data from FlyBase (FBgn0053530)

Term	Ontology	Evide
extracellular region	C	ISS

Acp53Ea, **Accessory gland-specific peptide 53Ea**

gene from *Drosophila melanogaster*, data from FlyBase (FBgn0015584)

Term	Ontology	Evide
physiological process	P	NAS
post-mating behavior	P	NAS
sperm competition	P	TAS
sperm displacement	P	NAS
extracellular region	C	NAS
hormone activity	F	NAS

Acp53Eb, **Accessory gland protein 53Eb**

gene from *Drosophila melanogaster*, data from FlyBase (FBgn0024500)

Term	Ontology	Evide
sperm displacement	P	NAS

Arp53D, **Actin-related protein 53D**

gene from *Drosophila melanogaster*, data from FlyBase (FBgn0011743)

Term	Ontology	Evide
cytoskeleton organization and biogenesis	P	ISS ISS

GOFish Tool

http://llama.med.harvard.edu - GoFish v1.11alpha - Microsoft Internet Explorer

Quit

File Organism Help

Gene Ontology Browser

- molecular_function
- cellular_component
- biological_process
  - cell communication
  - development
  - physiological processes
  - behavior
  - cell growth and/or maintenance
  - death

GO term viewer

- biological_process
  - death

Selected GO terms

A. apoptosis regulator
B. death

A & B

Remove GoFish!

Search GO terms

Search

- GO term word(s)
- Start(s) of GO term word(s)
- Start of GO term
- GO term ID(s)

Search results

Ranked gene products

	QS	Name	A	B	Descriptio
1	1	Bax	1	1	Bcl2-assc
2	1	Bcl2	1	1	B-cell leuk
3	1	Apaf1	1	1	apoptotic
4	1	Bcl10	1	1	B-cell leuk
5	1	Bak1	1	1	BCL2-ant
6	1	Bag3	1	1	Bcl2-assc
7	1	Bid	1	1	BH3 intera
8	1	Fadd	1	1	Fas (TNF
9	1	Bad	1	1	Bcl-assoc
10	1	Cradd	1	1	CASP2 ar
11	1	Cflar	1	1	CASP8 ar
12	1	Dffb	1	1	DNA fragr
13	1	Dffa	1	1	DNA fragr
14	1	Bcl2l	1	1	Bcl2-like
15	1	Birc1b	1	1	baculovira
16	1	Birc1a	1	1	baculovira
17	1	Birc1f	1	1	baculovira
18	1	Birc1e	1	1	baculovira

Gene product viewer

MGD name: Bcl2-associated X protein
MGD ID: MGI:99702
Marker symbol: Bax
Chromosome: VII

Internet

QuickGO: GO Browser - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites

Address http://www.ebi.ac.uk/ego/ Go Links

Google gene ontology Search 2419 blocked Check AutoLink AutoFill Options gene ontology

EMBL-EBI
European Bioinformatics Institute

Get Nucleotide sequences for Go ? Site search Go ?

Site Map EBI Database Queries

EBI Home | About EBI | Groups | Services | Toolbox | Databases | Downloads | Submissions

QuickGO

[Remove menu]

QuickGO GO Browser

- QuickGO home
- Search
- GO Annotation home
- Documentation
- Browser FAQ

QuickGO GO Browser

Quick links

A fast Gene Ontology browser.

QuickGO GO Browser

Search GO

Search GO term names/synonyms | Search all ontologies

Query Example: **dimerization**

- Get help with searching
- More information about this browser
- Get more information about the GOA project at the EBI
- Read the user manual
- InterPro
- GOA Proteomes

Please note: due to a bug in the image maps for the graphical term ancestry you are recommended to navigate term ancestry using the denormalized tree view.

Summary

QuickGO is a fast web based browser of the Gene Ontology data (see geneontology.org) based at the EBI, as well as the annotation of GO to UniProt and InterPro generated by the GOA project. It integrates into InterPro, providing links between the two data sets that are navigable via the web. Various search facilities also exist.

See the documentation index.

Normal	Printer Friendly	Text	Simple HTML	XML	Curator View

Internet

Onto-Express (OE)

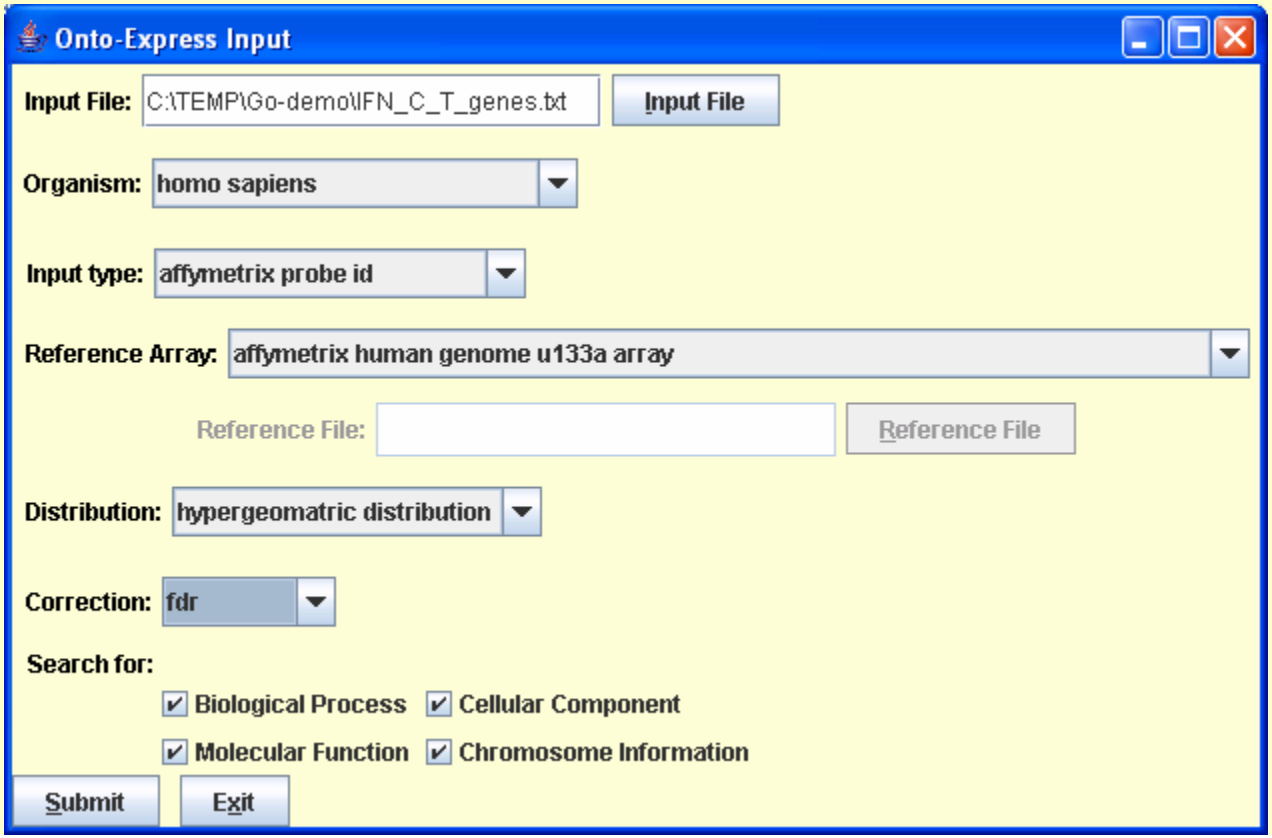
http://vortex.cs.wayne.edu/ontoexpress/servlet/UserInfo

Intelligent Systems and Bioinformatics Laboratory, Wayne State University

- Automatically translate gene lists of differentially regulated genes into functional profiles
- Functional profiles: biochemical function, biological process, cellular role, cellular component, molecular function and chromosome location.
- Statistical significance values are calculated for each category.

Onto-Express (OE)

- Login (c:\temp\go-demo)
- Run Onto-express
- Input:
  - Input file: interested gene list (209) from microarray analysis
  - Organism: (homo sapiens)
  - Input type: (affymetrix probe id)
  - Reference Array: (affymetrix human genome u133a array)
  - Distribution:
  - Correction:
  - Search for:



Onto-Express Results - C:\TEMP\Go-demo\IFN_C_T_genes.txt

Display

Display: Biological Process

Sort by: Name

Search

Function: OR Total >= :

select results OR p-value <= :

Search | Clear | Search Input

Legend

User Interactions:
- Unselected Function
- Synchronized Function
- Selected Function
- Searched Function

Functional Catagories Observed:
- More Than Expected
- Less Than Expected
- Same As Expected

Gene Regulation:
- Positive
- Negative
- No Change

Save Onto-Express Results

Save | Save as GIF image

Program

Draw Selected | Run Onto-Design | Run Onto-Compare

Tree View | Synchronized View | Synchronized Pie Chart | Single Gene View | Flat View | Flat Pie Chart

P-Value	Corrected P-Value	Total		
0.02087	0.02846	1	0.48%	activation of MAPK
0.00277	0.00954	1	0.48%	activation of MAPKKK
0.03821	0.04291	1	0.48%	acute-phase response
0.03608	0.04305	1	0.48%	adenylate cyclase activation
0.00131	0.00612	2	0.95%	amino acid biosynthesis
0.01479	0.02428	2	0.95%	amino acid transport
0.02226	0.02997	2	0.95%	angiogenesis
1.1E-4	0.00105	7	3.33%	anti-apoptosis
0.05685	0.05527	1	0.48%	antigen presentation
1.0E-5	2.1E-4	4	1.9%	antigen presentation, endogenous antigen
0.0	0.0	9	4.29%	antigen presentation, exogenous antigen
6.0E-5	6.3E-4	4	1.9%	antigen processing, endogenous antigen via MHC class I
0.0	0.0	9	4.29%	antigen processing, exogenous antigen via MHC class II
0.10309	0.08523	2	0.95%	antimicrobial humoral response (sensu Vertebrata)
8.0E-5	8.3E-4	12	5.71%	apoptosis
0.00157	0.00677	1	0.48%	arginine biosynthesis
0.02999	0.03661	1	0.48%	aromatic compound metabolism
0.11328	0.09293	1	0.48%	ATP synthesis coupled proton transport
0.01055	0.01831	1	0.48%	B cell activation
0.39949	0.23969	4	1.9%	biological process unknown
0.26249	0.17226	1	0.48%	blood coagulation
0.04954	0.0515	1	0.48%	calcium ion homeostasis
0.02611	0.03449	3	1.43%	calcium ion transport
0.00514	0.01403	1	0.48%	calcium-dependent cell-cell adhesion
0.27237	0.17708	1	0.48%	carbohydrate metabolism

Onto-Express Results - C:\TEMP\Go-demo\IFN_C_T_genes.txt

Display

Display: Biological Process

Sort by: Name

Search

Function: OR Total >= :

select results OR p-value <= :

Search Clear Search Input

Legend

User Interactions:
- Unselected Function
- Synchronized Function
- Selected Function
- Searched Function

Functional Catagories Observed:
- More Than Expected
- Less Than Expected
- Same As Expected

Gene Regulation:
- Positive
- Negative
- No Change

Save Onto-Express Results

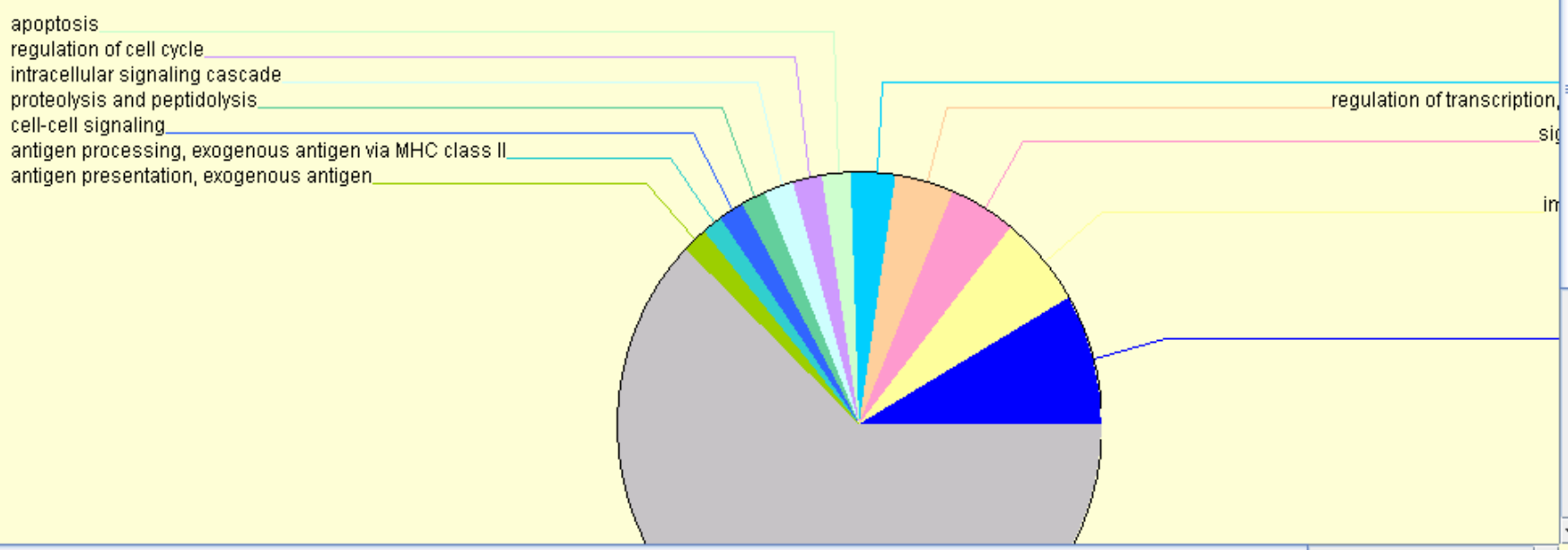
Save Save as GIF image

Program

Draw Selected Run Onto-Design Run Onto-Compare

Tree View | Synchronized View | Synchronized Pie Chart | Single Gene View | Flat View | Flat Pie Chart

Functional Category:
Number of Genes:
(Right click on each pie for additional options.)



MGI Gene Ontology GO_Slim Chart Tool

This GO_Slim Chart Tool bins the genes in your list according to MGI GO Slim definitions to help you discover common.

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

```
MGI:1914689
MGI:1918961
MGI:1919580
MGI:1921585
MGI:1335098
```

Browse...

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:

Search Reset

Uses TermFinder (0.5) implementation of Gavin Sherlock, Stanford Microarray Database.

Microsoft Excel - sample_results.txt

File Edit View Insert Format Tools Data Window Help Acrobat

R9C5 =

	1	2	3	4	5
1	Ontology version				
2	!date: Wed Aug 06 11:46:23 BST 2003				
3	!version: $Revision: 2.866 $				
4	Annotations version				
5	!software version: $Revision: 1.12 $				
6	!date: 08/05/2003 $				
7					
8	Any warnings will be listed here:				
9					
10	Results for Biological Process Bins:			All MGI	
11	Process	Count	Percentage	Count	Percentage
12	cell adhesion	0	0.00000	284	0.02020
13	cell-cell signaling	0	0.00000	133	0.00946
14	cell cycle and proliferat	23	0.29487	379	0.02696
15	death	3	0.03846	193	0.01373
16	cell organization and bio	0	0.00000	460	0.03272
17	protein metabolism	8	0.10256	1182	0.08407
18	DNA metabolism	78	1.00000	239	0.01700
19	RNA metabolism	9	0.11538	1012	0.07198
20	other metabolic processes	0	0.00000	1717	0.12212
21	stress response	78	1.00000	328	0.02333
22	transport	4	0.05128	1021	0.07262
23	developmental processes	2	0.02564	862	0.06131
24	signal transduction	2	0.02564	1131	0.08044
25	other biological processe	0	0.00000	2641	0.18784
26	all biological processes	78	1.00000	14060	1.00000

Note: overrepresentation of genes in bins for "DNA metabolism" and "stress response" compared with MGI overall