Using the Gene Ontology for Data Analysis

Judith Blake, Ph.D.

Mouse Genome Informatics
The Jackson Laboratory
Bar Harbor, Maine, USA

Ontologies for Molecular Biology

- "Ontologies provide controlled, consistent vocabularies to describe concepts and relationships, thereby enabling knowledge sharing" (Gruber 1993)

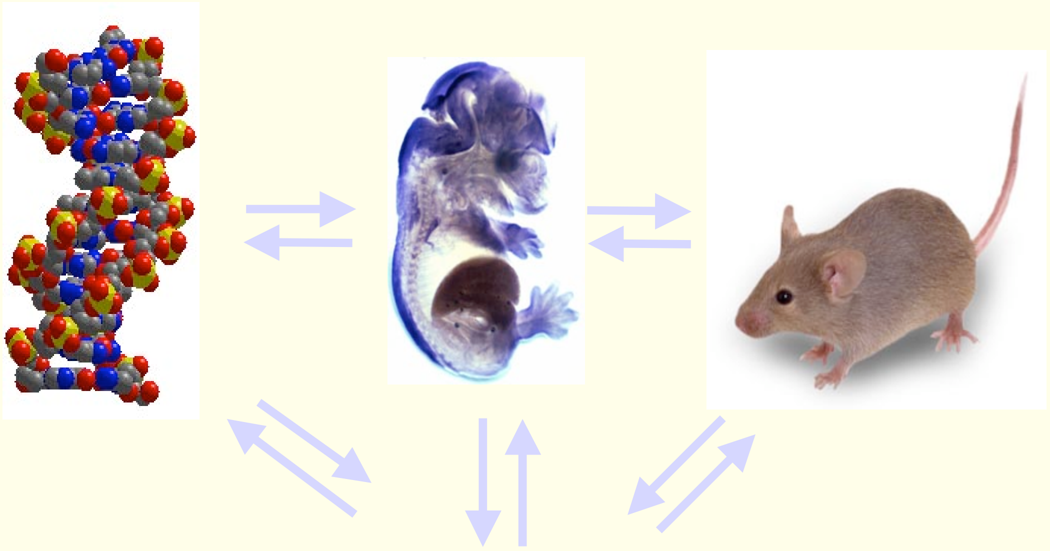
- Gene Ontologies (GO)
  - Ontologies for molecular biology domains developed and supported by the Gene Ontology Consortium for gene and gene product annotations for all organisms

Support Complex Queries

- Show me all *genes* involved in *cell adhesion* that are expressed in the *somites.*
- Show me all *genes* involved in *mesoderm formation* in *fly* and *mouse* that also show *cytokine activity*.
- For this set of genes, what aspects of *function* and/or *cellular localization* do they share

Mouse Genome Informatics (MGI)

the community informatics resource for the laboratory mouse



Genotype **Expression** **Phenotype**

Function

Objective:

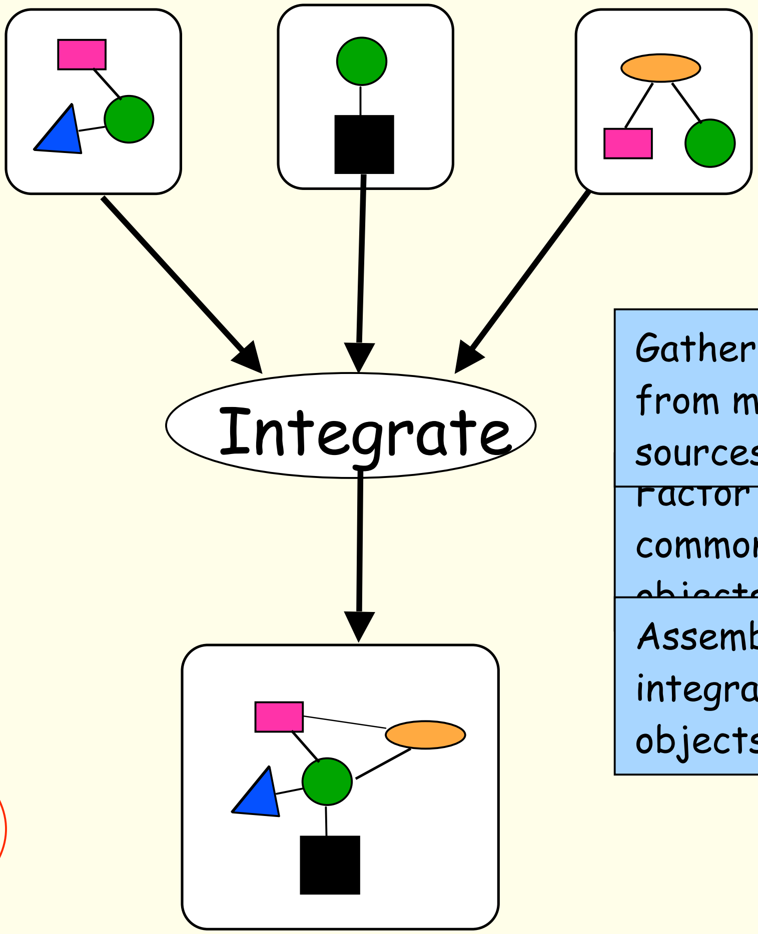
Facilitate the use of the mouse as a model for human biology by furthering our understanding of the relationship between genotype and phenotype.

Common Issues for Model Organism Databases

- Data Integration of Heterogeneous Data Sets
  - From Genotype to Phenotype
  - Experimental and Consensus Views
- Incorporation of Large Datasets
  - Whole genome annotation pipelines
  - Large scale mutagenesis projects
- Computational vs. Literature-based Data Collection and Evaluation
- Data Mining and Hypothesis Generation
  - extraction of new knowledge

Data Integration for Objects

- Within MGI
  - Genes
  - Sequence
  - Expression
  - Literature
  - Alleles
  - Phenotypes
- Between MGI and others
  - Via shared sequence annotations……SwissProt, LocusLink, RIKEN
  - Via shared semantic conceptualizations ……*Drosophila*, *Arabidopsis*, etc.

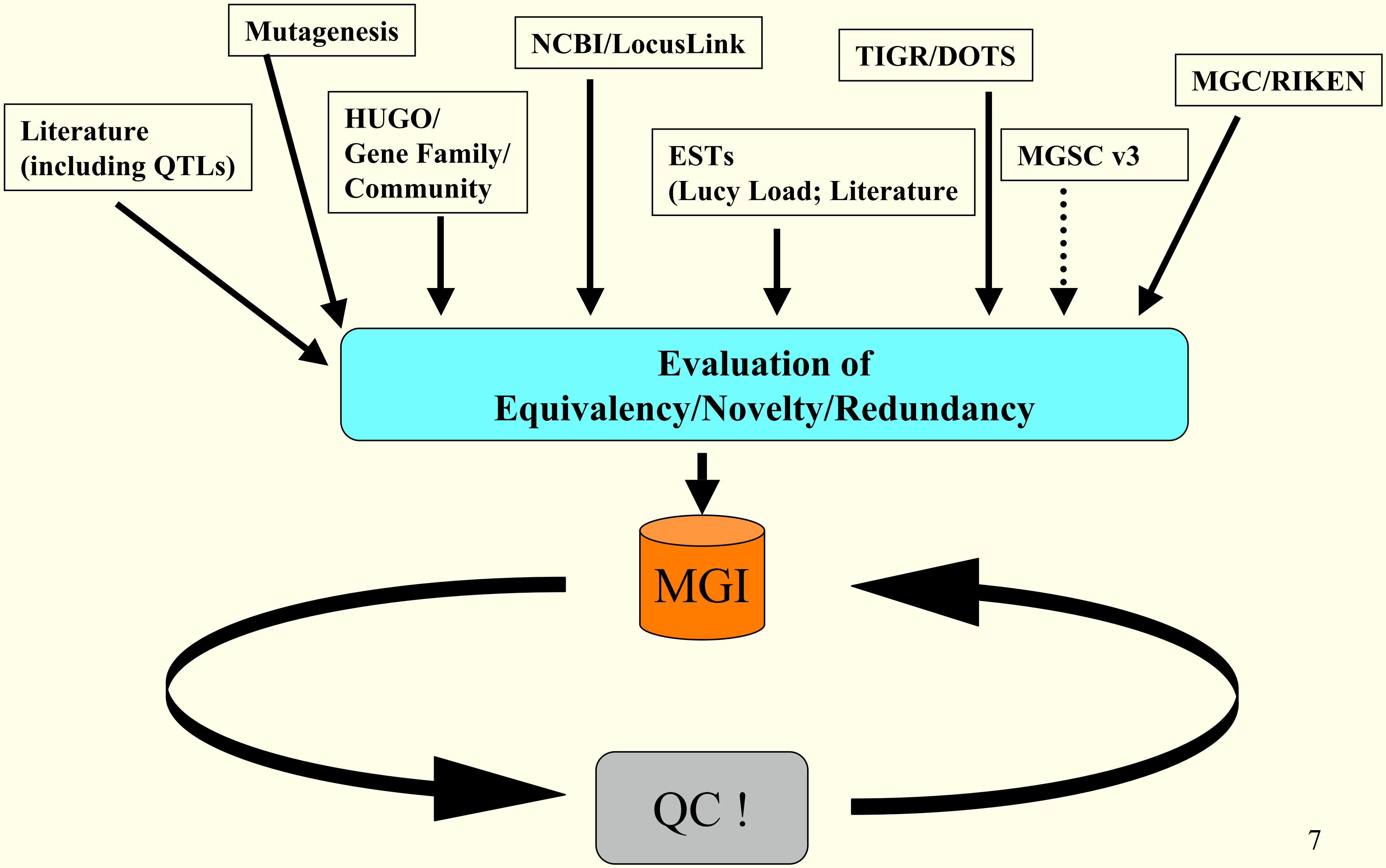


Gather data from multiple sources

Factor out common objects

Assemble integrated objects

Sources of New Genes and Loci For MGI



Literature (including QTLs)

Mutagenesis

HUGO/ Gene Family/ Community

NCBI/LocusLink

ESTs (Lucy Load; Literature

TIGR/DOTS

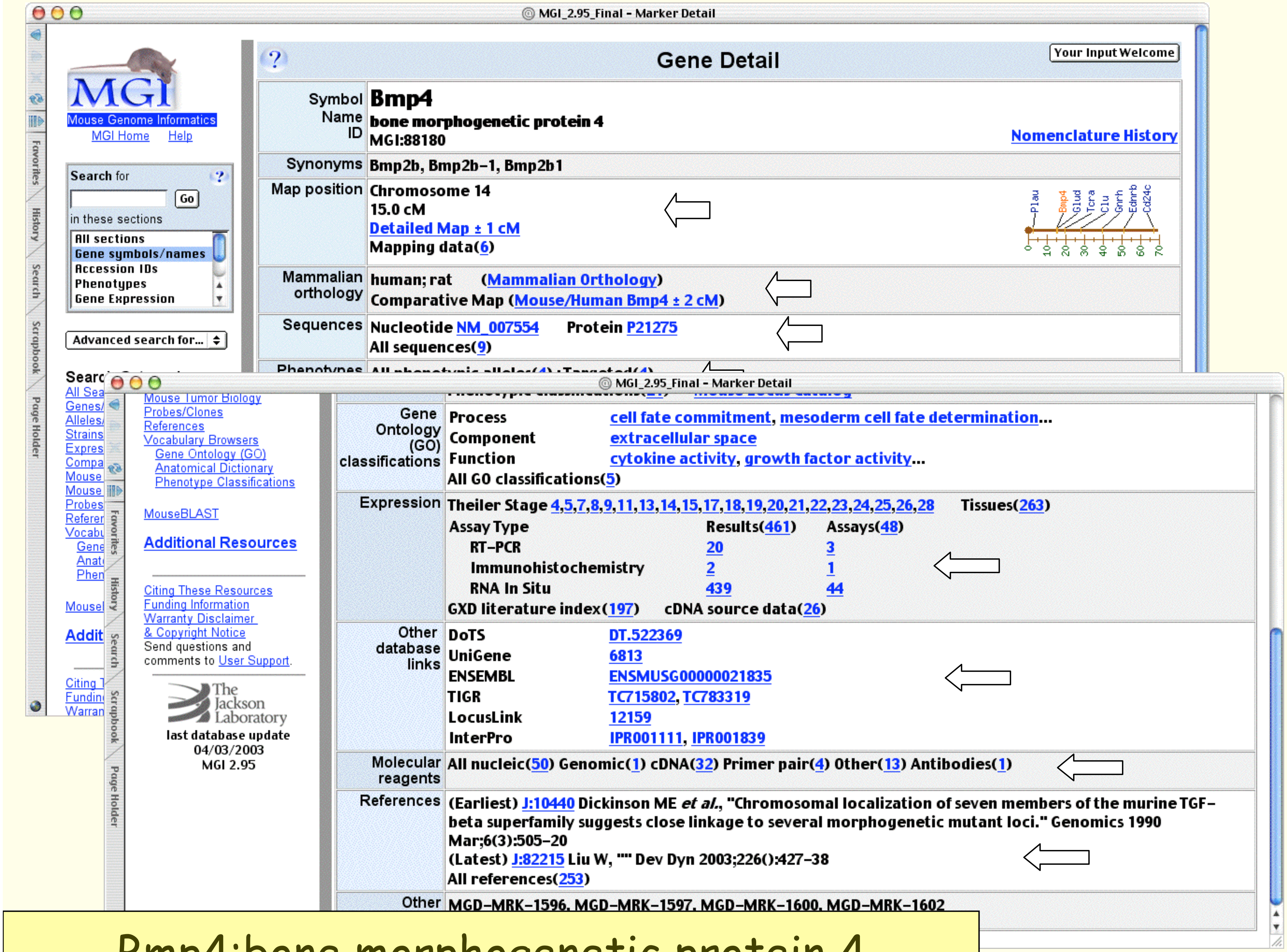
MGSC v3

MGC/RIKEN

Evaluation of Equivalency/Novelty/Redundancy

MGI

QC !



Send questions and comments to User Support.

The Jackson Laboratory
last database update
04/03/2003
MGI 2.95

Gene Detail

Symbol	**Bmp4**
Name	bone morphogenetic protein 4
ID	MGI:88180
Synonyms	Bmp2b, Bmp2b-1, Bmp2b1
Map position	Chromosome 14 15.0 cM Detailed Map ± 1 cM Mapping data(6)
Mammalian orthology	human; rat (Mammalian Orthology) Comparative Map (Mouse/Human Bmp4 ± 2 cM)
Sequences	Nucleotide NM_007554 Protein P21275 All sequences(9)
Gene Ontology (GO) classifications	Process: cell fate commitment, mesoderm cell fate determination... Component: extracellular space Function: cytokine activity, growth factor activity... All GO classifications(5)
Expression	Theiler Stage 4,5,7,8,9,11,13,14,15,17,18,19,20,21,22,23,24,25,26,28 Tissues(263)
Other database links	DoTS: DT.522369 UniGene: 6813 ENSEMBL: ENSMUSG00000021835 TIGR: TC715802, TC783319 LocusLink: 12159 InterPro: IPR001111, IPR001839
Molecular reagents	All nucleic(50) Genomic(1) cDNA(32) Primer pair(4) Other(13) Antibodies(1)
References	(Earliest) J:10440 Dickinson ME *et al.*, "Chromosomal localization of seven members of the murine TGF-beta superfamily suggests close linkage to several morphogenetic mutant loci." Genomics 1990 Mar;6(3):505-20 (Latest) J:82215 Liu W, "" Dev Dyn 2003;226():427-38 All references(253)
Other	MGD-MRK-1596, MGD-MRK-1597, MGD-MRK-1600, MGD-MRK-1602

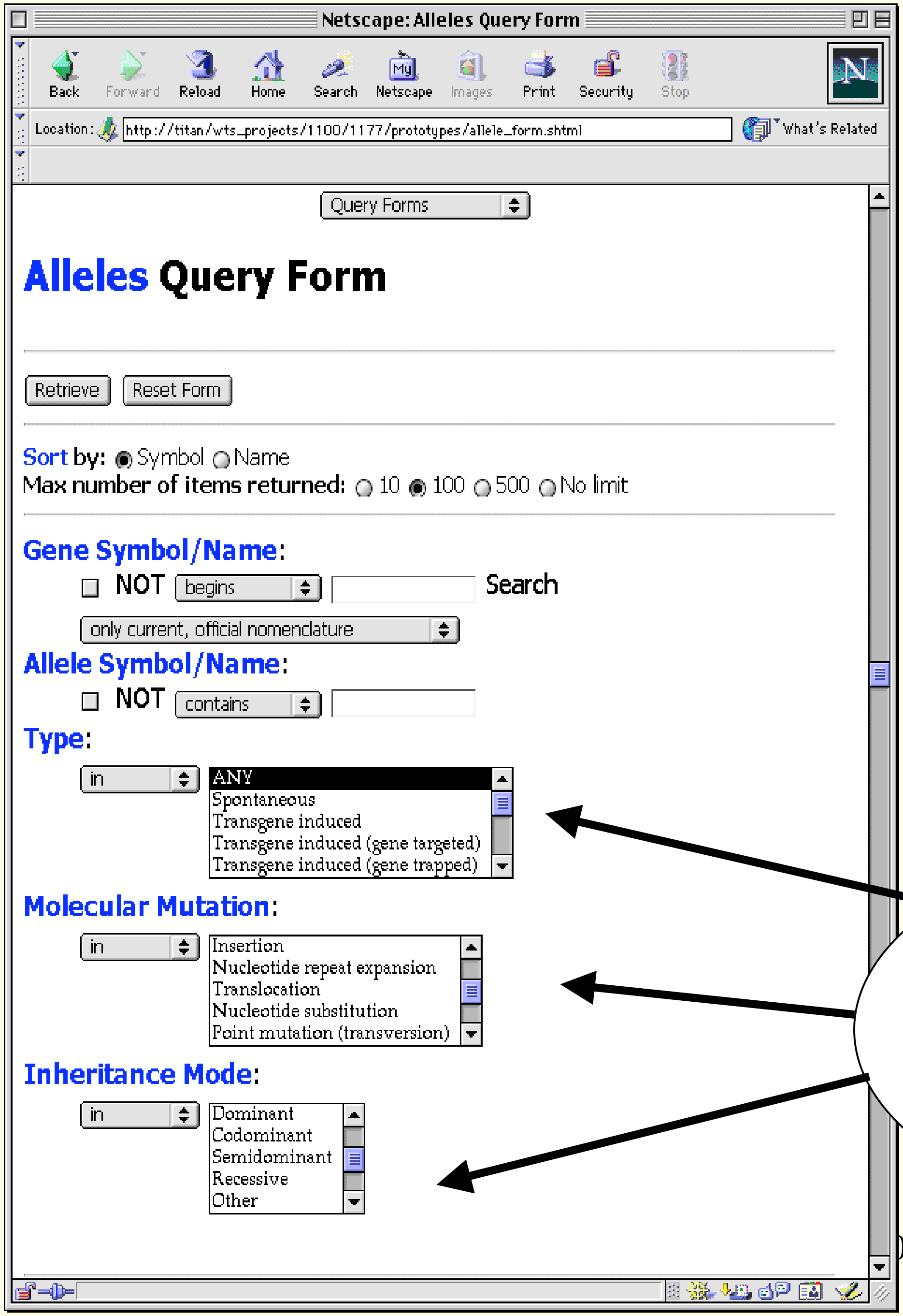
Assay Type	Results(461)	Assays(48)
RT-PCR	20	3
Immunohistochemistry	2	1
RNA In Situ	439	44

GXD literature index(197) cDNA source data(26)

Bmp4:bone morphogenetic protein 4

Semantic Integration of Shared Concepts

- Uniform Data Encoding
- Searchability
- Analysis and Comparison
- Complex Queries



Location: http://titan/wts_projects/1100/1177/prototypes/allele_form.shtml

Alleles Query Form

Sort by: ● Symbol ○ Name
Max number of items returned: ○ 10 ● 100 ○ 500 ○ No limit

Gene Symbol/Name:
☐ NOT begins
only current, official nomenclature

Allele Symbol/Name:
☐ NOT contains

Type:
in
ANY
Spontaneous
Transgene induced
Transgene induced (gene targeted)
Transgene induced (gene trapped)

Molecular Mutation:
in
Insertion
Nucleotide repeat expansion
Translocation
Nucleotide substitution
Point mutation (transversion)

Inheritance Mode:
in
Dominant
Codominant
Semidominant
Recessive
Other

Controlled Vocabularies for Annotation and Queries of Alleles

Multiple Keyword (C.V.) Sets in MGI

- Gene Nomenclature
- Gene/Marker Type
- Allele Type
- Assay Type
  - Expression
  - Mapping
- Molecular Mutation
- Inheritance Mode
- Evidence Codes
- Tissue Types
- Cell Lines
- Units
  - Cytogenetic
  - Molecular
- ES Cell Line
- Strain Nomenclature

Gene Names

keyword list

- Unique, rule-based names with synonyms
- Flat file
- Structure embedded in symb
- Gene grouping by name not c

- Primary for communicatio
- Strong community input
- Shared with Human and Rat

Where's *Cct1* ??

***Cat2*, chaperonin subunit 2**

***Cat3*, chaperonin subunit 3**

***Cat4*, chaperonin subunit 4**

***Cat5*, chaperonin subunit 5**

Cat1*, current symbol *Tcp1

But, keyword lists are not enough

- **Sheer number of terms too much to remember and sort**
  - Need standardized, stable, carefully defined terms
  - Need to describe different levels of detail
  - So…defined terms need to be related in a hierarchy

- **With structured vocabularies/hierarchies**
  - Parent/child relationships exist between terms
  - Increased depth -> Increased resolution
  - Can annotate data at appropriate level
  - May query at appropriate level

- **All model organisms database and genome annotation systems have same issues**

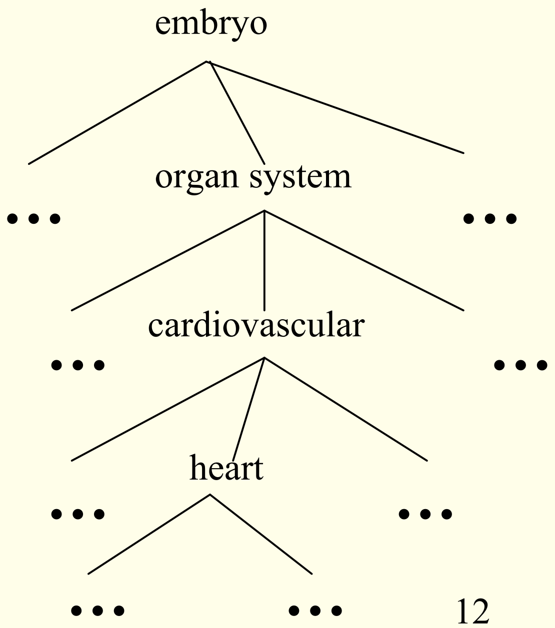
Anatomy keywords

Organ system

Cardiovascular system

Heart

Anatomy Hierarchy

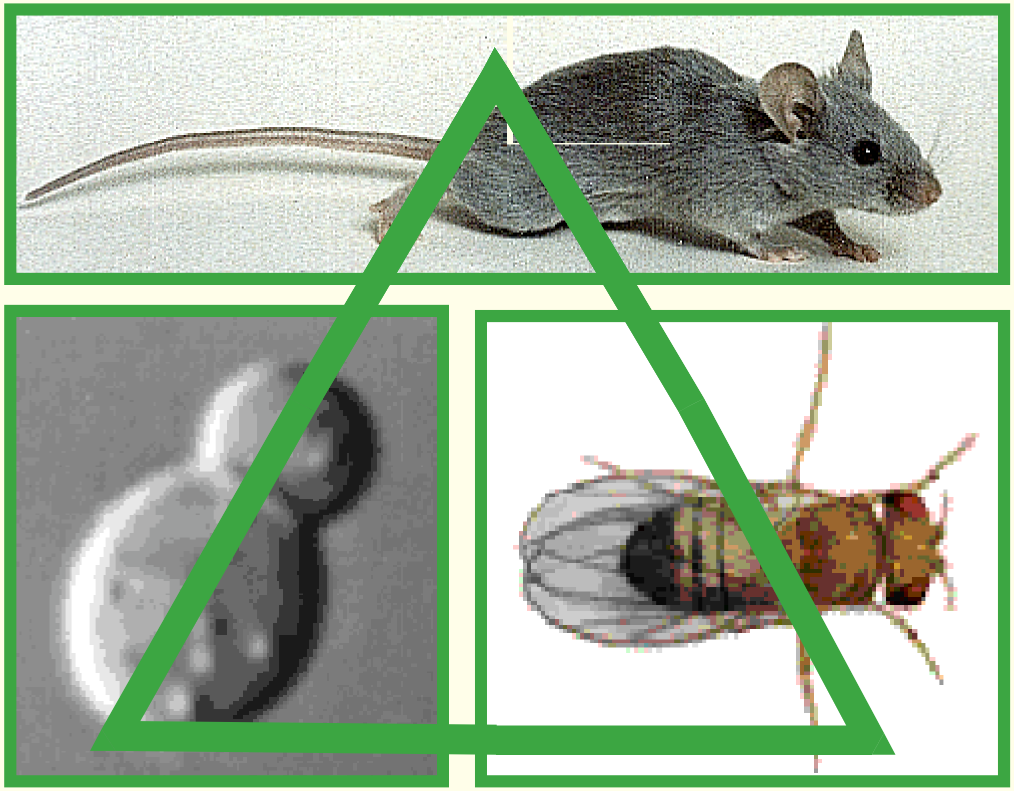


And thus, ‘we’ started the GO

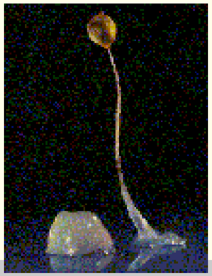
- Formed to develop a shared language adequate for the annotation of molecular characteristics across organisms.
- Seeks to achieve a mutual understanding of the definition and meaning of any word used. thus we are able to support cross-database queries.
- Members agree to provide database access via these common terms to gene product annotations and associated sequences.

GO began with recognized common need

- describing molecular biology of genes & gene products
- practical solution for implementation & use
- unifying, expandable, organism independent vocabularies



www.geneontology.org



DictyBase

The Wellcome Trust Sanger Institute

The *Schizosaccharomyces pombe* Genome Sequencing Project

The Pathogen Group

MGI Mouse Genome Informatics

FlyBase

BDGP

SGD

GO GENE ONTOLOGY™ CONSORTIUM

http://www.geneontology.org

InterPro

tair

SIB swissprot EBI

Genome KnowledgeBase

IncyteGenomics For Life.

GRAMENE

WormBase

compugen

ZFIN

AstraZeneca

TJ

TIGR THE INSTITUTE FOR GENOMIC RESEARCH

RGD RAT GENOME DATABASE

The GO vocabularies

- Molecular Function:

  What a product 'does', precise activity

- Biological Process

  Biological objective, accomplished via one or more ordered assemblies of functions

- Cellular Component

  'is located in' ('is a subcomponent of')

GO Project Goals:

1. Design structured vocabularies describing aspects of molecular biology

2. Support annotation of gene products using vocabulary terms

3. Provide database access via these common terms to gene product annotations and associated sequences

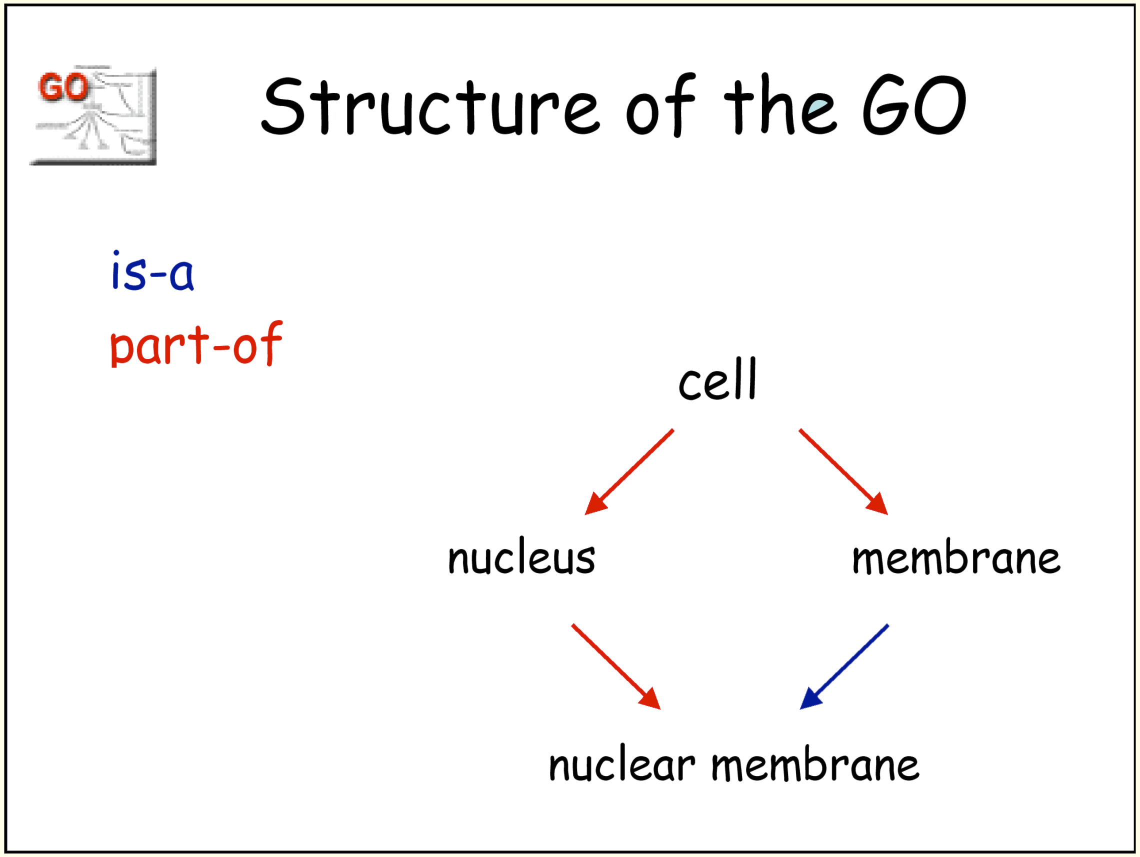
The Key Decisions:

- The vocabulary itself requires a serious and ongoing effort.
- Every concept must be carefully defined.
- The minimal data structure is a directed acyclic graph.
- All resources and annotations will be made publicly available to the community in a variety of formats (open source)

What GO is NOT:

- Not a way to unify biological databases
- Not a dictated standard
- Not a database of *gene products*, protein domains, or motifs
- Does not define evolutionary relationships

1. Build Vocabularies (ontologies)



Structure of the GO

is-a

part-of

cell

nucleus

membrane

nuclear membrane

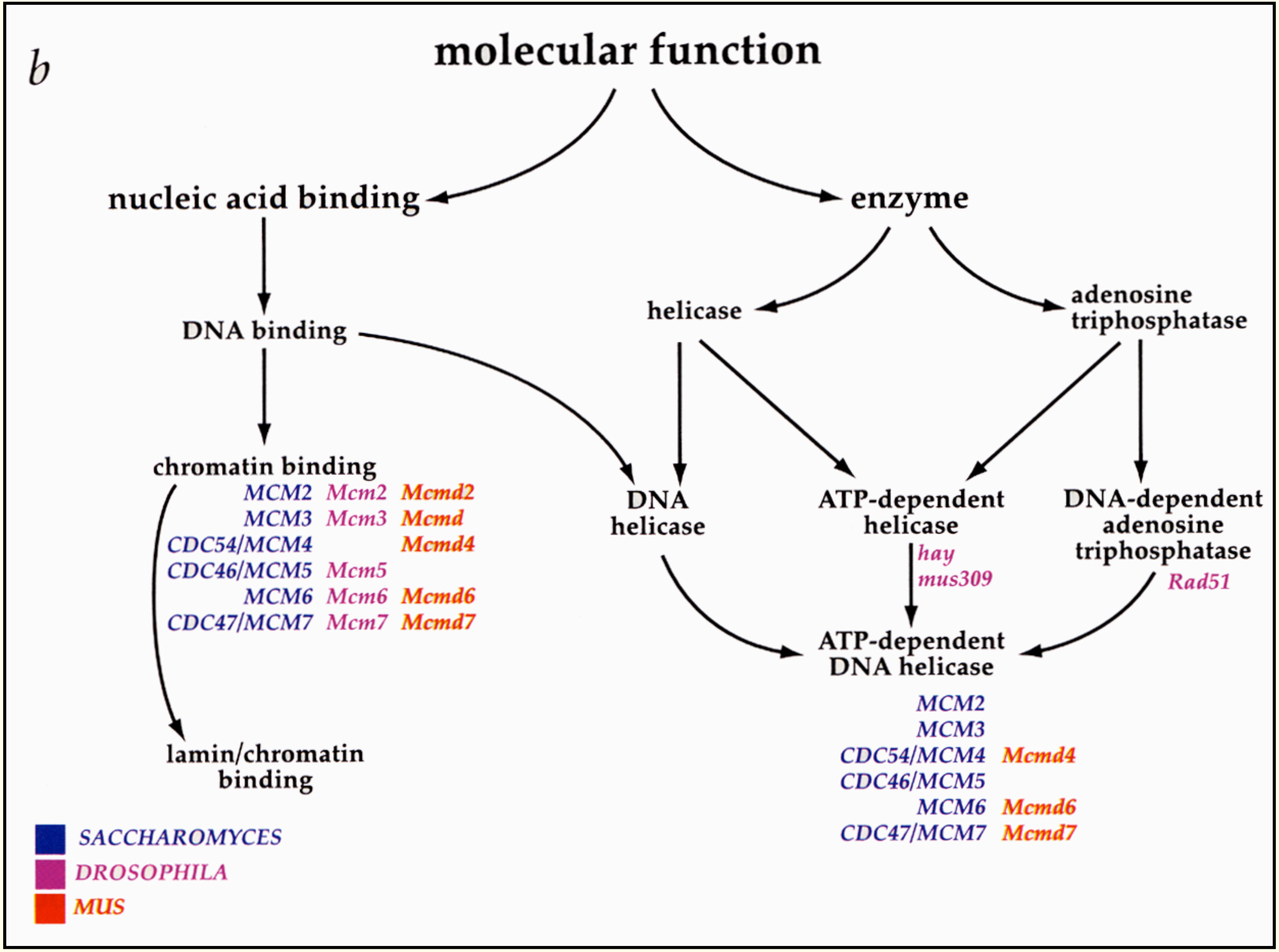
a. Directed acyclic graph(DAG): each child may have one or more parents

b. Relationships between terms defined

c. All terms are defined, accession ID associated with definition

d. True Path, all attributes of children must hold for all parents

An example of molecular function



b

molecular function

nucleic acid binding

enzyme

DNA binding

helicase

adenosine triphosphatase

chromatin binding

MCM2 *Mcm2* *Mcmd2*
MCM3 *Mcm3* *Mcmd*
CDC54/MCM4 *Mcmd4*
CDC46/MCM5 *Mcm5*
MCM6 *Mcm6* *Mcmd6*
CDC47/MCM7 *Mcm7* *Mcmd7*

DNA helicase

ATP-dependent helicase

hay
mus309

DNA-dependent adenosine triphosphatase

Rad51

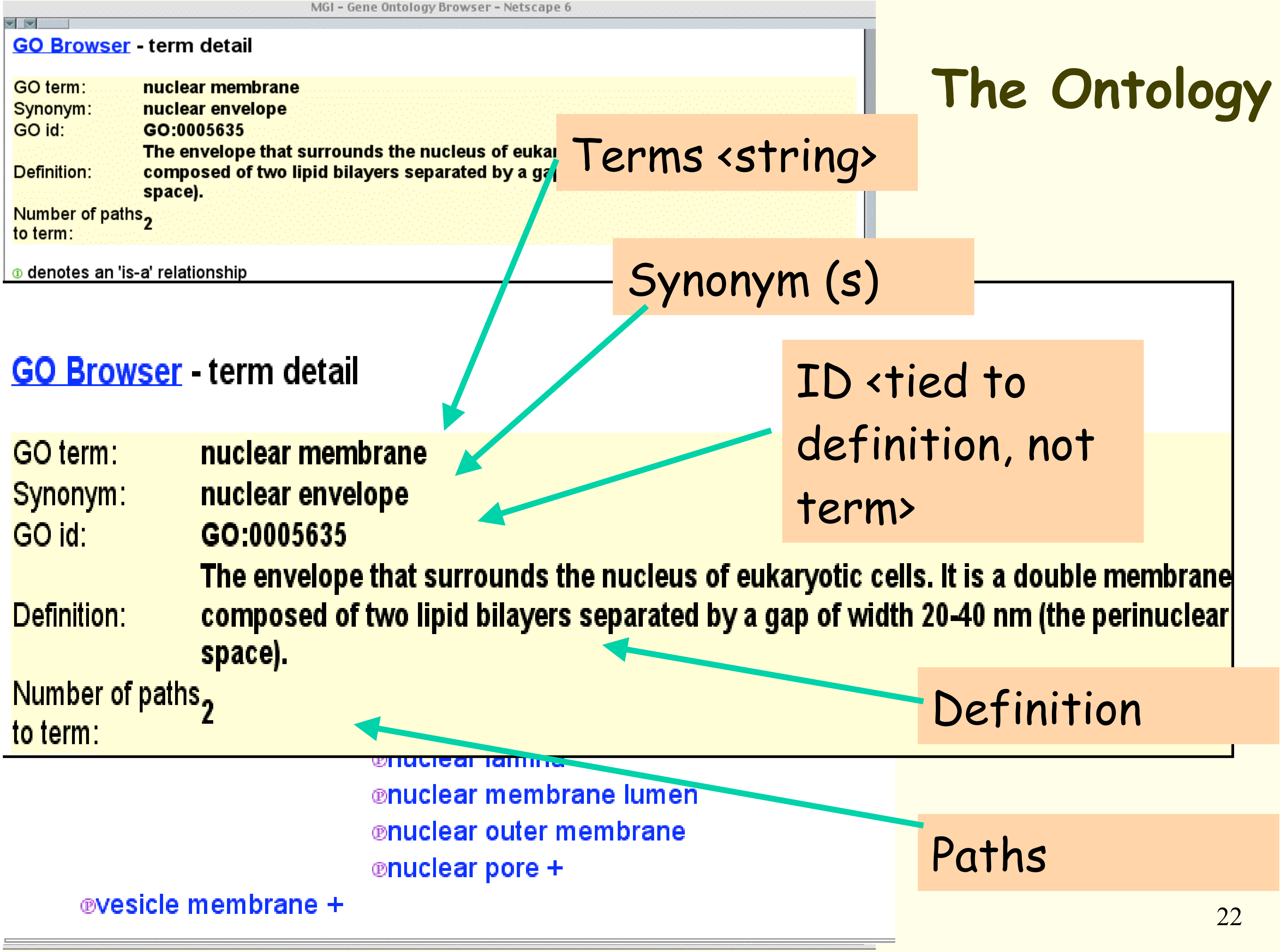
lamin/chromatin binding

ATP-dependent DNA helicase

MCM2
MCM3
CDC54/MCM4 *Mcmd4*
CDC46/MCM5
MCM6 *Mcmd6*
CDC47/MCM7 *Mcmd7*

🟦 *SACCHAROMYCES*
🟪 *DROSOPHILA*
🟥 *MUS*

The Ontology



GO Browser - term detail

GO term:	**nuclear membrane**
Synonym:	**nuclear envelope**
GO id:	**GO:0005635**
Definition:	**The envelope that surrounds the nucleus of eukar composed of two lipid bilayers separated by a ga space).**
Number of paths to term:	**2**

ⓘ denotes an 'is-a' relationship

GO Browser - term detail

GO term:	**nuclear membrane**
Synonym:	**nuclear envelope**
GO id:	**GO:0005635**
Definition:	**The envelope that surrounds the nucleus of eukaryotic cells. It is a double membrane composed of two lipid bilayers separated by a gap of width 20-40 nm (the perinuclear space).**
Number of paths to term:	**2**

ⓟnuclear lamina
ⓟnuclear membrane lumen
ⓟnuclear outer membrane
ⓟnuclear pore +
ⓟvesicle membrane +

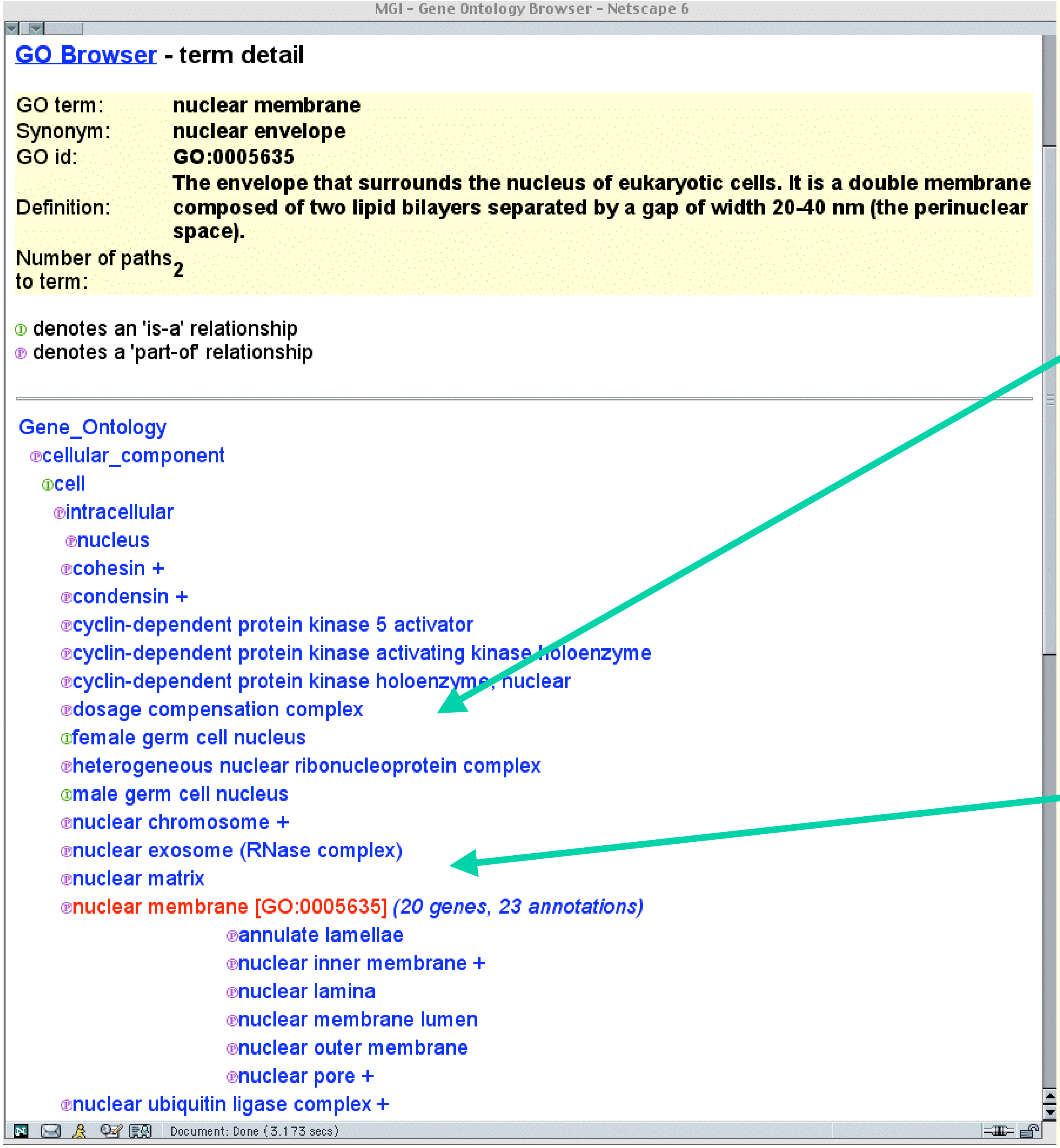
Terms <string>

Synonym (s)

ID <tied to definition, not term>

Definition

Paths



GO Browser - term detail

GO term:	**nuclear membrane**
Synonym:	**nuclear envelope**
GO id:	**GO:0005635**
Definition:	**The envelope that surrounds the nucleus of eukaryotic cells. It is a double membrane composed of two lipid bilayers separated by a gap of width 20-40 nm (the perinuclear space).**
Number of paths to term:	**2**

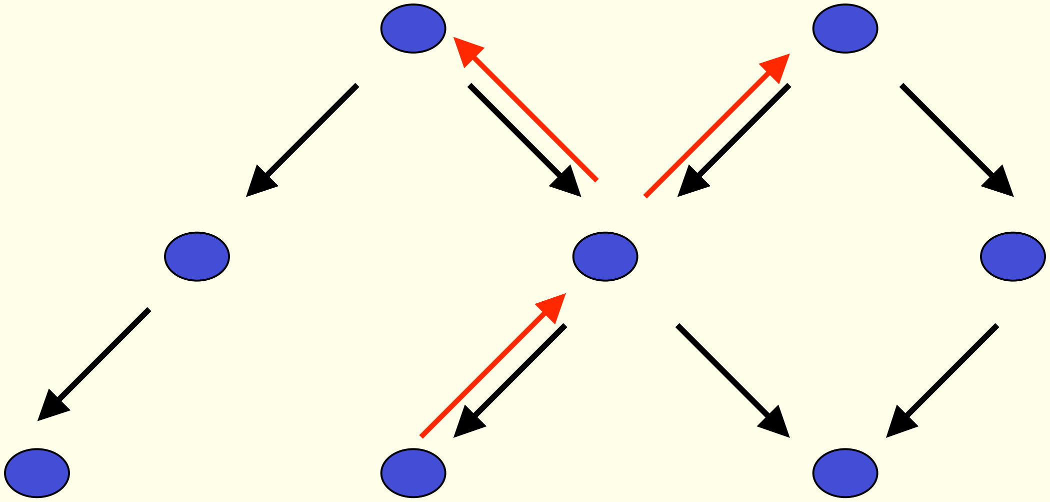
ⓘ denotes an 'is-a' relationship
ⓟ denotes a 'part-of' relationship

- Gene_Ontology
  - ⓟcellular_component
    - ⓘcell
      - ⓟintracellular
        - ⓟnucleus
          - ⓟcohesin +
          - ⓟcondensin +
          - ⓟcyclin-dependent protein kinase 5 activator
          - ⓟcyclin-dependent protein kinase activating kinase holoenzyme
          - ⓟcyclin-dependent protein kinase holoenzyme, nuclear
          - ⓟdosage compensation complex
          - ⓘfemale germ cell nucleus
          - ⓟheterogeneous nuclear ribonucleoprotein complex
          - ⓘmale germ cell nucleus
          - ⓟnuclear chromosome +
          - ⓟnuclear exosome (RNase complex)
          - ⓟnuclear matrix
          - ⓟnuclear membrane [GO:0005635] *(20 genes, 23 annotations)*
            - ⓟannulate lamellae
            - ⓟnuclear inner membrane +
            - ⓟnuclear lamina
            - ⓟnuclear membrane lumen
            - ⓟnuclear outer membrane
            - ⓟnuclear pore +
          - ⓟnuclear ubiquitin ligase complex +

A "female germ cell nucleus" *is-a* instance of a nucleus

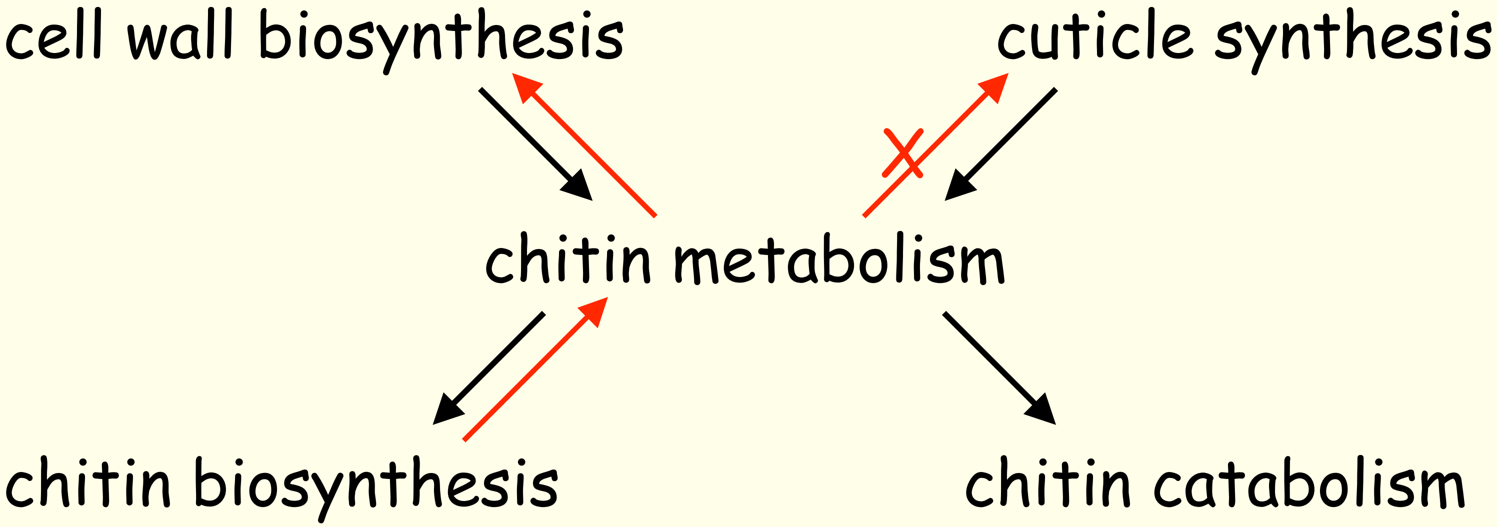
The "nuclear matrix" is *part-of* the nucleus

The True Path Rule



Every path from a node back to the root must be biologically accurate

The True Path Rule



cell wall biosynthesis

cuticle synthesis

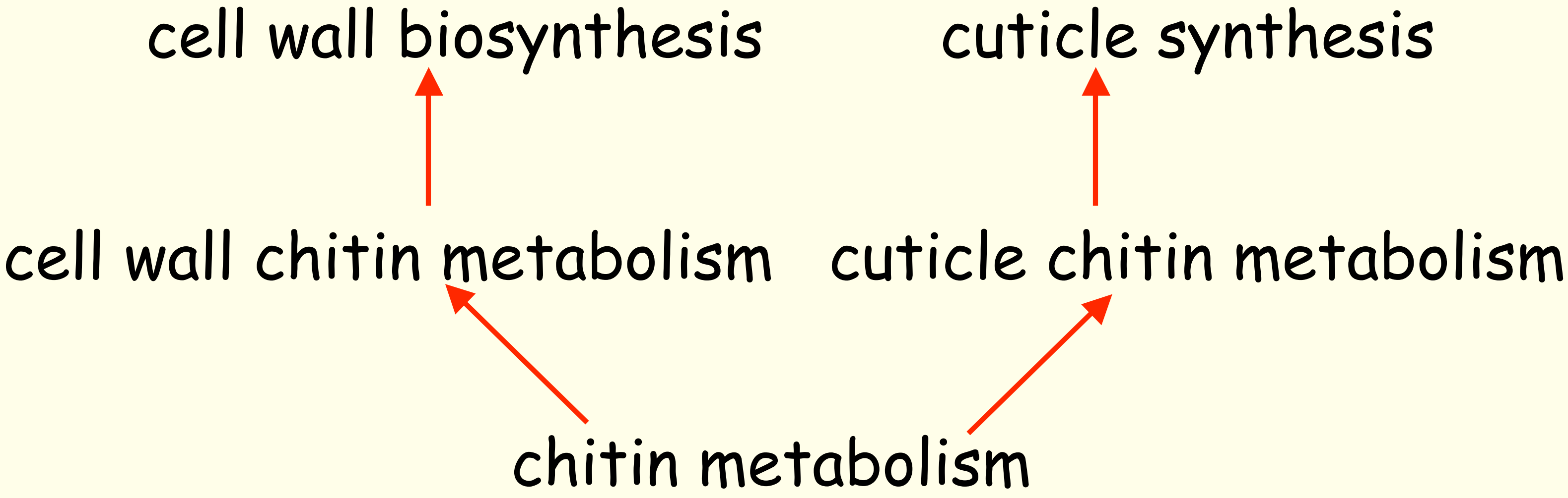
chitin metabolism

chitin biosynthesis

chitin catabolism

chitin metabolism: before revision

The True Path Rule

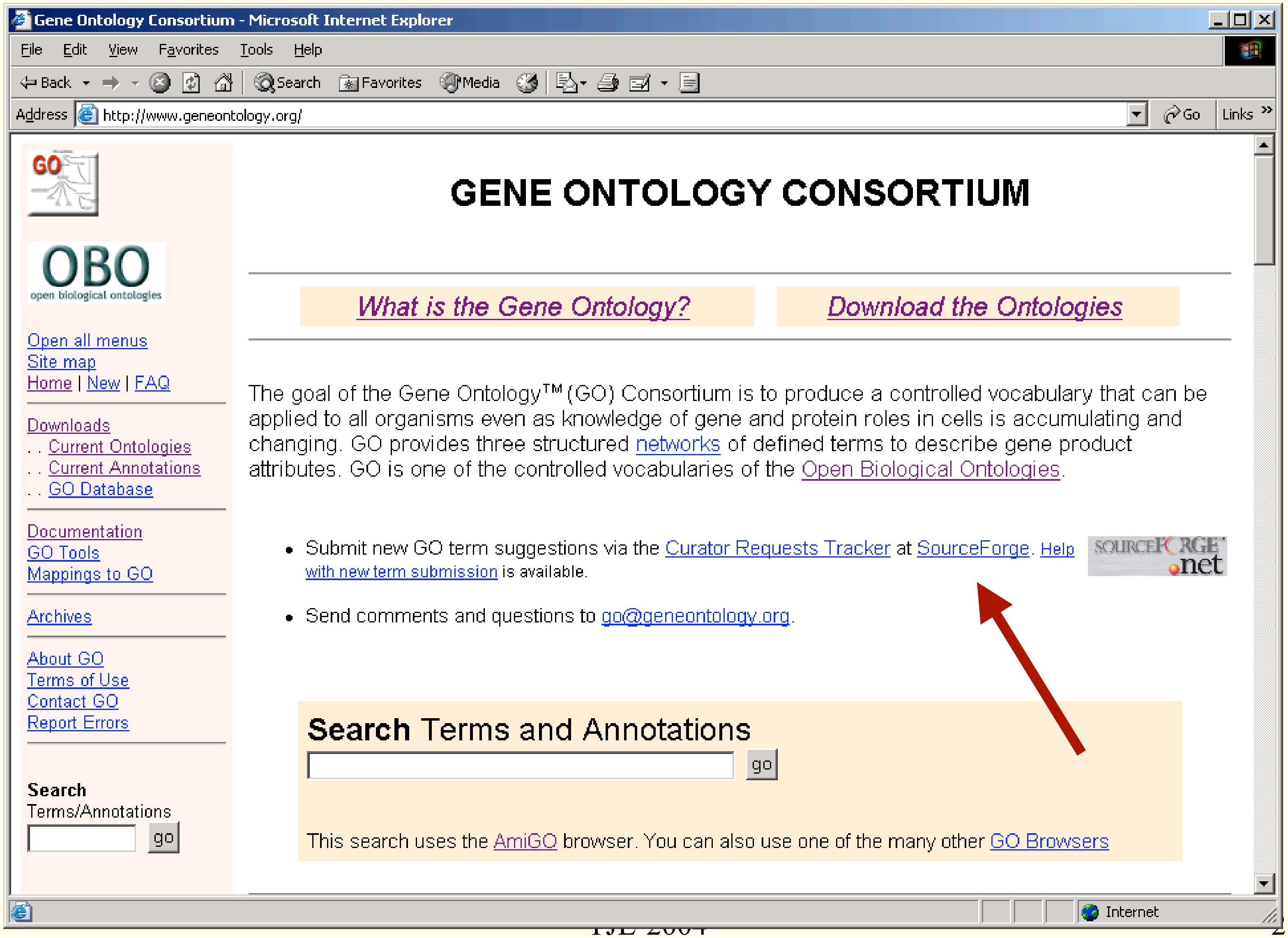


chitin metabolism: after revision

Current GO Projects

- Build Vocabularies
  - > 17,500 terms, 95% defined
  - 'Interest Groups' (*cell motility, protein modification*)
- Update Vocabularies
  - Add to UMLS / MeSH system
    - 'activity' term for function
  - SourceForge site for community input and tracking
- Add attributes (slots) for terms
  - Ex. 'DNA-binding with ……..
- Implement in DAML + OIL/OWL, etc.

GO Web Site:www.geneontology.org



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

Open all menus
Site map
Home | New | FAQ

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

Documentation
GO Tools
Mappings to GO

Archives

About GO
Terms of Use
Contact GO
Report Errors

Search
Terms/Annotations
go

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.

SOURCEFORGE .net

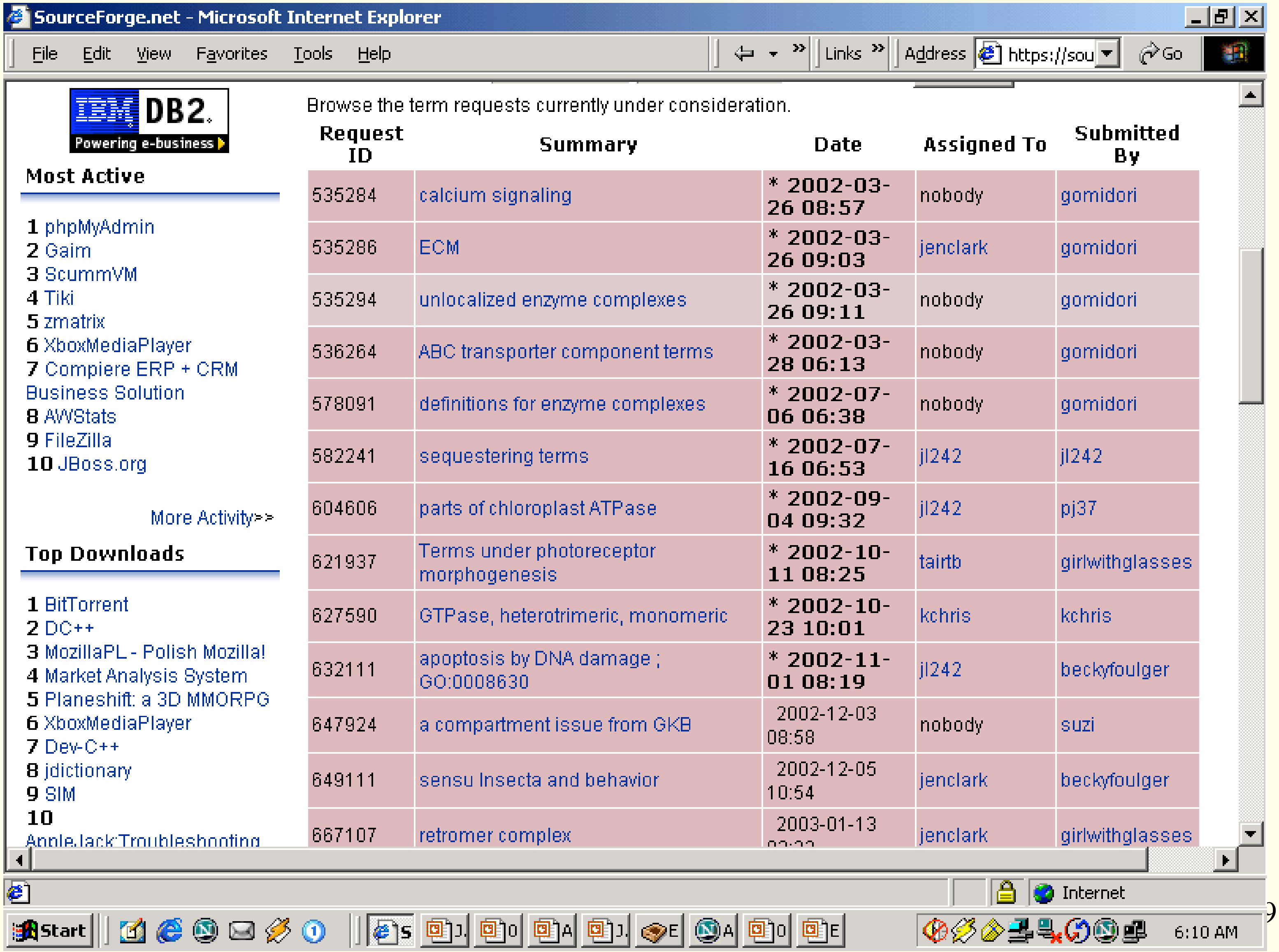
Search Terms and Annotations

go

This search uses the AmiGO browser. You can also use one of the many other GO Browsers

GO SourceForge Site

(for suggestions, corrections, interest groups)



Most Active

1 phpMyAdmin
2 Gaim
3 ScummVM
4 Tiki
5 zmatrix
6 XboxMediaPlayer
7 Compiere ERP + CRM Business Solution
8 AWStats
9 FileZilla
10 JBoss.org

More Activity>>

Top Downloads

1 BitTorrent
2 DC++
3 MozillaPL - Polish Mozilla!
4 Market Analysis System
5 Planeshift: a 3D MMORPG
6 XboxMediaPlayer
7 Dev-C++
8 jdictionary
9 SIM
10 AppleJackTroubleshooting

Browse the term requests currently under consideration.

Request ID	Summary	Date	Assigned To	Submitted By
535284	calcium signaling	* 2002-03-26 08:57	nobody	gomidori
535286	ECM	* 2002-03-26 09:03	jenclark	gomidori
535294	unlocalized enzyme complexes	* 2002-03-26 09:11	nobody	gomidori
536264	ABC transporter component terms	* 2002-03-28 06:13	nobody	gomidori
578091	definitions for enzyme complexes	* 2002-07-06 06:38	nobody	gomidori
582241	sequestering terms	* 2002-07-16 06:53	jl242	jl242
604606	parts of chloroplast ATPase	* 2002-09-04 09:32	jl242	pj37
621937	Terms under photoreceptor morphogenesis	* 2002-10-11 08:25	tairtb	girlwithglasses
627590	GTPase, heterotrimeric, monomeric	* 2002-10-23 10:01	kchris	kchris
632111	apoptosis by DNA damage ; GO:0008630	* 2002-11-01 08:19	jl242	beckyfoulger
647924	a compartment issue from GKB	2002-12-03 08:58	nobody	suzi
649111	sensu Insecta and behavior	2002-12-05 10:54	jenclark	beckyfoulger
667107	retromer complex	2003-01-13	jenclark	girlwithglasses

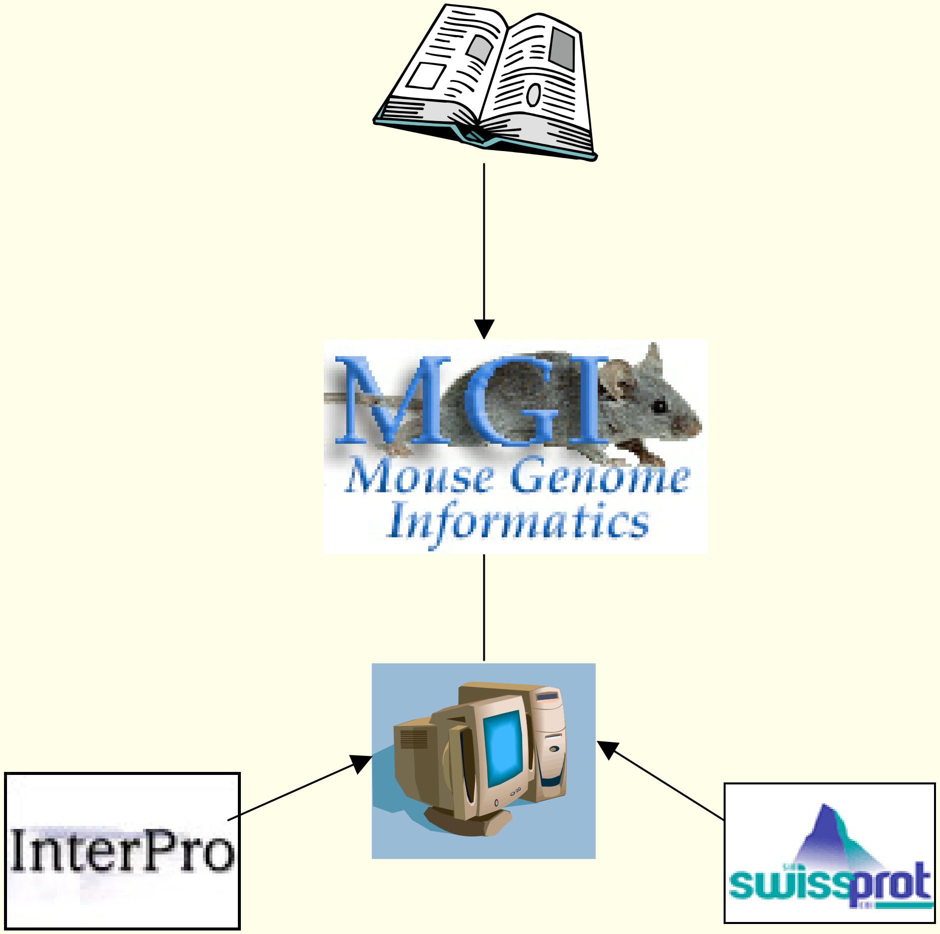
GO Goals - 2
Annotate genes & gene products to GO

- Use mouse (domain) biological expertise (literature)
- For each GO association provide
  - Evidence statement
  - Citation/attribution
- Annotate to finest granularity known experimentally
- Annotate 'NOT' value when determined

- Committed to GO repository weekly
- Committed to MGI ftp site nightly
- Incorporated into NCBI/LocusLink daily
- (Human_GO now in LL from EBI_SP_GOA)

GO Curation Strategies in MODs

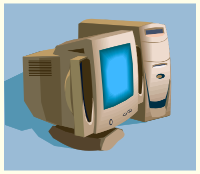
- Manual Curation
  - Emphasis on Primary Literature
  - Over 80,000 references
  - Five curators
- Computational
  - Collaborations between InterPro and SwissProt to integrate objects and assign GO terms
  - E.C. mappings
  - RIKEN pipeline

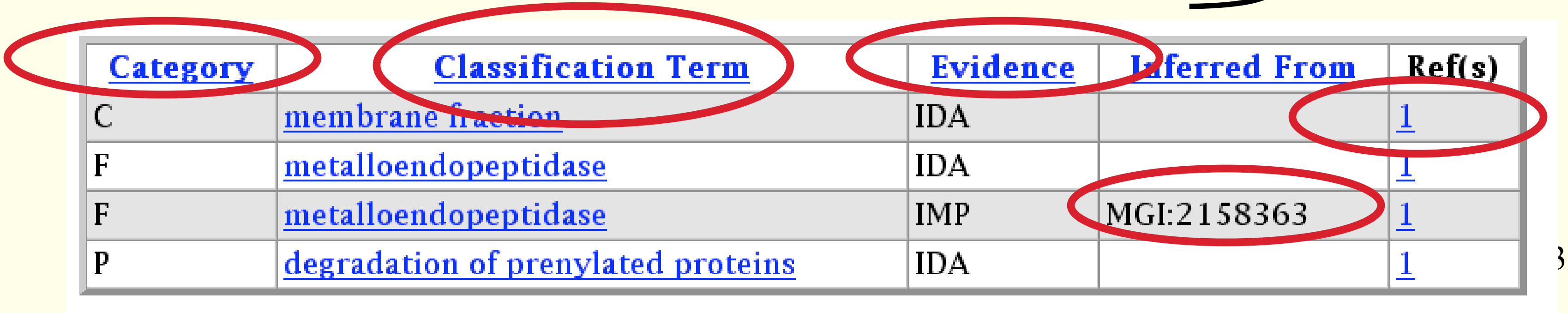


GO term associations supported by evidence

- ISS: Inferred from sequence or structural similarity
- IDA: Inferred from direct assay
- IPI: Inferred from physical interaction
- TAS: Traceable author statement
- IMP: Inferred from mutant phenotype
- IGI: Inferred from genetic interaction
- IEP: Inferred from expression pattern
- ND: no data available

- IEA: Inferred from electronic annotation

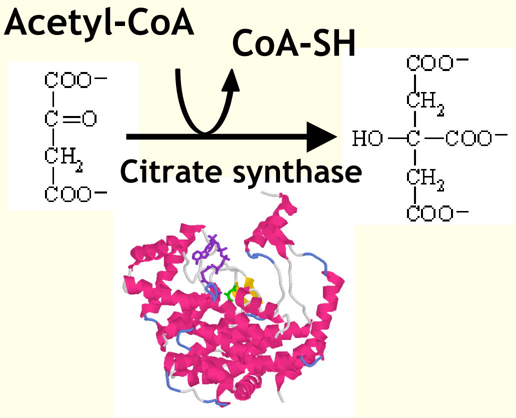




Category	Classification Term	Evidence	Inferred From	Ref(s)
C	membrane fraction	IDA		1
F	metalloendopeptidase	IDA		1
F	metalloendopeptidase	IMP	MGI:2158363	1
P	degradation of prenylated proteins	IDA		1

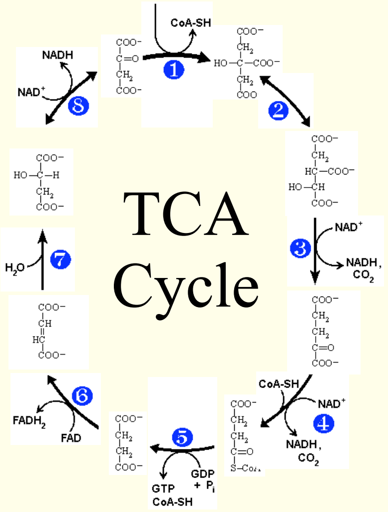
GO @ MGI

Function



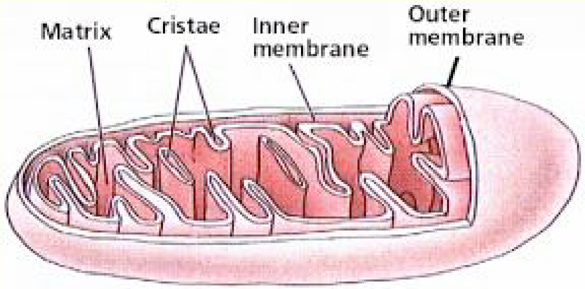
12,893 genes
30,308 annotations

Biological Process



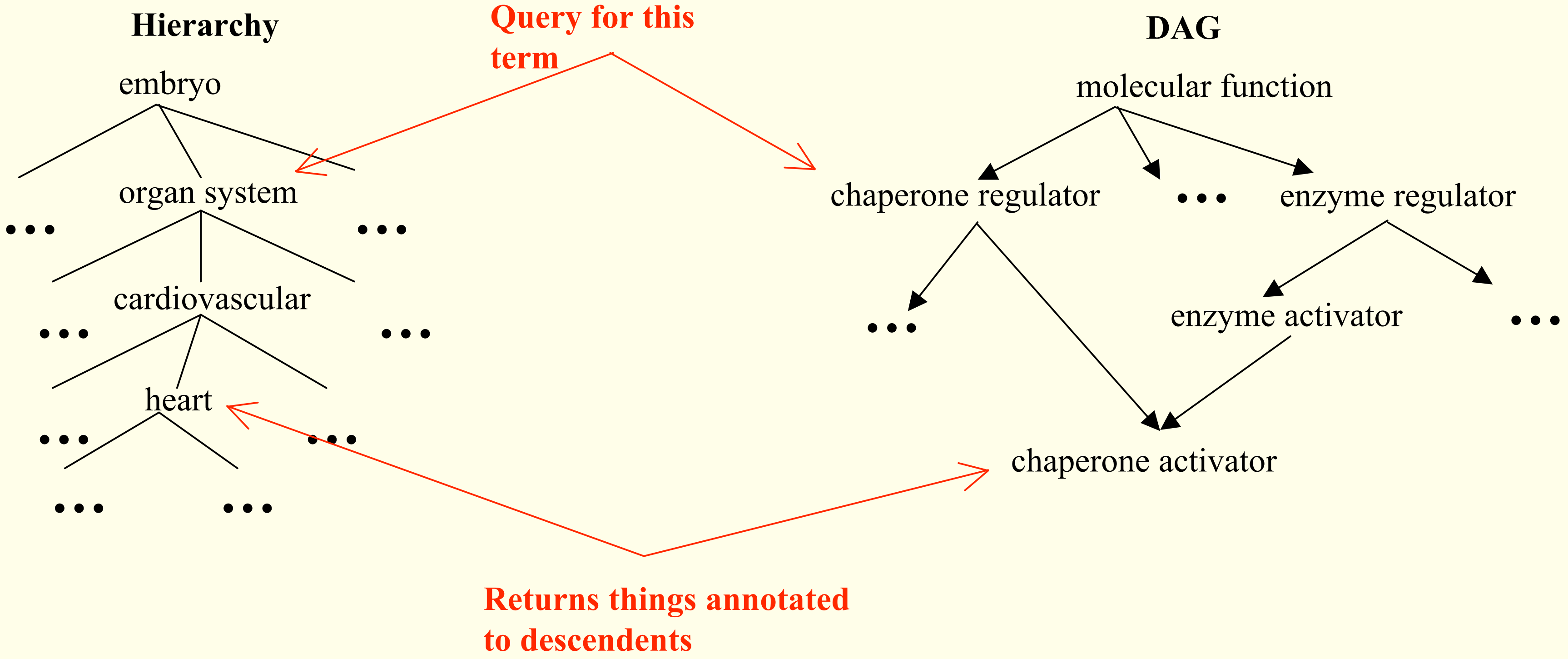
11,262 genes
21,895 annotations

Cellular Component



11,460 genes
20,049 annotations

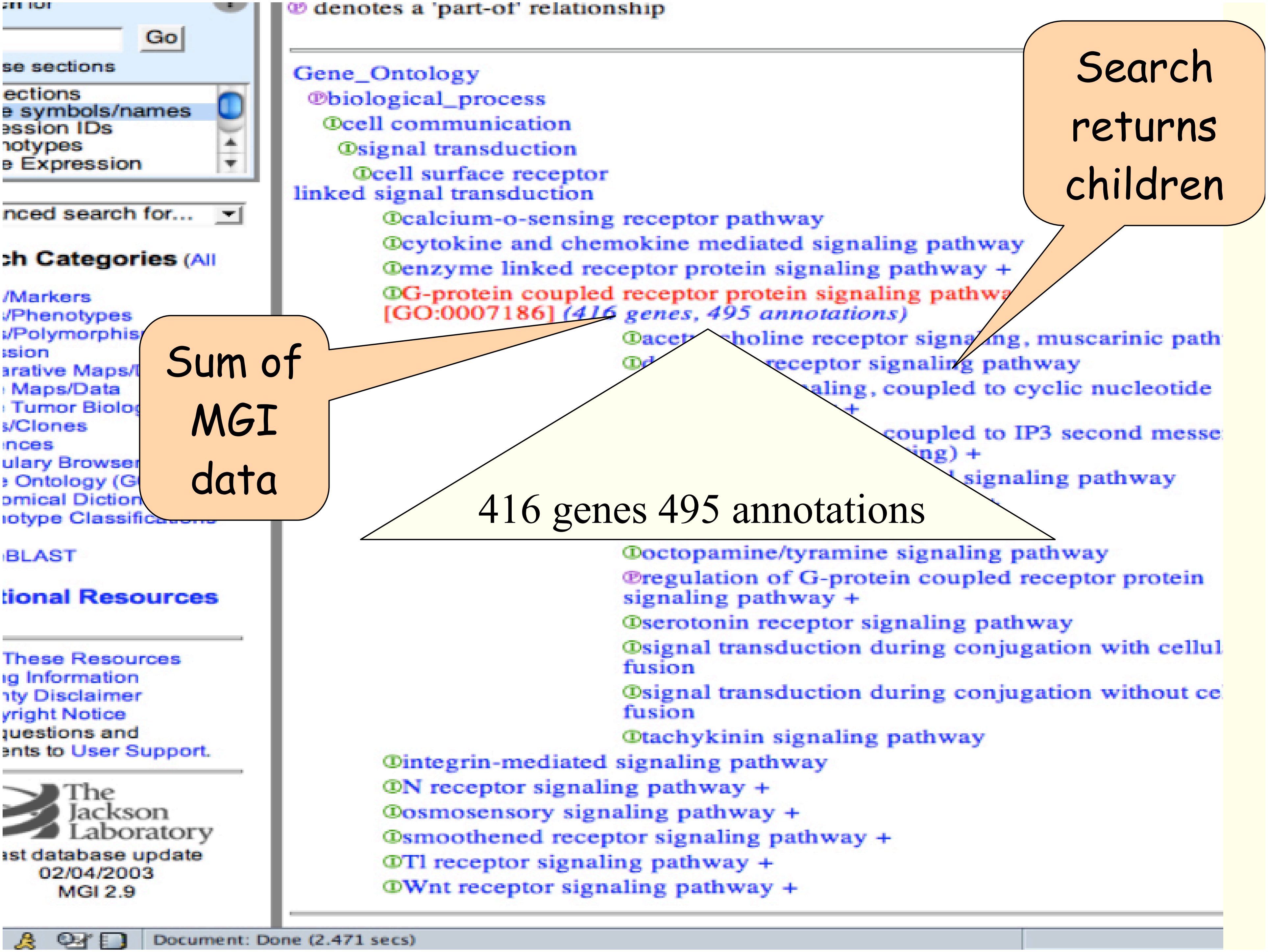
General Implementations for Vocabularies



Query for this term

Returns things annotated to descendents

1. **Annotate at appropriate level, query at appropriate level**
2. **Queries for higher level terms include annotations to lower level terms**


Ⓟ denotes a 'part-of' relationship

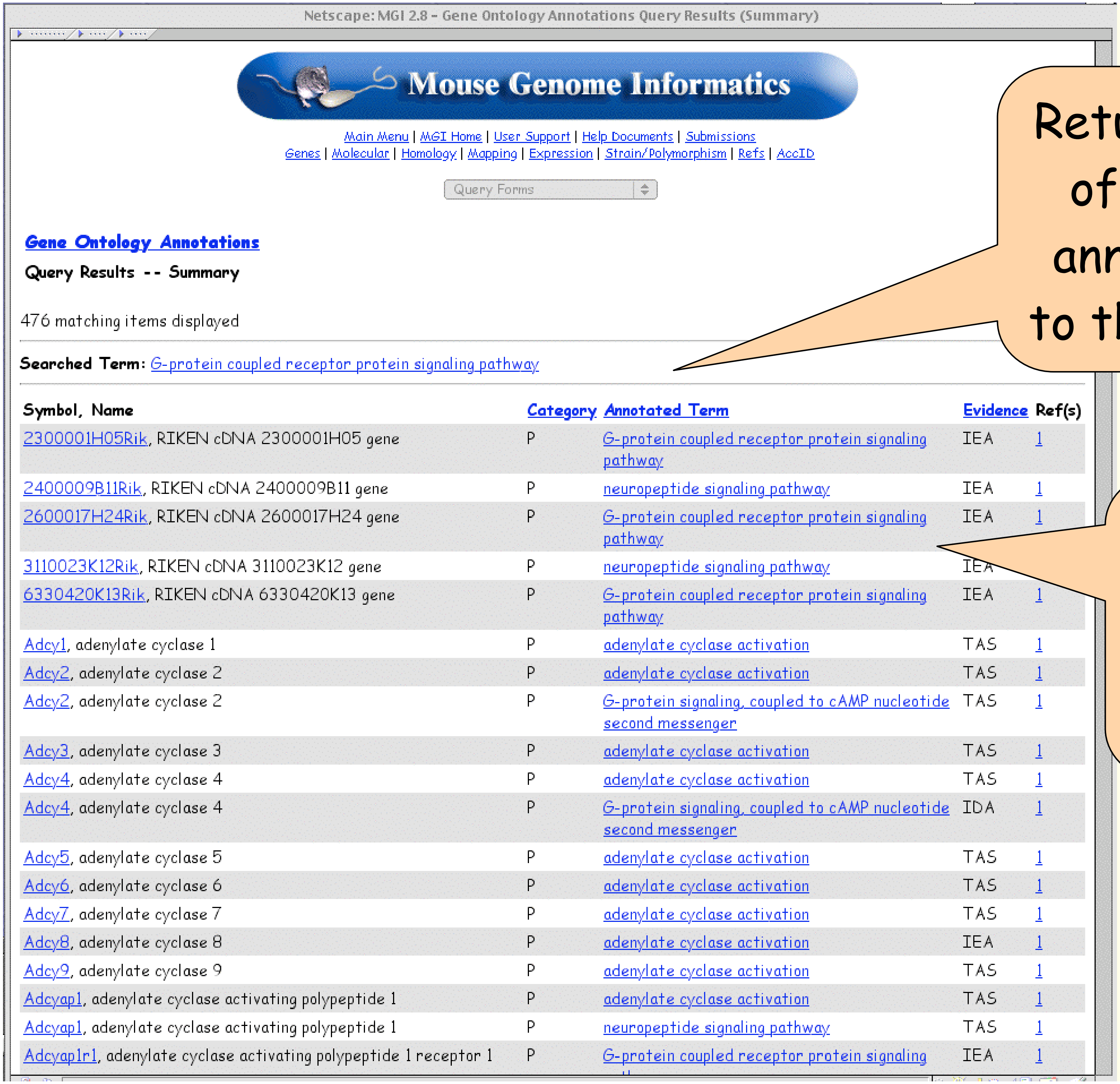
Gene_Ontology
Ⓟbiological_process
Ⓘcell communication
Ⓘsignal transduction
Ⓘcell surface receptor linked signal transduction
Ⓘcalcium-o-sensing receptor pathway
Ⓘcytokine and chemokine mediated signaling pathway
Ⓘenzyme linked receptor protein signaling pathway +
ⒾG-protein coupled receptor protein signaling pathwa
[GO:0007186] *(416 genes, 495 annotations)*
Ⓘacet holine receptor signa ing, muscarinic path
Ⓘd receptor signaling pathway
aling, coupled to cyclic nucleotide
+
coupled to IP3 second messe
ing) +
signaling pathway
Ⓘoctopamine/tyramine signaling pathway
Ⓟregulation of G-protein coupled receptor protein signaling pathway +
Ⓘserotonin receptor signaling pathway
Ⓘsignal transduction during conjugation with cellul fusion
Ⓘsignal transduction during conjugation without ce fusion
Ⓘtachykinin signaling pathway
Ⓘintegrin-mediated signaling pathway
ⒾN receptor signaling pathway +
Ⓘosmosensory signaling pathway +
Ⓘsmoothened receptor signaling pathway +
ⒾTl receptor signaling pathway +
ⒾWnt receptor signaling pathway +

Document: Done (2.471 secs)

Search returns children

Sum of MGI data

416 genes 495 annotations



Gene Ontology Annotations

Query Results -- Summary

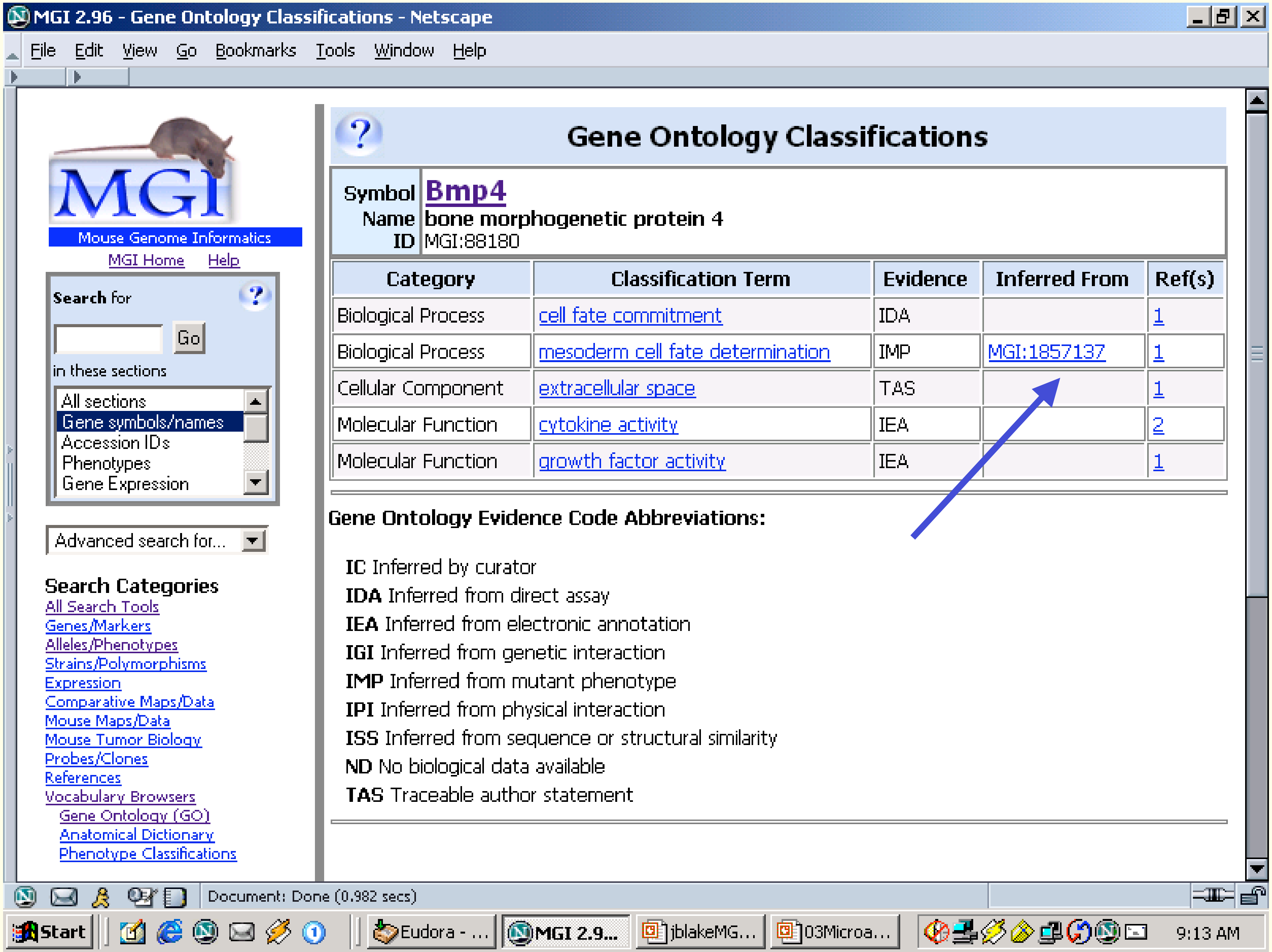
476 matching items displayed

Searched Term: G-protein coupled receptor protein signaling pathway

Symbol, Name	Category	Annotated Term	Evidence	Ref(s)
2300001H05Rik, RIKEN cDNA 2300001H05 gene	P	G-protein coupled receptor protein signaling pathway	IEA	1
2400009B11Rik, RIKEN cDNA 2400009B11 gene	P	neuropeptide signaling pathway	IEA	1
2600017H24Rik, RIKEN cDNA 2600017H24 gene	P	G-protein coupled receptor protein signaling pathway	IEA	1
3110023K12Rik, RIKEN cDNA 3110023K12 gene	P	neuropeptide signaling pathway	IEA	
6330420K13Rik, RIKEN cDNA 6330420K13 gene	P	G-protein coupled receptor protein signaling pathway	IEA	1
Adcy1, adenylate cyclase 1	P	adenylate cyclase activation	TAS	1
Adcy2, adenylate cyclase 2	P	adenylate cyclase activation	TAS	1
Adcy2, adenylate cyclase 2	P	G-protein signaling, coupled to cAMP nucleotide second messenger	TAS	1
Adcy3, adenylate cyclase 3	P	adenylate cyclase activation	TAS	1
Adcy4, adenylate cyclase 4	P	adenylate cyclase activation	TAS	1
Adcy4, adenylate cyclase 4	P	G-protein signaling, coupled to cAMP nucleotide second messenger	IDA	1
Adcy5, adenylate cyclase 5	P	adenylate cyclase activation	TAS	1
Adcy6, adenylate cyclase 6	P	adenylate cyclase activation	TAS	1
Adcy7, adenylate cyclase 7	P	adenylate cyclase activation	TAS	1
Adcy8, adenylate cyclase 8	P	adenylate cyclase activation	IEA	1
Adcy9, adenylate cyclase 9	P	adenylate cyclase activation	TAS	1
Adcyap1, adenylate cyclase activating polypeptide 1	P	adenylate cyclase activation	TAS	1
Adcyap1, adenylate cyclase activating polypeptide 1	P	neuropeptide signaling pathway	TAS	1
Adcyap1r1, adenylate cyclase activating polypeptide 1 receptor 1	P	G-protein coupled receptor protein signaling	IEA	1

Returns set of genes annotated to this term

New Genes with functional annotations



Gene Ontology Classifications

Symbol Bmp4
Name bone morphogenetic protein 4
ID MGI:88180

Category	Classification Term	Evidence	Inferred From	Ref(s)
Biological Process	cell fate commitment	IDA		1
Biological Process	mesoderm cell fate determination	IMP	MGI:1857137	1
Cellular Component	extracellular space	TAS		1
Molecular Function	cytokine activity	IEA		2
Molecular Function	growth factor activity	IEA		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

Model Organism Database

- Annotated to finest level of knowledge
- Public representation specific to database
- Regular contribution of GO annotations to common resource

Sequence-Computational Sets

- Computational rather than experimental annotation
- Dependent on existing knowledge and accuracy of existing annotations

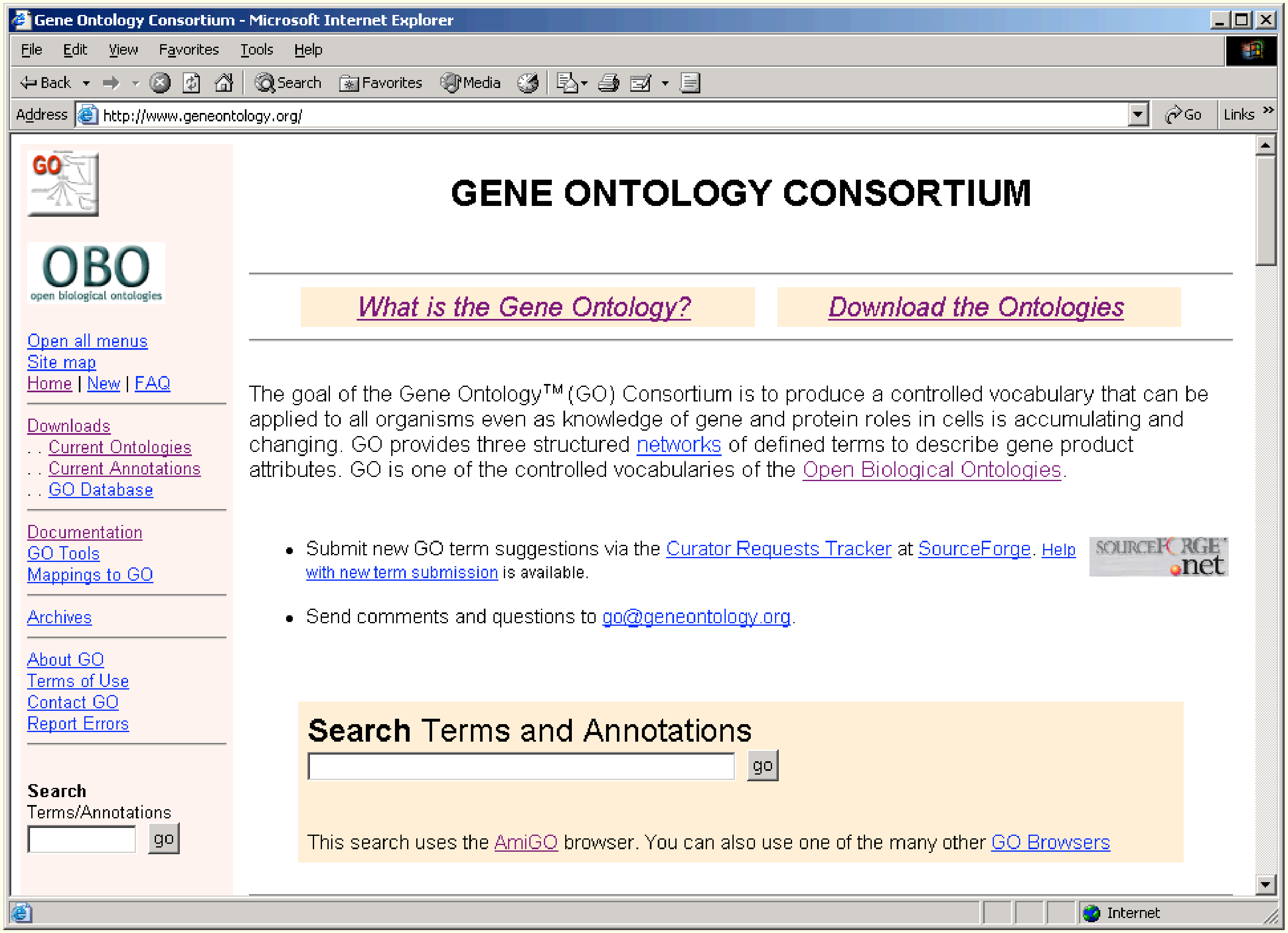
3. Implement and Support Common GO Resource

- Contribution of data files and documentations to GO site
- www.geneontology.org
- Bibiography: 60 annotation papers, 12 statistics pubs, 9 browsers
- Development of curation tools and browsers
- Cross-species search tools
- All Open Source, publicly available

How do I use the GO

- Getting the GO and GO_Association Files
- Data Mining
  - My Favorite Gene
  - By GO
  - By Sequence
- Analysis of Data
  - Clustering
  - Binning
- Other Tools

GO Web Site:www.geneontology.org



Open all menus
Site map
Home | New | FAQ

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

Documentation
GO Tools
Mappings to GO

Archives

About GO
Terms of Use
Contact GO
Report Errors

Search
Terms/Annotations
go

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.

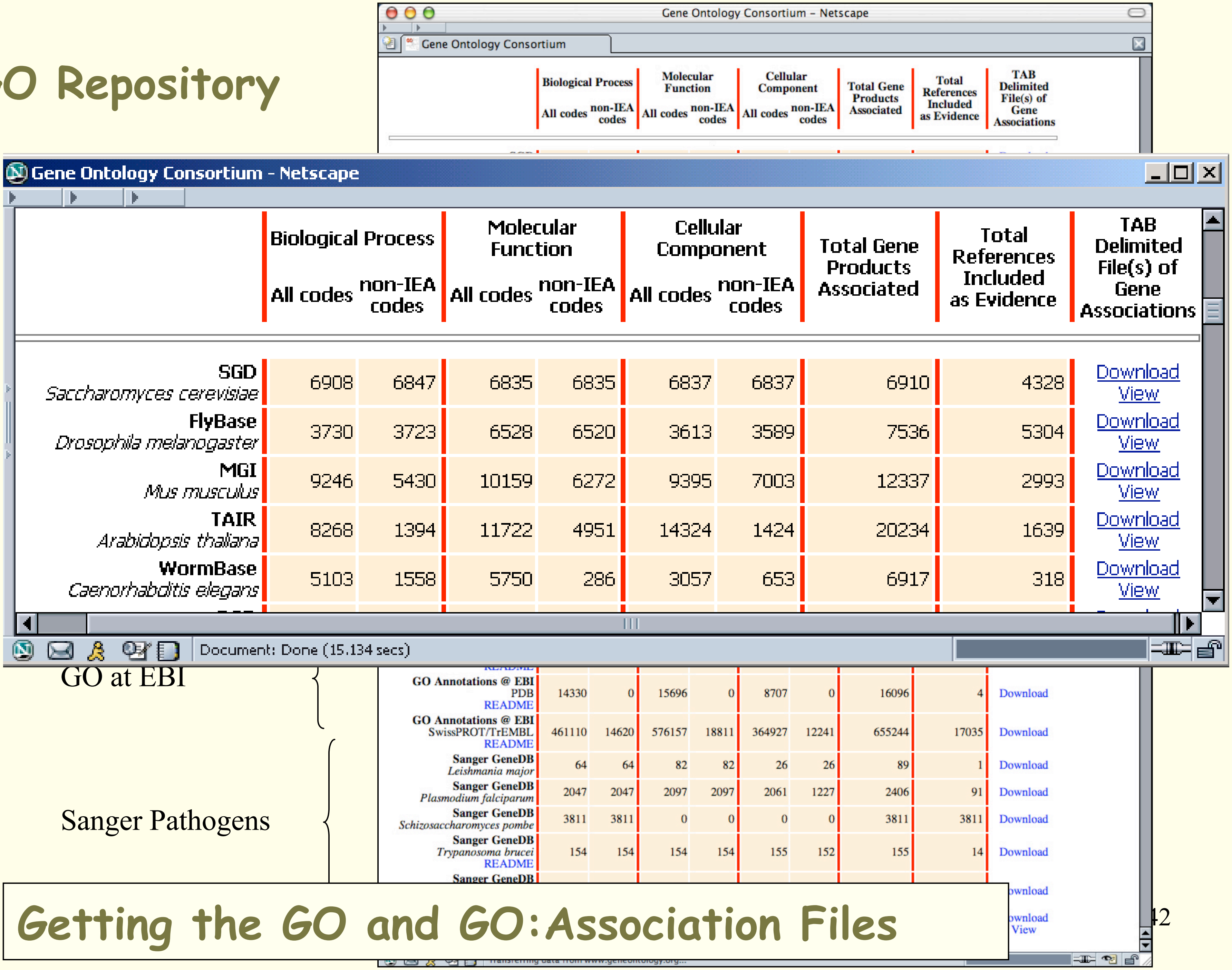
SOURCEFORGE.net

Search Terms and Annotations

go

This search uses the AmiGO browser. You can also use one of the many other GO Browsers

GO Repository



Gene Ontology Consortium - Netscape

	Biological Process		Molecular Function		Cellular Component		Total Gene Products Associated	Total References Included as Evidence	TAB Delimited File(s) of Gene Associations
	All codes	non-IEA codes	All codes	non-IEA codes	All codes	non-IEA codes			
SGD *Saccharomyces cerevisiae*	6908	6847	6835	6835	6837	6837	6910	4328	Download View
FlyBase *Drosophila melanogaster*	3730	3723	6528	6520	3613	3589	7536	5304	Download View
MGI *Mus musculus*	9246	5430	10159	6272	9395	7003	12337	2993	Download View
TAIR *Arabidopsis thaliana*	8268	1394	11722	4951	14324	1424	20234	1639	Download View
WormBase *Caenorhabditis elegans*	5103	1558	5750	286	3057	653	6917	318	Download View

Document: Done (15.134 secs)

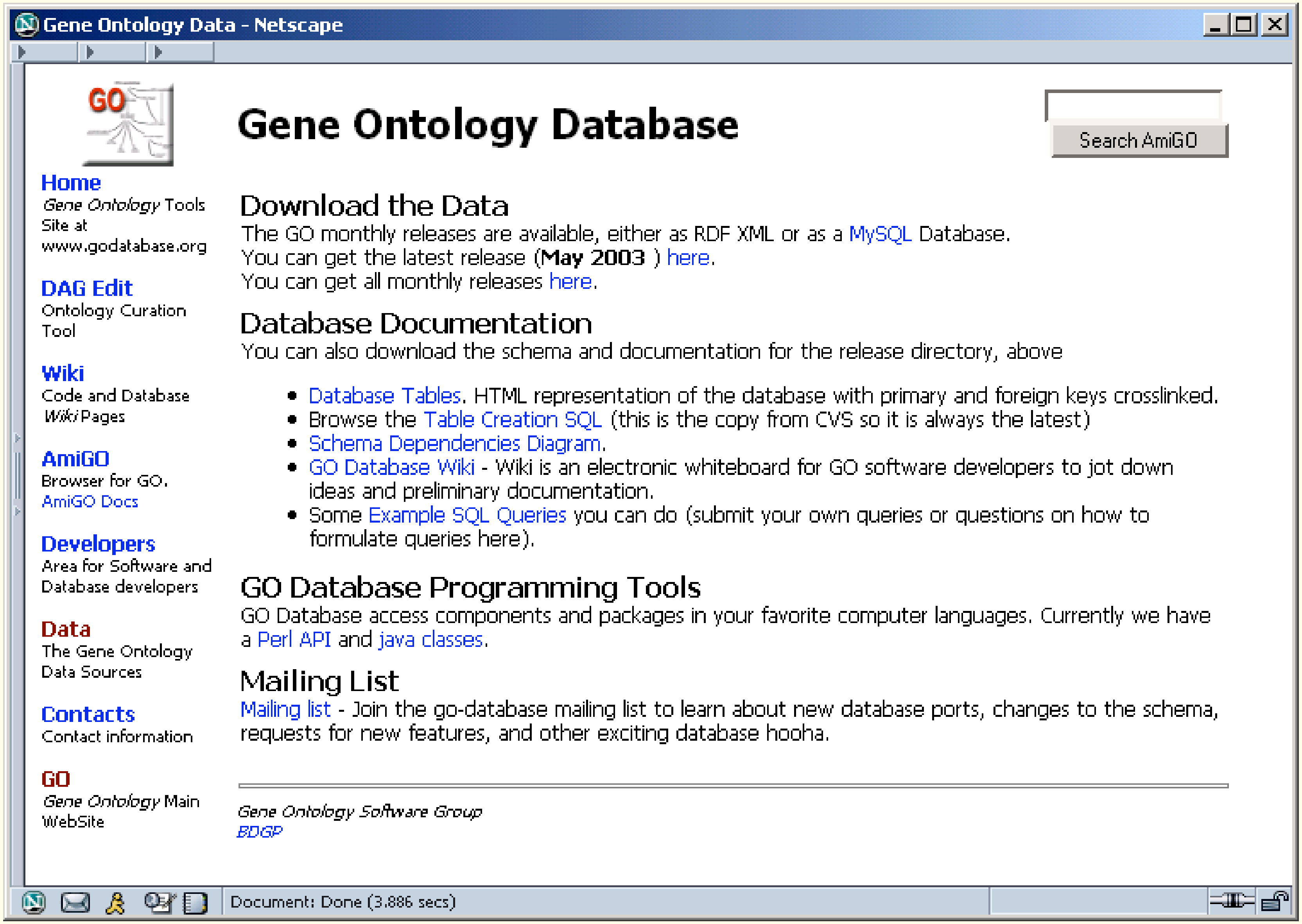
GO at EBI

Sanger Pathogens

	Biological Process		Molecular Function		Cellular Component		Total Gene Products Associated	Total References Included as Evidence	TAB Delimited File(s) of Gene Associations
	All codes	non-IEA codes	All codes	non-IEA codes	All codes	non-IEA codes			
GO Annotations @ EBI PDB README	14330	0	15696	0	8707	0	16096	4	Download
GO Annotations @ EBI SwissPROT/TrEMBL README	461110	14620	576157	18811	364927	12241	655244	17035	Download
Sanger GeneDB *Leishmania major*	64	64	82	82	26	26	89	1	Download
Sanger GeneDB *Plasmodium falciparum*	2047	2047	2097	2097	2061	1227	2406	91	Download
Sanger GeneDB *Schizosaccharomyces pombe*	3811	3811	0	0	0	0	3811	3811	Download
Sanger GeneDB *Trypanosoma brucei* README	154	154	154	154	155	152	155	14	Download
Sanger GeneDB									

Getting the GO and GO:Association Files

The GO Database



Home
Gene Ontology Tools Site at www.godatabase.org

DAG Edit
Ontology Curation Tool

Wiki
Code and Database *Wiki* Pages

AmiGO
Browser for GO.
AmiGO Docs

Developers
Area for Software and Database developers

Data
The Gene Ontology Data Sources

Contacts
Contact information

GO
Gene Ontology Main WebSite

Gene Ontology Database

Download the Data

The GO monthly releases are available, either as RDF XML or as a MySQL Database.
You can get the latest release (**May 2003**) here.
You can get all monthly releases here.

Database Documentation

You can also download the schema and documentation for the release directory, above

- Database Tables. HTML representation of the database with primary and foreign keys crosslinked.
- Browse the Table Creation SQL (this is the copy from CVS so it is always the latest)
- Schema Dependencies Diagram.
- GO Database Wiki - Wiki is an electronic whiteboard for GO software developers to jot down ideas and preliminary documentation.
- Some Example SQL Queries you can do (submit your own queries or questions on how to formulate queries here).

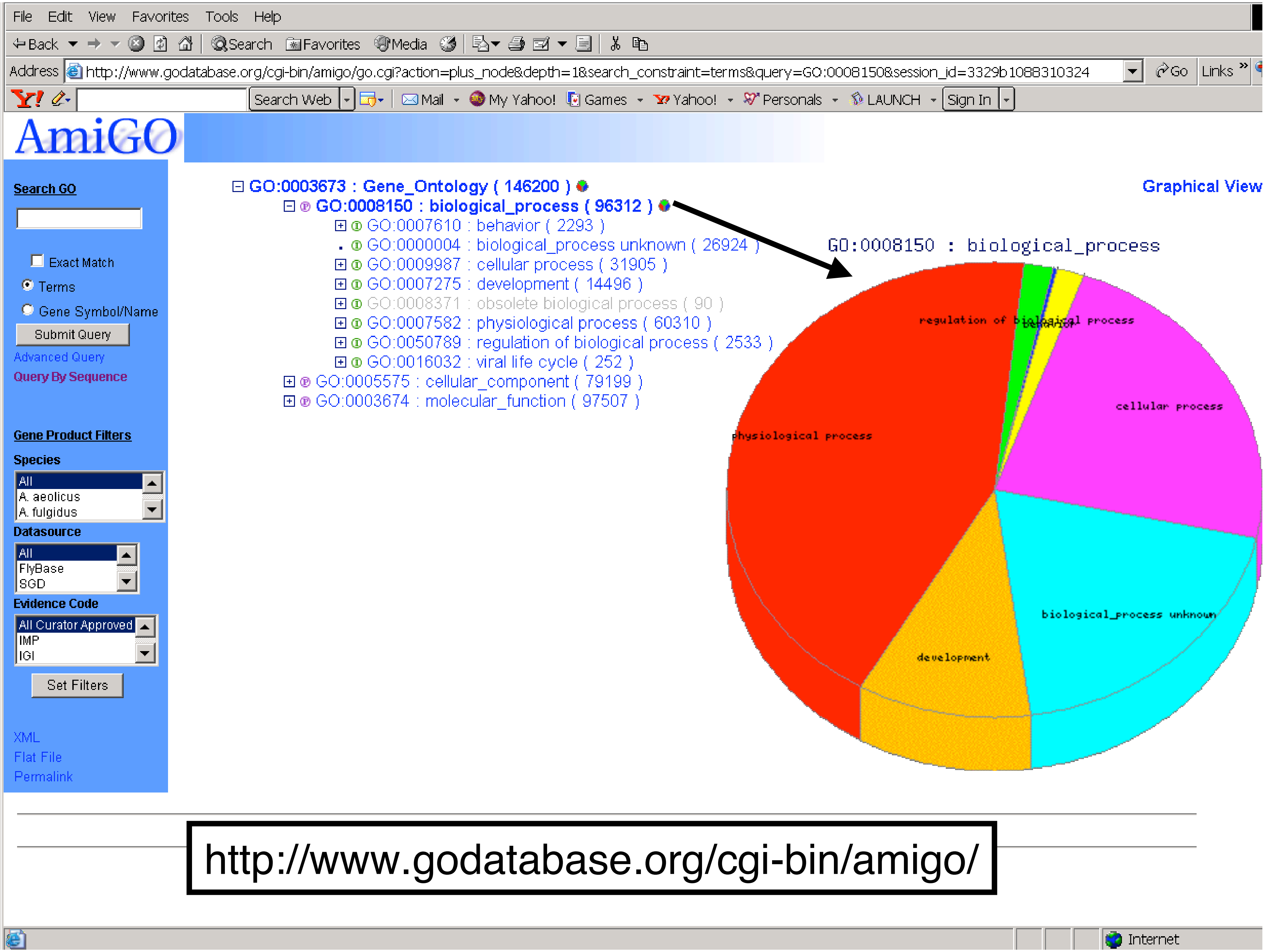
GO Database Programming Tools

GO Database access components and packages in your favorite computer languages. Currently we have a Perl API and java classes.

Mailing List

Mailing list - Join the go-database mailing list to learn about new database ports, changes to the schema, requests for new features, and other exciting database hooha.

Gene Ontology Software Group
BDGP



http://www.godatabase.org/cgi-bin/amigo/go.cgi?action=plus_node&depth=1&search_constraint=terms&query=GO:0008150&session_id=3329b1088310324

Species: All, A. aeolicus, A. fulgidus

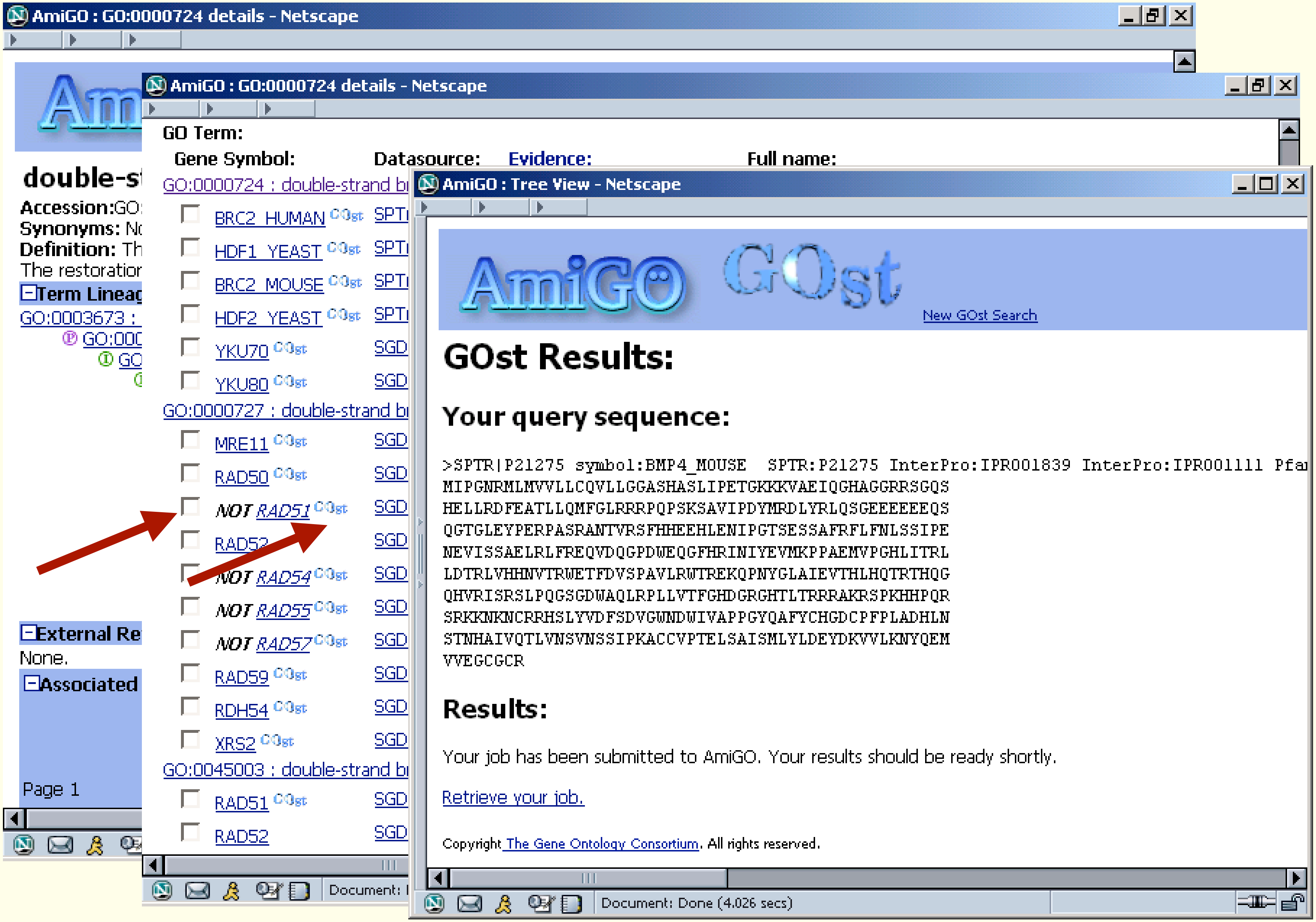
Datasource: All, FlyBase, SGD

Evidence Code: All Curator Approved, IMP, IGI

- GO:0003673 : Gene_Ontology (146200)
  - GO:0008150 : biological_process (96312)
    - GO:0007610 : behavior (2293)
    - GO:0000004 : biological_process unknown (26924)
    - GO:0009987 : cellular process (31905)
    - GO:0007275 : development (14496)
    - GO:0008371 : obsolete biological process (90)
    - GO:0007582 : physiological process (60310)
    - GO:0050789 : regulation of biological process (2533)
    - GO:0016032 : viral life cycle (252)
  - GO:0005575 : cellular_component (79199)
  - GO:0003674 : molecular_function (97507)

http://www.godatabase.org/cgi-bin/amigo/

Querying the GO



double-s

Accession:GO:

Synonyms: N

Definition: Th

The restoration

⊟**Term Lineag**

GO:0003673 :

Ⓟ GO:000

Ⓘ GO

⊟**External Re**

None.

⊟**Associated**

Page 1

GO Term:

Gene Symbol:	**Datasource:**	**Evidence:**	**Full name:**
GO:0000724 : double-strand b			
☐ BRC2_HUMAN GOst	SPTI		
☐ HDF1_YEAST GOst	SPTI		
☐ BRC2_MOUSE GOst	SPTI		
☐ HDF2_YEAST GOst	SPTI		
☐ YKU70 GOst	SGD		
☐ YKU80 GOst	SGD		
GO:0000727 : double-strand b			
☐ MRE11 GOst	SGD		
☐ RAD50 GOst	SGD		
☐ *NOT RAD51* GOst	SGD		
☐ RAD52	SGD		
☐ *NOT RAD54* GOst	SGD		
☐ *NOT RAD55* GOst	SGD		
☐ *NOT RAD57* GOst	SGD		
☐ RAD59 GOst	SGD		
☐ RDH54 GOst	SGD		
☐ XRS2 GOst	SGD		
GO:0045003 : double-strand b			
☐ RAD51 GOst	SGD		
☐ RAD52	SGD		

Document: I

New GOst Search

GOst Results:

Your query sequence:

```
>SPTR|P21275 symbol:BMP4_MOUSE  SPTR:P21275 InterPro:IPR001839 InterPro:IPR001111 Pfa
MIPGNRMLMVVLLCQVLLGGASHASLIPETGKKKVAEIQGHAGGRRSGQS
HELLRDFEATLLQMFGLRRRPQPSKSAVIPDYMRDLYRLQSGEEEEEEQS
QGTGLEYPERPASRANTVRSFHHEEHLENIPGTSESSAFRFLFNLSSIPE
NEVISSAELRLFREQVDQGPDWEQGFHRINIYEVMKPPAEMVPGHLITRL
LDTRLVHHNVTRWETFDVSPAVLRWTREKQPNYGLAIEVTHLHQTRTHQG
QHVRISRSLPQGSGDWAQLRPLLVTFGHDGRGHTLTRRRAKRSPKHHPQR
SRKKNKNCRRHSLYVDFSDVGWNDWIVAPPGYQAFYCHGDCPFPLADHLN
STNHAIVQTLVNSVNSSIPKACCVPTELSAISMLYLDEYDKVVLKNYQEM
VVEGCGCR
```

Results:

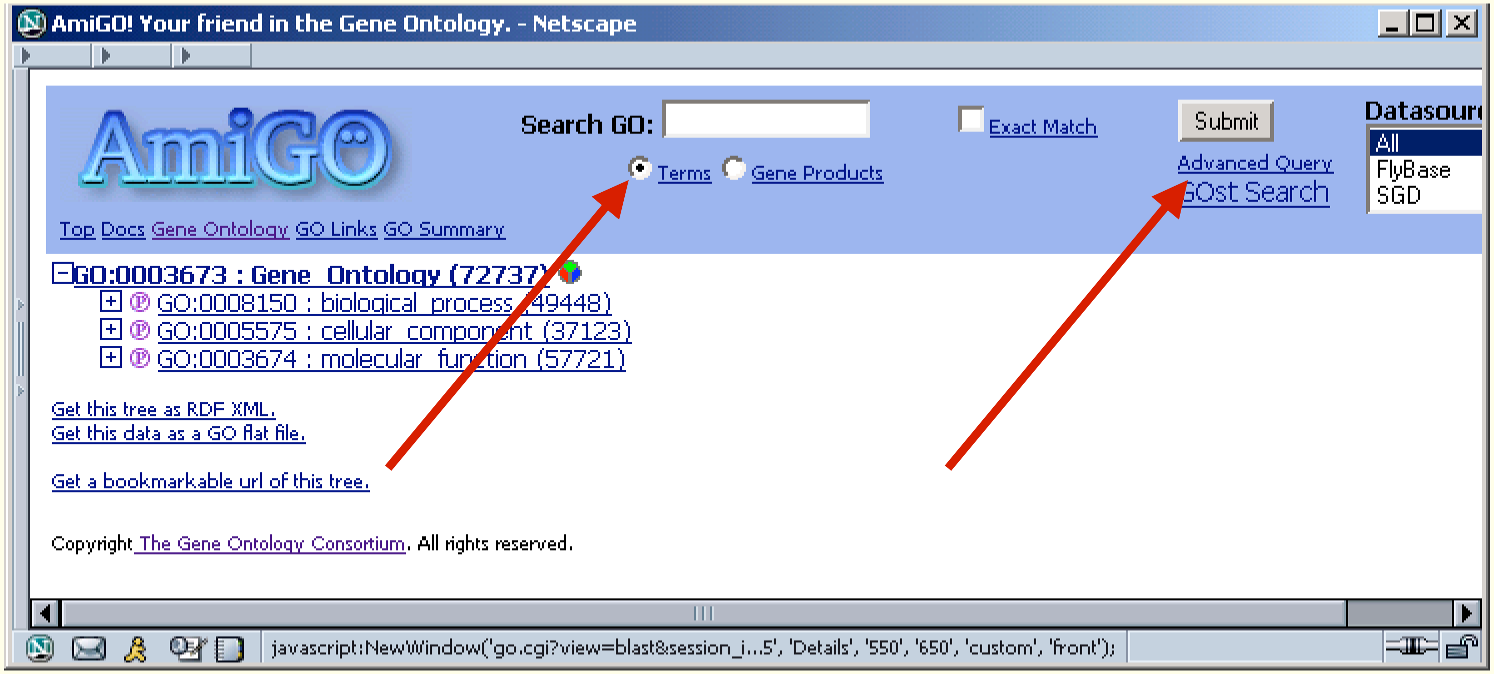
Your job has been submitted to AmiGO. Your results should be ready shortly.

Retrieve your job.

Copyright The Gene Ontology Consortium. All rights reserved.

Document: Done (4.026 secs)

Query GO – by concept



AmiGO! Your friend in the Gene Ontology. - Netscape

Search GO: ☐ Exact Match

● Terms ○ Gene Products

Submit

Advanced Query

GOst Search

Datasour

All
FlyBase
SGD

Top Docs Gene Ontology GO Links GO Summary

- GO:0003673 : Gene Ontology (72737)
  - GO:0008150 : biological process (49448)
  - GO:0005575 : cellular component (37123)
  - GO:0003674 : molecular function (57721)

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

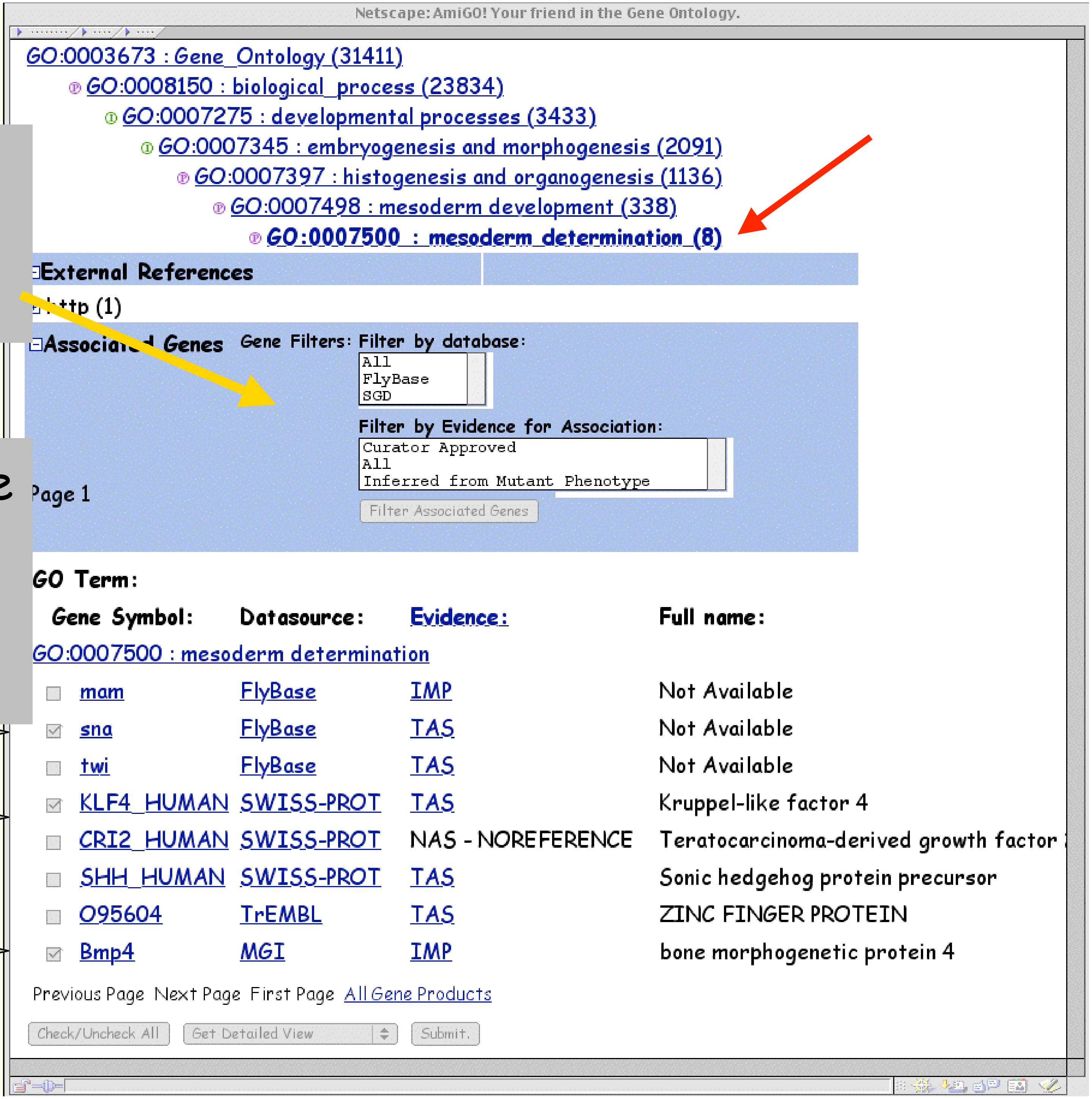
Get a bookmarkable url of this tree.

Copyright The Gene Ontology Consortium. All rights reserved.

javascript:NewWindow('go.cgi?view=blast&session_i...5', 'Details', '550', '650', 'custom', 'front');

Filter queries by organism or evidence

Select sequence based on functional annotation



GO:0003673 : Gene_Ontology (31411)
- GO:0008150 : biological_process (23834)
  - GO:0007275 : developmental processes (3433)
    - GO:0007345 : embryogenesis and morphogenesis (2091)
      - GO:0007397 : histogenesis and organogenesis (1136)
        - GO:0007498 : mesoderm development (338)
          - **GO:0007500 : mesoderm determination (8)**

External References

http (1)

Associated Genes **Gene Filters:** **Filter by database:**

All
FlyBase
SGD

Filter by Evidence for Association:

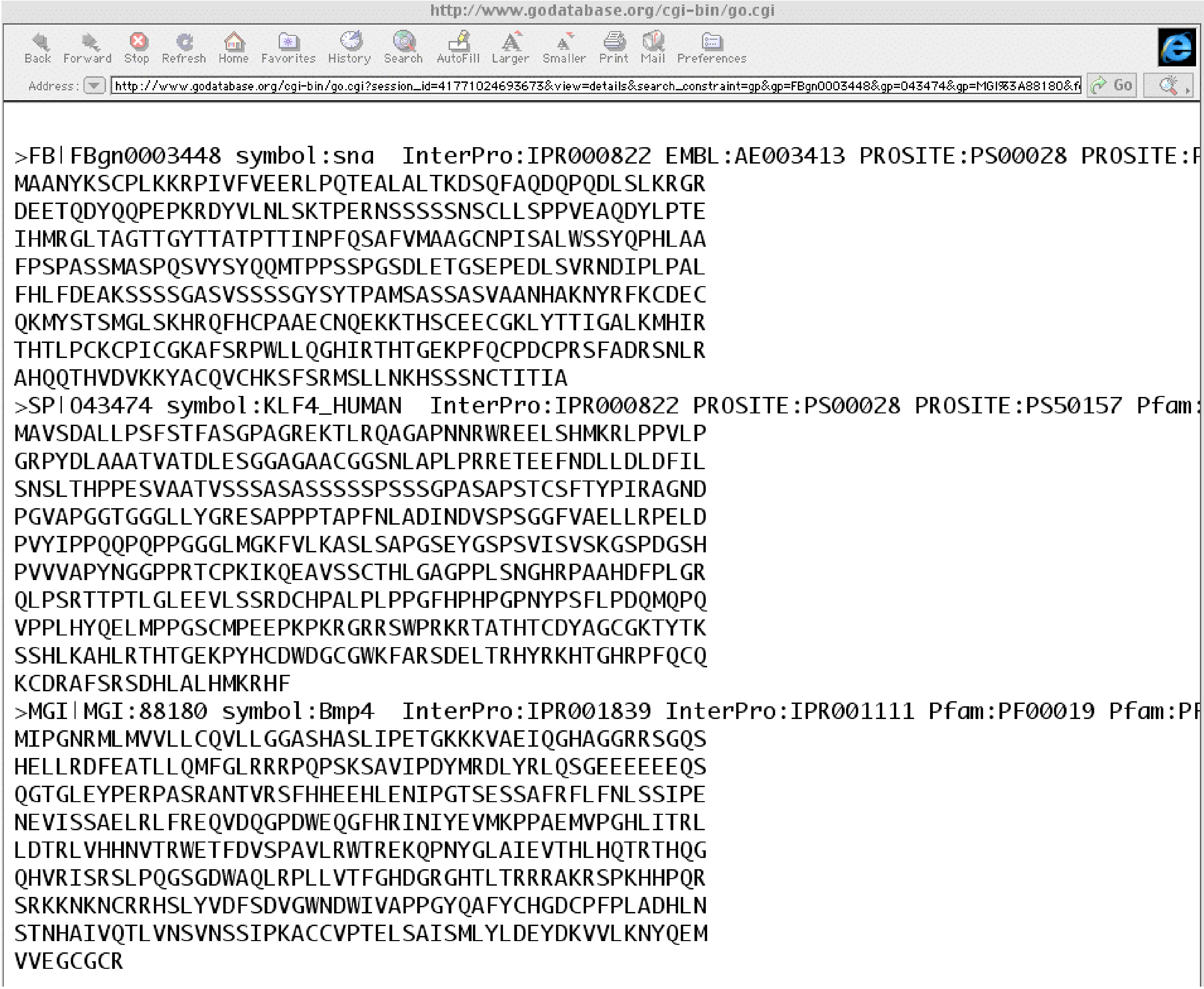
Curator Approved
All
Inferred from Mutant Phenotype

Page 1

GO Term:

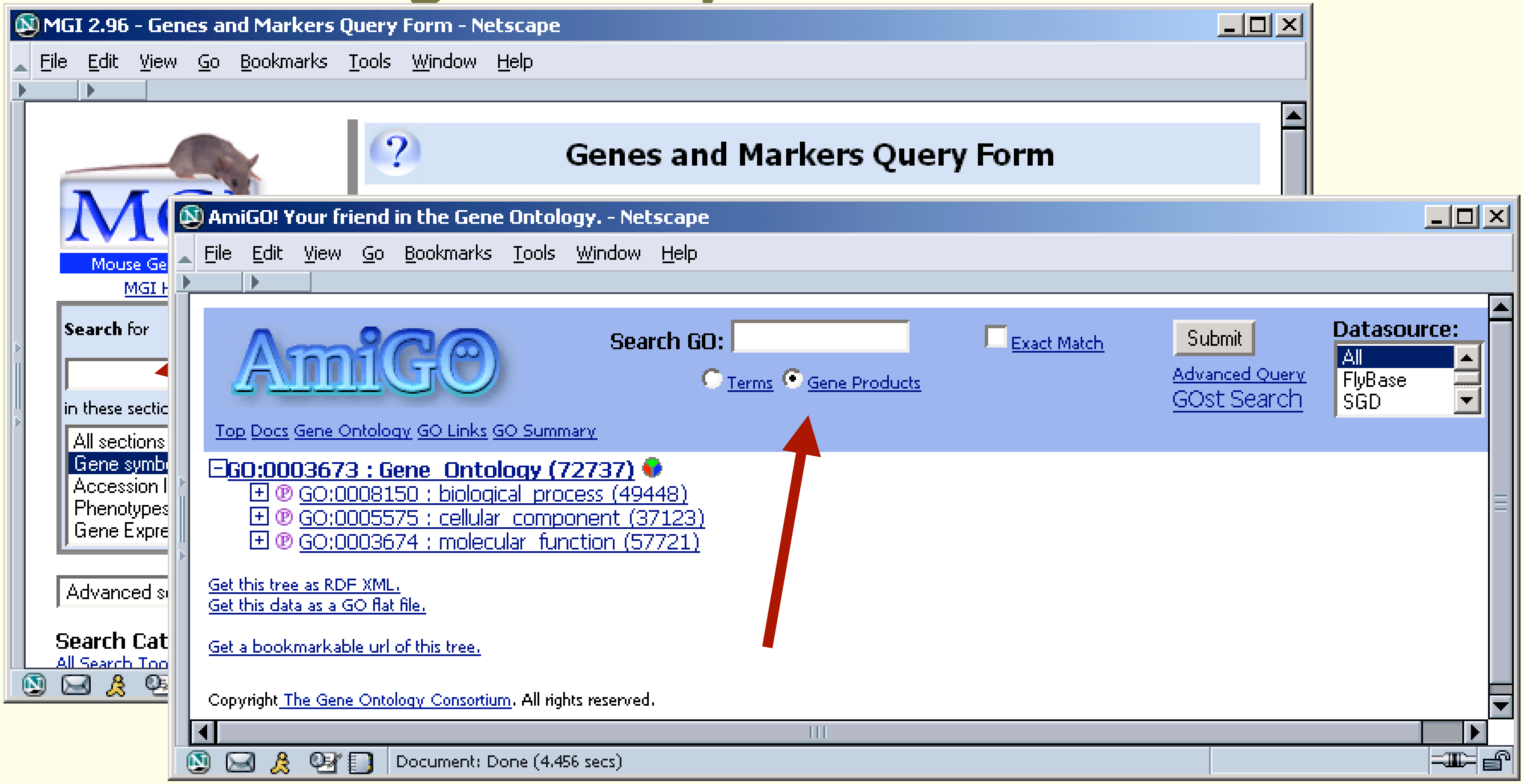
	Gene Symbol:	**Datasource:**	**Evidence:**	**Full name:**
GO:0007500 : mesoderm determination				
☐	mam	FlyBase	IMP	Not Available
☑	sna	FlyBase	TAS	Not Available
☐	twi	FlyBase	TAS	Not Available
☑	KLF4_HUMAN	SWISS-PROT	TAS	Kruppel-like factor 4
☐	CRI2_HUMAN	SWISS-PROT	NAS - NOREFERENCE	Teratocarcinoma-derived growth factor
☐	SHH_HUMAN	SWISS-PROT	TAS	Sonic hedgehog protein precursor
☐	O95604	TrEMBL	TAS	ZINC FINGER PROTEIN
☑	Bmp4	MGI	IMP	bone morphogenetic protein 4

Previous Page Next Page First Page All Gene Products



```
>FB|FBgn0003448 symbol:sna  InterPro:IPR000822 EMBL:AE003413 PROSITE:PS00028 PROSITE:
MAANYKSCPLKKRPIVFVEERLPQTEALALTKDSQFAQDQPQDLSLKRGR
DEETQDYQQPEPKRDYVLNLSKTPERNSSSSSNSCLLSPPVEAQDYLPTE
IHMRGLTAGTTGYTTATPTTINPFQSAFVMAAGCNPISALWSSYQPHLAA
FPSPASSMASPQSVYSYQQMTPPSSPGSDLETGSEPEDLSVRNDIPLPAL
FHLFDEAKSSSSGASVSSSSGYSYTPAMSASSASVAANHAKNYRFKCDEC
QKMYSTSMGLSKHRQFHCPAAECNQEKKTHSCEECGKLYTTIGALKMHIR
THTLPCKCPICGKAFSRPWLLQGHIRTHTGEKPFQCPDCPRSFADRSNLR
AHQQTHVDVKKYACQVCHKSFSRMSLLNKHSSSNCTITIA
>SP|043474 symbol:KLF4_HUMAN  InterPro:IPR000822 PROSITE:PS00028 PROSITE:PS50157 Pfam:
MAVSDALLPSFSTFASGPAGREKTLRQAGAPNNRWREELSHMKRLPPVLP
GRPYDLAAATVATDLESGGAGAACGGSNLAPLPRRETEEFNDLLDLDFIL
SNSLTHPPESVAATVSSSASASSSSSPSSSGPASAPSTCSFTYPIRAGND
PGVAPGGTGGGLLYGRESAPPPTAPFNLADINDVSPSGGFVAELLRPELD
PVYIPPQQPQPPGGGLMGKFVLKASLSAPGSEYGSPSVISVSKGSPDGSH
PVVVAPYNGGPPRTCPKIKQEAVSSCTHLGAGPPLSNGHRPAAHDFPLGR
QLPSRTTPTLGLEEVLSSRDCHPALPLPPGFHPHPGPNYPSFLPDQMQPQ
VPPLHYQELMPPGSCMPEEPKPKRGRRSWPRKRTATHTCDYAGCGKTYTK
SSHLKAHLRTHTGEKPYHCDWDGCGWKFARSDELTRHYRKHTGHRPFQCQ
KCDRAFSRSDHLALHMKRHF
>MGI|MGI:88180 symbol:Bmp4  InterPro:IPR001839 InterPro:IPR001111 Pfam:PF00019 Pfam:P
MIPGNRMLMVVLLCQVLLGGASHASLIPETGKKKVAEIQGHAGGRRSGQS
HELLRDFEATLLQMFGLRRRPQPSKSAVIPDYMRDLYRLQSGEEEEEEQS
QGTGLEYPERPASRANTVRSFHHEEHLENIPGTSESSAFRFLFNLSSIPE
NEVISSAELRLFREQVDQGPDWEQGFHRINIYEVMKPPAEMVPGHLITRL
LDTRLVHHNVTRWETFDVSPAVLRWTREKQPNYGLAIEVTHLHQTRTHQG
QHVRISRSLPQGSGDWAQLRPLLVTFGHDGRGHTLTRRRAKRSPKHHPQR
SRKKNKNCRRHSLYVDFSDVGWNDWIVAPPGYQAFYCHGDCPFPLADHLN
STNHAIVQTLVNSVNSSIPKACCVPTELSAISMLYLDEYDKVVLKNYQEM
VVEGCGCR
```

Using GO…my favorite Gene



MGI 2.96 - Genes and Markers Query Form - Netscape

Genes and Markers Query Form

Search for

in these sectic

- All sections
- Gene symb
- Accession I
- Phenotypes
- Gene Expre

Advanced s

Search Cat

All Search Too

AmiGO! Your friend in the Gene Ontology. - Netscape

Search GO:

○ Terms ● Gene Products

☐ Exact Match

Datasource:

- All
- FlyBase
- SGD

Top Docs Gene Ontology GO Links GO Summary

- GO:0003673 : Gene_Ontology (72737)
  - GO:0008150 : biological_process (49448)
  - GO:0005575 : cellular_component (37123)
  - GO:0003674 : molecular_function (57721)

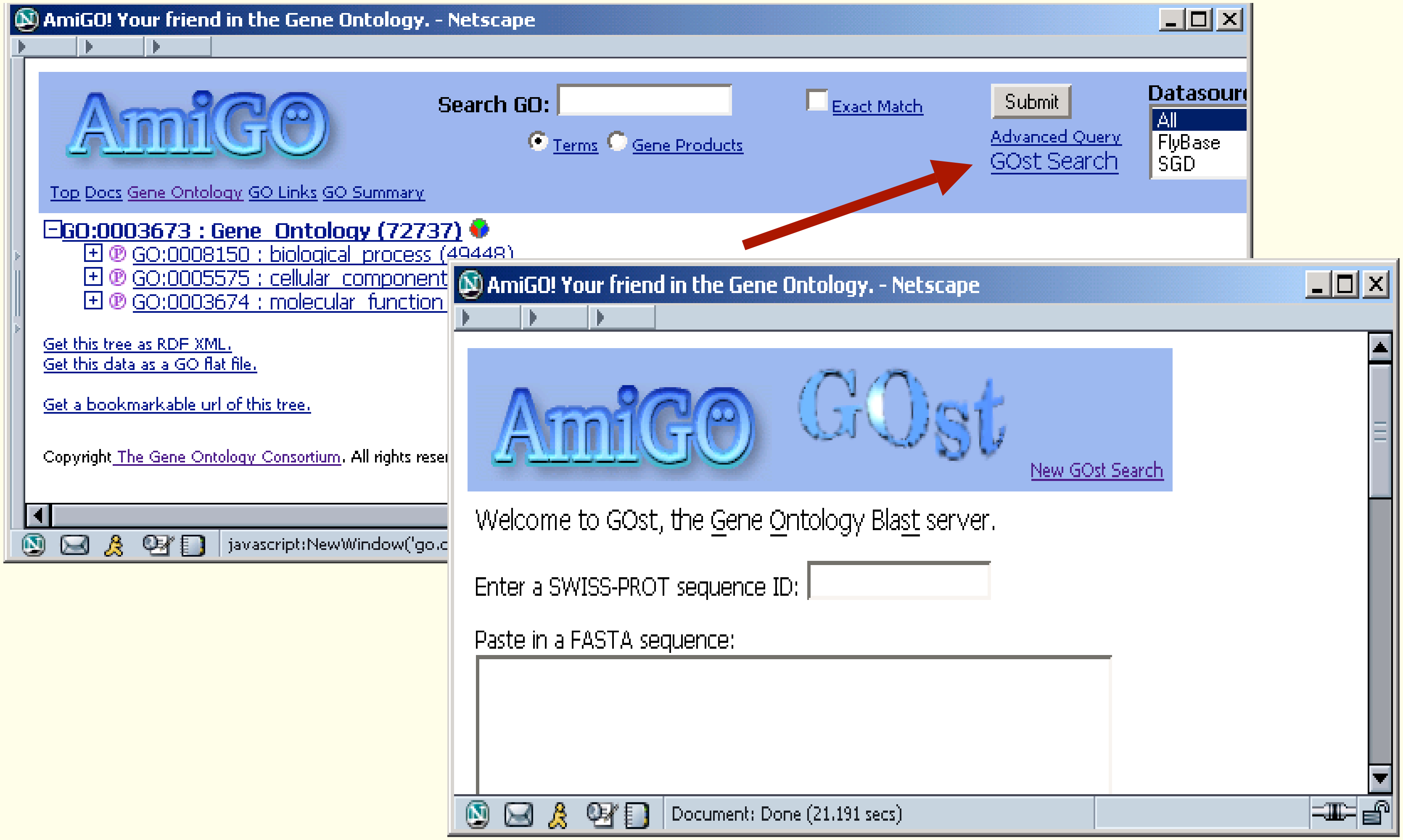
Get this tree as RDF XML.
Get this data as a GO flat file.

Get a bookmarkable url of this tree.

Copyright The Gene Ontology Consortium. All rights reserved.

Document: Done (4.456 secs)

Data Mining by Sequence



AmiGO! Your friend in the Gene Ontology. - Netscape

Search GO:

☐ Exact Match

● Terms ○ Gene Products

Advanced Query

GOst Search

Datasource: All, FlyBase, SGD

Top Docs Gene Ontology GO Links GO Summary

GO:0003673 : Gene Ontology (72737)

- GO:0008150 : biological process (49448)
- GO:0005575 : cellular component
- GO:0003674 : molecular function

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

Get a bookmarkable url of this tree.

Copyright The Gene Ontology Consortium. All rights reser

javascript:NewWindow('go.c

AmiGO! Your friend in the Gene Ontology. - Netscape

New GOst Search

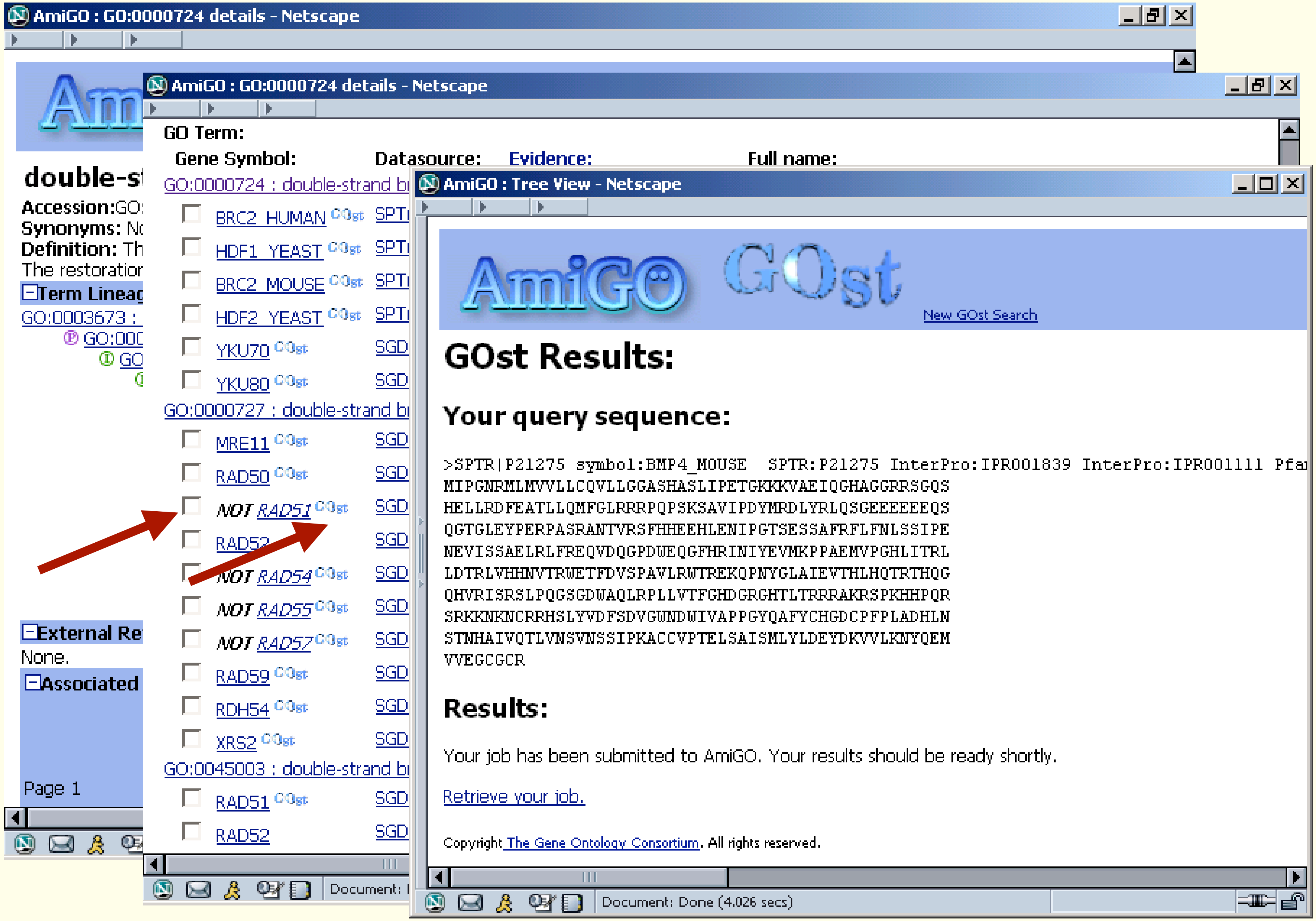
Welcome to GOst, the Gene Ontology Blast server.

Enter a SWISS-PROT sequence ID:

Paste in a FASTA sequence:

Document: Done (21.191 secs)

Data Mining by Concept-Sequence



GOst Results:

Your query sequence:

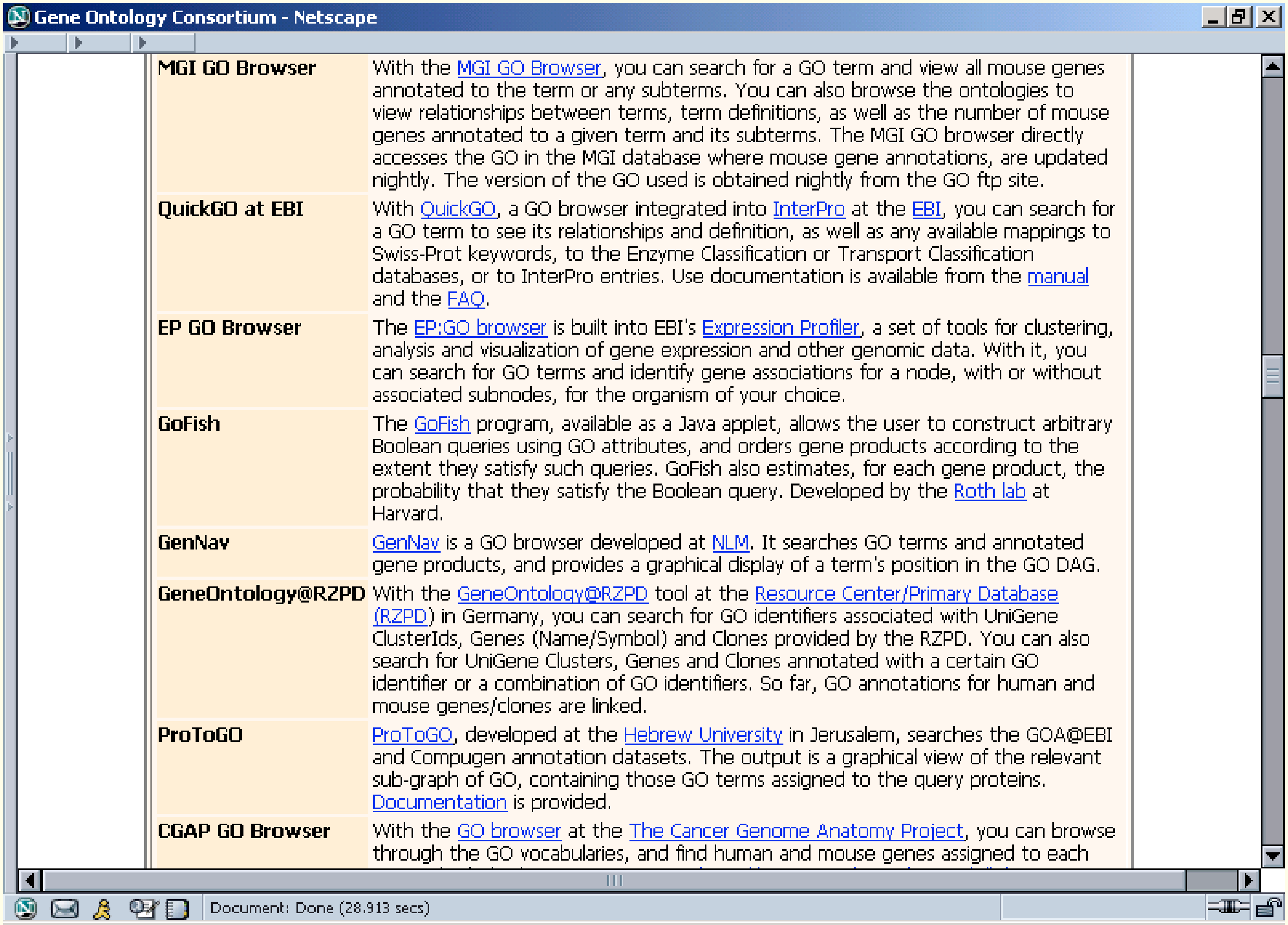
```
>SPTR|P21275 symbol:BMP4_MOUSE  SPTR:P21275 InterPro:IPR001839 InterPro:IPR001111 Pfa
MIPGNRMLMVVLLCQVLLGGASHASLIPETGKKKVAEIQGHAGGRRSGQS
HELLRDFEATLLQMFGLRRRPQPSKSAVIPDYMRDLYRLQSGEEEEEEQS
QGTGLEYPERPASRANTVRSFHHEEHLENIPGTSESSAFRFLFNLSSIPE
NEVISSAELRLFREQVDQGPDWEQGFHRINIYEVMKPPAEMVPGHLITRL
LDTRLVHHNVTRWETFDVSPAVLRWTREKQPNYGLAIEVTHLHQTRTHQG
QHVRISRSLPQGSGDWAQLRPLLVTFGHDGRGHTLTRRRAKRSPKHHPQR
SRKKNKNCRRHSLYVDFSDVGWNDWIVAPPGYQAFYCHGDCPFPLADHLN
STNHAIVQTLVNSVNSSIPKACCVPTELSAISMLYLDEYDKVVLKNYQEM
VVEGCGCR
```

Results:

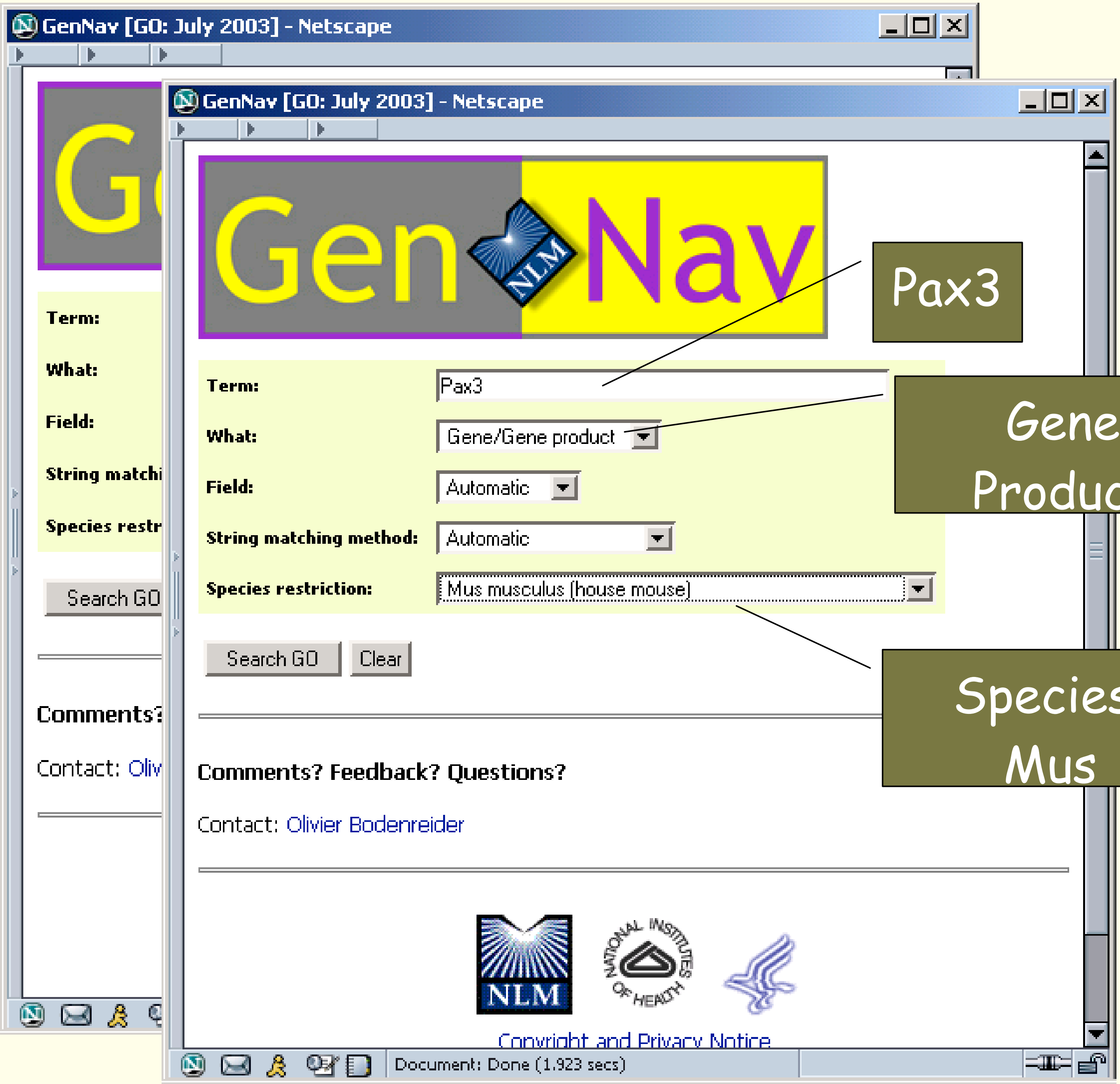
Your job has been submitted to AmiGO. Your results should be ready shortly.

Retrieve your job.

Other GO Browsers



MGI GO Browser	With the MGI GO Browser, you can search for a GO term and view all mouse genes annotated to the term or any subterms. You can also browse the ontologies to view relationships between terms, term definitions, as well as the number of mouse genes annotated to a given term and its subterms. The MGI GO browser directly accesses the GO in the MGI database where mouse gene annotations, are updated nightly. The version of the GO used is obtained nightly from the GO ftp site.
QuickGO at EBI	With QuickGO, a GO browser integrated into InterPro at the EBI, you can search for a GO term to see its relationships and definition, as well as any available mappings to Swiss-Prot keywords, to the Enzyme Classification or Transport Classification databases, or to InterPro entries. Use documentation is available from the manual and the FAQ.
EP GO Browser	The EP:GO browser is built into EBI's Expression Profiler, a set of tools for clustering, analysis and visualization of gene expression and other genomic data. With it, you can search for GO terms and identify gene associations for a node, with or without associated subnodes, for the organism of your choice.
GoFish	The GoFish program, available as a Java applet, allows the user to construct arbitrary Boolean queries using GO attributes, and orders gene products according to the extent they satisfy such queries. GoFish also estimates, for each gene product, the probability that they satisfy the Boolean query. Developed by the Roth lab at Harvard.
GenNav	GenNav is a GO browser developed at NLM. It searches GO terms and annotated gene products, and provides a graphical display of a term's position in the GO DAG.
GeneOntology@RZPD	With the GeneOntology@RZPD tool at the Resource Center/Primary Database (RZPD) in Germany, you can search for GO identifiers associated with UniGene ClusterIds, Genes (Name/Symbol) and Clones provided by the RZPD. You can also search for UniGene Clusters, Genes and Clones annotated with a certain GO identifier or a combination of GO identifiers. So far, GO annotations for human and mouse genes/clones are linked.
ProToGO	ProToGO, developed at the Hebrew University in Jerusalem, searches the GOA@EBI and Compugen annotation datasets. The output is a graphical view of the relevant sub-graph of GO, containing those GO terms assigned to the query proteins. Documentation is provided.
CGAP GO Browser	With the GO browser at the The Cancer Genome Anatomy Project, you can browse through the GO vocabularies, and find human and mouse genes assigned to each



Term: Pax3

What: Gene/Gene product

Field: Automatic

String matching method: Automatic

Species restriction: Mus musculus (house mouse)

Comments? Feedback? Questions?

Contact: Olivier Bodenreider

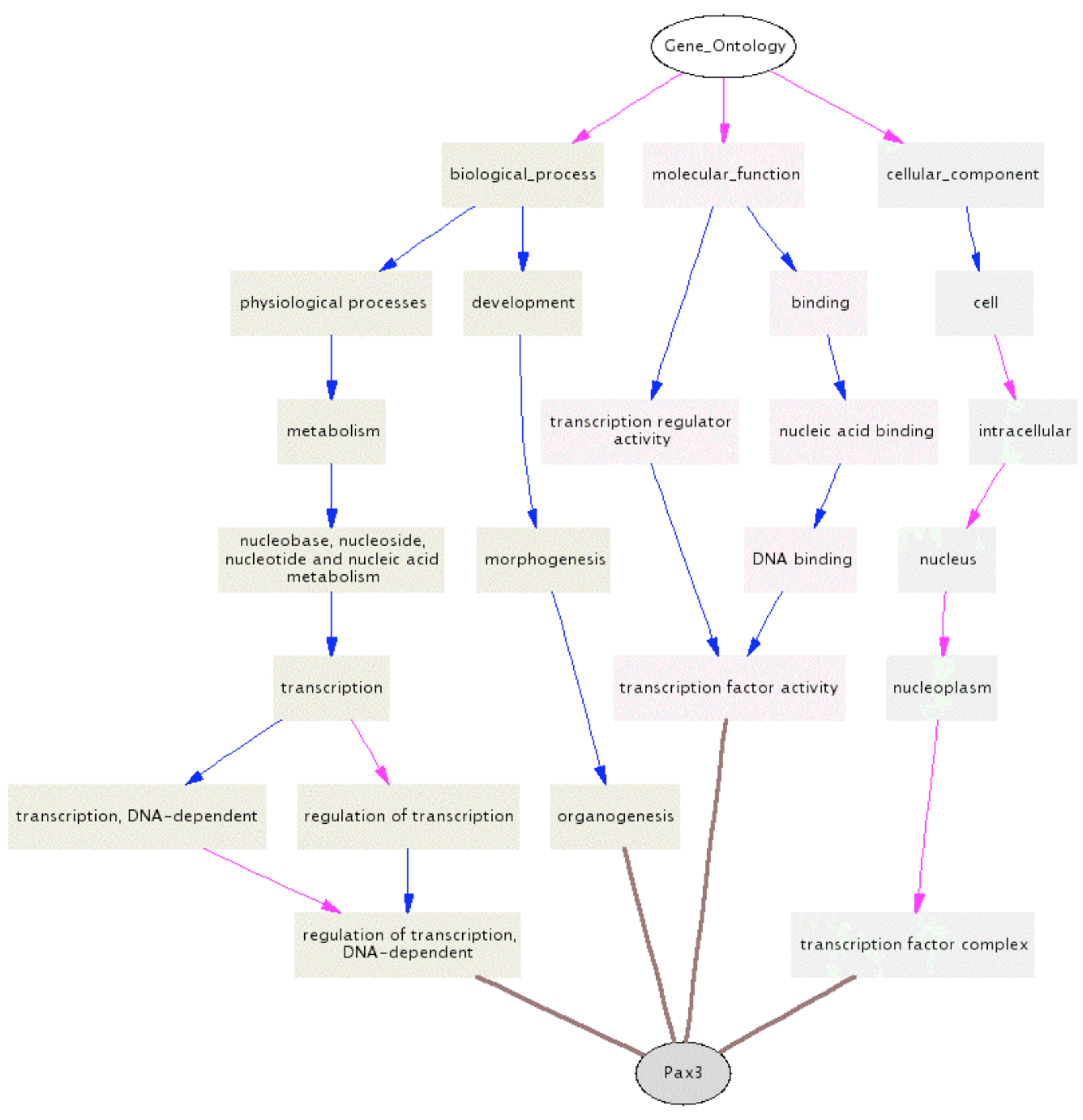
Copyright and Privacy Notice

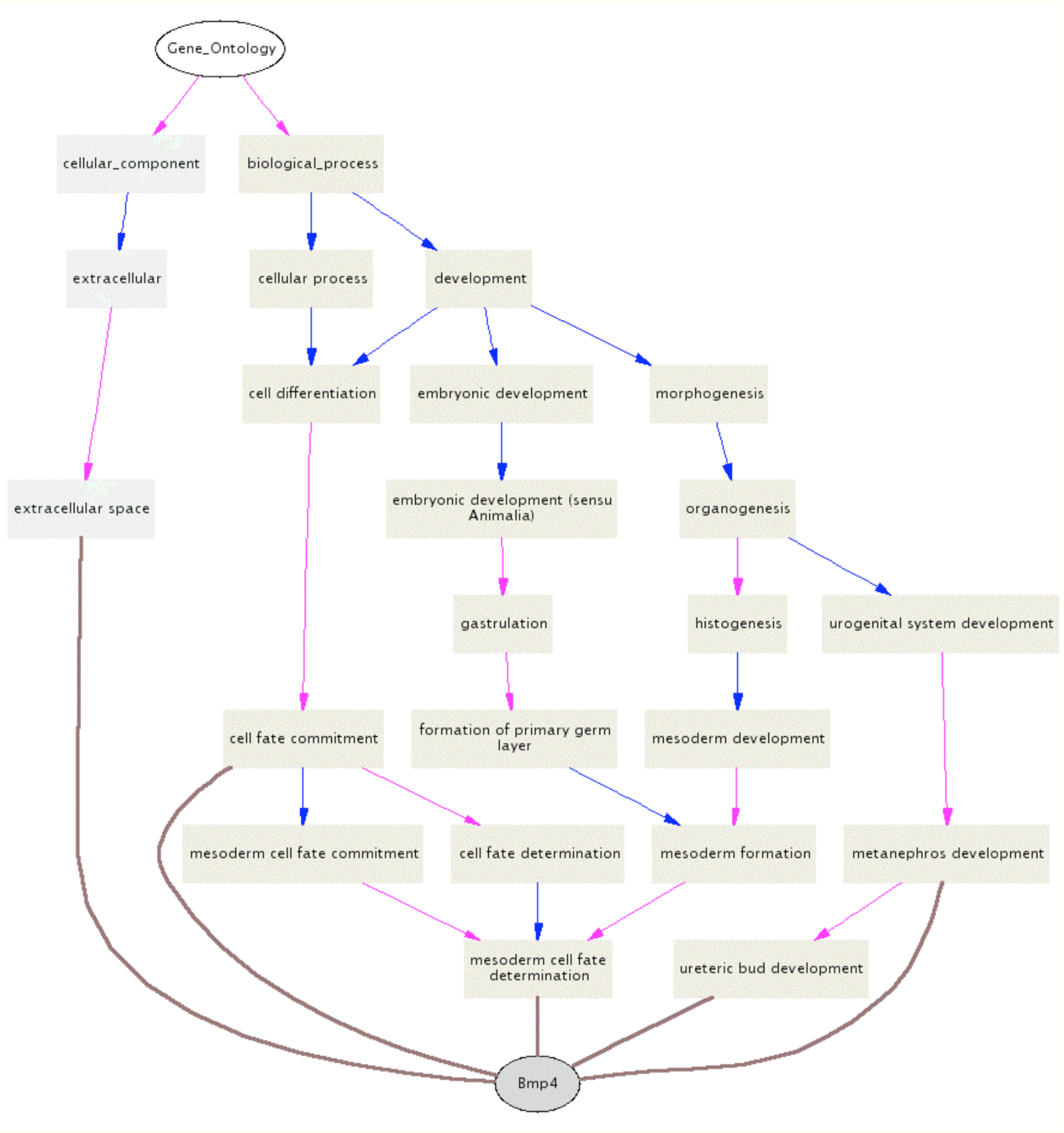
Document: Done (1.923 secs)

Pax3

Gene Product

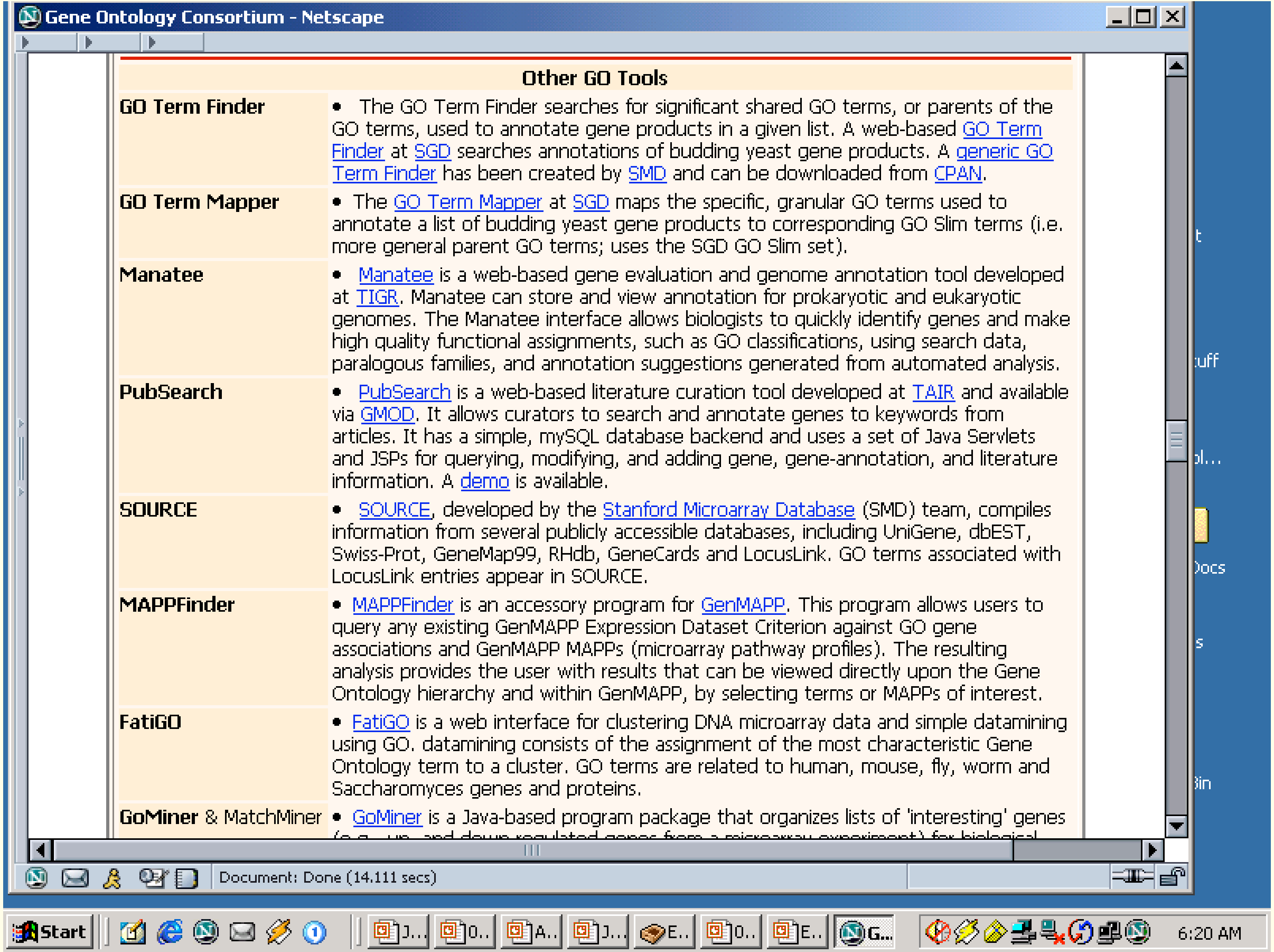
Species: Mus


Gene_Ontology
biological_process
molecular_function
cellular_component
physiological processes
development
binding
cell
metabolism
transcription regulator activity
nucleic acid binding
intracellular
nucleobase, nucleoside, nucleotide and nucleic acid metabolism
morphogenesis
DNA binding
nucleus
transcription
transcription factor activity
nucleoplasm
transcription, DNA-dependent
regulation of transcription
organogenesis
regulation of transcription, DNA-dependent
transcription factor complex
Pax3


Gene_Ontology
cellular_component
biological_process
extracellular
cellular process
development
cell differentiation
embryonic development
morphogenesis
extracellular space
embryonic development (sensu Animalia)
organogenesis
gastrulation
histogenesis
urogenital system development
cell fate commitment
formation of primary germ layer
mesoderm development
mesoderm cell fate commitment
cell fate determination
mesoderm formation
metanephros development
mesoderm cell fate determination
ureteric bud development
Bmp4

GO Tools

14 applications contributed

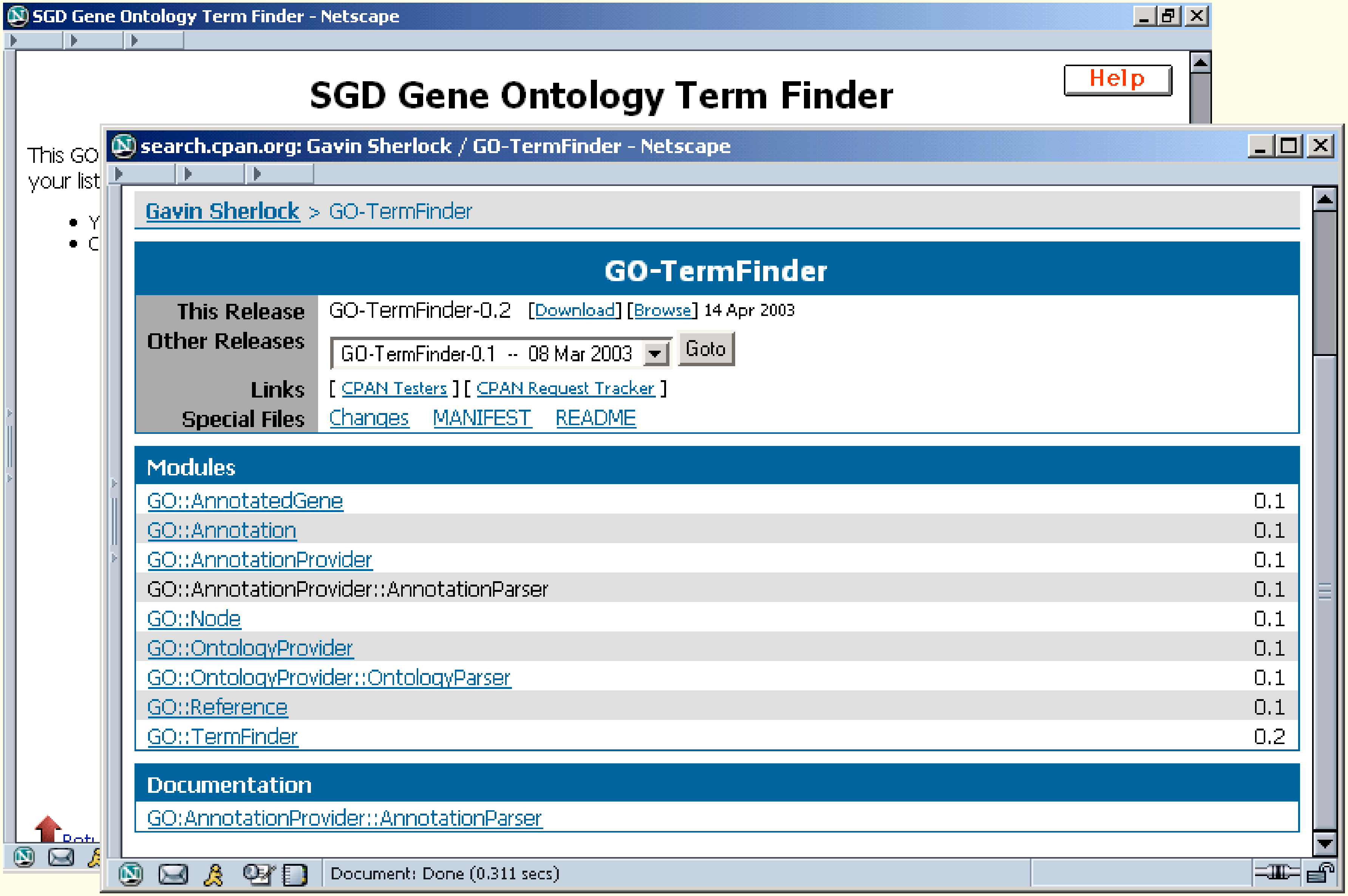


Other GO Tools	
GO Term Finder	- The GO Term Finder searches for significant shared GO terms, or parents of the GO terms, used to annotate gene products in a given list. A web-based GO Term Finder at SGD searches annotations of budding yeast gene products. A generic GO Term Finder has been created by SMD and can be downloaded from CPAN.
GO Term Mapper	- The GO Term Mapper at SGD maps the specific, granular GO terms used to annotate a list of budding yeast gene products to corresponding GO Slim terms (i.e. more general parent GO terms; uses the SGD GO Slim set).
Manatee	- Manatee is a web-based gene evaluation and genome annotation tool developed at TIGR. Manatee can store and view annotation for prokaryotic and eukaryotic genomes. The Manatee interface allows biologists to quickly identify genes and make high quality functional assignments, such as GO classifications, using search data, paralogous families, and annotation suggestions generated from automated analysis.
PubSearch	- PubSearch is a web-based literature curation tool developed at TAIR and available via GMOD. It allows curators to search and annotate genes to keywords from articles. It has a simple, mySQL database backend and uses a set of Java Servlets and JSPs for querying, modifying, and adding gene, gene-annotation, and literature information. A demo is available.
SOURCE	- SOURCE, developed by the Stanford Microarray Database (SMD) team, compiles information from several publicly accessible databases, including UniGene, dbEST, Swiss-Prot, GeneMap99, RHdb, GeneCards and LocusLink. GO terms associated with LocusLink entries appear in SOURCE.
MAPPFinder	- MAPPFinder is an accessory program for GenMAPP. This program allows users to query any existing GenMAPP Expression Dataset Criterion against GO gene associations and GenMAPP MAPPs (microarray pathway profiles). The resulting analysis provides the user with results that can be viewed directly upon the Gene Ontology hierarchy and within GenMAPP, by selecting terms or MAPPs of interest.
FatiGO	- FatiGO is a web interface for clustering DNA microarray data and simple datamining using GO. datamining consists of the assignment of the most characteristic Gene Ontology term to a cluster. GO terms are related to human, mouse, fly, worm and Saccharomyces genes and proteins.
GoMiner & MatchMiner	- GoMiner is a Java-based program package that organizes lists of 'interesting' genes

Analysis of Data: Clustering

- GO Term Finder
  - Searches for significant shared GO terms
  - Gavin Sherlock, Stanford Microarray Database
- VLAD
  - Web interface for clustering DNA microarray data
  - Limited organismal coverage
    - Human, mouse, fly, worm, yeast
- MAPPFinder
  - Accessory program for GenMAPP (MicroArrayPathwayProfiles)

GO Term Finder



SGD Gene Ontology Term Finder

This GO
your list

- Y
- C

Gavin Sherlock > GO-TermFinder

GO-TermFinder

This Release	GO-TermFinder-0.2 [Download] [Browse] 14 Apr 2003
Other Releases	GO-TermFinder-0.1 -- 08 Mar 2003
Links	[CPAN Testers] [CPAN Request Tracker]
Special Files	Changes MANIFEST README

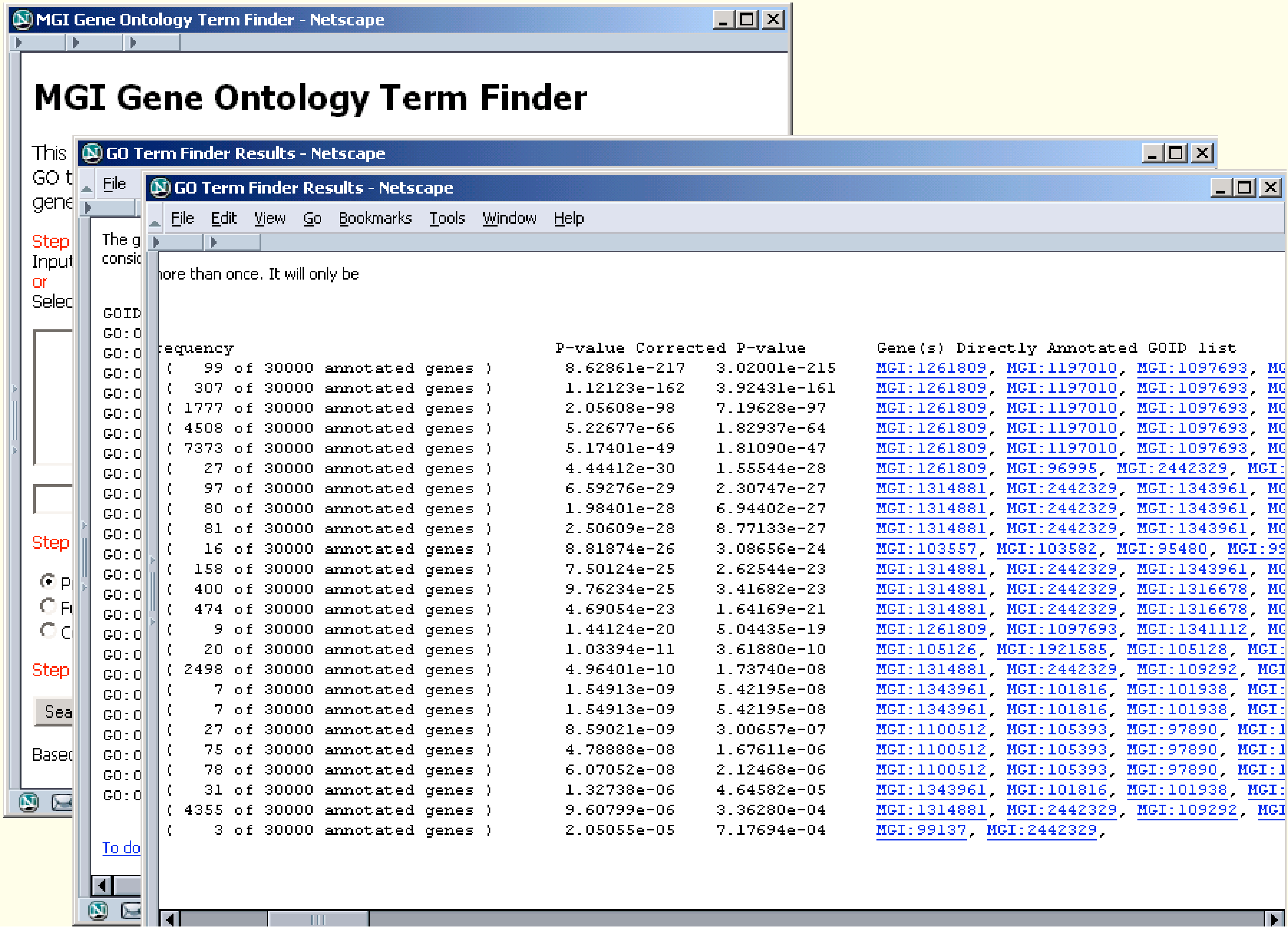
Modules

GO::AnnotatedGene	0.1
GO::Annotation	0.1
GO::AnnotationProvider	0.1
GO::AnnotationProvider::AnnotationParser	0.1
GO::Node	0.1
GO::OntologyProvider	0.1
GO::OntologyProvider::OntologyParser	0.1
GO::Reference	0.1
GO::TermFinder	0.2

Documentation

GO:AnnotationProvider::AnnotationParser

GO Tools at MGI



MGI Gene Ontology Term Finder

This
GO t
gene

Step
Input
or
Selec

Step

● P
○ F
○ C

Step

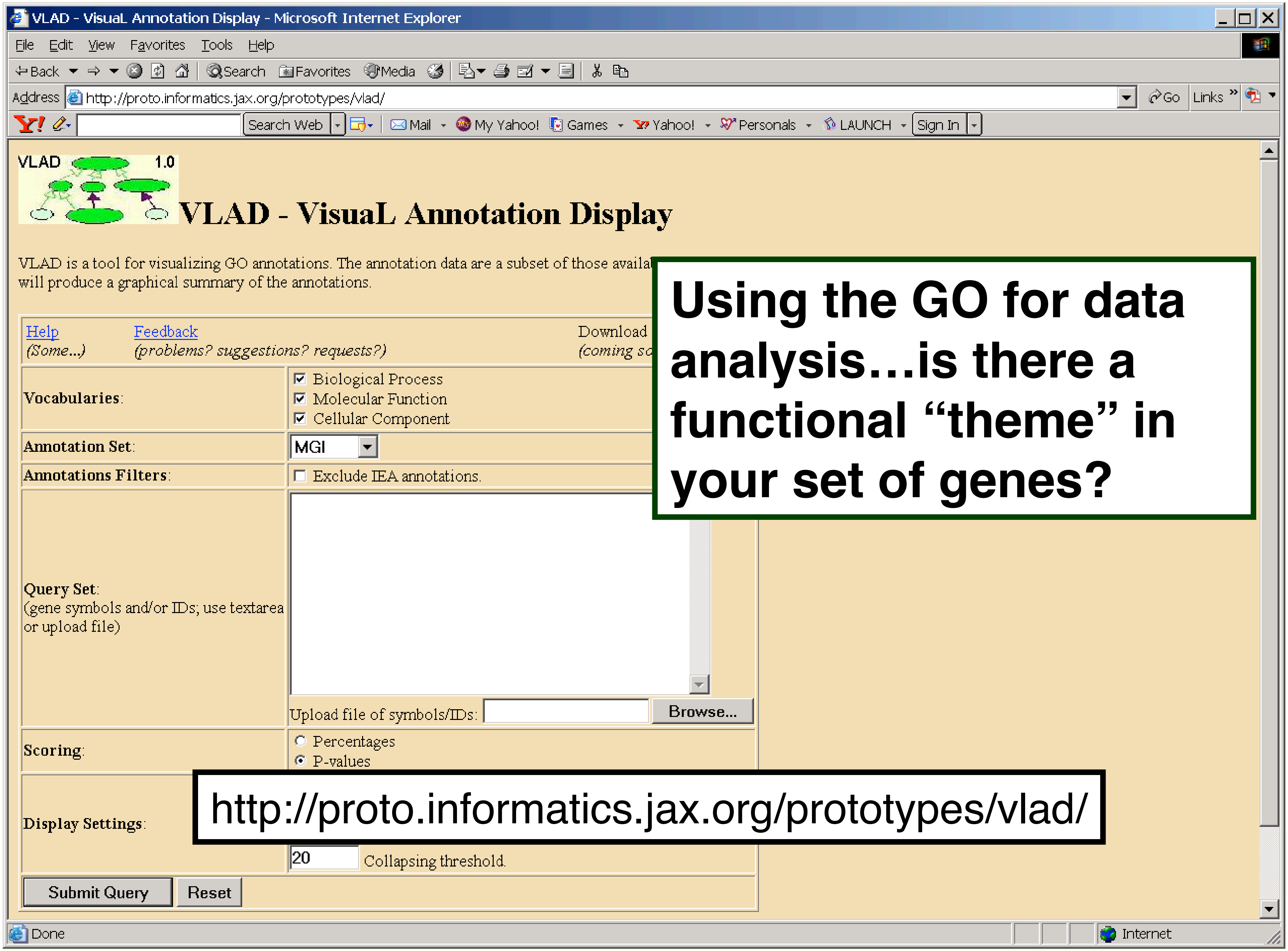
Based

The g
consid

GOID

ore than once. It will only be

```
requency                              P-value Corrected P-value     Gene(s) Directly Annotated GOID list
(   99 of 30000 annotated genes )      8.62861e-217   3.02001e-215   MGI:1261809, MGI:1197010, MGI:1097693, MG
(  307 of 30000 annotated genes )      1.12123e-162   3.92431e-161   MGI:1261809, MGI:1197010, MGI:1097693, MG
( 1777 of 30000 annotated genes )      2.05608e-98    7.19628e-97    MGI:1261809, MGI:1197010, MGI:1097693, MG
( 4508 of 30000 annotated genes )      5.22677e-66    1.82937e-64    MGI:1261809, MGI:1197010, MGI:1097693, MG
( 7373 of 30000 annotated genes )      5.17401e-49    1.81090e-47    MGI:1261809, MGI:1197010, MGI:1097693, MG
(   27 of 30000 annotated genes )      4.44412e-30    1.55544e-28    MGI:1261809, MGI:96995, MGI:2442329, MGI:
(   97 of 30000 annotated genes )      6.59276e-29    2.30747e-27    MGI:1314881, MGI:2442329, MGI:1343961, MG
(   80 of 30000 annotated genes )      1.98401e-28    6.94402e-27    MGI:1314881, MGI:2442329, MGI:1343961, MG
(   81 of 30000 annotated genes )      2.50609e-28    8.77133e-27    MGI:1314881, MGI:2442329, MGI:1343961, MG
(   16 of 30000 annotated genes )      8.81874e-26    3.08656e-24    MGI:103557, MGI:103582, MGI:95480, MGI:99
(  158 of 30000 annotated genes )      7.50124e-25    2.62544e-23    MGI:1314881, MGI:2442329, MGI:1343961, MG
(  400 of 30000 annotated genes )      9.76234e-25    3.41682e-23    MGI:1314881, MGI:2442329, MGI:1316678, MG
(  474 of 30000 annotated genes )      4.69054e-23    1.64169e-21    MGI:1314881, MGI:2442329, MGI:1316678, MG
(    9 of 30000 annotated genes )      1.44124e-20    5.04435e-19    MGI:1261809, MGI:1097693, MGI:1341112, MG
(   20 of 30000 annotated genes )      1.03394e-11    3.61880e-10    MGI:105126, MGI:1921585, MGI:105128, MGI:
( 2498 of 30000 annotated genes )      4.96401e-10    1.73740e-08    MGI:1314881, MGI:2442329, MGI:109292, MGI
(    7 of 30000 annotated genes )      1.54913e-09    5.42195e-08    MGI:1343961, MGI:101816, MGI:101938, MGI:
(    7 of 30000 annotated genes )      1.54913e-09    5.42195e-08    MGI:1343961, MGI:101816, MGI:101938, MGI:
(   27 of 30000 annotated genes )      8.59021e-09    3.00657e-07    MGI:1100512, MGI:105393, MGI:97890, MGI:1
(   75 of 30000 annotated genes )      4.78888e-08    1.67611e-06    MGI:1100512, MGI:105393, MGI:97890, MGI:1
(   78 of 30000 annotated genes )      6.07052e-08    2.12468e-06    MGI:1100512, MGI:105393, MGI:97890, MGI:1
(   31 of 30000 annotated genes )      1.32738e-06    4.64582e-05    MGI:1343961, MGI:101816, MGI:101938, MGI:
( 4355 of 30000 annotated genes )      9.60799e-06    3.36280e-04    MGI:1314881, MGI:2442329, MGI:109292, MGI
(    3 of 30000 annotated genes )      2.05055e-05    7.17694e-04    MGI:99137, MGI:2442329,
```



Address: http://proto.informatics.jax.org/prototypes/vlad/

VLAD - VisuaL Annotation Display

VLAD is a tool for visualizing GO annotations. The annotation data are a subset of those availa
will produce a graphical summary of the annotations.

Help *(Some...)*	Feedback *(problems? suggestions? requests?)*	Download *(coming so*

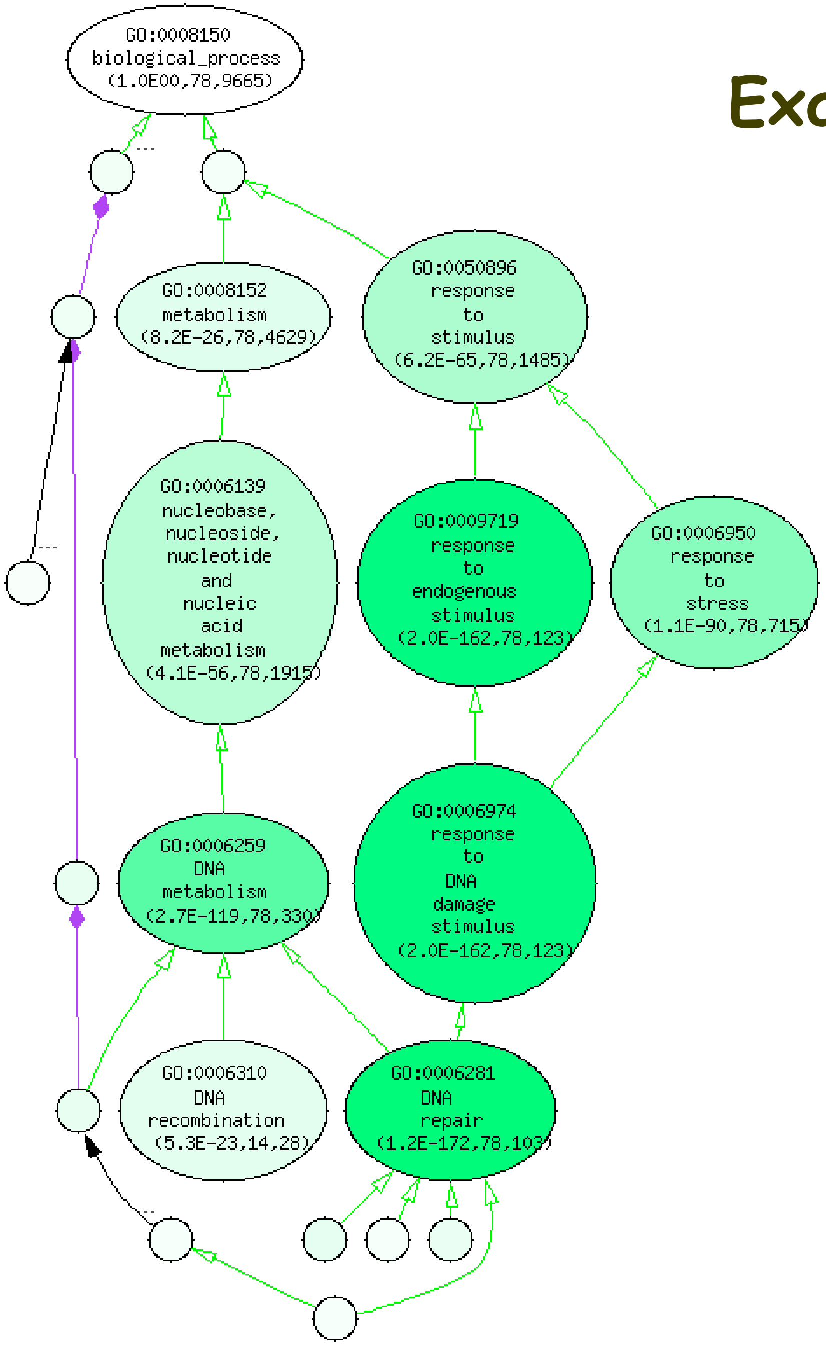
Vocabularies:	☑ Biological Process ☑ Molecular Function ☑ Cellular Component
Annotation Set:	MGI
Annotations Filters:	☐ Exclude IEA annotations.
Query Set: (gene symbols and/or IDs; use textarea or upload file)	Upload file of symbols/IDs: Browse...
Scoring:	○ Percentages ● P-values
Display Settings:	20 Collapsing threshold.

Submit Query Reset

Using the GO for data analysis…is there a functional “theme” in your set of genes?

http://proto.informatics.jax.org/prototypes/vlad/

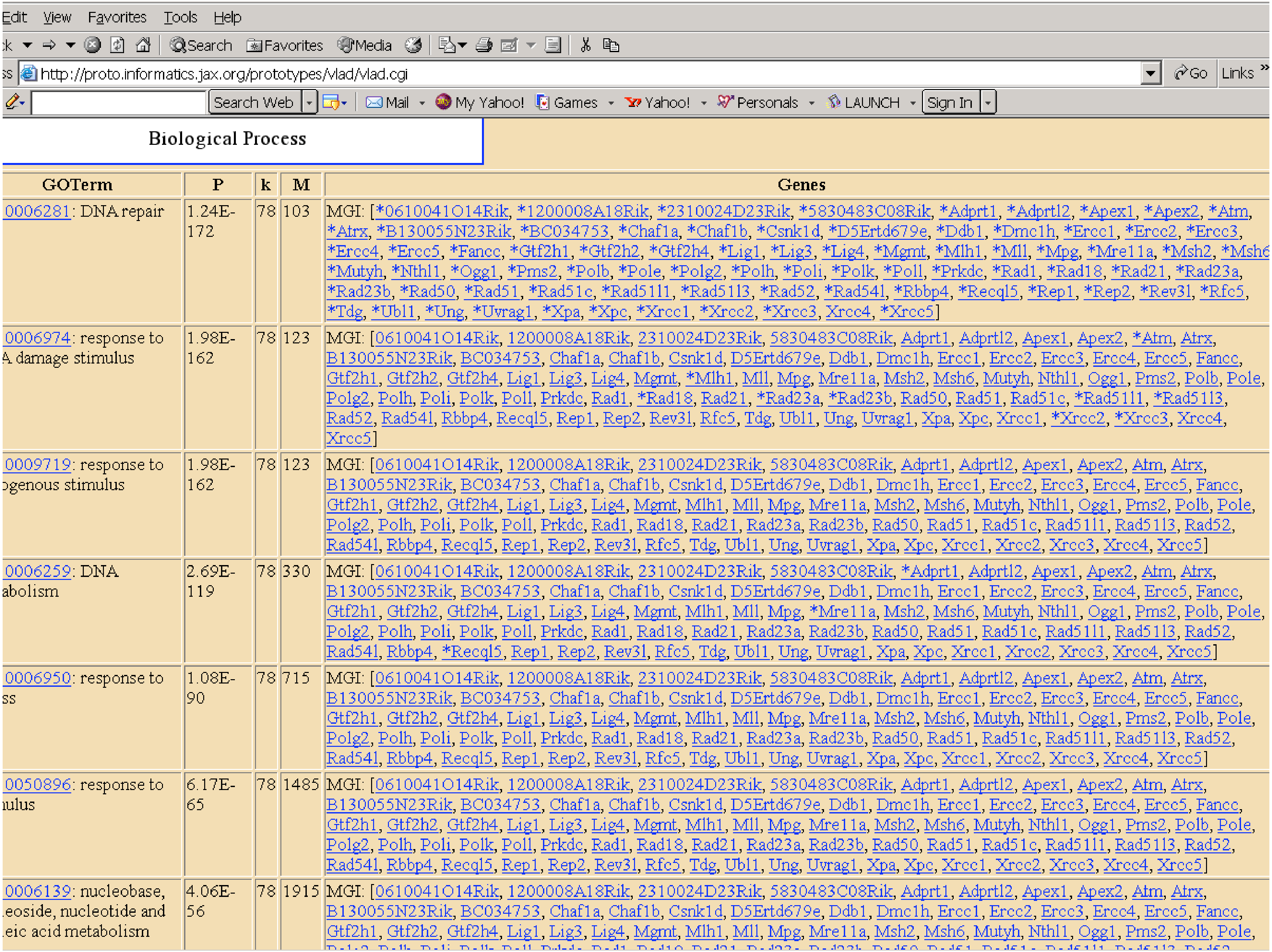
Example of VLAD Output



Biological Process

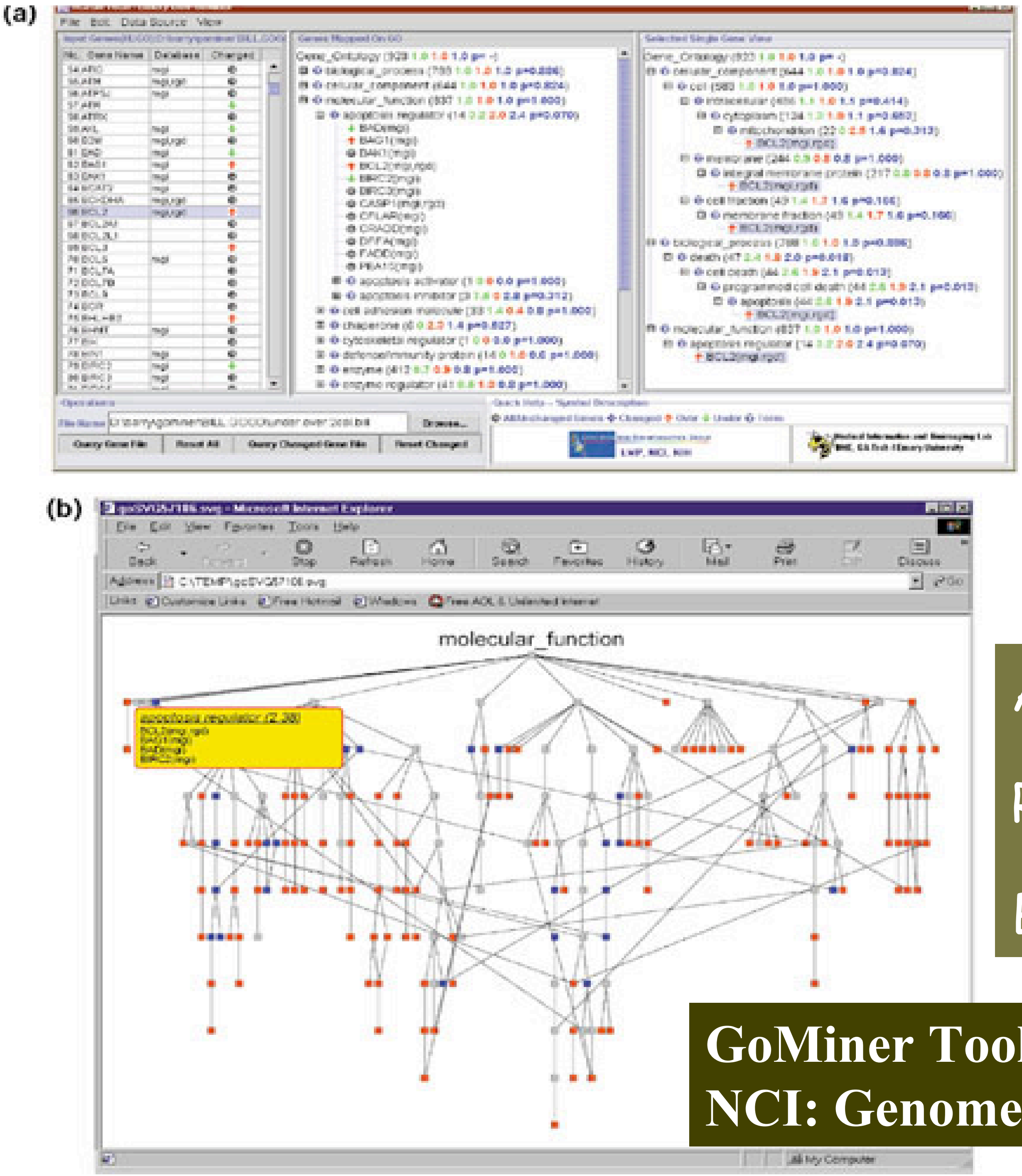
Compare annotations associated with the test set to the entire universe of GO annotations….

<u>DNA Repair</u> seems to be a common theme.



Biological Process

GOTerm	P	k	M	Genes
0006281: DNA repair	1.24E-172	78	103	MGI: [*0610041O14Rik, *1200008A18Rik, *2310024D23Rik, *5830483C08Rik, *Adprt1, *Adprtl2, *Apex1, *Apex2, *Atm, *Atrx, *B130055N23Rik, *BC034753, *Chaf1a, *Chaf1b, *Csnk1d, *D5Ertd679e, *Ddb1, *Dmc1h, *Ercc1, *Ercc2, *Ercc3, *Ercc4, *Ercc5, *Fancc, *Gtf2h1, *Gtf2h2, *Gtf2h4, *Lig1, *Lig3, *Lig4, *Mgmt, *Mlh1, *Mll, *Mpg, *Mre11a, *Msh2, *Msh6 *Mutyh, *Nthl1, *Ogg1, *Pms2, *Polb, *Pole, *Polg2, *Polh, *Poli, *Polk, *Poll, *Prkdc, *Rad1, *Rad18, *Rad21, *Rad23a, *Rad23b, *Rad50, *Rad51, *Rad51c, *Rad51l1, *Rad51l3, *Rad52, *Rad54l, *Rbbp4, *Recql5, *Rep1, *Rep2, *Rev3l, *Rfc5, *Tdg, *Ubl1, *Ung, *Uvrag1, *Xpa, *Xpc, *Xrcc1, *Xrcc2, *Xrcc3, Xrcc4, *Xrcc5]
0006974: response to A damage stimulus	1.98E-162	78	123	MGI: [0610041O14Rik, 1200008A18Rik, 2310024D23Rik, 5830483C08Rik, Adprt1, Adprtl2, Apex1, Apex2, *Atm, Atrx, B130055N23Rik, BC034753, Chaf1a, Chaf1b, Csnk1d, D5Ertd679e, Ddb1, Dmc1h, Ercc1, Ercc2, Ercc3, Ercc4, Ercc5, Fancc, Gtf2h1, Gtf2h2, Gtf2h4, Lig1, Lig3, Lig4, Mgmt, *Mlh1, Mll, Mpg, Mre11a, Msh2, Msh6, Mutyh, Nthl1, Ogg1, Pms2, Polb, Pole, Polg2, Polh, Poli, Polk, Poll, Prkdc, Rad1, *Rad18, Rad21, *Rad23a, *Rad23b, Rad50, Rad51, Rad51c, *Rad51l1, *Rad51l3, Rad52, Rad54l, Rbbp4, Recql5, Rep1, Rep2, Rev3l, Rfc5, Tdg, Ubl1, Ung, Uvrag1, Xpa, Xpc, Xrcc1, *Xrcc2, *Xrcc3, Xrcc4, Xrcc5]
0009719: response to ogenous stimulus	1.98E-162	78	123	MGI: [0610041O14Rik, 1200008A18Rik, 2310024D23Rik, 5830483C08Rik, Adprt1, Adprtl2, Apex1, Apex2, Atm, Atrx, B130055N23Rik, BC034753, Chaf1a, Chaf1b, Csnk1d, D5Ertd679e, Ddb1, Dmc1h, Ercc1, Ercc2, Ercc3, Ercc4, Ercc5, Fancc, Gtf2h1, Gtf2h2, Gtf2h4, Lig1, Lig3, Lig4, Mgmt, Mlh1, Mll, Mpg, Mre11a, Msh2, Msh6, Mutyh, Nthl1, Ogg1, Pms2, Polb, Pole, Polg2, Polh, Poli, Polk, Poll, Prkdc, Rad1, Rad18, Rad21, Rad23a, Rad23b, Rad50, Rad51, Rad51c, Rad51l1, Rad51l3, Rad52, Rad54l, Rbbp4, Recql5, Rep1, Rep2, Rev3l, Rfc5, Tdg, Ubl1, Ung, Uvrag1, Xpa, Xpc, Xrcc1, Xrcc2, Xrcc3, Xrcc4, Xrcc5]
0006259: DNA abolism	2.69E-119	78	330	MGI: [0610041O14Rik, 1200008A18Rik, 2310024D23Rik, 5830483C08Rik, *Adprt1, Adprtl2, Apex1, Apex2, Atm, Atrx, B130055N23Rik, BC034753, Chaf1a, Chaf1b, Csnk1d, D5Ertd679e, Ddb1, Dmc1h, Ercc1, Ercc2, Ercc3, Ercc4, Ercc5, Fancc, Gtf2h1, Gtf2h2, Gtf2h4, Lig1, Lig3, Lig4, Mgmt, Mlh1, Mll, Mpg, *Mre11a, Msh2, Msh6, Mutyh, Nthl1, Ogg1, Pms2, Polb, Pole, Polg2, Polh, Poli, Polk, Poll, Prkdc, Rad1, Rad18, Rad21, Rad23a, Rad23b, Rad50, Rad51, Rad51c, Rad51l1, Rad51l3, Rad52, Rad54l, Rbbp4, *Recql5, Rep1, Rep2, Rev3l, Rfc5, Tdg, Ubl1, Ung, Uvrag1, Xpa, Xpc, Xrcc1, Xrcc2, Xrcc3, Xrcc4, Xrcc5]
0006950: response to ss	1.08E-90	78	715	MGI: [0610041O14Rik, 1200008A18Rik, 2310024D23Rik, 5830483C08Rik, Adprt1, Adprtl2, Apex1, Apex2, Atm, Atrx, B130055N23Rik, BC034753, Chaf1a, Chaf1b, Csnk1d, D5Ertd679e, Ddb1, Dmc1h, Ercc1, Ercc2, Ercc3, Ercc4, Ercc5, Fancc, Gtf2h1, Gtf2h2, Gtf2h4, Lig1, Lig3, Lig4, Mgmt, Mlh1, Mll, Mpg, Mre11a, Msh2, Msh6, Mutyh, Nthl1, Ogg1, Pms2, Polb, Pole, Polg2, Polh, Poli, Polk, Poll, Prkdc, Rad1, Rad18, Rad21, Rad23a, Rad23b, Rad50, Rad51, Rad51c, Rad51l1, Rad51l3, Rad52, Rad54l, Rbbp4, Recql5, Rep1, Rep2, Rev3l, Rfc5, Tdg, Ubl1, Ung, Uvrag1, Xpa, Xpc, Xrcc1, Xrcc2, Xrcc3, Xrcc4, Xrcc5]
0050896: response to ulus	6.17E-65	78	1485	MGI: [0610041O14Rik, 1200008A18Rik, 2310024D23Rik, 5830483C08Rik, Adprt1, Adprtl2, Apex1, Apex2, Atm, Atrx, B130055N23Rik, BC034753, Chaf1a, Chaf1b, Csnk1d, D5Ertd679e, Ddb1, Dmc1h, Ercc1, Ercc2, Ercc3, Ercc4, Ercc5, Fancc, Gtf2h1, Gtf2h2, Gtf2h4, Lig1, Lig3, Lig4, Mgmt, Mlh1, Mll, Mpg, Mre11a, Msh2, Msh6, Mutyh, Nthl1, Ogg1, Pms2, Polb, Pole, Polg2, Polh, Poli, Polk, Poll, Prkdc, Rad1, Rad18, Rad21, Rad23a, Rad23b, Rad50, Rad51, Rad51c, Rad51l1, Rad51l3, Rad52, Rad54l, Rbbp4, Recql5, Rep1, Rep2, Rev3l, Rfc5, Tdg, Ubl1, Ung, Uvrag1, Xpa, Xpc, Xrcc1, Xrcc2, Xrcc3, Xrcc4, Xrcc5]
0006139: nucleobase, eoside, nucleotide and eic acid metabolism	4.06E-56	78	1915	MGI: [0610041O14Rik, 1200008A18Rik, 2310024D23Rik, 5830483C08Rik, Adprt1, Adprtl2, Apex1, Apex2, Atm, Atrx, B130055N23Rik, BC034753, Chaf1a, Chaf1b, Csnk1d, D5Ertd679e, Ddb1, Dmc1h, Ercc1, Ercc2, Ercc3, Ercc4, Ercc5, Fancc, Gtf2h1, Gtf2h2, Gtf2h4, Lig1, Lig3, Lig4, Mgmt, Mlh1, Mll, Mpg, Mre11a, Msh2, Msh6, Mutyh, Nthl1, Ogg1, Pms2, Polb, Pole,



Color indicates up/down regulation

Apotosis Regulator

Red: up by 1.5 fold

Blue: down 1.5 fold

GoMiner Tool, John Weinstein *et al*, NCI: Genome Biol. 4 (R28) 2003

Analysis of Data - Binning

- GO_Slims
  - High-level sets of terms
  - Can be specific for specific datasets
- Comparative GO_Slims
  - Data analysis

Molecular Function Bins

(MGI-RIKEN example)

1.) **defense/immunity protein**: defense/immunity protein

2.) **cytoskeletal protein**: cytoskeletal regulator OR motor OR structural constituent of cytoskeleton OR structural constituent of eye lens OR structural constituent of muscle OR cytoskeletal binding protein

3.) **transcription regulator**: transcription regulator

4.) **cell adhesion molecule**: cell adhesion molecule

5.) **ligand binding or carrier**: ligand binding or carrier

6.) **ligand**: ligand

7.) **receptor**: receptor

8.) **other signal transduction molecule**: signal transducer EXCLUDING (ligand OR receptor)

9.) **enzyme**: enzyme

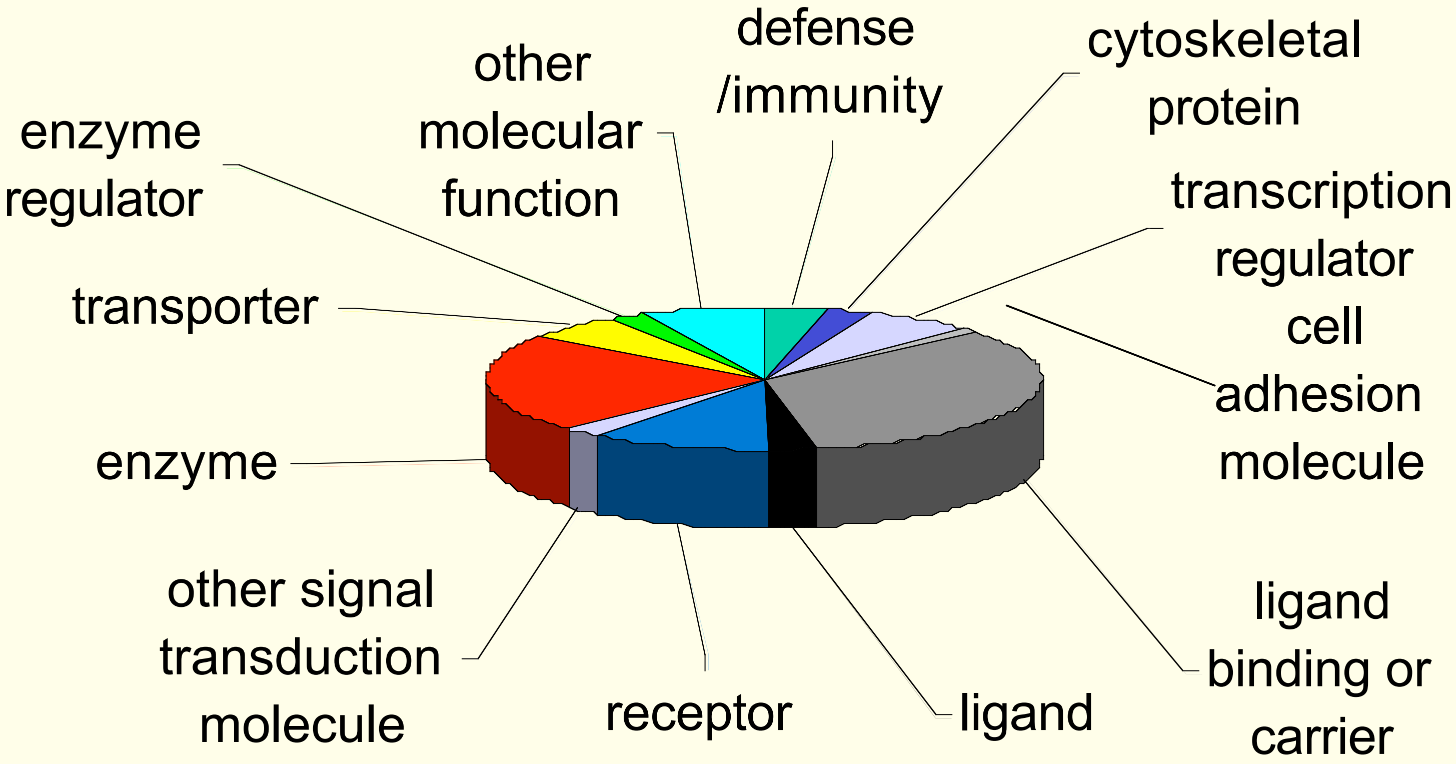
10.) **transporter**: transporter

11.) **enzyme regulator**: enzyme regulator

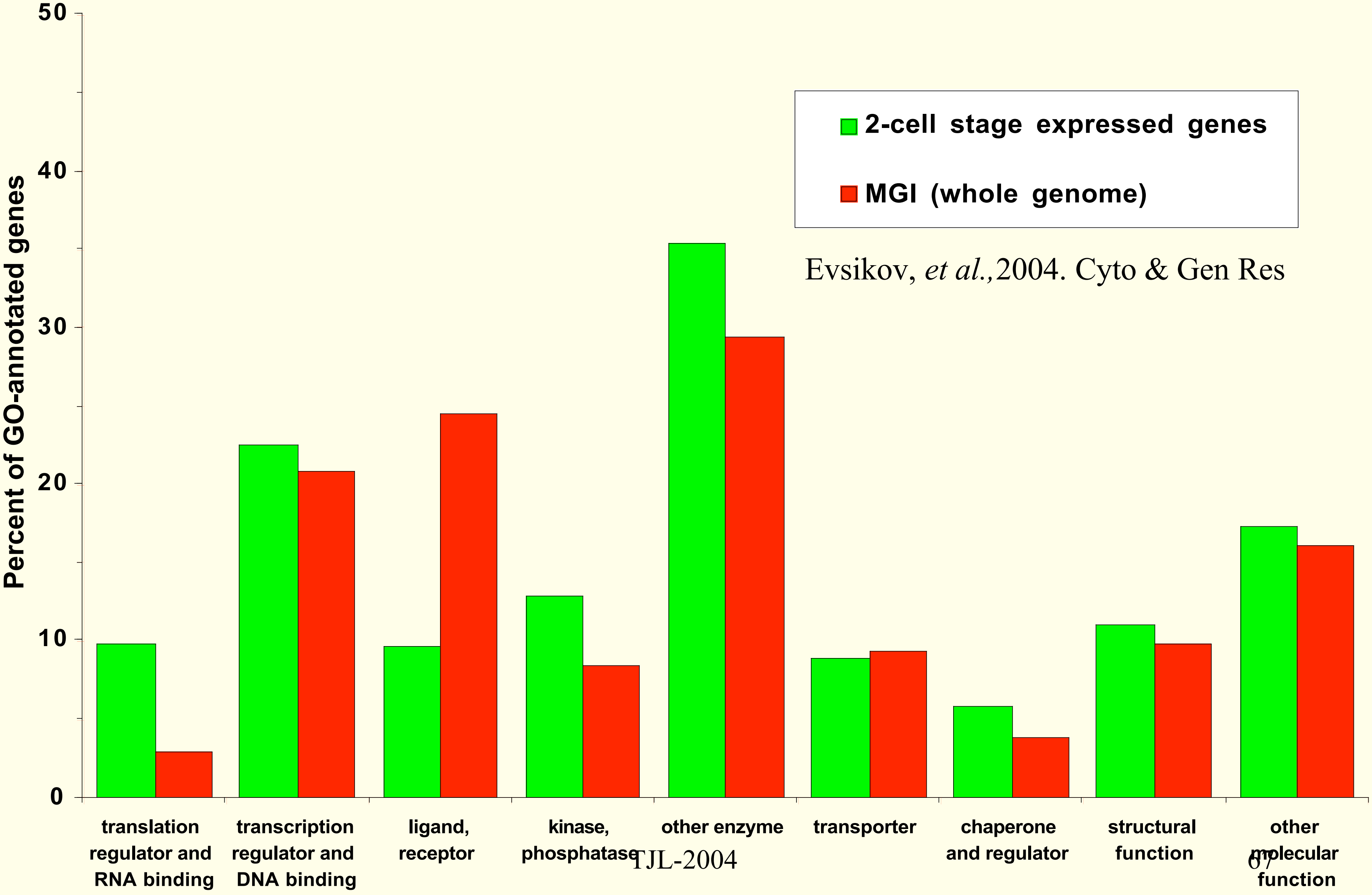
12.) **other molecular function**: NOT (1-11)

Molecular Function Ontology

(MGI annotations)



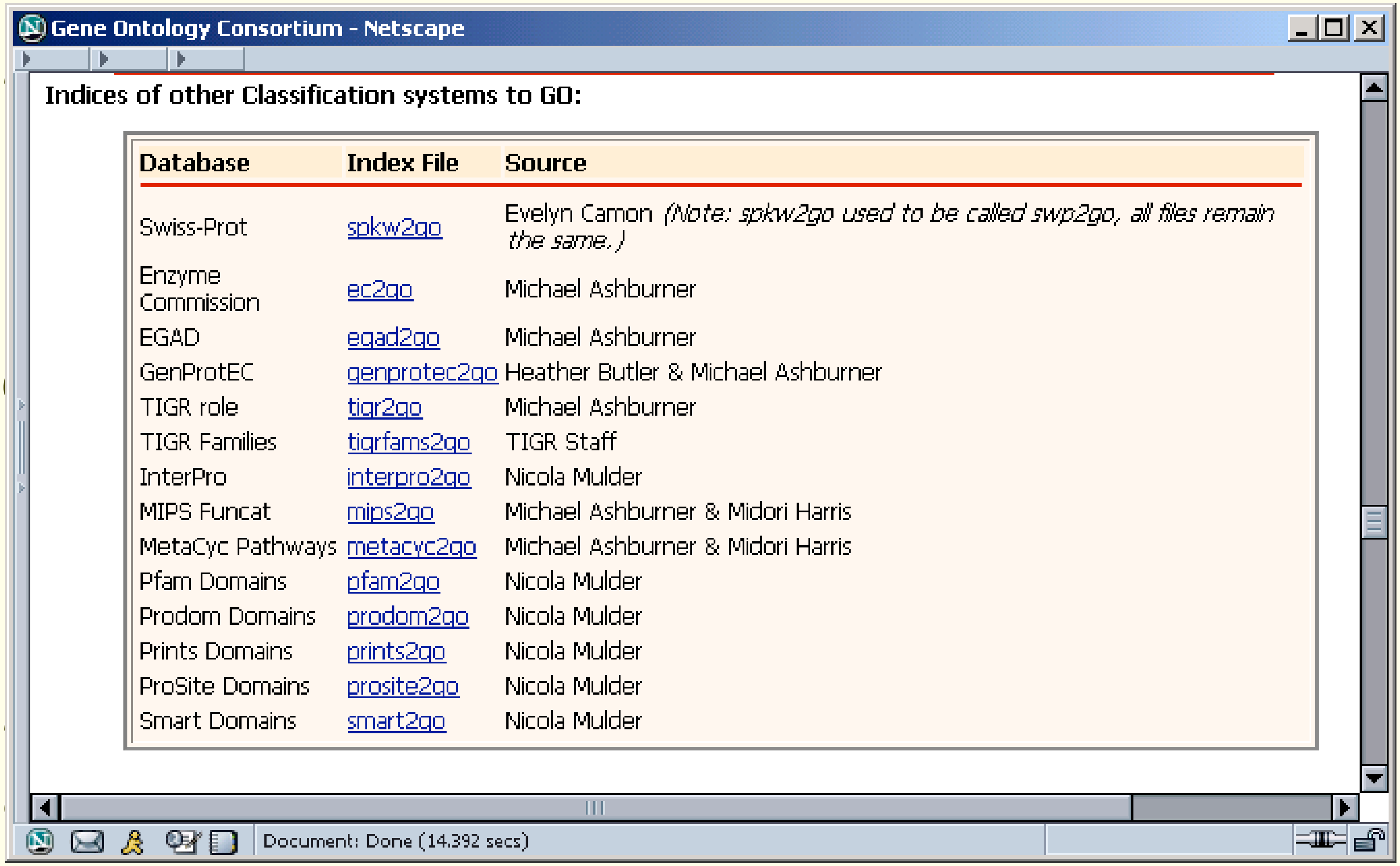
Molecular functions: 2 cell stage expressed genes



Evsikov, *et al.,*2004. Cyto & Gen Res

Other GO Tools

- Indices of other Classifications to GO

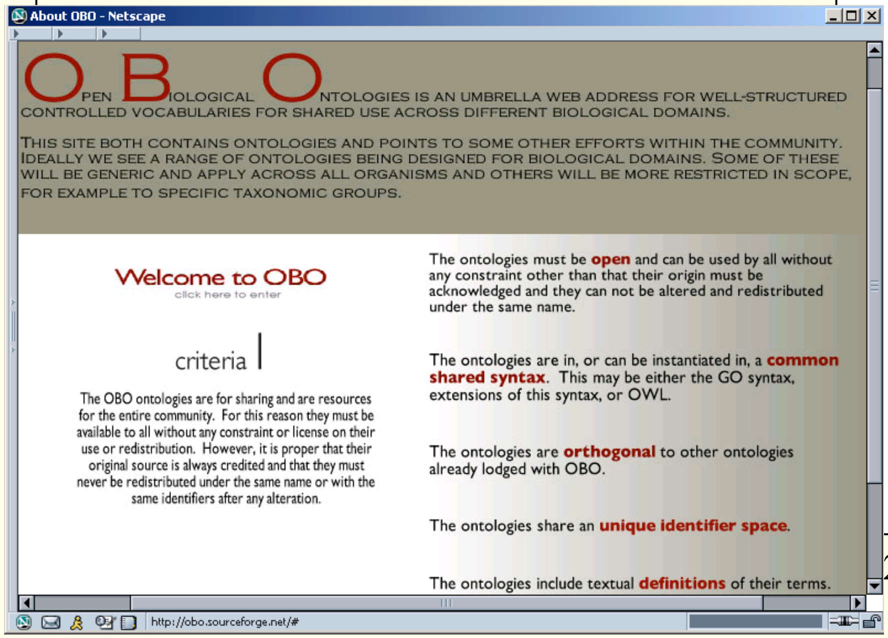


Indices of other Classification systems to GO:

Database	Index File	Source
Swiss-Prot	spkw2go	Evelyn Camon *(Note: spkw2go used to be called swp2go, all files remain the same.)*
Enzyme Commission	ec2go	Michael Ashburner
EGAD	egad2go	Michael Ashburner
GenProtEC	genprotec2go	Heather Butler & Michael Ashburner
TIGR role	tigr2go	Michael Ashburner
TIGR Families	tigrfams2go	TIGR Staff
InterPro	interpro2go	Nicola Mulder
MIPS Funcat	mips2go	Michael Ashburner & Midori Harris
MetaCyc Pathways	metacyc2go	Michael Ashburner & Midori Harris
Pfam Domains	pfam2go	Nicola Mulder
Prodom Domains	prodom2go	Nicola Mulder
Prints Domains	prints2go	Nicola Mulder
ProSite Domains	prosite2go	Nicola Mulder
Smart Domains	smart2go	Nicola Mulder

Extending the paradigm

- OBO – Open Biological Ontologies
  - -Open and are in GO syntax or DAML+OIL
  - -Orthogonal to existing ontologies to facilitate combinatorial approaches
  - -Share unique identifie



OPEN BIOLOGICAL ONTOLOGIES IS AN UMBRELLA WEB ADDRESS FOR WELL-STRUCTURED CONTROLLED VOCABULARIES FOR SHARED USE ACROSS DIFFERENT BIOLOGICAL DOMAINS.

THIS SITE BOTH CONTAINS ONTOLOGIES AND POINTS TO SOME OTHER EFFORTS WITHIN THE COMMUNITY. IDEALLY WE SEE A RANGE OF ONTOLOGIES BEING DESIGNED FOR BIOLOGICAL DOMAINS. SOME OF THESE WILL BE GENERIC AND APPLY ACROSS ALL ORGANISMS AND OTHERS WILL BE MORE RESTRICTED IN SCOPE, FOR EXAMPLE TO SPECIFIC TAXONOMIC GROUPS.

Welcome to OBO
click here to enter

criteria

The OBO ontologies are for sharing and are resources for the entire community. For this reason they must be available to all without any constraint or license on their use or redistribution. However, it is proper that their original source is always credited and that they must never be redistributed under the same name or with the same identifiers after any alteration.

The ontologies must be **open** and can be used by all without any constraint other than that their origin must be acknowledged and they can not be altered and redistributed under the same name.

The ontologies are in, or can be instantiated in, a **common shared syntax**. This may be either the GO syntax, extensions of this syntax, or OWL.

The ontologies are **orthogonal** to other ontologies already lodged with OBO.

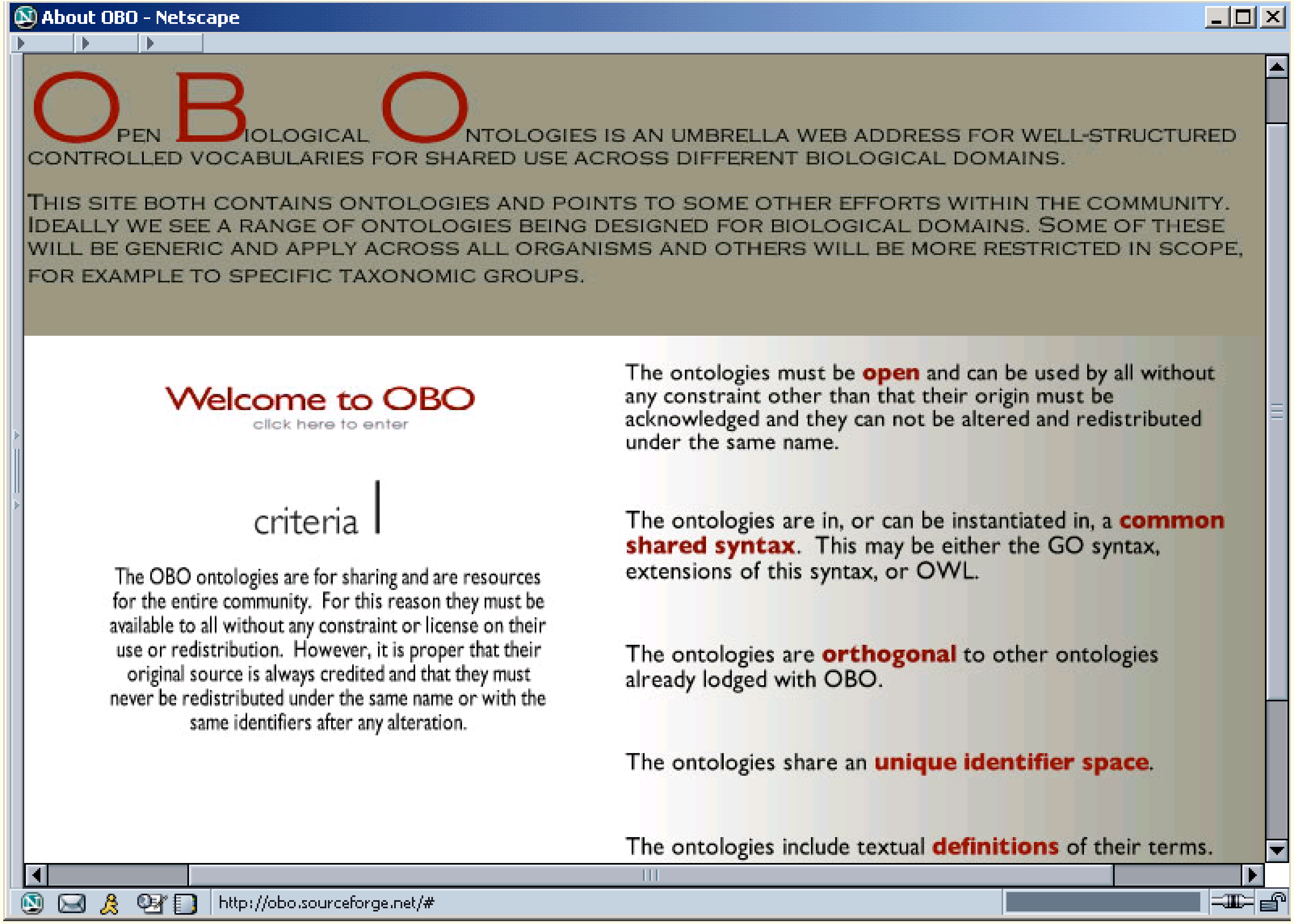
The ontologies share an **unique identifier space**.

The ontologies include textual **definitions** of their terms.

- •Anatomies
- •Cell Types
- •Sequence Attributes
- •Temporal Attributes
- •Phenotypes
- •Diseases

http:obo.sf.net

Open Biological Ontologies - OBO



OPEN BIOLOGICAL ONTOLOGIES IS AN UMBRELLA WEB ADDRESS FOR WELL-STRUCTURED CONTROLLED VOCABULARIES FOR SHARED USE ACROSS DIFFERENT BIOLOGICAL DOMAINS.

THIS SITE BOTH CONTAINS ONTOLOGIES AND POINTS TO SOME OTHER EFFORTS WITHIN THE COMMUNITY. IDEALLY WE SEE A RANGE OF ONTOLOGIES BEING DESIGNED FOR BIOLOGICAL DOMAINS. SOME OF THESE WILL BE GENERIC AND APPLY ACROSS ALL ORGANISMS AND OTHERS WILL BE MORE RESTRICTED IN SCOPE, FOR EXAMPLE TO SPECIFIC TAXONOMIC GROUPS.

Welcome to OBO
click here to enter

criteria

The OBO ontologies are for sharing and are resources for the entire community. For this reason they must be available to all without any constraint or license on their use or redistribution. However, it is proper that their original source is always credited and that they must never be redistributed under the same name or with the same identifiers after any alteration.

The ontologies must be **open** and can be used by all without any constraint other than that their origin must be acknowledged and they can not be altered and redistributed under the same name.

The ontologies are in, or can be instantiated in, a **common shared syntax**. This may be either the GO syntax, extensions of this syntax, or OWL.

The ontologies are **orthogonal** to other ontologies already lodged with OBO.

The ontologies share an **unique identifier space**.

The ontologies include textual **definitions** of their terms.

Sequence Ontology - SO

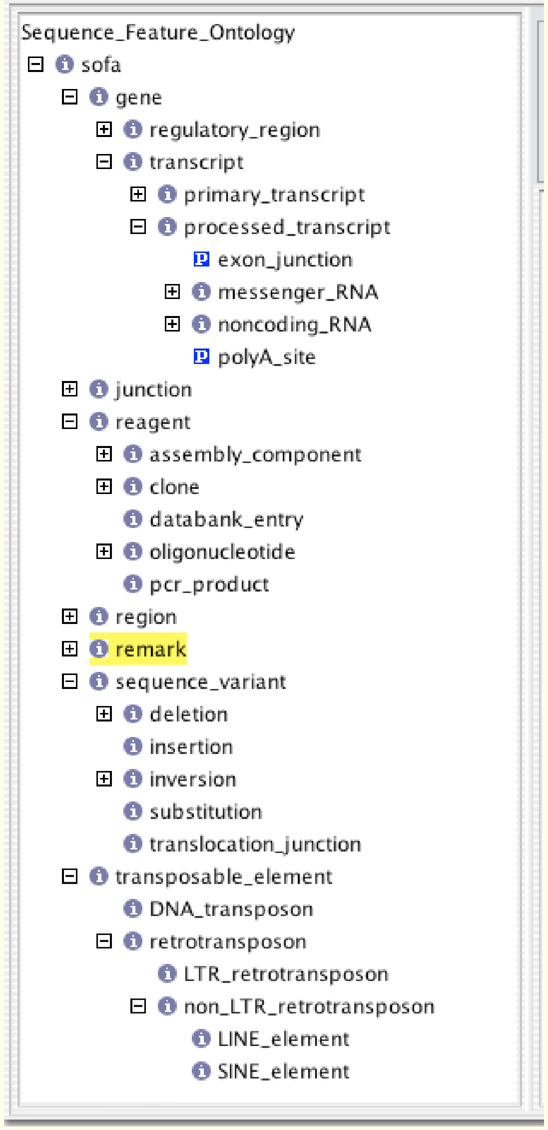
Provides a structured controlled vocabulary for sequence annotation, for the exchange of annotation data and for the description of sequence objects in databases. S
minimal version of SO for use in data exchange.

- **OBO relationship types**
- **MESH**
- genomic & proteomic
  - **gene structure and variation**
  - gene product
    - gene product name
    - **molecular function**
    - **biological process**
    - **cellular component**
    - protein
      - **protein domain**
      - **protein covalent bond**
      - **protein-protein interaction**
- biochemical
  - **biochemical substance**
  - cell signalling
  - **physical-chemical methods and properties**
- developmental timeline
  - plant development
    - **Arabidopsis development**
    - **Rice development**
  - animal development
    - **Human anatomy and development**
    - **Mus anatomy and development**
    - **Zebrafish anatomy and development**

contact	**SONG developers**
home	**SO home page**
name space	SO (subset is SOFA)
status	Discussion release.
ontology	**so.ontology**
definitions	**so.definition**
qualifiers	
format	**DAG-Edit**

http://song.sourceforge.net/

Sequence Ontology (SO)



Sequence_Feature_Ontology
- ⊟ sofa
  - ⊟ gene
    - ⊞ regulatory_region
    - ⊟ transcript
      - ⊞ primary_transcript
      - ⊟ processed_transcript
        - exon_junction
        - ⊞ messenger_RNA
        - ⊞ noncoding_RNA
        - polyA_site
  - ⊞ junction
  - ⊟ reagent
    - ⊞ assembly_component
    - ⊞ clone
    - databank_entry
    - ⊞ oligonucleotide
    - pcr_product
  - ⊞ region
  - ⊞ remark
  - ⊟ sequence_variant
    - ⊞ deletion
    - insertion
    - ⊞ inversion
    - substitution
    - translocation_junction
  - ⊟ transposable_element
    - DNA_transposon
    - ⊟ retrotransposon
      - LTR_retrotransposon
      - ⊟ non_LTR_retrotransposon
        - LINE_element
        - SINE_element

- A structured controlled vocabulary for the description of primary annotations of nucleic acid sequence
- Can provide structured representations of these primary annotations within genome and model organism databases
- Supports exchange and comparative analysis between information systems

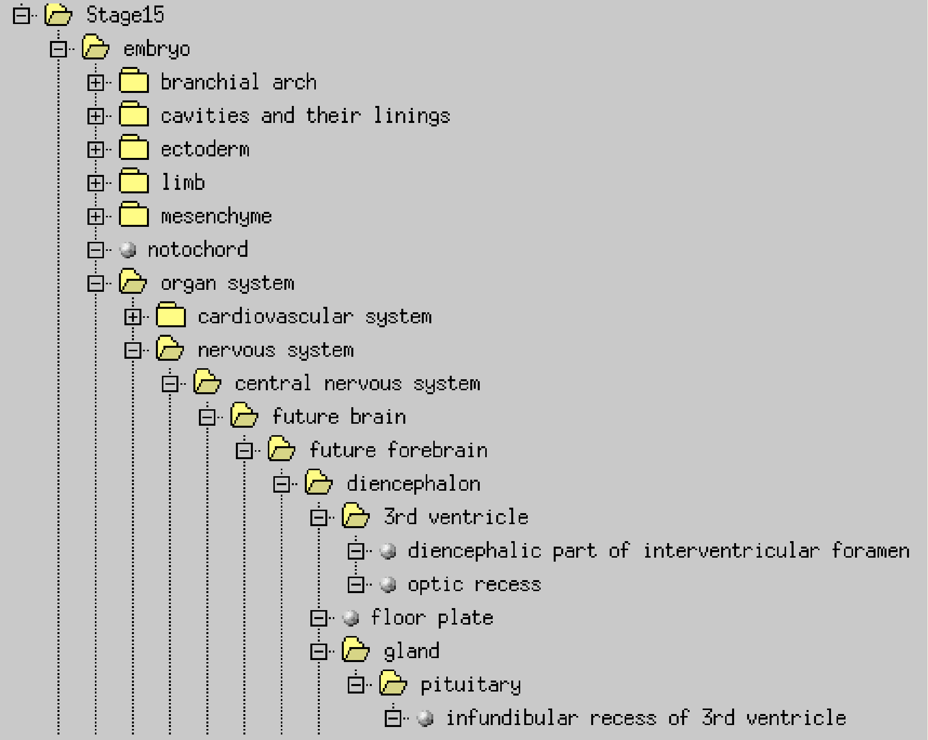
Structured Vocabularies in MGI

- Gene Index (nomenclature)
- Anatomies
- GO:
  - Molecular function,
  - Biological process,
  - Cellular component
- SO – Sequence Ontology
- Phenotypes – MP
- Disease Models

Mouse Anatomies

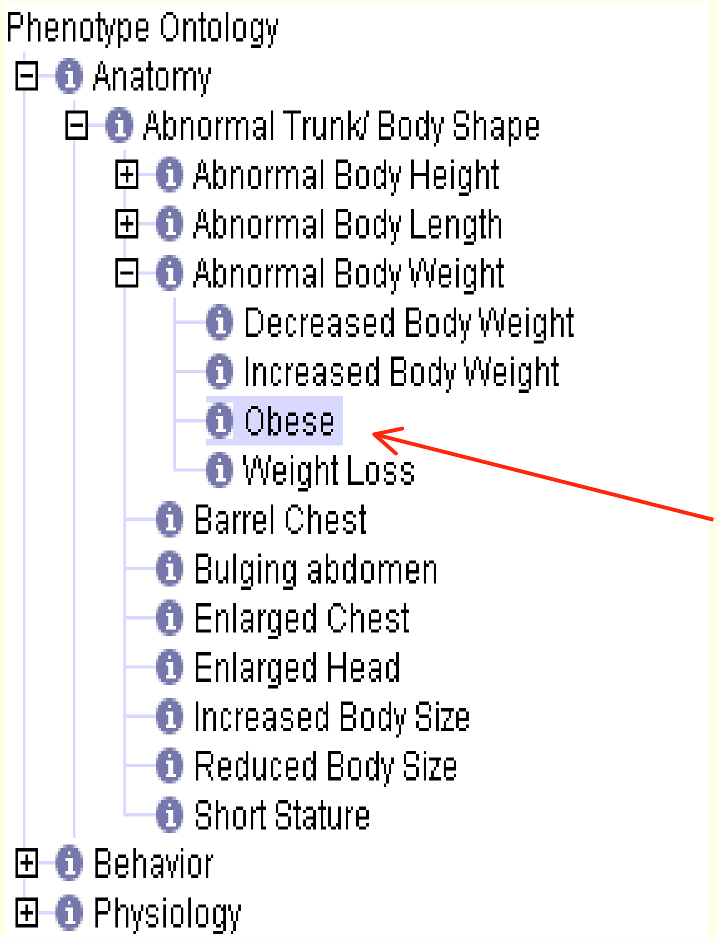
hierarchies

- Developmental and Adult
- Core Vocabularies for Biology
- Incorporate time and lineage components



- Stage15
  - embryo
    - branchial arch
    - cavities and their linings
    - ectoderm
    - limb
    - mesenchyme
    - notochord
    - organ system
      - cardiovascular system
      - nervous system
        - central nervous system
          - future brain
            - future forebrain
              - diencephalon
                - 3rd ventricle
                  - diencephalic part of interventricular foramen
                  - optic recess
                - floor plate
                - gland
                  - pituitary
                    - infundibular recess of 3rd ventricle

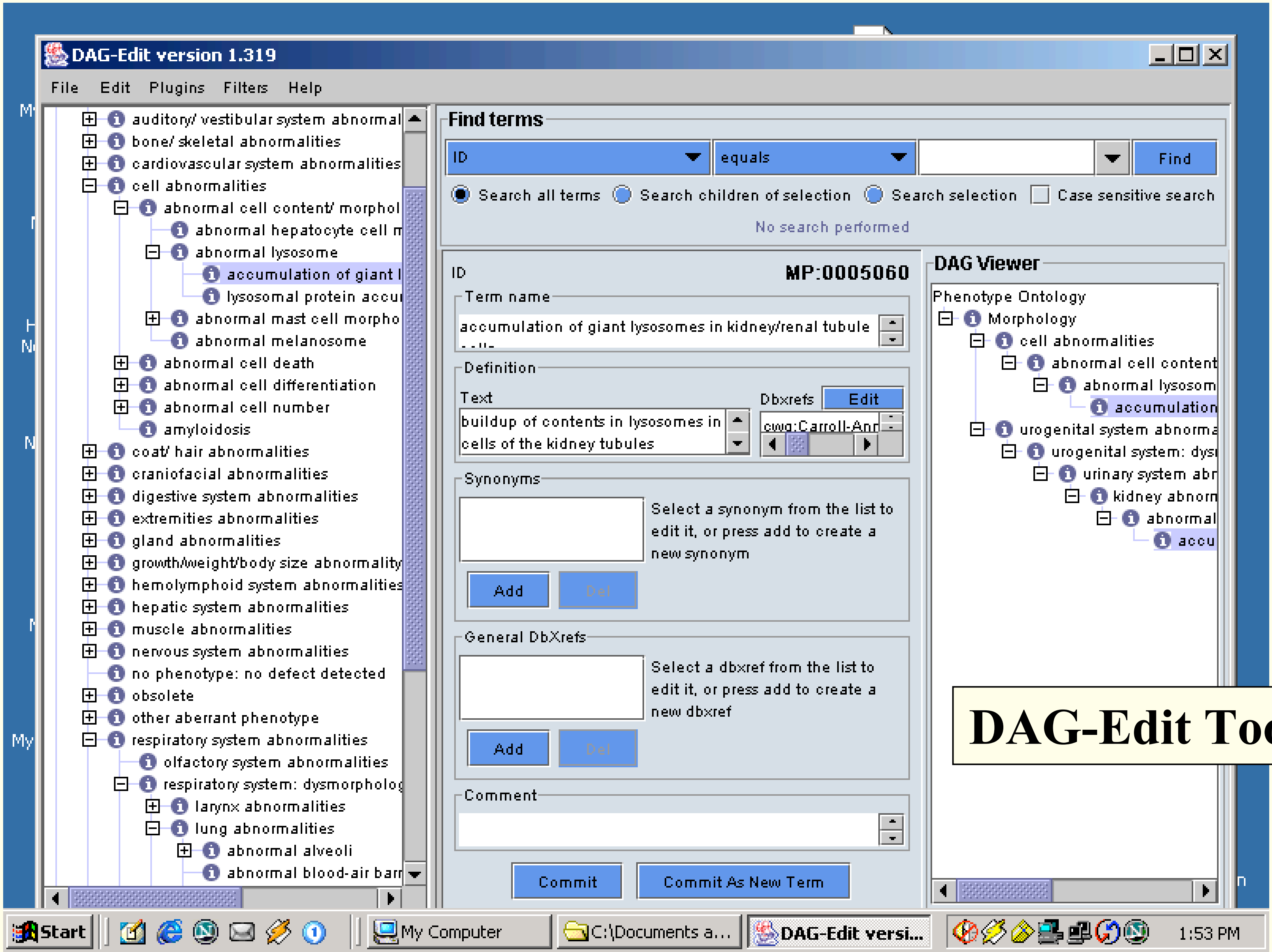
Phenotypes and Diseases



Phenotype Ontology
- ⊟ Anatomy
  - ⊟ Abnormal Trunk/ Body Shape
    - ⊞ Abnormal Body Height
    - ⊞ Abnormal Body Length
    - ⊟ Abnormal Body Weight
      - Decreased Body Weight
      - Increased Body Weight
      - Obese
      - Weight Loss
    - Barrel Chest
    - Bulging abdomen
    - Enlarged Chest
    - Enlarged Head
    - Increased Body Size
    - Reduced Body Size
    - Short Stature
- ⊞ Behavior
- ⊞ Physiology

Mammalian Mutant Phenotype: A controlled, defined vocabulary of anatomical, behavioral and physiological traits used to describe mouse mutant phenotypes.

Mammalian Phenotype Ontology



- ⊞ auditory/ vestibular system abnormal
- ⊞ bone/ skeletal abnormalities
- ⊞ cardiovascular system abnormalities
- ⊟ cell abnormalities
  - ⊟ abnormal cell content/ morphol
    - abnormal hepatocyte cell m
    - ⊟ abnormal lysosome
      - accumulation of giant l
      - lysosomal protein accu
    - ⊞ abnormal mast cell morpho
    - abnormal melanosome
  - ⊞ abnormal cell death
  - ⊞ abnormal cell differentiation
  - ⊞ abnormal cell number
  - amyloidosis
- ⊞ coat/ hair abnormalities
- ⊞ craniofacial abnormalities
- ⊞ digestive system abnormalities
- ⊞ extremities abnormalities
- ⊞ gland abnormalities
- ⊞ growth/weight/body size abnormality
- ⊞ hemolymphoid system abnormalities
- ⊞ hepatic system abnormalities
- ⊞ muscle abnormalities
- ⊞ nervous system abnormalities
- no phenotype: no defect detected
- ⊞ obsolete
- ⊞ other aberrant phenotype
- ⊟ respiratory system abnormalities
  - olfactory system abnormalities
  - ⊟ respiratory system: dysmorpholog
    - ⊞ larynx abnormalities
    - ⊟ lung abnormalities
      - ⊞ abnormal alveoli
      - abnormal blood-air barr

ID ▾ equals ▾

● Search all terms ○ Search children of selection ○ Search selection ☐ Case sensitive search

No search performed

ID MP:0005060

Term name: accumulation of giant lysosomes in kidney/renal tubule

Definition

Text: buildup of contents in lysosomes in cells of the kidney tubules

Dbxrefs: cwg:Carroll-Anr

Synonyms: Select a synonym from the list to edit it, or press add to create a new synonym

General DbXrefs: Select a dbxref from the list to edit it, or press add to create a new dbxref

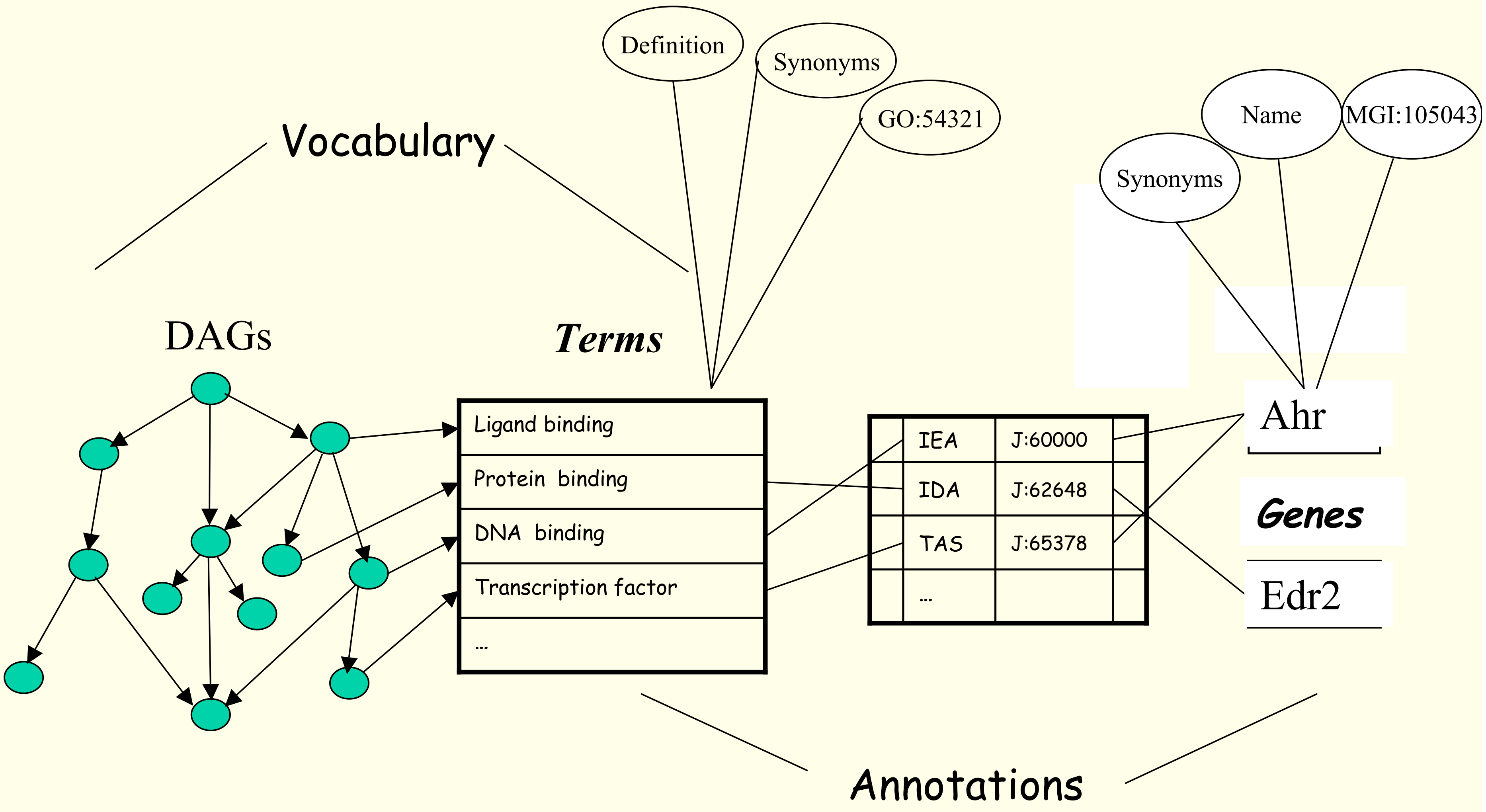
Comment

DAG Viewer

- Phenotype Ontology
  - ⊟ Morphology
    - ⊟ cell abnormalities
      - ⊟ abnormal cell content
        - ⊟ abnormal lysosom
          - accumulation
    - ⊟ urogenital system abnorma
      - ⊟ urogenital system: dysi
        - ⊟ urinary system abr
          - ⊟ kidney abnorm
            - ⊟ abnormal
              - accu

DAG-Edit Tool

Vocabulary Implementation in MGI



Vocabulary

Definition

Synonyms

GO:54321

Synonyms

Name

MGI:105043

DAGs

Terms

Ligand binding
Protein binding
DNA binding
Transcription factor
…

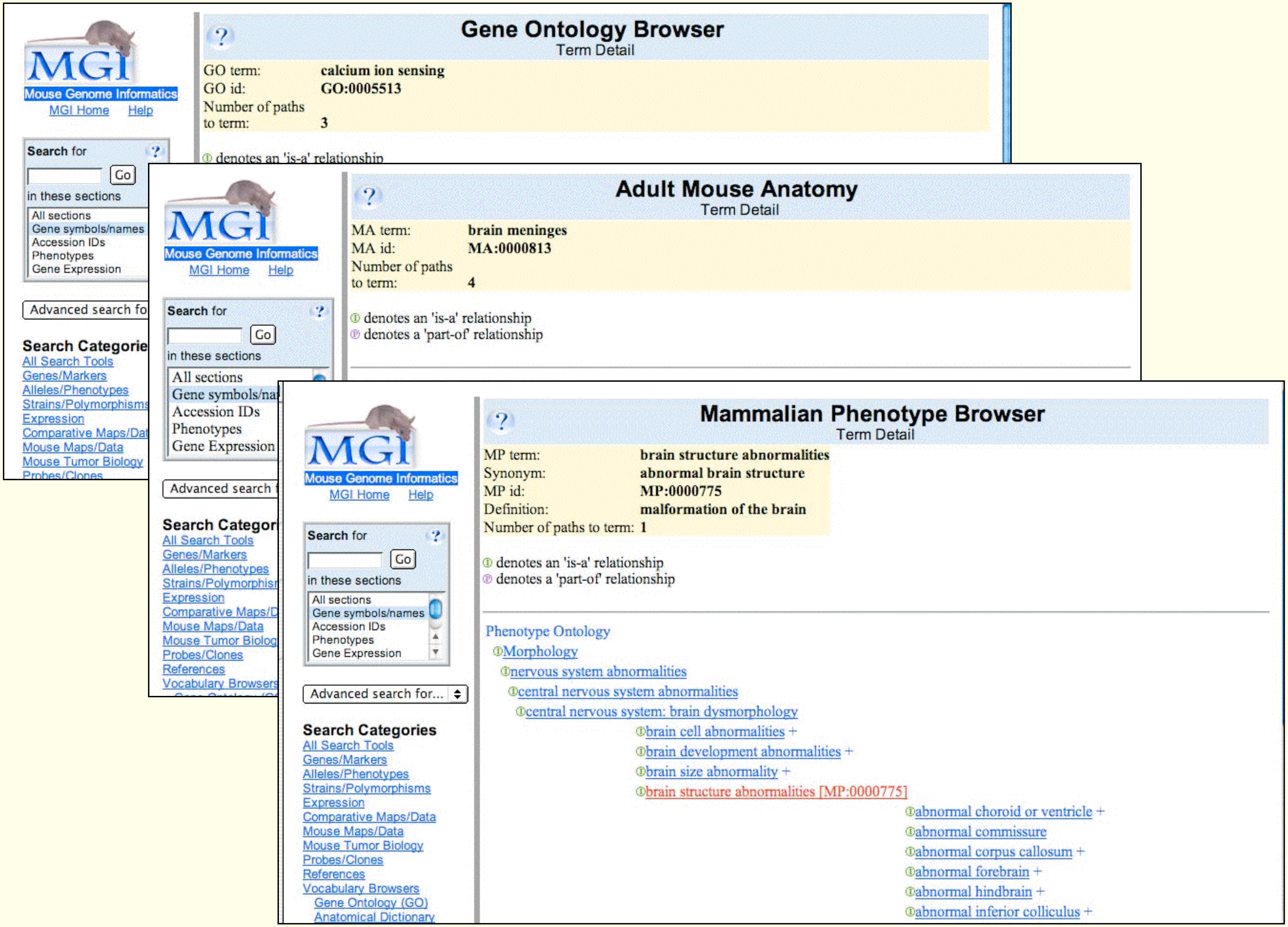
	IEA	J:60000	
	IDA	J:62648	
	TAS	J:65378	
	…		

Ahr

Genes

Edr2

Annotations



Gene Ontology Browser

Term Detail

GO term: **calcium ion sensing**
GO id: **GO:0005513**
Number of paths to term: **3**

Ⓘ denotes an 'is-a' relationship

Adult Mouse Anatomy

Term Detail

MA term: **brain meninges**
MA id: **MA:0000813**
Number of paths to term: **4**

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

Mammalian Phenotype Browser

Term Detail

MP term: **brain structure abnormalities**
Synonym: **abnormal brain structure**
MP id: **MP:0000775**
Definition: **malformation of the brain**
Number of paths to term: **1**

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

Phenotype Ontology
- ⒾMorphology
  - Ⓘnervous system abnormalities
    - Ⓘcentral nervous system abnormalities
      - Ⓘcentral nervous system: brain dysmorphology
        - Ⓘbrain cell abnormalities +
        - Ⓘbrain development abnormalities +
        - Ⓘbrain size abnormality +
        - Ⓘbrain structure abnormalities [MP:0000775]
          - Ⓘabnormal choroid or ventricle +
          - Ⓘabnormal commissure
          - Ⓘabnormal corpus callosum +
          - Ⓘabnormal forebrain +
          - Ⓘabnormal hindbrain +
          - Ⓘabnormal inferior colliculus +

People

Summary

- Ontologies support semantic integration for functional genomics and promote broader access to knowledge

- The Gene Ontology project precipitated a generalized implementation for ontologies for molecular biology

- Bio-ontologies and other annotation standards facilitate development of logic inference systems

Acknowledgments - MGI

<u>MGI Ontologies</u>

Martin Ringwald -Anatomies
Janan Eppig - Phenotype
Carol Bult - Sequence
Joel Richardson – Ontology Theory
Jim Kadin – Generic Ontology Tools

Lori Corbani
Josh Winslow
Jon Beal
Richard Baldarelli

Lois Maltais
Rebecca Corey

<u>Gene Ontologies</u>

David Hill
Harold Drabkin
Mary Dolan
Li Ni

<u>Anatomies</u>

Collaboration with Edinburgh University
    Jonathan Bard
    David Duncan
Terry Hayamizu

<u>Mammalian Phenotype</u>

Cindy Smith
Carroll Goldsmith

Acknowledgments - GO

EBI/GO & FlyBase
Michael Ashbruner
Midori Harris
Jane Lomax
Amelia Ireland
Rebecca Folgar
Jennifer Clark

Berkeley-BDGP
Suzanna Lewis
John Richter
Chris Mungall
Karen Eilbeck

Stanford-SGD
Mike Cherry
Karen Christie
Eurie Hong

Gramene
Lincoln Stein
Pankaj Jaiswal
Doreen Ware

Jackson Lab –MGI
David Hill
Harold Drabkin
Martin Ringwald
Joel Richardson
Li Ni
Mary Dolan

EBI-SWISS-PROT
Rolf Apweiler
Evelyn Cameron
Daniel Barrell

CalTech-WormBase
Paul Sternberg
Kimberly Van Auken

UChicago-DictyBase
Rex Chisholm
Pascale Gaudet

UWisc-RGD
Simon Twillinger
Victoria

Carnegie Institute-TAIR
Sue Rhee
Chandra Theesfeld
Tanya Beradini

Sanger Center-PathogenGenomes
Matt Berriman
Val Wood

TIGR-Microbial Genomes,Arabidopsis
Michelle Gwynn
Linda Hannick

Gene Ontology Consortium

www.geneontology.org

Open Biological Ontologies

obo.sf.net

Mouse Genome Informatics

www.informatics.jax.org

The GO Consortium including OBO is supported by the National Genome Research Institute (NHGRI) and by the European Union RTD Programme

The Mouse Genome Informatics program is support by NIH/NHGRI (MGD), NIH/NICHD(GXD) and DOE(MGS)

The Jackson Laboratory