

# GS01 0163

## Analysis of Microarray Data

Keith Baggerly and Kevin Coombes  
Section of Bioinformatics

Department of Biostatistics and Applied Mathematics  
UT M. D. Anderson Cancer Center

[kabagg@mdanderson.org](mailto:kabagg@mdanderson.org)

[kcoombes@mdanderson.org](mailto:kcoombes@mdanderson.org)

19 October 2004

# Lecture 14: Biology: Functions and Networks

- GeneOntology
- GoMiner
- Regulatory networks and metabolic pathways

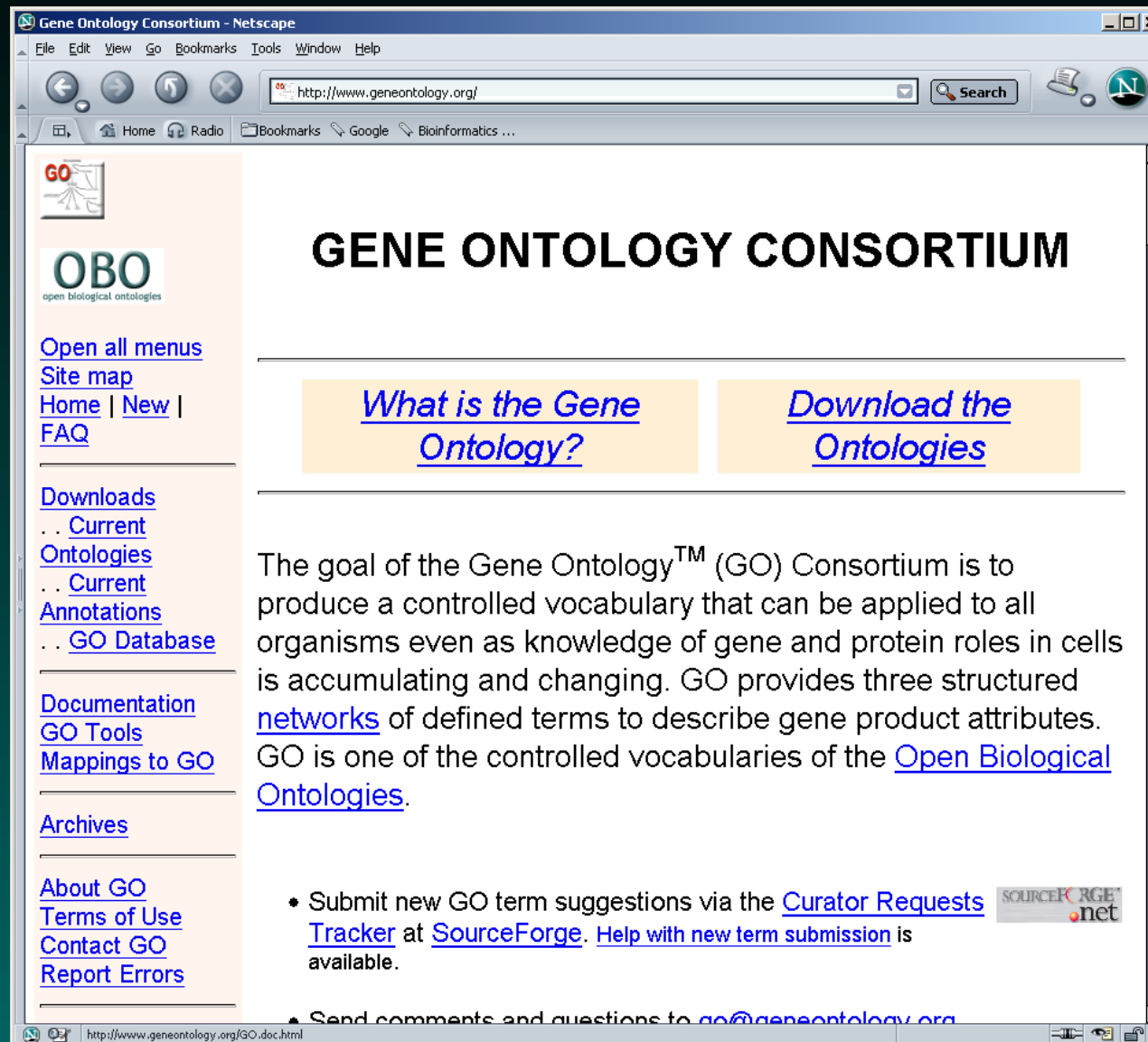
# GeneOntology

GeneOntology (GO) uses controlled vocabularies to create a directed acyclic graph (DAG; a generalized tree) that describes the kinds of functions or properties that a gene might have.

GeneOntology consists fundamentally of two pieces:

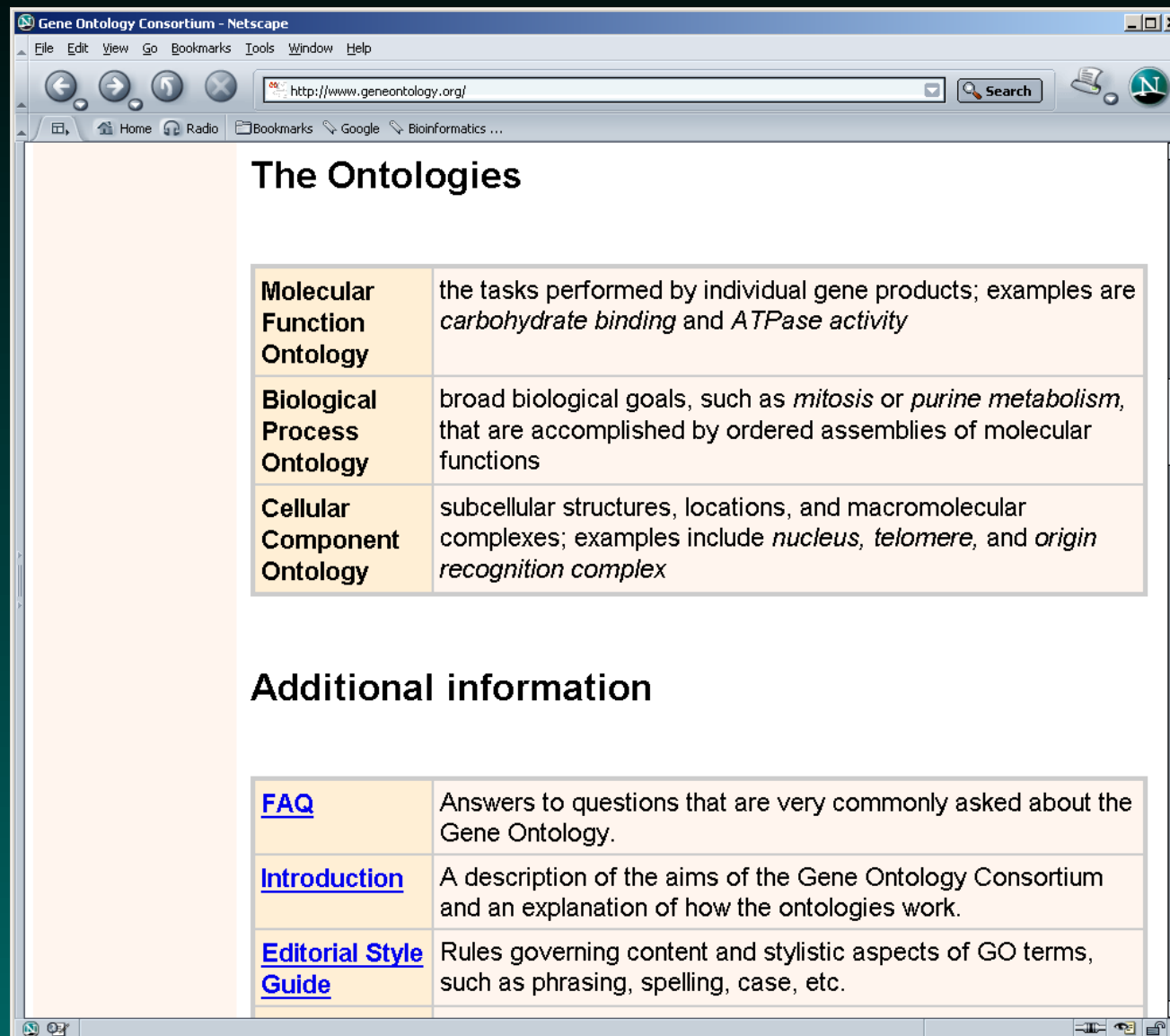
1. The DAG that describes functions and relations between them
2. Annotations that describe which genes actually have which functions

# http://www.geneontology.org



The screenshot shows a Netscape browser window displaying the Gene Ontology Consortium website. The browser's address bar shows the URL <http://www.geneontology.org/>. The website features a navigation menu on the left with links for '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', and 'Report Errors'. The main content area is titled 'GENE ONTOLOGY CONSORTIUM' and includes two prominent yellow buttons: 'What is the Gene Ontology?' and 'Download the Ontologies'. Below these buttons, a paragraph explains the consortium's goal: to produce a controlled vocabulary for gene and protein roles. It mentions 'networks of defined terms' and 'Open Biological Ontologies'. A bulleted list at the bottom provides instructions on how to submit new GO term suggestions via the 'Curator Requests Tracker' at SourceForge, with a 'Help with new term submission' link. A 'SOURCEFORGE.net' logo is visible next to this text. The browser's status bar at the bottom indicates the current page is <http://www.geneontology.org/GO.doc.html>.

# GeneOntology: The top level



The screenshot shows a Netscape browser window displaying the Gene Ontology Consortium website. The address bar shows the URL <http://www.geneontology.org/>. The page content is as follows:

## The Ontologies

<b>Molecular Function Ontology</b>	the tasks performed by individual gene products; examples are <i>carbohydrate binding</i> and <i>ATPase activity</i>
<b>Biological Process Ontology</b>	broad biological goals, such as <i>mitosis</i> or <i>purine metabolism</i> , that are accomplished by ordered assemblies of molecular functions
<b>Cellular Component Ontology</b>	subcellular structures, locations, and macromolecular complexes; examples include <i>nucleus</i> , <i>telomere</i> , and <i>origin recognition complex</i>

## Additional information

<a href="#">FAQ</a>	Answers to questions that are very commonly asked about the Gene Ontology.
<a href="#">Introduction</a>	A description of the aims of the Gene Ontology Consortium and an explanation of how the ontologies work.
<a href="#">Editorial Style Guide</a>	Rules governing content and stylistic aspects of GO terms, such as phrasing, spelling, case, etc.

## GeneOntology annotations in LocusLink

You can find the GeneOntology annotations for individual genes in LocusLink. For genes with known functions, the LocusLink page will contain a section titled “GeneOntology”, which contains a list of the known functions for that gene.

Every GO annotation asserts that a specific gene has a specific function. As part of the design of GO, each assertion is itself annotated to explain the kinds of evidence the assertion is based on, as well as the organization or individual that supplied the annotation.

# GO annotations of the androgen receptor

**LocusLink Report - Netscape**

File Edit View Go Bookmarks Tools Window Help

http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=367

Home Radio Bookmarks Google Bioinformatics ...

**LocusLink Home**

AR Index:  
Top of Page  
Nomenclature  
Overview  
Function  
Relationships  
Map  
RefSeq  
Related  
Seqs  
Links

**LocusLink:**  
Collaborators  
Download  
FAQ  
Help  
Statistics

**RefSeq:**  
About  
Download  
FAQ  
Statistics

**Gene Ontology™:**

Term	Evidence	Source	Pub
• <a href="#">androgen binding</a>	NAS	<a href="#">GOA</a>	pm
• <a href="#">androgen receptor activity</a>	IEA	<a href="#">GOA</a>	pm
• <a href="#">androgen receptor activity</a>	NAS	<a href="#">GOA</a>	pm
• <a href="#">cell proliferation</a>	NAS	<a href="#">GOA</a>	pm
• <a href="#">cell-cell signaling</a>	TAS	<a href="#">GOA</a>	pm
• <a href="#">nucleus</a>	NR	<a href="#">GOA</a>	
• <a href="#">nucleus</a>	IEA	<a href="#">GOA</a>	
• <a href="#">prostate gland development</a>	NAS	<a href="#">GOA</a>	pm
• <a href="#">protein dimerization activity</a>	NAS	<a href="#">GOA</a>	pm
• <a href="#">receptor activity</a>	IEA	<a href="#">GOA</a>	
• <a href="#">regulation of transcription, DNA-dependent</a>	IEA	<a href="#">GOA</a>	
• <a href="#">sex differentiation</a>	NAS	<a href="#">GOA</a>	pm
• <a href="#">signal transduction</a>	TAS	<a href="#">GOA</a>	pm
• <a href="#">steroid binding</a>	IEA	<a href="#">GOA</a>	
• <a href="#">transcription factor activity</a>	IEA	<a href="#">GOA</a>	

# http://www.ebi.ac.uk/GOA/

Gene Ontology Annotation @ EBI - Netscape

http://www.ebi.ac.uk/GOA/

EMBL-EBI European Bioinformatics Institute

GOA DATABASE

GOA @EBI

GOA is a project run by the European Bioinformatics Institute that aims to provide assignments of gene products to the Gene Ontology (GO) resource.

The goal of the Gene Ontology Consortium is to produce a dynamic controlled vocabulary that can be applied to all organisms, even while knowledge of gene and protein roles in cells is still accumulating and changing. In the GOA project, this vocabulary will be applied to a non-redundant set of proteins described in the UniProt Resource (Swiss-Prot/TrEMBL/PIR-PSD) and Ensembl databases that collectively provide complete proteomes for Homo sapiens and other organisms.

In the first stage of this project, GO assignments have been applied to a data set representing the human proteome by a combination of electronic mappings and manual curation. Subsequently, GO assignments for all complete and incomplete proteomes that exist in UniProt have been

Access GOA

Search GOA via SRS

Go

GO

The EBI's Gene Ontology Consortium pages. GO is an international consortium of scientists with the editorial office based at the EBI.

UniProt/ Swiss-Prot

- GOA Home
- Introduction
- Contents of Current Release
- Data Searching and Retrieval
- Forthcoming Changes
- GOA News
- Feedback



# GO annotations of the androgen receptor

**LocusLink Report - Netscape**

File Edit View Go Bookmarks Tools Window Help

http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=367

Home Radio Bookmarks Google Bioinformatics ...

**LocusLink Home**

AR Index:  
Top of Page  
Nomenclature  
Overview  
Function  
Relationships  
Map  
RefSeq  
Related  
Seqs  
Links

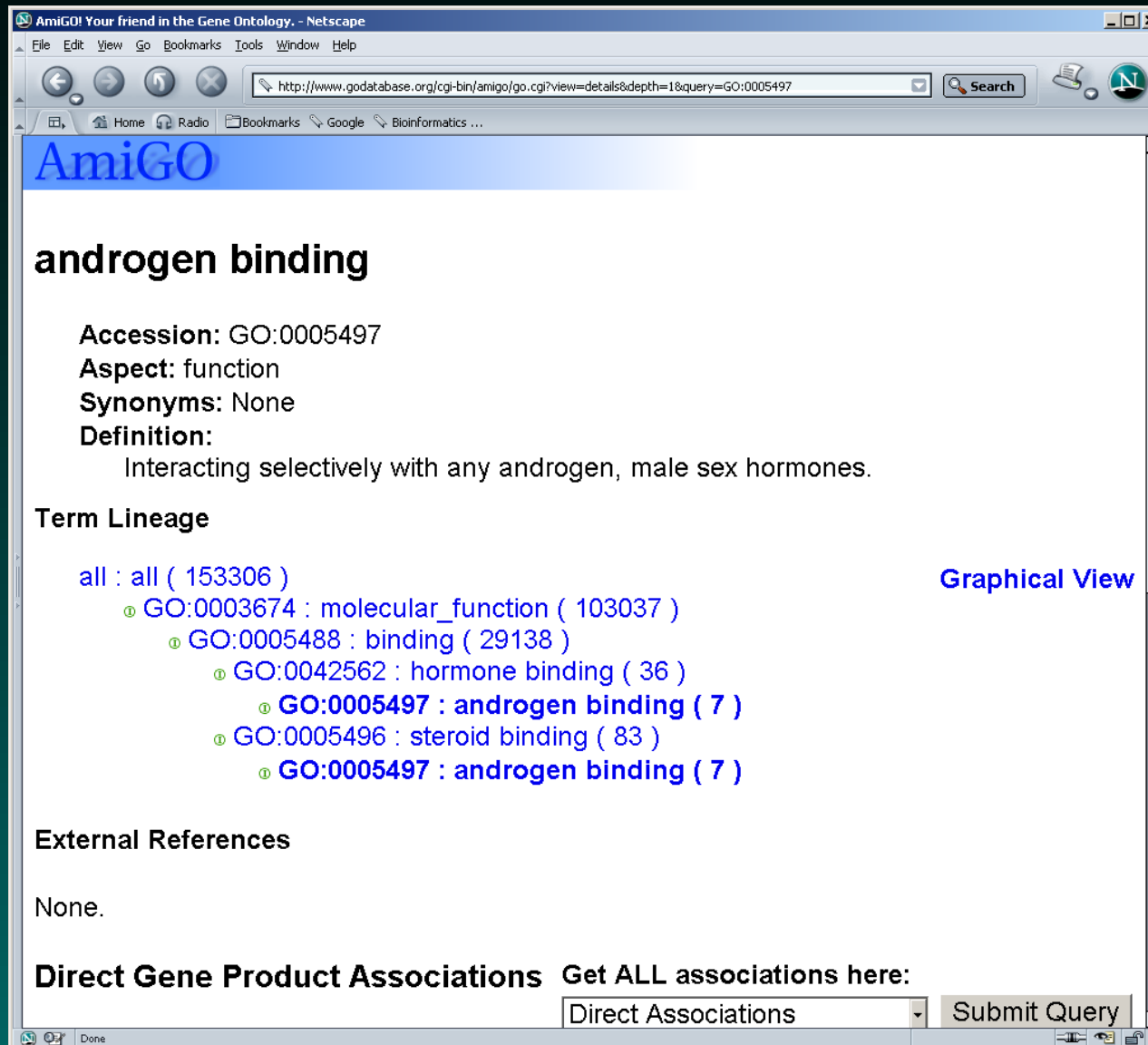
**LocusLink:**  
Collaborators  
Download  
FAQ  
Help  
Statistics

**RefSeq:**  
About  
Download  
FAQ  
Statistics

**Gene Ontology™:**

Term	Evidence	Source	Pub
• <a href="#">androgen binding</a>	NAS	<a href="#">GOA</a>	pm
• <a href="#">androgen receptor activity</a>	IEA	<a href="#">GOA</a>	pm
• <a href="#">androgen receptor activity</a>	NAS	<a href="#">GOA</a>	pm
• <a href="#">cell proliferation</a>	NAS	<a href="#">GOA</a>	pm
• <a href="#">cell-cell signaling</a>	TAS	<a href="#">GOA</a>	pm
• <a href="#">nucleus</a>	NR	<a href="#">GOA</a>	
• <a href="#">nucleus</a>	IEA	<a href="#">GOA</a>	
• <a href="#">prostate gland development</a>	NAS	<a href="#">GOA</a>	pm
• <a href="#">protein dimerization activity</a>	NAS	<a href="#">GOA</a>	pm
• <a href="#">receptor activity</a>	IEA	<a href="#">GOA</a>	
• <a href="#">regulation of transcription, DNA-dependent</a>	IEA	<a href="#">GOA</a>	
• <a href="#">sex differentiation</a>	NAS	<a href="#">GOA</a>	pm
• <a href="#">signal transduction</a>	TAS	<a href="#">GOA</a>	pm
• <a href="#">steroid binding</a>	IEA	<a href="#">GOA</a>	
• <a href="#">transcription factor activity</a>	IEA	<a href="#">GOA</a>	

# GO browsing



AmiGO! Your friend in the Gene Ontology. - Netscape

http://www.godatabase.org/cgi-bin/amigo/go.cgi?view=details&depth=1&query=GO:0005497

## AmiGO

### androgen binding

**Accession:** GO:0005497  
**Aspect:** function  
**Synonyms:** None  
**Definition:**  
Interacting selectively with any androgen, male sex hormones.

#### Term Lineage

all : all ( 153306 ) [Graphical View](#)

- GO:0003674 : molecular\_function ( 103037 )
  - GO:0005488 : binding ( 29138 )
    - GO:0042562 : hormone binding ( 36 )
      - GO:0005497 : androgen binding ( 7 )**
      - GO:0005496 : steroid binding ( 83 )
        - GO:0005497 : androgen binding ( 7 )

#### External References

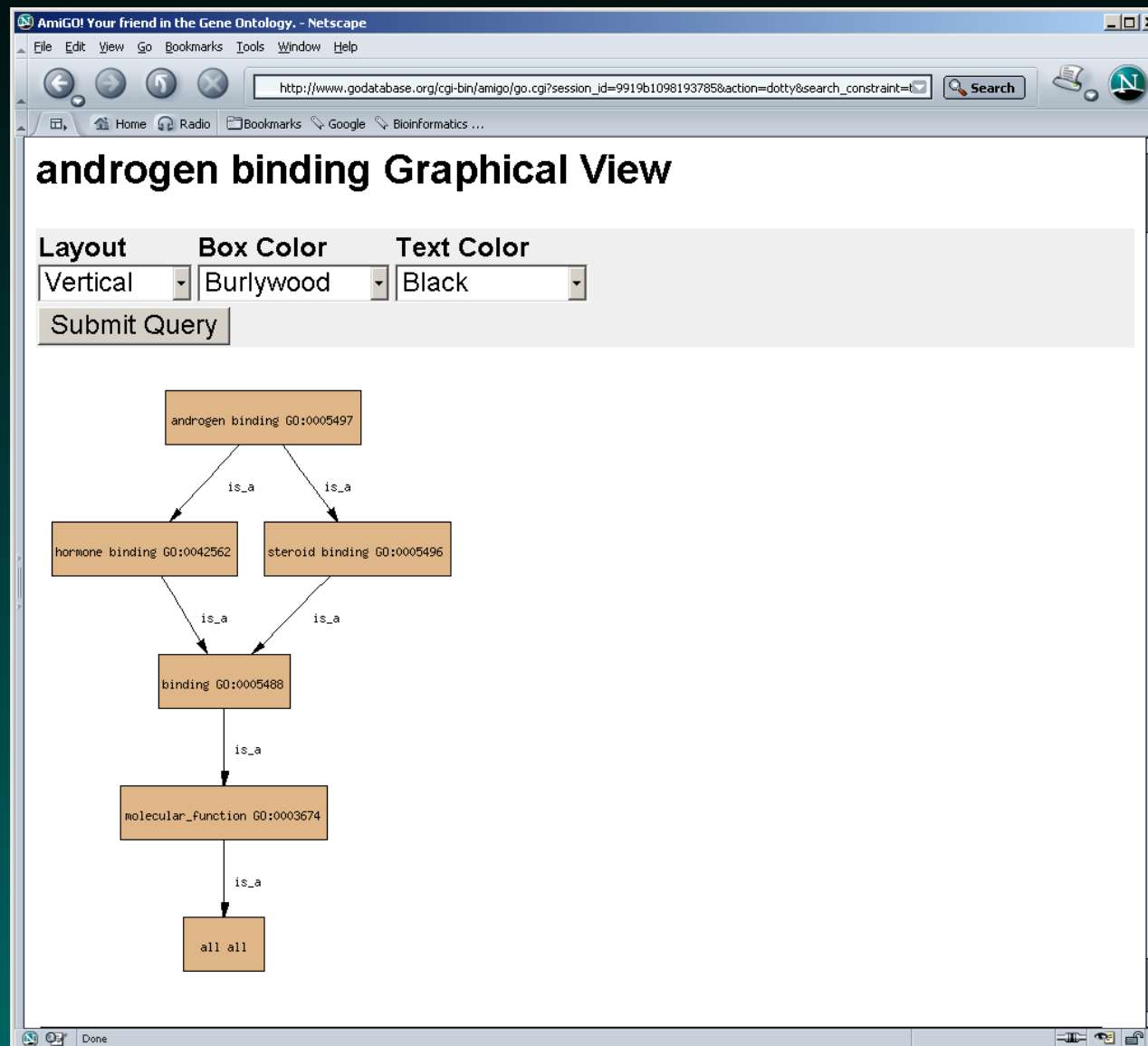
None.

#### Direct Gene Product Associations

Get ALL associations here:

Direct Associations

# GO browsing



## Edges are relationships

Edges in the DAG represent two kinds of relationships:

**is\_a** : Used when the child node is a special case of the parent node. For example, `hormone binding` **is\_a** kind of `binding`.

**part\_of** : Used when the child node is a component of the parent node. For example, a `membrane` is **part\_of** a `cell`

Genes may be annotated into different levels of the hierarchy, depending on how detailed the evidence is. In general, a gene not only has the function corresponding to the node with direct annotation, but also has every property at parent nodes up through the hierarchy.

# GO annotations of the androgen receptor

**LocusLink Report - Netscape**

File Edit View Go Bookmarks Tools Window Help

http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=367

Home Radio Bookmarks Google Bioinformatics ...

**LocusLink Home**

AR Index:  
Top of Page  
Nomenclature  
Overview  
Function  
Relationships  
Map  
RefSeq  
Related  
Seqs  
Links

**LocusLink:**  
Collaborators  
Download  
FAQ  
Help  
Statistics

**RefSeq:**  
About  
Download  
FAQ  
Statistics

**Gene Ontology™:**

Term	Evidence	Source	Pub
• <a href="#">androgen binding</a>	NAS	<a href="#">GOA</a>	pm
• <a href="#">androgen receptor activity</a>	IEA	<a href="#">GOA</a>	pm
• <a href="#">androgen receptor activity</a>	NAS	<a href="#">GOA</a>	pm
• <a href="#">cell proliferation</a>	NAS	<a href="#">GOA</a>	pm
• <a href="#">cell-cell signaling</a>	TAS	<a href="#">GOA</a>	pm
• <a href="#">nucleus</a>	NR	<a href="#">GOA</a>	
• <a href="#">nucleus</a>	IEA	<a href="#">GOA</a>	
• <a href="#">prostate gland development</a>	NAS	<a href="#">GOA</a>	pm
• <a href="#">protein dimerization activity</a>	NAS	<a href="#">GOA</a>	pm
• <a href="#">receptor activity</a>	IEA	<a href="#">GOA</a>	
• <a href="#">regulation of transcription, DNA-dependent</a>	IEA	<a href="#">GOA</a>	
• <a href="#">sex differentiation</a>	NAS	<a href="#">GOA</a>	pm
• <a href="#">signal transduction</a>	TAS	<a href="#">GOA</a>	pm
• <a href="#">steroid binding</a>	IEA	<a href="#">GOA</a>	
• <a href="#">transcription factor activity</a>	IEA	<a href="#">GOA</a>	

## GeneOntology: Evidence Codes

**IDA** : **inferred from direct assay**; indicates that the annotation is based on a paper describing an experiment that directly tested this function for this gene

**TAS** : **traceable author statement**; based on a review article or textbook that includes references to the original experiments

**IMP** : **inferred from mutant phenotype**; based on experiments involving mutations, knockouts, antisense, etc.

**IPI** : **inferred from physical interaction**; based on assays (like co-immunoprecipitation) that demonstrate physical interactions between the gene in question and other gene products

**IGI** : **inferred from genetic interaction**; based on experiments (such as synthetic lethals, suppressors, functional complementation) that show a genetic interaction between the gene in question and another gene

**ISS** : **inferred from sequence or structure similarity**; based on BLAST results that have been reviewed for accuracy by a curator

**IEP** : **inferred from expression pattern**; based on Northern, Westerns, or microarray experiments that reveal information about the timing or location of expression

**NAS** : **non-traceable author statement**; statements in papers (abstract, introduction, discussion) that a curator cannot trace to another publication

**IEA** : inferred from electronic annotation; based on sequence similarity searches or database records that have not been reviewed by a curator

**IC** : inferred by curator; even though no direct evidence is available, the property can reasonably be inferred by the curator. For example, it is reasonable to infer from direct evidence of “transcription factor activity” that the gene product is found in the nucleus

**ND** : no biological data available; only used for annotations to “unknown”

**NR** : not recorded; used only for annotations created before curators started adding evidence codes



## Quality of evidence

The evidence codes fall into a rough hierarchy indicating how strongly the annotation of function should be believed.

1. IDA, TAS
2. IMP, IPI, IGI
3. ISS, IEP
4. NAS
5. IEA
6. IC

# GoMiner



GoMiner Home Page - Netscape

File Edit View Go Bookmarks Tools Window Help

http://discover.nci.nih.gov/gominer/ Search

Home Radio Bookmarks Google Bioinformatics ...

**GoMINER** Build:138

Genomics and Bioinformatics Group  
LMP, NCI, NIH

Medical Informatics & Bioimaging Lab,  
BME GA Tech/  
Emory University

Home Getting Started Requirements Installation Command Line Database FAQ Citing Mirrors Credits

GoMiner is a tool for biological interpretation of 'omic' data – including data from gene expression microarrays. Omic experiments often generate lists of dozens or hundreds of genes that differ in expression between samples, raising the question

*What does it all mean biologically?*

To answer this question, GoMiner leverages the [Gene Ontology \(GO\)](#) to identify the biological processes, functions and components represented in these lists. Instead of analyzing microarray results with a gene-by-gene approach, GoMiner classifies the genes into biologically coherent categories and assesses these categories. The insights gained through GoMiner can generate hypotheses to guide additional research.

**To get started using GoMiner**

- Read the [Instructions](#), and verify that your environment satisfies the system requirements
- Download the GoMiner program file, [gominer.jar](#)
- Read the [Quick Start](#) and try out the [sample files](#)

http://discover.nci.nih.gov/gominer/index.jsp

# http://discover.nci.nih.gov/gominer



The screenshot shows a Netscape browser window titled "GoMiner Installation - Netscape". The address bar contains the URL "http://discover.nci.nih.gov/gominer/installation.jsp". The page features a blue header with the "GoMINER" logo (Build:138) and the text "Genomics and Bioinformatics Group, LMP, NCI, NIH" and "Medical Informatics & Bioimaging Lab, BME GA Tech/ Emory University". A navigation menu includes links for Home, Getting Started, Requirements, Installation (highlighted in red), Command Line, Database, FAQ, Citing, Mirrors, and Credits. The main content area is titled "Core Application" and lists several steps for installation. Below this is a section for "Install Helper Applications" with a sub-section for "Adobe SVG Viewer".

**GoMINER** Build:138

Genomics and Bioinformatics Group  
LMP, NCI, NIH

Medical Informatics & Bioimaging Lab,  
BME GA Tech/  
Emory University

Home Getting Started Requirements **Installation** Command Line Database FAQ Citing Mirrors Credits

## Core Application

- Verify that your machine satisfies the [system requirements](#)
- Download the GoMiner program file, [gominer.jar](#)
- Install the optional components (if any) listed below you want to use.
- Either double-click the jar file, or run "java -jar gominer.jar" from the command line
- For large input files (>10,000 genes) run "java -Xms256M -Xmx256M -jar gominer.jar"  
This will allocate more memory to run the application.
- For optimal performance, [install a local copy](#) of the server and database.

## Install Helper Applications

The first two helper applications are web browser components, and provide additional visualization features.

### Adobe SVG Viewer

# GoMiner: Getting Started

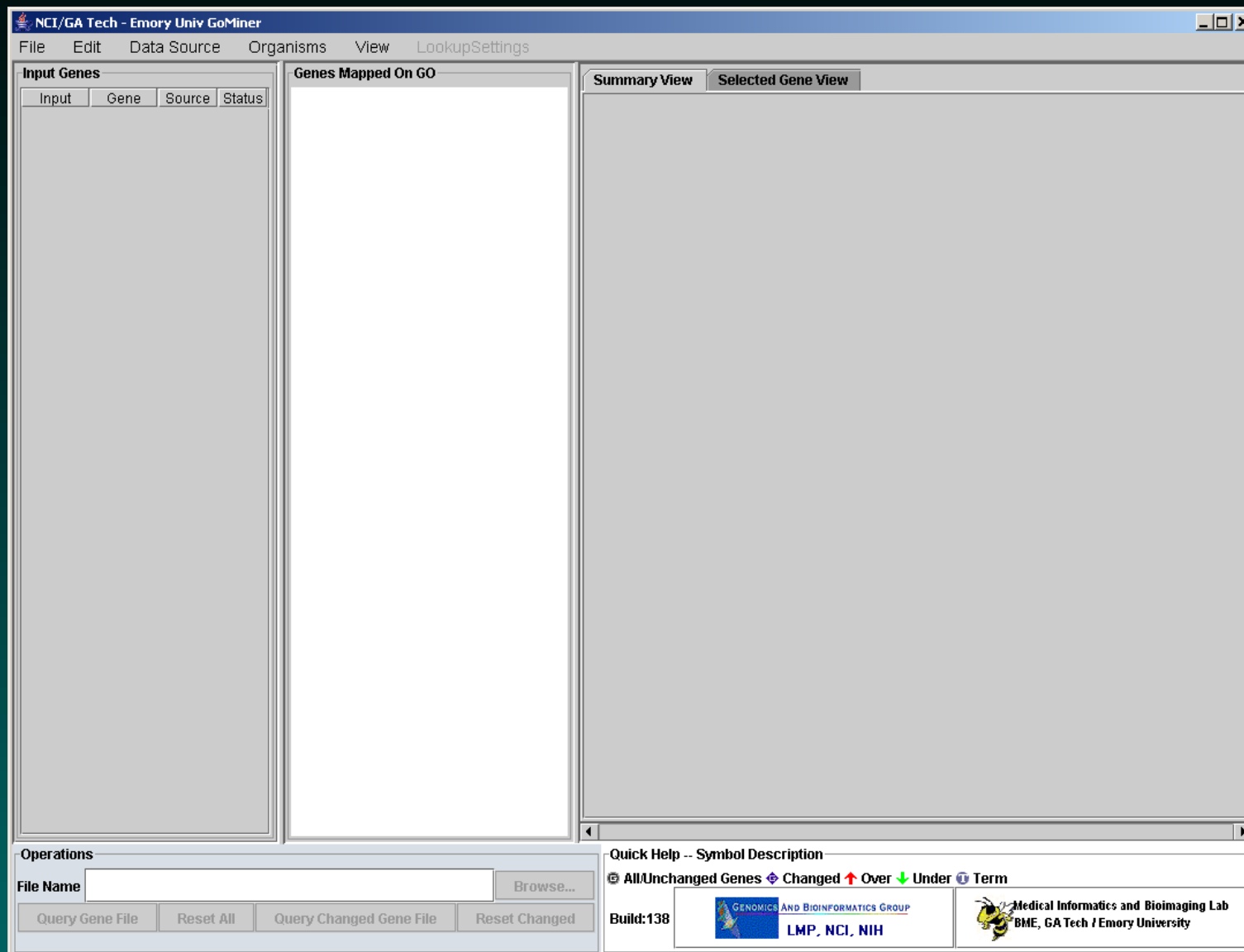
You need a machine with

- Java 1.3 or higher
- Windows 98 or higher, Mac OS X or higher, Solaris, Linux, or FreeBSD
- High-speed internet access

Download the GoMiner Java code, install it, and double-click on it to start the program.

Then go to “File” – > “Load GO Terms” and click “OK”.

# GoMiner Start



# GoMiner: GO terms loaded

The screenshot displays the GoMiner web application interface. The window title is "NCI/GA Tech - Emory Univ GoMiner". The menu bar includes "File", "Edit", "Data Source", "Organisms", "View", and "LookupSettings".

The main interface is divided into three panels:

- Input Genes:** A table with columns "Input", "Gene", "Source", and "Status". It is currently empty.
- Genes Mapped On GO:** A list of GO terms with expand/collapse icons and a "top" link. The terms are:
  - biological\_process
  - cellular\_component
  - molecular\_function
  - obsolete\_component
  - obsolete\_function
  - obsolete\_process
- Summary View / Selected Gene View:** A large empty area for displaying results.

At the bottom, there is an "Operations" section with a "File Name" input field and a "Browse..." button. Below this are four buttons: "Query Gene File", "Reset All", "Query Changed Gene File", and "Reset Changed".

On the right side of the bottom panel, there is a "Quick Help -- Symbol Description" section with a radio button for "All/Unchanged Genes" and a "Term" button. Below this, it says "Build:138" and features logos for the "GENOMICS AND BIOINFORMATICS GROUP" (LMP, NCI, NIH) and the "Medical Informatics and Bioimaging Lab" (BME, GA Tech / Emory University).

# GoMiner as GO browser

NCI/GA Tech - Emory Univ GoMiner

File Edit Data Source Organisms View LookupSettings

**Input Genes**

Input	Gene	Source	Status

**Genes Mapped On GO**

- top
- biological\_process
  - behavior
  - biological\_process unknown
  - cellular\_process
  - development
  - physiological process
    - cellular physiological process
    - coagulation
    - death
      - aging
      - cell death
      - tissue death
    - extracellular structure organization ar
    - germination
    - homeostasis
    - metabolism
    - organismal physiological process
    - pathogenesis
    - regulation of physiological process
    - response to stimulus
      - response to biotic stimulus
      - response to circadian rhythm
      - response to endogenous stimulus
      - response to external stimulus
      - response to stress
    - secretion
    - seed dormancy
  - regulation of biological process
  - viral life cycle
- cellular\_component
- molecular\_function
- obsolete\_component
- obsolete\_function
- obsolete\_process

**Operations**

File Name  Browse...

Query Gene File Reset All Query Changed Gene File Reset Changed

**Quick Help -- Symbol Description**

All/Unchanged Genes Changed Over Under Term

Build:138

GENOMICS AND BIOINFORMATICS GROUP  
LMP, NCI, NIH

Medical Informatics and Bioimaging Lab  
BME, GA Tech / Emory University

## Getting array data into GoMiner

1. Go to “Data Source” and select “UniProt (Hs)” to restrict to human gene annotations
2. Need a file containing a list of all genes in the experiment, one HUGO symbol per line. Use the “Browse” button, and then click “Query Gene File” to load this information. This may take some time...
3. Need a file containing a list of genes that changed. Can be one HUGO symbol per line. Optionally, you can include a second column with 1 (overexpressed) or -1 (underexpressed). Use “Browse” and “Query Changed Gene File” to load this data.

Note: GeneLink or Source can convert from various gene ids to HUGO symbols.



# GoMiner with array gene list loaded

**Input Genes**

Input	Gene	Source	Status
YWHAE	143E_...	UniProt	⊕
SFN	143S_...	UniProt	⊕
PPP2R...	2A5A_...	UniProt	⊕
PPP2R...	2A5B_...	UniProt	⊕
PPP2R...	2A5D_...	UniProt	⊕
PPP2R...	2A5E_...	UniProt	⊕
PPP2R...	2A5G_...	UniProt	⊕
PPP2R...	2AAA_...	UniProt	⊕
PPP2R...	2AAB_...	UniProt	⊕
PPP2R...	2ABA_...	UniProt	⊕
PPP2R...	2ABB_...	UniProt	⊕
HLA-DMA	2DMA_...	UniProt	⊕
HLA-D...	2DMB_...	UniProt	⊕
HLA-DOA	2DOA_...	UniProt	⊕
HLA-DRA	2DRA_...	UniProt	⊕
BH3BP2	3BP2_...	UniProt	⊕
SLC3A2	4F2_H_...	UniProt	⊕
A2M	A2MG_...	UniProt	⊕
ACTN1	AAC1_...	UniProt	⊕
PRKAB1	AAKB_...	UniProt	⊕
PRKAG1	AAKG_...	UniProt	⊕
ATBF1	ABF1_...	UniProt	⊕
ABL1	ABL1_...	UniProt	⊕
ABL2	ABL2_...	UniProt	⊕
ABR	ABR_H_...	UniProt	⊕
ACY1	ACY1_...	UniProt	⊕
ADAM17	AD17_...	UniProt	⊕
ADA	ADA_H_...	UniProt	⊕
ADD3	ADDG_...	UniProt	⊕
ADH6	ADH6_...	UniProt	⊕
ADK	ADK_H_...	UniProt	⊕
AOX1	ADO_H_...	UniProt	⊕
ADSS	ADSS	⊕	
SLC25A5	ADT2_...	UniProt	⊕
MLLT2	AF4_H_...	UniProt	⊕
GLA	AGAL_...	UniProt	⊕
ANGPT1	AGP1_...	UniProt	⊕
ANGPT2	AGP2_...	UniProt	⊕
AHR	AHR_H_...	UniProt	⊕

**Genes Mapped On GO**

- top (1299)
  - biological\_process (1245)
    - behavior (8)
    - biological\_process\_unknown (27)
    - cellular\_process (847)
    - development (220)
    - physiological process (1139)
      - IGF2\_HUMAN (IGF2) - (UniProt)
      - IGFA\_HUMAN (IGF1) - (UniProt)
      - O43200 (TSHR) - (UniProt)
      - PGH1\_HUMAN (PTGS1) - (UniProt)
      - PGH2\_HUMAN (PTGS2) - (UniProt)
      - REL1\_HUMAN (RLN1) - (UniProt)
      - RLF\_HUMAN (RLF) - (UniProt)
    - cellular physiological process (568)
    - coagulation (16)
    - death (123)
      - aging (2)
        - cell death (122)
          - cell aging (1)
          - cytolysis (3)
          - programmed cell death (120)
            - apoptosis (120)
              - regulation of programmed cell de.
        - extracellular structure organization and b
        - homeostasis (13)
        - metabolism (823)
        - organismal physiological process (254)
        - pathogenesis (3)
        - regulation of physiological process (239)
        - response to stimulus (359)
        - secretion (2)
        - regulation of biological process (403)
        - viral life cycle (8)
      - cellular\_component (1070)
      - molecular\_function (1215)
        - obsolete\_component
        - obsolete\_function
        - obsolete\_process

**Summary View Selected Gene View**

Category Name	P-Chng	P-Undr	P-Ovr	Tot	Chng	Undr	Ovr	Category ID
ATP-dependent hel...	1.0000	1.0000	1.0000	11	0	0	0	GO:00080...
transcription elong...	1.0000	1.0000	1.0000	1	0	0	0	GO:00080...
protein C-terminus ...	1.0000	1.0000	1.0000	2	0	0	0	GO:00080...
microtubule binding	1.0000	1.0000	1.0000	3	0	0	0	GO:00080...
regulation of heart r...	1.0000	1.0000	1.0000	1	0	0	0	GO:00080...
circulation	1.0000	1.0000	1.0000	9	0	0	0	GO:00080...
beta-catenin binding	1.0000	1.0000	1.0000	1	0	0	0	GO:00080...
chemokine activity	1.0000	1.0000	1.0000	18	0	0	0	GO:00080...
oligopeptide transp...	1.0000	1.0000	1.0000	1	0	0	0	GO:00151...
peptide transporter ...	1.0000	1.0000	1.0000	2	0	0	0	GO:00151...
L-amino acid trans...	1.0000	1.0000	1.0000	1	0	0	0	GO:00151...
acidic amino acid tr...	1.0000	1.0000	1.0000	1	0	0	0	GO:00151...
amino acid transpo...	1.0000	1.0000	1.0000	2	0	0	0	GO:00151...
hexose transporter ...	1.0000	1.0000	1.0000	2	0	0	0	GO:00151...
monosaccharide tr...	1.0000	1.0000	1.0000	2	0	0	0	GO:00151...
carbohydrate trans...	1.0000	1.0000	1.0000	3	0	0	0	GO:00151...
nitric oxide metabol...	1.0000	1.0000	1.0000	4	0	0	0	GO:00462...
sodium ion transpo...	1.0000	1.0000	1.0000	1	0	0	0	GO:00150...
hydrogen ion trans...	1.0000	1.0000	1.0000	3	0	0	0	GO:00150...
monovalent inorga...	1.0000	1.0000	1.0000	3	0	0	0	GO:00150...
ion transporter activ...	1.0000	1.0000	1.0000	4	0	0	0	GO:00150...
protein phosphatas...	1.0000	1.0000	1.0000	3	0	0	0	GO:00150...
thrombin receptor a...	1.0000	1.0000	1.0000	1	0	0	0	GO:00150...
glutathione disulfid...	1.0000	1.0000	1.0000	1	0	0	0	GO:00150...
peptide disulfide ox...	1.0000	1.0000	1.0000	1	0	0	0	GO:00150...
disulfide oxidoredu...	1.0000	1.0000	1.0000	4	0	0	0	GO:00150...
protein transport	1.0000	1.0000	1.0000	26	0	0	0	GO:00150...
Cajal body	1.0000	1.0000	1.0000	1	0	0	0	GO:00150...
coreceptor activity	1.0000	1.0000	1.0000	6	0	0	0	GO:00150...
glucuronosyltransf...	1.0000	1.0000	1.0000	2	0	0	0	GO:00150...
nuclear organizatio...	1.0000	1.0000	1.0000	25	0	0	0	GO:00069...
organelle organizat...	1.0000	1.0000	1.0000	32	0	0	0	GO:00069...
unfolded protein re...	1.0000	1.0000	1.0000	1	0	0	0	GO:00069...
alcohol catabolism	1.0000	1.0000	1.0000	4	0	0	0	GO:00461...
response to unfold...	1.0000	1.0000	1.0000	5	0	0	0	GO:00069...
ER-nuclear signali...	1.0000	1.0000	1.0000	1	0	0	0	GO:00069...
response to lipid hy...	1.0000	1.0000	1.0000	1	0	0	0	GO:00069...
phenol metabolism	1.0000	1.0000	1.0000	2	0	0	0	GO:00189...

**Operations**

File Name: C:\Source\GoMinerExample\total.gene Browse...

**Quick Help -- Symbol Description**

All/Unchanged Genes
  Changed
  Over
  Under
  Term

Build:138

# GoMiner with changed gene list loaded

NCI/GA Tech - Emory Univ GoMiner

File Edit Data Source Organisms View LookupSettings

**Input Genes**

Input	Gene	Source	Status
YWHAE	143E_...	UniProt	⊖
SFN	143S_...	UniProt	⊖
PPP2R...	2A5A_...	UniProt	⊖
PPP2R...	2A5B_...	UniProt	⊖
PPP2R...	2A5D_...	UniProt	⊖
PPP2R...	2A5E_...	UniProt	⊖
PPP2R...	2A5G_...	UniProt	⊖
PPP2R...	2AAA_...	UniProt	⊖
PPP2R...	2AAB_...	UniProt	⊖
PPP2R...	2ABA_...	UniProt	⊖
PPP2R...	2ABB_...	UniProt	⊖
HLA-DMA	2DMA_...	UniProt	⊖
HLA-D...	2DMB_...	UniProt	⊖
HLA-DOA	2DOA_...	UniProt	⊖
HLA-DRA	2DRA_...	UniProt	⬇
BH3BP2	3BP2_...	UniProt	⊖
SLC3A2	4F2_H_...	UniProt	⊖
A2M	A2MG_...	UniProt	⊖
ACTN1	AAC1_...	UniProt	⊖
PRKAB1	AAKB_...	UniProt	⊖
PRKAG1	AAKG_...	UniProt	⊖
ATBF1	ABF1_...	UniProt	⊖
ABL1	ABL1_...	UniProt	⊖
ABL2	ABL2_...	UniProt	⊖
ABR	ABR_H_...	UniProt	⊖
ACY1	ACY1_...	UniProt	⊖
ADAM17	AD17_...	UniProt	⊖
ADA	ADA_H_...	UniProt	⊖
ADD3	ADDG_...	UniProt	⊖
ADH6	ADH6_...	UniProt	⊖
ADK	ADK_H_...	UniProt	⊖
AOX1	ADO_H_...	UniProt	⊖
ADSS	ADSS_...	UniProt	⊖
SLC25A5	ADT2_...	UniProt	⊖
MLLT2	AF4_H_...	UniProt	⊖
GLA	AGAL_...	UniProt	⊖
ANGPT1	AGP1_...	UniProt	⊖
ANGPT2	AGP2_...	UniProt	⊖
AHR	AHR_H_...	UniProt	⬆

**Genes Mapped On GO**

- ⊖ (1299 1.00 p=1.00 1.00 p=1.00 1.00 p=1.00)
  - ⊖ biological\_process (1245 1.03 p=0.17 1.01 p=0.48 1.02 p=0.17)
    - ⊖ biological\_process unknown (27 1.30 p=0.46 0.53 p=0.86 0.88 p=0.1)
    - ⊖ cellular\_process (847 0.99 p=0.58 0.97 p=0.69 0.98 p=0.67)
      - ⊖ development (220 0.96 p=0.62 1.12 p=0.35 1.04 p=0.43)
      - ⊖ physiological\_process (1139 1.08 p=0.04 1.05 p=0.11 1.06 p=0.01)
        - ⊖ cellular\_physiological\_process (568 1.02 p=0.48 0.99 p=0.57 1.0)
        - ⊖ coagulation (16 1.10 p=0.61 0.90 p=0.69 0.99 p=0.62)
        - ⊖ death (123 1.28 p=0.26 1.17 p=0.34 1.22 p=0.20)
          - ⊖ cell\_death (122 1.29 p=0.25 1.18 p=0.33 1.23 p=0.19)
            - ⊖ cytolysis (3 0.00 p=1.00 4.81 p=0.19 2.64 p=0.33)
            - ⊖ programmed\_cell\_death (120 1.32 p=0.24 1.08 p=0.45 1.1)
            - ⊖ apoptosis (120 1.32 p=0.24 1.08 p=0.45 1.19 p=0.24)
              - ⊖ regulation\_of\_programmed\_cell\_death (77 1.14 p=0.45 0.9)
      - ⊖ homeostasis (13 0.00 p=1.00 5.55 p=0.00 3.05 p=0.02)
      - ⊖ metabolism (823 0.90 p=0.91 1.14 p=0.04 1.03 p=0.33)
      - ⊖ organismal\_physiological\_process (254 1.87 p=0.00 0.91 p=0.71)
      - ⊖ regulation\_of\_physiological\_process (239 1.10 p=0.38 1.39 p=0.05 1)
      - ⊖ response\_to\_stimulus (359 1.47 p=0.01 1.09 p=0.34 1.26 p=0.02)
    - ⊖ regulation\_of\_biological\_process (403 1.18 p=0.18 1.25 p=0.06 1.22 p=0.06)
    - ⊖ viral\_life\_cycle (8 2.19 p=0.38 1.80 p=0.44 1.98 p=0.27)
  - ⊖ cellular\_component (1070 0.97 p=0.78 0.97 p=0.78 0.97 p=0.84)
  - ⊖ molecular\_function (1215 0.95 p=0.96 0.91 p=1.00 0.93 p=1.00)
  - ⊖ obsolete\_component
  - ⊖ obsolete\_function
  - ⊖ obsolete\_process

**Summary View Selected Gene View**

Category Name	P-Chng	P-Undr	P-Ovr	Tot	Chng	Undr	Ovr	Category ID
cytoplasmic sequ...	0.0002	0.0178	0.0260	4	4	2	2	GO:00429...
negative regulati...	0.0002	0.0178	0.0260	4	4	2	2	GO:00429...
transcription fact...	0.0002	0.0178	0.0260	4	4	2	2	GO:00429...
regulation of transc...	0.0002	0.0178	0.0260	4	4	2	2	GO:00429...
regulation of protei...	0.0002	0.0178	0.0260	4	4	2	2	GO:00423...
regulation of nucleo...	0.0002	0.0178	0.0260	4	4	2	2	GO:00468...
chemokine activity	0.0008	0.0782	0.0060	18	8	3	5	GO:00080...
G-protein-coupled r...	0.0008	0.0782	0.0060	18	8	3	5	GO:00016...
chemokine recepto...	0.0008	0.0782	0.0060	18	8	3	5	GO:00423...
chemotaxis	0.0012	0.0547	0.0112	37	12	5	7	GO:00069...
taxis	0.0012	0.0547	0.0112	37	12	5	7	GO:00423...
response to wound...	0.0015	0.0227	0.0296	75	19	9	10	GO:00096...
response to chemi...	0.0018	0.0814	0.0097	54	15	6	9	GO:00422...
response to pathog...	0.0030	0.2972	0.0055	6	4	1	3	GO:00096...
regulation of transp...	0.0030	0.0414	0.0593	6	4	2	2	GO:00510...
immune response	0.0033	0.0002	0.4695	207	39	24	15	GO:00069...
response to pest, p...	0.0036	0.0178	0.0743	123	26	13	13	GO:00096...
extracellular space	0.0038	0.0039	0.2217	47	13	8	5	GO:00056...
protein threonine/tyr...	0.0063	0.0558	0.0794	7	4	2	2	GO:00047...
MAP kinase kinase ...	0.0063	0.0558	0.0794	7	4	2	2	GO:00047...
response to pathog...	0.0063	0.3374	0.0092	7	4	1	3	GO:00428...
antigen processing	0.0070	0.0001	1.0000	15	6	6	0	GO:00303...
antigen presentation	0.0070	0.0001	1.0000	15	6	6	0	GO:00198...
MHC class II recept...	0.0074	0.0024	0.5475	11	5	4	1	GO:00450...
response to extern...	0.0075	0.0400	0.0743	123	25	12	13	GO:00096...
defense response	0.0088	0.0008	0.4993	225	40	24	16	GO:00069...
response to biotic s...	0.0089	0.0013	0.4397	246	43	25	18	GO:00096...
inflammatory respo...	0.0096	0.1695	0.0232	52	13	5	8	GO:00069...
innate immune res...	0.0096	0.1695	0.0232	52	13	5	8	GO:00450...
physiological proce...	0.0099	0.0372	0.1127	1139	153	70	83	GO:00075...
metal ion homeost...	0.0114	1.0000	0.0008	12	5	0	5	GO:00068...
cell ion homeostasi...	0.0114	1.0000	0.0008	12	5	0	5	GO:00068...
di-, tri-valent inorga...	0.0114	1.0000	0.0008	12	5	0	5	GO:00300...
cation homeostasi...	0.0114	1.0000	0.0008	12	5	0	5	GO:00300...
ion homeostasi...	0.0114	1.0000	0.0008	12	5	0	5	GO:00508...
response to abiotic ...	0.0119	0.1597	0.0309	65	15	6	9	GO:00096...
transforming growt...	0.0159	1.0000	0.0048	2	2	0	2	GO:00306...
NF-kappaB-nucleu...	0.0159	0.1107	0.1338	2	2	1	1	GO:00423...


**Operations**

File Name: C:\Source\GoMinerExample\undr.over.2col Browse...


**Quick Help -- Symbol Description**

All/Unchanged Genes
  Changed
  Over
  Under
  Term

Build:138



GENOMICS AND BIOINFORMATICS GROUP  
LMP, NCI, NIH



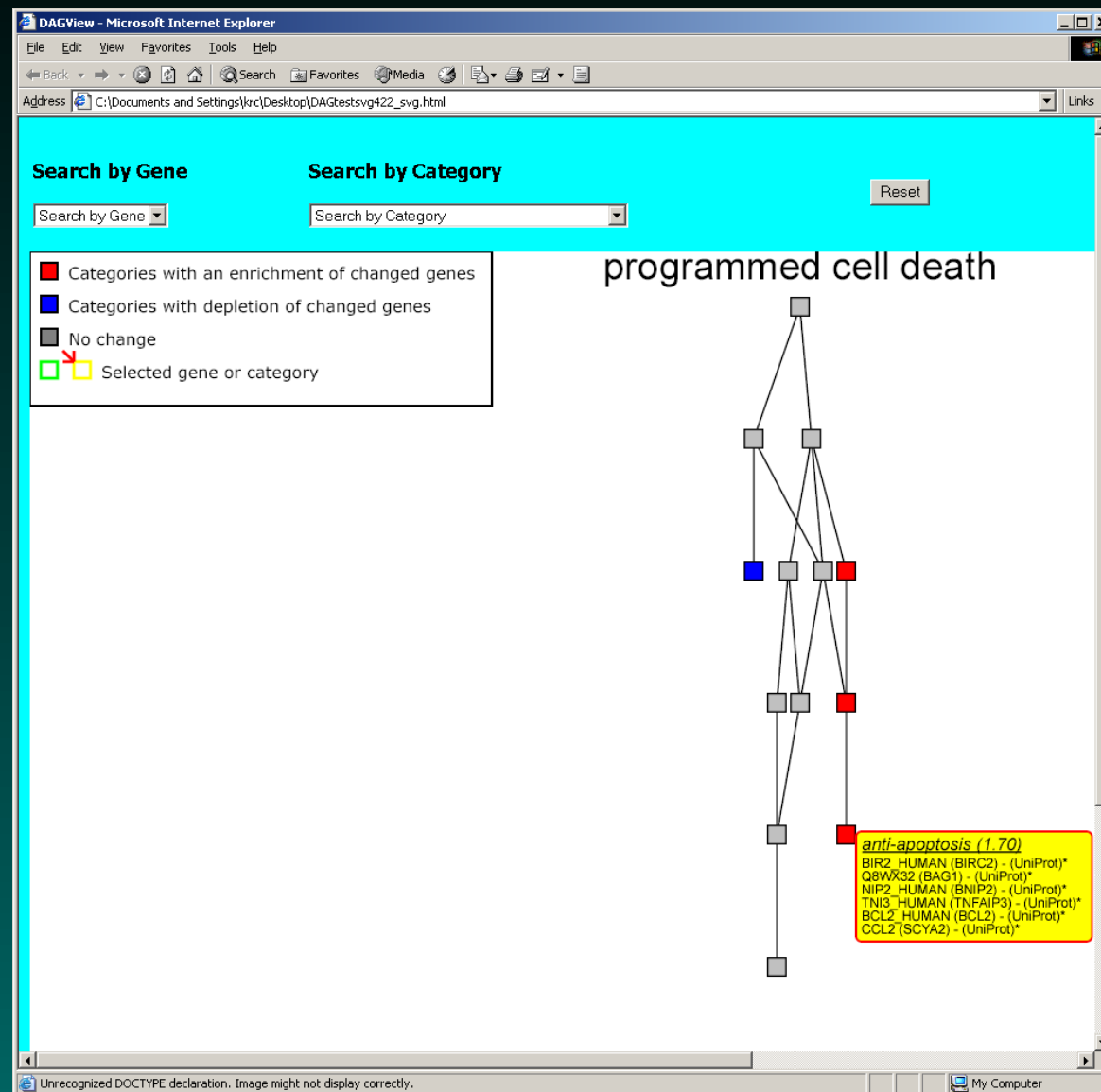
Medical Informatics and Bioimaging Lab  
BME, GA Tech / Emory University

# GoMiner subgraphs

The screenshot displays the NCI/GA Tech - Emory Univ GoMiner application. The interface is divided into several main sections:

- Input Genes:** A table on the left lists input genes with columns for Input, Gene, Source, and Status. Genes include YWHAE, SFN, PPP2R... 2A5A..., HLA-DMA, HLA-D..., HLA-DOA, HLA-DRA, SH3BP2, SLC3A2, A2M, ACTN1, PRKAB1, PRKAG1, ATBF1, ABL1, ABL2, ABR, ACY1, ADAM17, ADA, ADD3, ADH6, ADK, AOX1, ADSS, SLC25A5, MLLT2, GLA, ANGPT1, ANGPT2, and AHR.
- Genes Mapped On GO:** A central panel showing a hierarchical tree of GO terms. The root is 'biological\_process (1245 1.03 p=0.17 1.01 p=0.48 1.02 p=0.17)'. Other terms include 'cellular process', 'development', 'cellular physiological process', 'coagulation', 'death', 'cell death', 'cytolysis', 'programmed cell death', 'apoptosis', 'apoptotic program', 'regulation of apoptosis', 'regulation of programmed cell death', 'homeostasis', 'metabolism', 'organismal physiological process', 'regulation of physiological process', 'response to stimulus', and 'regulation of biological process'. A context menu is open over the 'apoptosis' term, offering options: 'Export summary data to text file', 'DAG of changed genes', 'Export DAG of changed genes to file', and 'Export Genes By Category'.
- Summary View / Selected Gene View:** A panel on the right showing a detailed view of the selected GO term and its associated genes. It lists terms like 'top (1299 1.00 p=1.00 1.00 p=1.00)', 'biological\_process', 'cellular process', 'cellular physiological process', 'cell death', 'physiological process', 'cellular physiological process', 'programmed cell death', 'death', and 'cell death' with associated p-values.
- Operations:** A bottom section with a 'File Name' field (C:\Source\GoMinerExample\...), a 'Browse...' button, and buttons for 'Query Gene File', 'Reset All', 'Query Changed Gene File', and 'Reset Changed'.
- Quick Help -- Symbol Description:** A legend for symbols used in the GO terms: All/Unchanged Genes (circle), Changed (square), Over (up arrow), Under (down arrow), and Term (circle with dot).
- Footer:** 'Build:138' and logos for 'GENOMICS AND BIOINFORMATICS GROUP LMP, NCI, NIH' and 'Medical Informatics and Bioimaging Lab BME, GA Tech / Emory University'.

# GoMiner subgraphs



## Intepreting GoMiner results

Enrichment is computed as

$$\frac{\text{changed genes in category} / \text{total genes in category}}{\text{changed genes on array} / \text{all genes on array}}$$

Statistical evidence of enrichment is based on a Fisher exact test.

## Example

With the sample files supplied from GoMiner, they have an array with 1399 genes, of which 177 change expression. A total of 37 genes on the array are annotated to the biological process category of `chemotaxis`, and 12 of those 37 genes changed expression. So, we have a  $2 \times 2$  contingency table that looks like

chemotaxis:	yes	no	all
Changed	12	165	177
Unchanged	25	1197	1222
Total	37	1362	1399

The p-value arising from a Fisher exact test with these values is 0.001226, which is the value reported by GoMiner for the enrichment of the `chemotaxis` category.

## Intepreting GoMiner results

The p-values from the Fisher test are not corrected for multiple testing, but they should be since one is potentially looking at all GO categories. The categories are not independent, so it is not clear how to correct for multiple testing.

If one filters the gene list from the array before testing differential expression (for example, by removing low expressing or low variance genes), should those genes be included in the “query gene file” for the experiment?

The Fisher exact test is not completely appropriate, since genes can have multiple overlapping annotations into the GO DAG.

No existing test exploits the quality of evidence for the GO annotations.

## Regulatory networks and metabolic pathways

The harder question to be asked about the results of a microarray experiment is: what pathways or networks are changed?

Distinction:

- **Pathways** typically refer to metabolic pathways (that process various small molecules) or signalling pathways that carry information from outside the cell into the nucleus.
- **Networks** typically refer to interactions that control the expression level of various gene products.

Pathways are characterized by direct physical interactions between proteins. Networks are characterized by indirect interactions between genetic DNA or RNA or protein.



## Pathway and network resources

- KEGG (Kyoto Encyclopedia of Genes and Genomes) is a publically available tool, probably best used for metabolic pathways and secondarily for signalling pathways.
- PathArt (Jubilant Biosystems) is a commercial tool for investigating networks and pathways. Underlying network is based on curated interactions from the literature.
- Ingenuity Pathway Analysis (Ingenuity) is another curated commercial tool for investigating networks and pathways.

# http://www.biocarta.com/genes/index.asp

The screenshot shows a Netscape browser window displaying the BioCarta website. The address bar shows the URL <http://www.biocarta.com/genes/index.asp>. The website header includes the BioCarta logo and navigation links: HOME, ABOUT, SUPPORT, CONTACT US, CAREERS, NEWS. Below the header is a secondary navigation bar with links for FEATURES, PATHWAYS, CUSTOM SERVICES, GENES, PRODUCTS, and LOGIN. The main content area is titled "PATHWAYS MAIN CATEGORIES" and features a small diagram of a pathway. The text describes the site's focus on dynamic graphical models of molecular relationships and its open-source approach. Below this is a "BROWSE PATHWAYS BY CATEGORY" section with buttons for various biological processes. At the bottom, there is a "SEARCH PATHWAYS BY TITLE" section with input fields for "Pathway Name" and "Gene Name", each followed by a "SEARCH" button.

**BIOCARTA** Charting Pathways of Life

HOME | ABOUT | SUPPORT | CONTACT US | CAREERS | NEWS

FEATURES | PATHWAYS | CUSTOM SERVICES | GENES | PRODUCTS | LOGIN

Interactive graphic models of molecular and cellular pathways

USA EUROPE JAPAN

## PATHWAYS MAIN CATEGORIES

Observe how genes interact in dynamic graphical models. Our online maps depict molecular relationships from areas of active research. In an "open source" approach, this community-fed forum constantly integrates emerging proteomic information from the scientific community. It also catalogs and summarizes important resources providing information for over 120,000 genes from multiple species. Find both classical pathways as well as current suggestions for new pathways.

### BROWSE PATHWAYS BY CATEGORY

- New Pathways
- Adhesion
- Apoptosis
- Cell Activation
- Cell Cycle Regulation
- Cell Signalling
- Cytokines/Chemokines
- Browse all Pathways
- Developmental Biology
- Expression
- Hematopoiesis
- Immunology
- Metabolism
- Neuroscience

### SEARCH PATHWAYS BY TITLE

Pathway Name  → SEARCH

Gene Name  → SEARCH

# You can't compute on pictures...

The screenshot shows a Netscape browser window with the address bar displaying `http://www.biocarta.com/pathfiles/h_p53Pathway.asp`. The page title is "BioCarta - Charting Pathways of Life - Netscape". The main heading is "PATHWAYS > p53 Signaling Pathway". Below the heading, it says "Submitted by: Craig Monell, PhD" and "Guru: [icon]". There are links for "COMMENT ON THIS PATHWAY", "DESCRIPTION", "CONTRIBUTORS", "SAVE THIS LINK", "SUBMIT", and "LEGEND".

The central diagram, titled "p53 Signaling Pathway", illustrates the molecular events following DNA damage. Key components include:
 

- Damaged DNA**: Initiates the pathway.
- ATM**: Kinase that phosphorylates P53.
- P53**: Central transcription factor that regulates MDM2, RB, and GADD45.
- MDM2**: Inhibits P53 and targets it for ubiquitination (Ub).
- RB**: Inhibits E2F, which is involved in cell cycle progression.
- GADD45**: Promotes DNA repair and cell cycle arrest.
- CDK4, CDK2, CYCLIN D1, CYCLIN E, PCNA**: Cell cycle regulators.
- Apoptosis**: Pathway involving BCL-2, BAX, and APAF-1.
- Ubiquitin Pathway**: Involves Ub and the PROTEASOME.

On the left sidebar, there are sections for "PRODUCT INDEX", "PRODUCT SEARCH", "SEARCH" (with "Contains" and "Exact" options), "Advanced Search", "PRODUCT HIGHLIGHT" (ON/OFF), "PROTEIN LIST", and "REQUEST A CATALOG".

On the right, there are icons for "This Pathway:" (a head icon) and "Other Species:" (a mouse icon).

At the bottom of the diagram area, it says "March 2000".

# http://nashua.case.edu/pathways

Pathways Database System - Home - Netscape

File Edit View Go Bookmarks Tools Window Help

http://nashua.case.edu/pathways/ Search

Home Radio Bookmarks Google Bioinformatics ...

**Center for Computational Genomics**

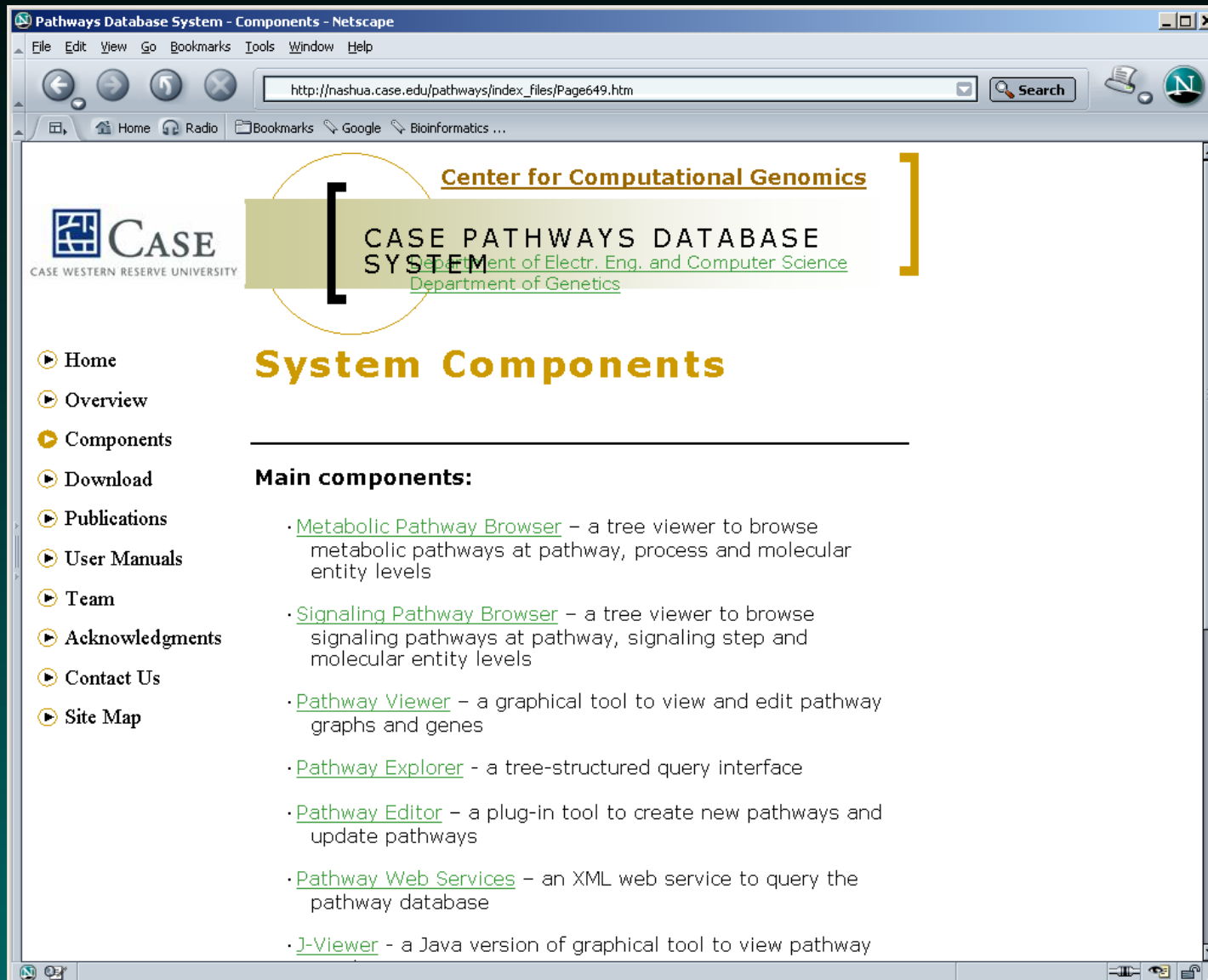
**CASE PATHWAYS DATABASE SYSTEM**  
 Department of Electr. Eng. and Computer Science  
 Department of Genetics

**CASE**  
 CASE WESTERN RESERVE UNIVERSITY

- ▶ Home
- ▶ Overview
- ▶ Components
- ▶ Download
- ▶ Publications
- ▶ User Manuals
- ▶ Team
- ▶ Acknowledgments
- ▶ Contact Us
- ▶ Site Map

Living organisms behave as complex systems that are flexible and adaptive to their surroundings. At the molecular level, organisms consist of intricate networks of molecular reactions, which are often called "biochemical pathways". In order to maintain, visualize, and ultimately, analyze organism functions that result from biochemical pathways, the Pathways Database System is being developed. The system contains a pathways database and the associated tools to store, compare, query, and visualize biochemical pathways. The aim is to develop an integrated database, and the associated tools to support computational analysis and visualization of biochemical pathways. The ultimate goal of the system is to describe, utilize and predict systems functions and behaviors of living organisms.

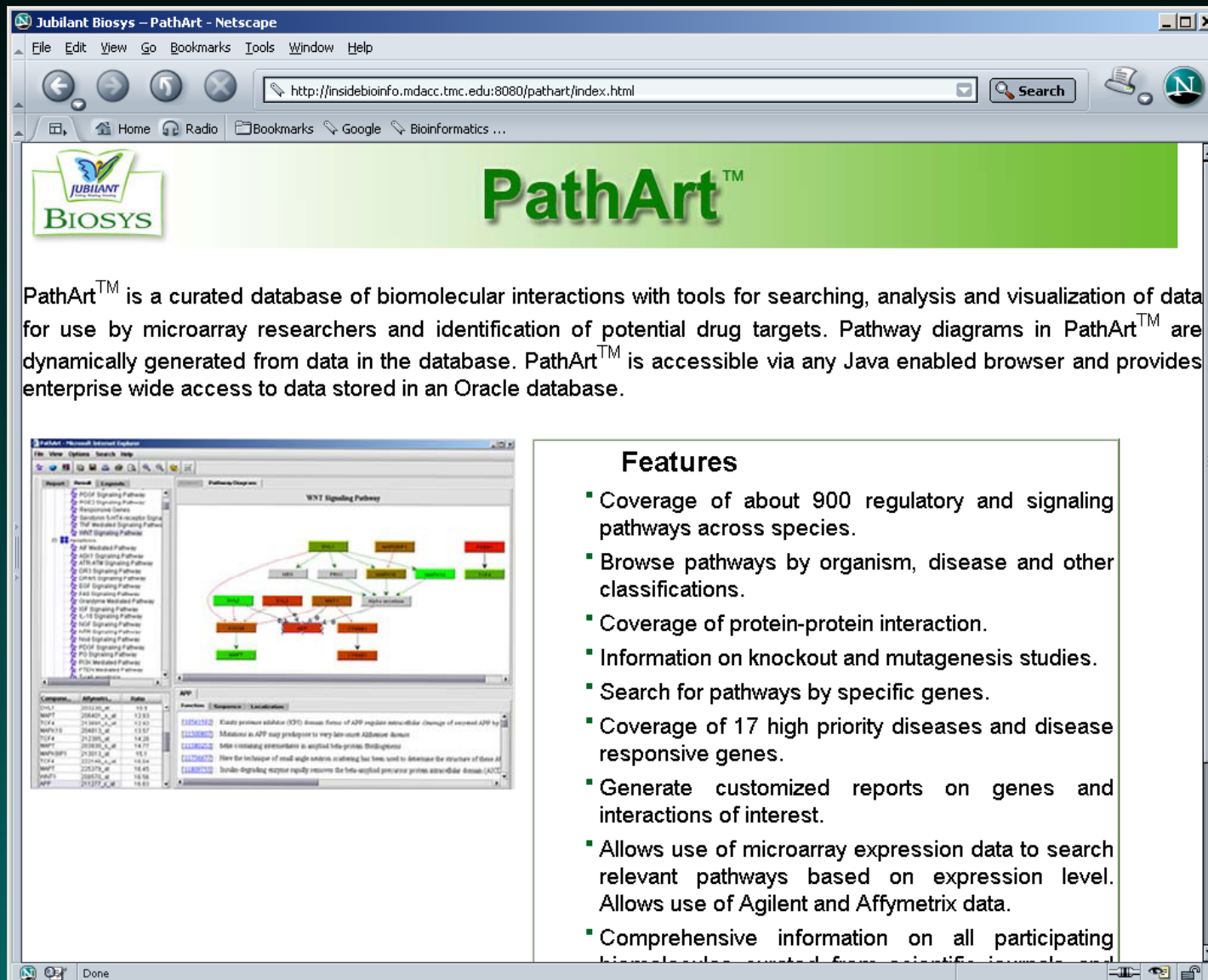
# Case-Western Pathway Browser



The screenshot shows a Netscape browser window displaying the Case-Western Pathway Browser website. The browser's address bar shows the URL [http://nashua.case.edu/pathways/index\\_files/Page649.htm](http://nashua.case.edu/pathways/index_files/Page649.htm). The website header includes the CASE logo (Case Western Reserve University) and the text "Center for Computational Genomics" and "CASE PATHWAYS DATABASE SYSTEM". Below the header, the page is titled "System Components". A left-hand navigation menu lists: Home, Overview, Components (highlighted), Download, Publications, User Manuals, Team, Acknowledgments, Contact Us, and Site Map. The main content area, under the heading "Main components:", lists several tools:

- [Metabolic Pathway Browser](#) – a tree viewer to browse metabolic pathways at pathway, process and molecular entity levels
- [Signaling Pathway Browser](#) – a tree viewer to browse signaling pathways at pathway, signaling step and molecular entity levels
- [Pathway Viewer](#) – a graphical tool to view and edit pathway graphs and genes
- [Pathway Explorer](#) – a tree-structured query interface
- [Pathway Editor](#) – a plug-in tool to create new pathways and update pathways
- [Pathway Web Services](#) – an XML web service to query the pathway database
- [J-Viewer](#) – a Java version of graphical tool to view pathway

# Jubilant PathArt



PathArt™ is a curated database of biomolecular interactions with tools for searching, analysis and visualization of data for use by microarray researchers and identification of potential drug targets. Pathway diagrams in PathArt™ are dynamically generated from data in the database. PathArt™ is accessible via any Java enabled browser and provides enterprise wide access to data stored in an Oracle database.

### Features

- Coverage of about 900 regulatory and signaling pathways across species.
- Browse pathways by organism, disease and other classifications.
- Coverage of protein-protein interaction.
- Information on knockout and mutagenesis studies.
- Search for pathways by specific genes.
- Coverage of 17 high priority diseases and disease responsive genes.
- Generate customized reports on genes and interactions of interest.
- Allows use of microarray expression data to search relevant pathways based on expression level. Allows use of Agilent and Affymetrix data.
- Comprehensive information on all participating biomolecules curated from scientific journals and